Fisheries and Oceans Canada / Pêches et Océans Canada
Ecosystems and Oceans Science / Sciences des écosystèmes et des océans

Canadian Science Advisory Secretariat (CSAS)

Research Document 2024/035

Gulf Region

Southern Gulf of St. Lawrence Scallop Population Model and Limit Reference Point Review

A. Harbicht, L. Landry, M. Niles

Fisheries and Oceans Canada
Gulf Fisheries Centre
343 Université Avenue, P.O. Box 5030
Moncton, NB, E1C 9B6

May 2024

Canada

Foreword

This series documents the scientific basis for the evaluation of aquatic resources and ecosystems in Canada. As such, it addresses the issues of the day in the time frames required and the documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Published by:

Fisheries and Oceans Canada
Canadian Science Advisory Secretariat
200 Kent Street
Ottawa ON K1A 0E6

http://www.dfo-mpo.gc.ca/csas-sccs/
csas-sccs@dfo-mpo.gc.ca

ISSN 1919-5044
ISBN 978-0-660-71854-5 Cat. No. Fs70-5/2024-035E-PDF

Correct citation for this publication:

Harbicht, A., Landry, L. and Niles, M. 2024. Southern Gulf of St. Lawrence Scallop Population Model and Limit Reference Point Review. DFO Can. Sci. Advis. Sec. Res. Doc. 2024/035. v + 96 p.

Aussi disponible en français :

Harbicht, A., Landry, L. and Niles, M. 2024. Modèle de population de pétoncles du sud du golfe du Saint-Laurent et révision du point de référence limite. Secr. can. des avis sci. du MPO. Doc. de rech. 2024/035. v + 102 p.

TABLE OF CONTENTS

ABSTRACT

In accordance with the Precautionary Approach Framework, harvested fish stocks require a harvest strategy be incorporated into their fishery management plans to prevent harm to stocks. This involves establishing a Limit Reference Point (LRP), which is often set to 40% of the biomass at the stock's maximum sustainable yield (B_{MSY}). The goal of this report is to determine an LRP for sea scallops (*Placopecten magellanicus*) in the Southern Gulf of St. Lawrence (sGSL, Northwest Atlantic Fisheries Organization Division 4T), a stock for which available data is insufficient to perform a standard size- or age-based stock assessment.

Despite over a century of scallop fishing in the sGSL and the development of a commercial fishery in the 1950s and 60s, fisheries independent data on the population is limited, and catch-at-age/size data is absent for the commercial fishery. This document compiles a historical time series of catch data spanning from 1923 to 2021 and effort (or its proxy), represented as annual counts of active fishing boats, spanning 1976 to 2021.

Nine well-known data-limited methods are then applied to these data series to derive estimates of stock status and biological reference points, facilitating the calculation of an LRP. Among the models tested, the Bayesian State Space model, JABBA, is identified as the most suitable choice based on its characteristics, assumptions, estimates, and ease of use. Estimates from this model indicate a B_{MSY} for sea scallops in the core area of the sGSL of 1377 tonnes (t) (meat weight), corresponding to an LRP of 551 t.

1. INTRODUCTION

1.1. CONTEXT

Canada's Sustainable Fisheries Framework (SFF) and amendments to Canada's Fisheries Act s.6.2(1) define the goals of the Fishery Decision-Making Framework incorporating the Precautionary Approach (PA) Policy applied to fish stocks that are targeted by commercial, recreational, or subsistence fisheries managed by Fisheries and Oceans Canada (DFO 2006, 2022). The PA Framework (DFO 2009) requires that a harvest strategy be incorporated into respective fisheries management plans to prevent harm to the stock and promote rebuilding when stock status is low. The main goal of the PA framework is to provide DFO with a method of managing fish stocks that is cautious when scientific knowledge is uncertain and which avoids causing serious harm to fish stocks or their ecosystem.

The PA Policy framework identifies the following primary components: biological reference points (BRPs) which delineate stock status zones: healthy, cautious, and critical and, a harvest strategy and harvest decision rules (Figure 1). The upper stock reference point (USR) marks the boundary between the healthy and cautious zones and the limit reference point (LRP) marks the boundary between the cautious and critical zones. The removal reference establishes the maximum removal rates of fish stocks in the healthy zone while the target biomass reference point (TRP), set to a biomass level equal to or greater than the USR, represents a stock abundance level that promotes long term sustainability as well as commercially viable harvest.

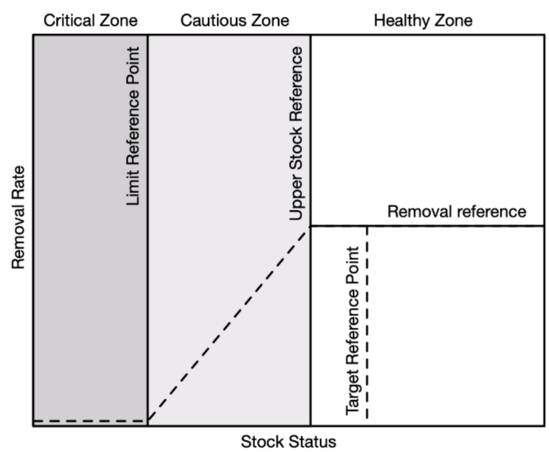

Figure 1. A visual representation of the various biological reference points outlined within the department of fisheries and ocean's precautionary approach framework (from DFO 2009).

The goal of the present research document is to develop BRPs, and primarily an LRP, for sea scallops (*Placopecten magellanicus*) in the southern Gulf of St. Lawrence (sGSL) Northwest Atlantic Fisheries Organization (NAFO) Division 4T, a key harvested stock for which data is limited and no formal biomass estimates currently exist. According to the PA framework, BRPs should be based on standard biomass and harvest metrics (e.g., fishing mortality or exploitation). For sea scallops in the sGSL, however, these metrics are not available through traditional age- or size-based stock assessment models and alternative methods suited to data-limited stocks are required.

1.2. SPECIES BIOLOGY

The sea scallop, a bivalve mollusc, inhabits the northwest Atlantic coastal waters of the Gulf of St. Lawrence, extending south to Cape Hatteras, North Carolina (USA, Posgay 1957). These epibenthic sedentary filter-feeders actively feed on a diverse diet comprising phytoplankton, small zooplankton, ciliates, detritus, and bacteria (Shumway et al. 1987). Their sedentary nature makes them susceptible to disturbances of sediment caused by dredging and fishing operations (Dickie and Medcof 1963; Caddy 1973).

Within the sGSL, scallops play a crucial role as prey for several commercially important species, including rock crabs, lobster, American plaice and yellowtail flounder. Simultaneously, they support a directed fishery, further emphasizing their ecological and economic importance in the region (Naidu and Meron 1986).

1.3. FISHERY

The scallop fishery in the sGSL is organized into four Scallop Fishing Areas (SFAs; Figure 2), with SFA 21 further subdivided into sub-zones A, B, and C. These six SFAs collectively support a multifaceted fishery, including a commercial fishery, a limited Indigenous Food, Social and Ceremonial fishery, and, to a lesser extent, a recreational dive fishery.

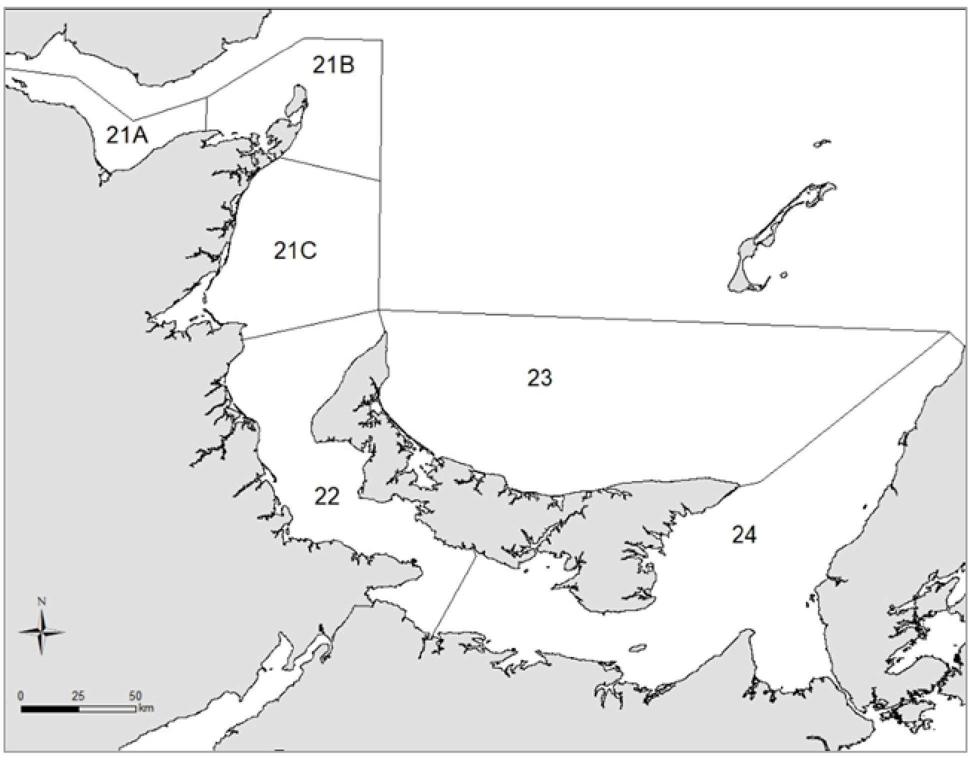

Figure 2. Scallop Fishing Areas (SFA) of the southern Gulf of St. Lawrence.

The commercial fishery operates competitively without quotas, primarily relying on input controls, such as limitations on the numbers of licences available. This number has remained relatively stable over time, fluctuating only between 768 and 773 licences from 2012 – 2016 (DFO 2019). Fishing seasons vary between 24 and 72 days, depending on the area. Over the past two decades, buffer zones have been implemented, creating refuges for scallops, though the primary goal of these buffer zones was to protect juvenile lobster habitat rather than scallops. Fishing vessels are restricted to a maximum length of 14 m, and scallop dredges, typically Digby-type, are restricted to a width of 4.88 or 6 m, with 82.6 or 88.9 mm rings depending on the SFA. While there is no minimum harvest size limits for the commercial fishery, harvest size is influenced by regulations on the ring size and meat count limits (39 – 44 meats per 500 g depending on the SFA).

Monitoring of commercial scallop fishing is done through sales slips from registered buyers and a mandatory logbook program for harvesters. Although the number of licences sold may remain relatively constant over time, the percentage of active licences varies among SFAs, with the highest in SFA 22 and the lowest in SFA 23 (42% and 6% respectively in 2016, Niles et al. 2021) and through time. The commercial scallop fishery is considered complementary, and effort fluctuates based on the value of alternative stocks (e.g., lobster) rather than the price of scallops, despite this stock's value varying considerably over the years ($0.57/kg in 1967 vs. $28.66/kg in 2010, Mallet 2010). This, coupled with the practice of harvesters purchasing "back-pocket" licences (Lanteigne and Davidson 1992), contributes to a potentially significant amount of latent fishing effort (Lanteigne and Davidson 1992).

In contrast, the recreational fishery is managed with a maximum daily limit (50 scallops per diver; 100 in SFA 24), a restricted fishing season (May 1st to Oct. 31), and a minimum size limit (shell heights over 102 mm). Recreational harvesters are also required to maintain logbook entries. Despite minimal catches, Overall, the extent the recreational fishery is minimal, with landings ranging from 0.02 - 0.19 tonnes (t) between 2003 and 2016, and fishing efforts essentially restricted to SFA 21 (Niles et al. 2021).

1.4. CORE HABITAT

Scallops form dense, localized aggregations known as beds, the prime target of commercial fisheries. In the sGSL, these beds are found at depths of 15 – 37 m (Dickie and MacInnes 1958), primarily on sand-gravel or gravel-pebble substrates, occasionally extending to mud-sand or rocky substrates. While scallops display flexibility in substrate use, the density and the condition of adults on a bed is typically higher on sand-gravel or gravel-pebble substrates (McDonald et al. 2021, 2022; Wilson et al. 2021).

Within the sGSL, there are three primary beds, all of which are located within the Northumberland straight (Figure 3) : the West Point and Cape Tormentine beds in SFA 22, and the Pictou bed in SFA 24 (DFO 2019). Historical knowledge of these beds dates back 70+ years (Chiasson 1951, 1952; Dickie 1951), and their distributions have remained relatively stable, as evidenced by spatial analysis of fishing effort described in daily fishing logbooks from 2001 to 2016 (Niles et al. 2021).

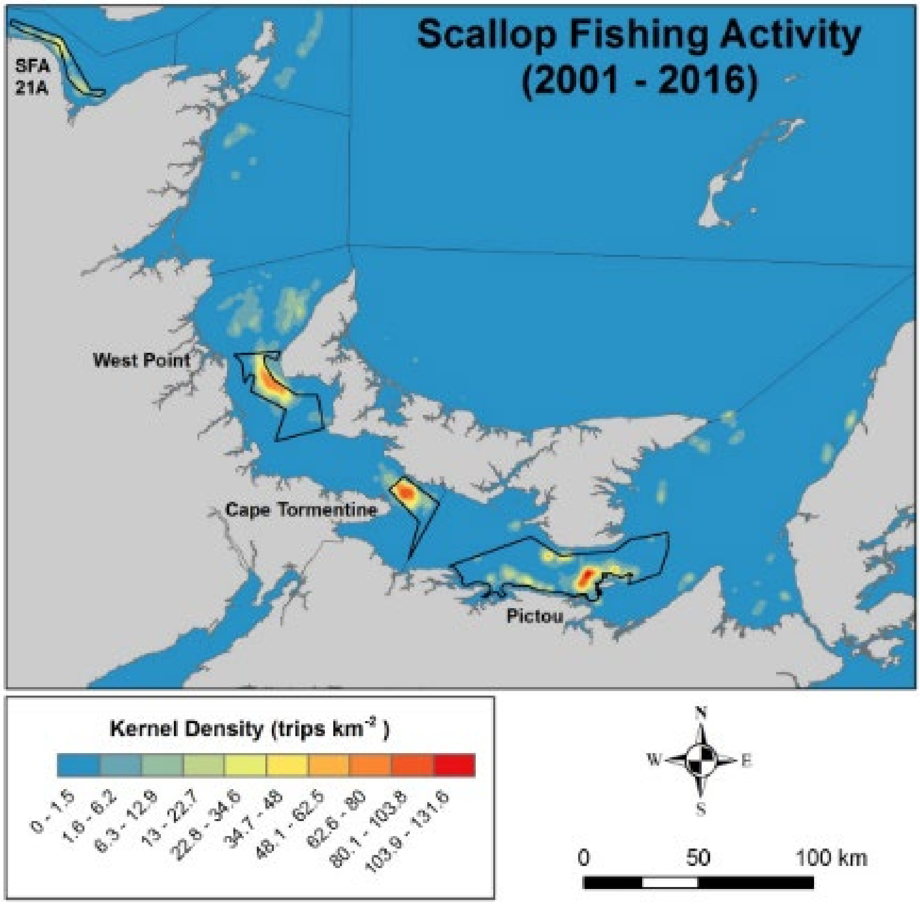

Figure 3. Density of commercial fishing in the sGSL for scallops expressed as the total number of daily trips between 2001 and 2016 (from Niles et al. 2021).

Catches have remained commercially viable on the West Point, Cape Tormentine, and Pictou beds throughout the history of the fishery and the vast majority of landings for the sGSL as a whole originate within the two SFAs that contain them, SFAs 22 and 24. Since the development of the commercial fishery, the proportion of total landings for the sGSL originating within SFAs 22 and 24 have ranged from a minimum of 64% in 1998 to a maximum of ~100% throughout the 1950s and much of the 60s (Figure 4). Additional beds exist within these SFAs, but their contribution to total landings is limited. Specifically, the two primary beds in SFA 22 contribute, on average, 78% of the total landings for this SFA, while the Pictou bed contributes, on average, 77% of the total landings from this SFA (Table 14 in Niles et al. 2021). As a result, SFAs 22 and 24 have been identified as the core scallop habitat within the sGSL, with reported landings primarily reflecting the input from the three primary beds above.

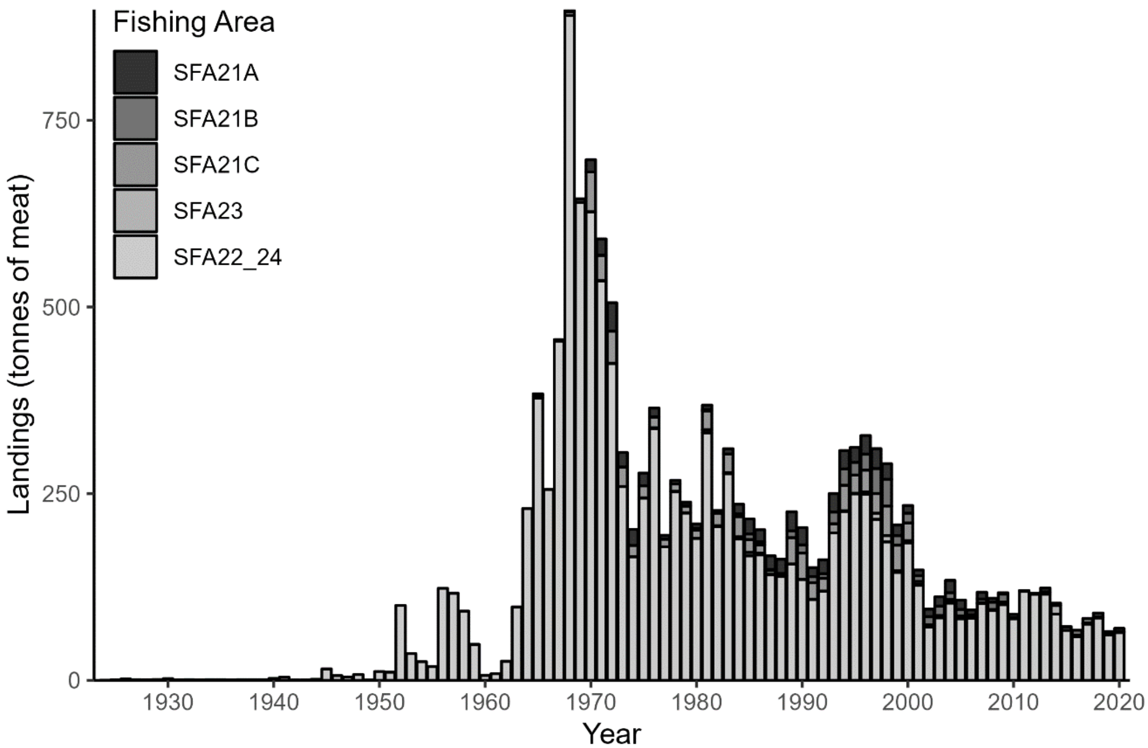

Figure 4: Annual scallop landings in the sGSL according to scallop fishing area (SFA). SFAs 22 and 24 represent the core habitat within the sGSL and are presented as a single unit.

Identifying core habitat is crucial for sedentary species like scallops, as these areas serve as the main demographic source within the larger metapopulation. Core populations act as a source to replenish less productive habitats following population declines due to mass die-offs or overfishing. By monitoring such core habitat, managers can assess the long-term stability of the larger population (Smith et al. 2015, 2017). In the sGSL, the three beds located in the SFAs 22 and 24 represent the core scallop habitat (Niles et al. 2021) and will be the focus of this research document.

1.4.1. Catch

This document compiles a comprehensive time series of catch data spanning from 1923 to 2021 from various sources. Further details on dataset compilation are available in appendix 1. It's important to note that early records for the scallop fishery may be incomplete, as initial data collection was reported only annually by county and province from 1923 to 1946. Communication among provincial regulators was inconsistently recorded during this period, impacting data accuracy. However, data quality significantly improved over time.

Starting in 1947, records were compiled annually by statistical district, and from 1967, monthly records were kept by landing port in a computerized system. By 1977, data were recorded at the level of individual sale transactions, specifying landing types as meat, roe, or live sales. Since 1982, sales slips have been categorized into commercial sales and local sales destined for personal consumption (supplementary B), along with irregular sales to registered buyers (supplementary A). For this study, all landings are considered in terms of kilograms of meat using a conversion factor of 8.3 (Lanteigne and Davidson 1991) for landings listed as live or roe. Due to the lack of precision in the estimated landings for personal use, supplementary B landings were excluded from the analyses below.

The early development of the scallop fishery in the sGSL was gradual and initially concentrated solely within the Northumberland straight. Significant growth occurred in the 1950s when offshore scallop trawlers from the Maritimes region extended their harvesting activities into the Gulf (Bourne 1964; Jamieson 1978). However, by 1957, new regulations restricted the fishery within the sGSL to the inshore fleet (vessels less than 20 m in total length), resulting in a brief drop in catches (Figure 4).

During this time, productive scallop beds were identified within the core area and elsewhere in the sGSL, leading to a rapid increase in catches, peaking at nearly 900 t of meat in 1968 before declining. By the mid-1970s, annual landings had decreased to approximately 300 t. Despite increased input from the SFA 21 areas during the 1990s, catches never again exceeded 370 t. In recent years, scallop catches from the sGSL core area have stabilized at around 80 t, with 66.50 t of scallop meat landed in 2021.

As the landings have been recorded digitally for each sales slip as of 1982, it is assumed that the accuracy of these records are relatively high. For this reason, when models accept an estimate of observation error around catch values, a CV of 0.1 was used for the entire time series to account for unrepresented catches for personal consumption (supplementary B) or general miss-reporting of landings. When possible, models that allow differing levels of observation error throughout time were given a catch CV of 0.2 prior to 1976.

1.4.2. Effort

Similar to the record-keeping for landings, the measurement of effort in the sea scallop fishery has undergone changes over the years, evolving towards increased precision. Initially, annual records lacked detailed information about effort, providing only the number of licences issued (Cardin 1924). Subsequent refinements included collecting counts of active licences (Jamieson 1978), and later, the number of fishing trips per season (Worms et al. 1986). Over time, the collection of effort-related data improved significantly.

Since 2003, harvesters have been mandated to maintain daily logbook entries, recording the number and duration of tows performed (DFO 1998; Davidson et al. 2007). Given the importance of effort indices, particularly for population models and catch per unit effort (CPUE) calculations, our objective was to construct the longest possible time series of effort indices.

The most accurate index of effort, the number of hours towed, is the product of the average tow duration and the number of tows per trip, and only available from 2003 onwards (Davidson et al. 2007). To avoid limiting our input effort data to a relatively short timeseries, we opted to use a metric with a much longer record: annual numbers of active fishing boats. Between 1985 and 2021, the number of active scallop fishing boats was determined from the number of unique Canadian Fishing Vessel (CFV) numbers present in the digitized sales slips. Prior to 1985, data on the number of active fishing vessels was estimated as the reported numbers of active licence holders, with a ratio of 1:0.98 with active boats based on 23 overlapping years. Using active licence holders as a proxy for active fishing boats extends the time series back to 1976.

Effort estimates prior to 1976 were derived from survey questionnaires, which had limitations in scope and accuracy. Participation in these surveys was voluntary, and harvesters often indicated active participation in every fishery listed. Hence, only effort data from 1976 onwards are considered here. The use of active boats per year as the index of effort is justified due to the increase level of information it provides for modelling purposes, 46 years rather than 19, and due to the strong positive correlation with the more accurate metric of hours towed (Pearson, $r = 0.94$, $p < 0.01$, Figure 5).

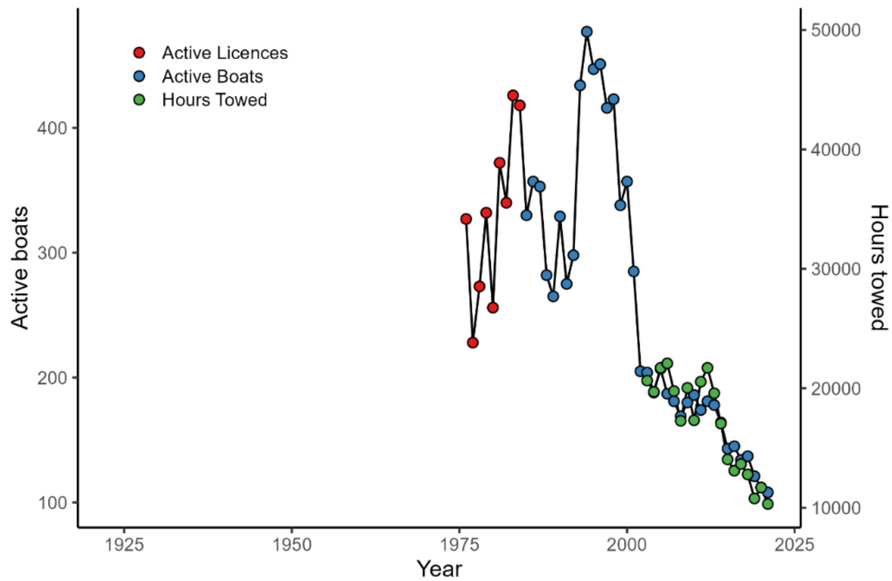

Figure 5. Effort data used in the estimation of an LRP for the sGSL scallop stock. Colors indicate the source of the effort index. The number of active licences was considered as a 1:1 proxy for the number of active boats. The annual total number of hours towed (green) is included to support the use of the highly correlated and longer timeseries of active boats per year.

1.4.3. CPUE

To facilitate modeling, a time series of catch per unit effort (CPUE) in the sGSL scallop fishery was calculated from 1976 to 2021. CPUE was derived by dividing the total catch (tonnes) within the core area over the number of active fishing boats. Although some changes, such as a reduction in maximum drag width occurred during this period, they were infrequent and generally applied uniformly across both SFAs. Since the effort and catches are aggregated across these two SFAs to produce a single timeseries, raw and unstandardized CPUE values were used for modelling purposes (Figure 6).

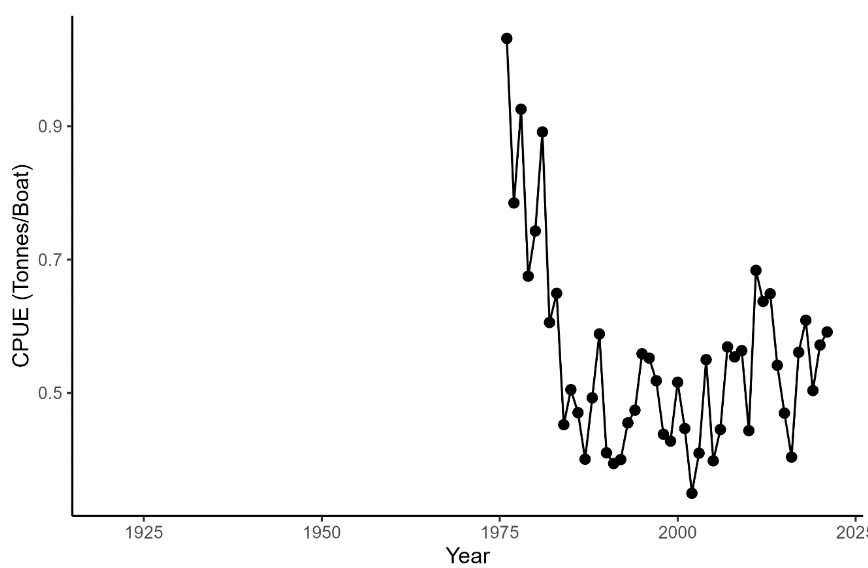

Figure 6. Catch per unit effort estimates from the scallop fishery in the sGSL based on total landings from the core area and the number of active fishing boats.

1.4.4. Abundance Estimates

At present, biomass estimates for the core area within the sGSL are available via two primary sources: an ongoing fishery-independent research survey and a depletion model based on cumulative landings and CPUE. Both methods aim to estimate the biomass present on the three main beds within the core area, namely West point, (WP), Cape Tormentine (CT), and Pictou (P).

Logbook records since 2003 provide data on the locations of fishing efforts within the core area. Valid locations, accounting for an average of 81% of total logbook records, are assumed to be representative of all fishing efforts. Areas with 20 fishing trips or more per km^2 from valid locations are used to classify the main beds. In 2021, 48.6% of all landings within core area originated within one of the three main beds. Assuming uniform catchability (q) across both major and marginal beds within the core area, landings are considered representative of the underlying biomass. A rough biomass estimate for the entire core area can then be extrapolated from biomass estimates on the beds using a conversion factor of 2.06 (1/0.485). However, if q is lower outside of the main beds, the resulting biomass estimate for the entire core area would be an underestimate.

A fishery-independent research survey conducted each fall since 2019 provides biomass estimates for the main beds. In 2021, estimates were 43 t for West Point, 40 t for Cape Tormentine, and 22 t for Pictou. Collectively, this biomass estimate of 105 t for the three beds corresponding to a core area estimate of 216 t using the conversion factor.

More recently, since 2022, parallel research surveys have been conducted in the spring, prior to the start of the annual scallop fishing season. These spring surveys suggest that biomass estimates on the beds may be as much as 39 - 65% higher earlier in the season (Niles, personal communication). This suggests that the survey biomass estimates in the core area may be underestimates, and that the true pre-fishery biomass levels could be as high as 356 t.

A Leslie depletion model applied to the landings of the three core beds in 2021, based on logbook locations, indicates biomass estimates of 19.4 t for West Point, 32.8 t for Cape Tormentine, and 24.7 t for Pictou beds. These results align with the October survey estimates and result in a collective biomass estimate of 76.85 t for the major beds and 158.1 t for the core area as a whole, using the conversion factor. Combining these results with survey estimates suggests biomass levels in the core area are likely range between 158.1 and 216 t, possibly reaching as high as 356 t in the spring, before the fishery begins.

1.5. MODELLING PARAMETERS

1.5.1. Sexual maturation

In the sGSL, sea scallops attain sexual maturity at shell heights greater than 70 mm, typically around 4 years old. However, they only contribute noticeably to population recruitment when gonadal growth exceeds somatic growth (~85 mm shell height, or age 5-6 in the sGSL), and egg production begins to increase exponentially with shell height (Beninger 1987; Bonardelli and Himmelman 1995). For this reason, an age of 5 will be used in subsequent models which require an age at maturity as an input variable.

1.5.2. Natural mortality (M)

The determination of the natural mortality rate (M) for a stock poses challenges due to its dynamic nature across time, ages, and environmental conditions. Direct measurement is often difficult, however, indirect estimation methods leverage more easily measurable demographic and growth parameters, and can provide a range of possible values. Cope and Hamel (2022) developed a tool integrating these methods, using von Bertalanffy growth parameters (L_{inf} = 14.68 cm, k = 0.133, t_0 = -0.947, Figure A1.12), a maximum age of 16 (from survey samples), and an age at 50% maturity of 5 (as mentioned above). The tool generated M estimates ranging from 0.2 to 0.42 with an average of 0.31 (n =12).

Additionally, direct estimation methods based on clapper ratios observed during research surveys and at-sea sampling (1982 – 2023), following the method of Merrill and Posgay (1964), produced M estimates ranging from 0.08 to 0.38, averaging 0.23 (n = 19). Combining all estimates yields a mean M of 0.26 (Figure 7). This average is slightly higher than observations in comparable scallop populations in the Bay of Fundy (Smith and Lundy 2002) or on Georges Bank (Hart and Chang 2022) where average M estimates were 0.19 and 0.23 respectively. For models requiring a prior distribution for M, a lognormal distribution was used with a mean of 0.26 and a standard deviation on the log scale of 0.15.

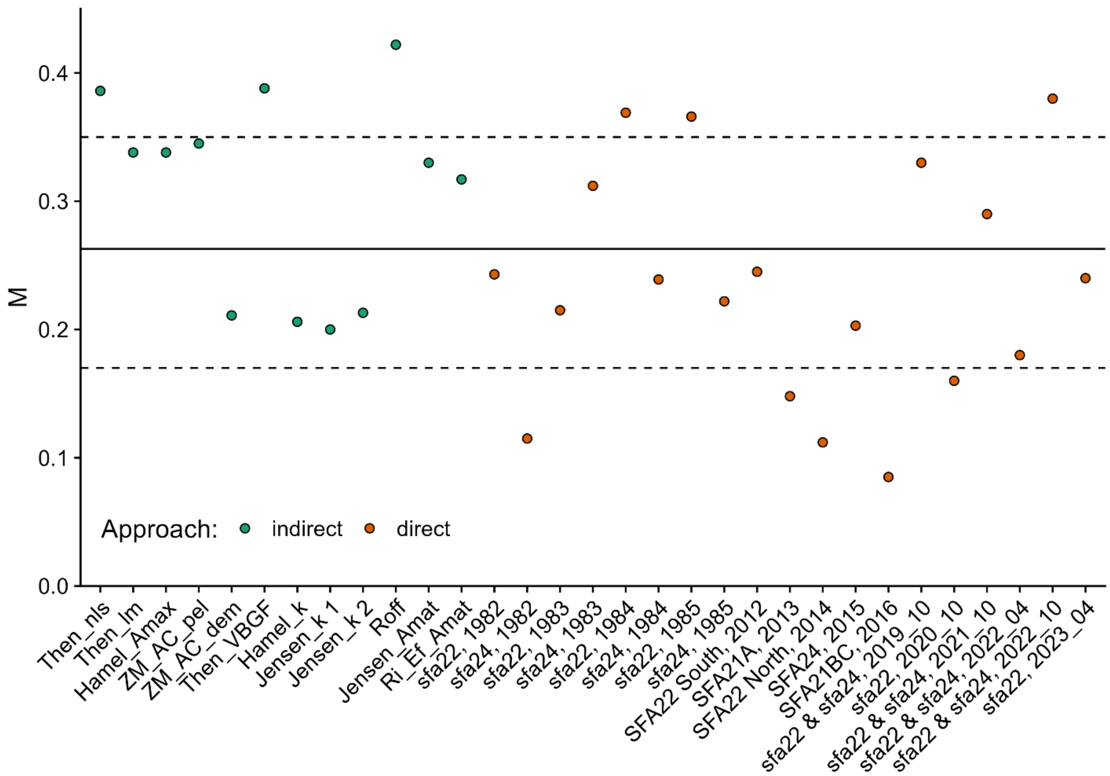

Figure 7: Natural mortality (M) estimates based on indirect and direct estimation methods. Indirect methods were implemented via the estimation tool described in Cope and Hamel (2022), direct estimates are from clapper ratios counted during research surveys or at-sea-sampling.

1.5.3. Carrying capacity (K)

Determining the carrying capacity of the core habitat within the sGSL presents challenges, however, a rough estimate of the range can be produced based on the catch history. Scallop populations, being sedentary and localized on densely populated beds with locations known to harvesters, may face considerable depletion within a single fishing season from particularly high fishing effort. Considering the largest annual landings from within the core area were ~900 t, if we assume the beds were nearly completely depleted that year, the lower limit for K can conservatively be set to 1000 t. Alternatively, if we assume the most intense fishing within this area only depleted the population by 1/6, then an upper limit for the possible range of K can be set to 6000 t. The population of scallops in the core area was therefore assumed to have an average value at the midpoint of 3000 t, and a standard deviation of 1000, following a normal distribution, producing the requisite range of possible K values.

1.5.4. Saturation levels (S)

The catch time series for the core area extends back to 1923, capturing the early stages of the fishery's development. Consequently, initial saturation levels are assumed to be at 1, signifying that the population was at its carrying capacity. Models incorporating priors for initial saturation levels (B_0/K) were assigned a highly confident prior with a mean of 1 (or 0.99 for the beta distribution) and a standard deviation of 0.01, producing a range of 0.95-1.

Recent estimates of biomass in the core area, described above, suggest a biomass ranging from 158 t (according to depletion models), or between 216 t and 356 t (according to survey results). Considering the estimated range for K of 1000 - 6000, current saturation levels (B_{2021}/K) are inferred to be between 0.026 and 0.36, with a stronger likelihood around 0.06. In modelling, a moderately confident prior was provided a mean of 0.05 and a standard deviation of 0.1 which corresponds to a range of 0.026 – 0.20.

For models requiring an intermediate saturation level prior, an uninformative prior of a mean of 0.6 and a standard deviation of 0.3 with a normal distribution was provided for 1975. This reflects the belief that the population was still above B_{MSY} at this time.

1.5.5. F_{MSY}/M

The ratio of the fishing removal rate at MSY (F_{MSY}) and the natural mortality rate (M) is a key indicator of a stock's resiliency, and its value varies among species. While Restrepo et al. (1998) suggest that M is often a conservative estimate of F_{MSY}, recent perspectives, such as Zhou et al. (2012), highlight that F_{MSY}/M is frequently below 1 for many fish, sharks, and rays. Conversely, bivalve molluscs, known for their high resiliency, typically exhibit F_{MSY}/M greater than 1. An example is the Georges Bank and Mid-Atlantic stocks, where the long-term average fishing mortality is 0.75, and the estimated natural mortality rate is only 0.1, resulting in an F_{MSY}/M of 7.5 (Hart 2006). With a stochastic yield-per-recruit model for the same stocks, Hart (2013) used deterministic M estimates of 0.12 and 0.15 based on clapper ratios, and found F_{MSY} estimates ranging from 0.33 under their baseline assumptions to 0.17 under low resiliency conditions. While lower than her previous estimates, this model's estimated F_{MSY}/M of between 2.75 and 1.13 are still in excess of 1. In light of these considerations, model priors for F_{MSY}/M were set to a mean of 2 in the lognormal distribution, with a standard deviation of 0.2.

1.5.6. B_{MSY}/K

The B_{MSY}/K ratio plays a crucial role in shaping the surplus production curve. For species adhering to a Schaefer (logistic) population growth model, this ratio is 0.5 (Schaefer 1954a). However, populations with a steeper population growth trajectory at low densities, often described by the Fox or more generalized Pella-Tomlinson model (Pella and Tomlinson 1969; Fox 1970), may exhibit different ratios.

Given the robust nature and high reproductive potential of sea scallops, it can be expected that B_{MSY}/K for this species falls within the range of 0.25 to 0.5, corresponding to a Pella-Tomlinson shape parameter between 1 and 2. For models requiring a prior for B_{MSY}/K, a mean of 0.38 in a normal distribution, with a standard deviation of 0.07 is employed. Alternatively, a mean Pella-Tomlinson shape parameter of 1.5 in a normal distribution, with a standard deviation of 0.25 is used.

1.5.7. Intrinsic rate of population growth (r)

Sea scallop, recognized for their high fecundity, rapid individual growth rate, low age at maturity, and longevity, are considered moderately resilient with notably high intrinsic growth rate (Smith and Rago 2004; Lidgard and Norden 2011). SeaLifeBase categorizes sea scallop r values with an average of 0.56 and ranging between 0.37 – 0.84. This range is consistent with estimates near 0.50 for the Georges Bank stock (Dvora Hart, personal communication).

For models requiring a prior for r, a normal distribution with a mean of 0.5 and a standard deviation of 0.1 is used. This corresponds to a range of 0.3 - 0.8. and is consistent with the documented resilience and growth characteristics of sea scallops while remaining sufficiently general.

2. INTERIM MANAGEMENT ADVICE

While the main focus of this document is to assess various models for generating biomass estimates in the sGSL scallop stock and establish a biomass limit corresponding to an LRP of $0.4B_{MSY}$, we acknowledge the unique challenges posed by stocks like the sGSL sea scallop. Despite the abundance of data collected over the years, much of these data do not align with the requirements of conventional stock assessment methods, making confident management decisions challenging.

It's important to recognize that models intended for such situations may not always yield optimal results, and there may be a need for in-depth discussions regarding their outcomes. To address this, we have incorporated less detailed methods that can provide potential interim removal limits that may be used while awaiting more robust assessment results.

2.1. SCALAR APPROACH

The scalar approach, developed by Restrepo et al. in 1998, simplifies the establishment of precautionary removal limits by using a historical catch pattern as a proxy for the maximum sustainable yield (MSY) and reducing this value by a scalar multiplier. The MSY_{proxy} is identified as the mean catch during a time period with relatively stable catch and effort levels and the scaling factor applied is chosen based on the whether the population is believed to be above, at, or below B_{MSY}. The authors suggest a range of slightly precautionary ($0.75MSY_{proxy}$), to highly precautionary ($0.25MSY_{proxy}$). While effective in the short term, this approach has limitations, particularly when dealing with variable effort or developing fisheries and may result in overly cautious removal limits.

In the sGSL sea scallop catch history for the core area, a period of relatively stable catches and effort occurred between 1976 and 1987, with an average landed catch of 221.9 t. Given the current believe that the scallop population is below B_{MSY}, a scalar of $0.25MSY_{proxy}$ was applied, resulting in a recommended removal limit of 55.5 t per year. Although historical catches consistently exceeded this removal limit, recent years (2019-2021) have approached or matched this limit.

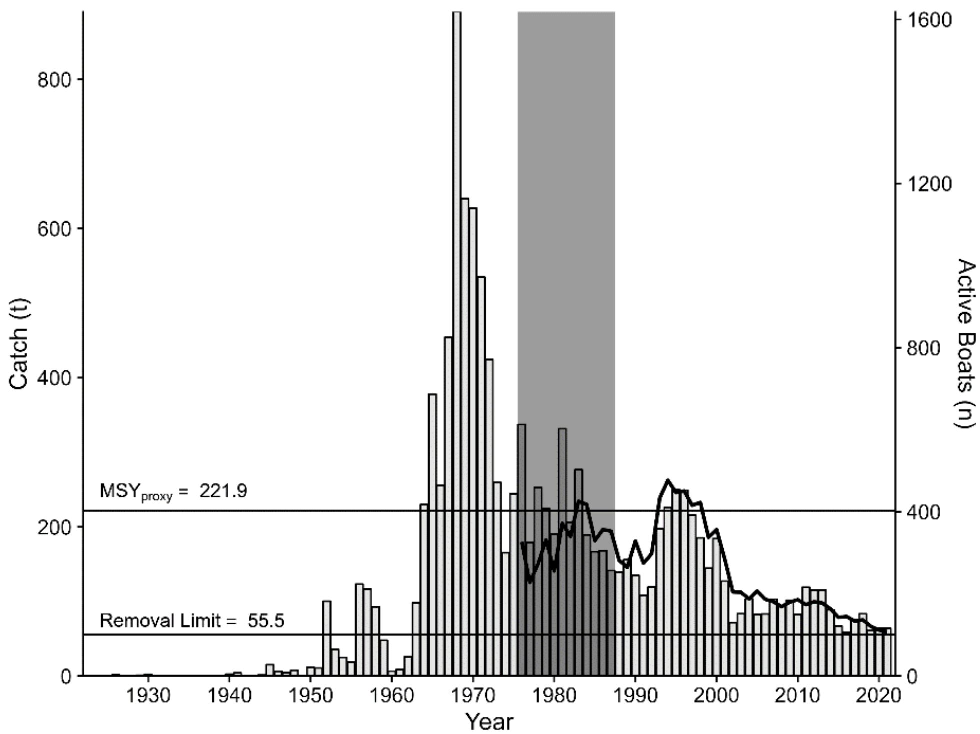

Figure 8. Catch (bars) and effort (line) for the core sea scallop habitat in the southern sGSL, comprised of scallop fishing areas 22 and 24. The area in grey represents a period with roughly stable catches and effort from which a proxy for MSY was calculated.

2.2. DEPLETION-CORRECTED AVERAGE CATCH (DCAC)

The depletion-corrected average catch (DCAC) method, developed by MacCall (2009), builds upon the potential-yield formula of Alverson and Pereyra (1969) and Gulland (1971). This approach incorporates the consideration of an initial windfall harvest as stock abundance decreases from B_0 towards B_{MSY}. The DCAC method calculates a sustainable yield (Y_{sust}) aimed at preventing further stock declines, assuming the stock is maintained near historical abundance levels.

To apply the DCAC method, estimates of M, B_{MSY}/B_0, F_{MSY}/M, and depletion (1-S) are required. Utilizing the provided estimates, and their ranges/distributions, the DCAC method yielded a median Y_{sust} value of 127.9 t per year (Figure 9). Historically, catches in the sGSL surpassed this sustainable removal limit. However, this trend changed in 2002 when landings fell below Y_{sust}, a pattern that has persisted to the end of the time series.

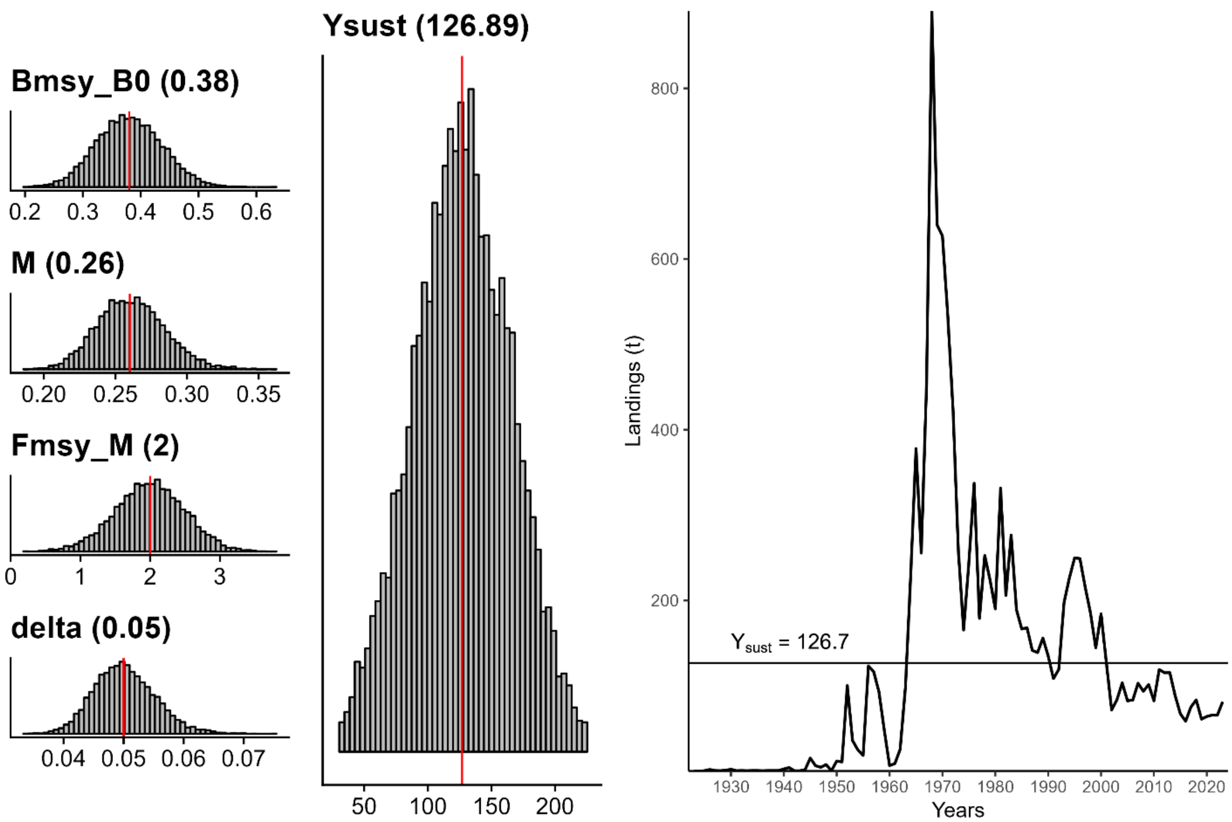

Figure 9: The input (left) and output (middle) parameter ranges for the Depletion-Corrected Average Catch method, and the mean sustainable yield plotted against the timeseries of total landings from the core area within the southern Gulf of St. Lawrence (right).

2.3. CATCH RATIO ANALYSIS

The catch ratio analysis method, developed by Froese and Kesner-Reyes (2002) and refined by Anderson et al. (2012), categorizes stocks based on temporal patterns in a smoothed catch time series (C). This method distinguishes between underdeveloped, developing, fully exploited, overexploited, and collapsed states using the ratio of $C:C_{max}$ (where C_{max} is the maximum landed catch) and thresholds derived from the RAM Legacy Stock Assessment Database (Ricard et al. 2012; re3data.org). In the present application, the smoother was set using a span of 0.3 as opposed to the authors' suggested default of 0.6 in order to more accurately capture the peak catch in the late 1960s.

When applied to the core area, the sGSL sea scallop stock was classified as developing until 1973, transitioning to a fully exploited state until 1983, and subsequently moving to an overexploited state (Figure 10). Based on the catch ratio and relative biomass relationship modelled in Anderson et al. (2012), populations are at approximately B_{MSY} when the catch ratio is ≈ 0.80. For scallops in the core area within the sGSL, this occurred in 1974, with an annual catch of 165.43 t. Using this as a proxy for MSY and applying the same scalar as above (0.25), a precautious removal limit of 41.36 t can be set.

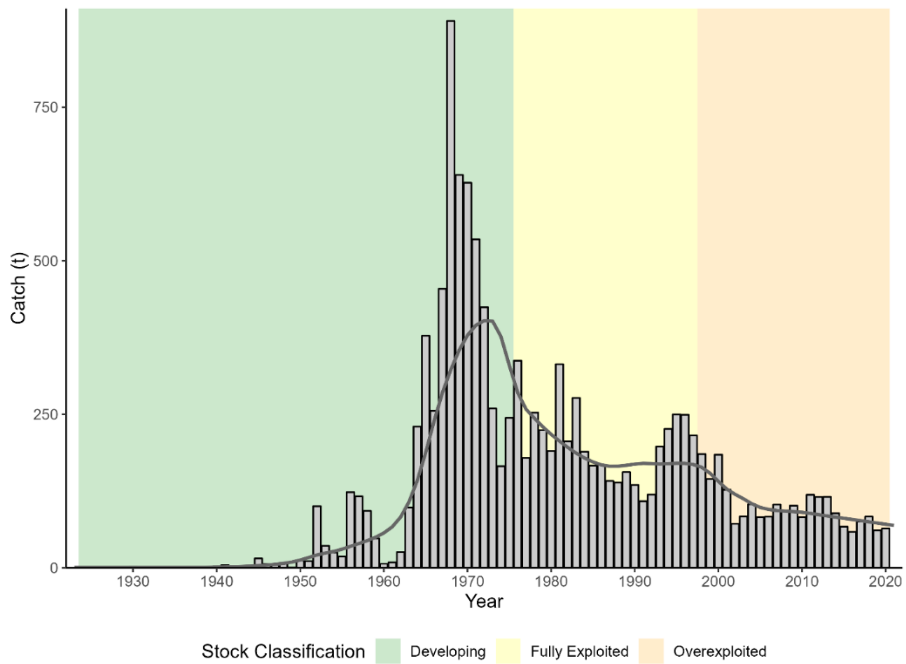

Figure 10. Sea scallop stock status plot based on the methods of Anderson et al. (2012) for the core area within the sGSL (scallop fishing areas 22 and 24).

3. LIMIT REFERENCE POINT

A limit reference point (LRP) is a critical threshold indicating when a stock's biomass is considered inadequate to sustain recruitment and reproductive capacity. Stocks below their LRP are susceptible to recruitment overfishing and significant harm may result to the stocks, their habitat, or other ecologically associated species. LRPs serve as operational control points in harvest strategies, triggering the implementation of rebuilding plans under policies like the Precautionary Approach (PA) Policy (DFO 2021) and Integrated Fishery Management Plans (IFMPs).

Selecting an LRP involves various methods such as formal stock assessments estimating spawning stock biomass at maximum sustainable yield (SSB_{MSY}, or B_{MSY}) and proxies derived from average biomass estimates during productive fishing periods when data are limited. Other approaches involve setting an LRP as a fraction of the maximum predicted recruitment, the lowest biomass level from which a stock has been observed to have recovered ($B_{recover}$), or a fisheries related metric such as F_{MSY}. Evaluating multiple LRP candidates estimated through different methods is crucial for enhancing confidence and identifying risks.

Estimates of B_{MSY} can be obtained from population models commonly used to assess fish stocks, such as surplus production models and delay-difference models. These models are combined with others describing the relationship between spawning stock size and expected recruitment, like the Beverton-Holt stock recruitment model. However, when data on stock abundance and/or composition are absent or limited, the use of formal stock assessment models to estimate B_{MSY} or F_{MSY} may be restricted. In such cases, biologists can resort to simpler models requiring less input data (catch only or catch and effort) which make assumptions about the stock to produce biomass estimates. These options provide flexibility when data limitations exist while still allowing for informed management decisions.

3.1. DATA-LIMITED STOCK STATUS

Recently, Boudreau and Duplisea (2022) introduced a categorization tool addressing uncertainty in developing LRPs for Canadian fish stocks with limited data. Their framework aligns broadly with the National Oceanic and Atmospheric Administration (NOAA) tiers 1 to 6 (Reuter et al. 2010; Newman et al. 2015; Punt et al. 2020) and the International Council for the Exploration of the Sea (ICES) classification levels 1 through 5 (ICES 2012, 2021). Regardless of categorization, all stocks are assumed to be commercially harvested, ensuring the availability of a time series of catches.

The four data-limited categories are:

- Category A: Stocks with abundance indices and catch-at-length.
- Category B: Stocks with abundance indices but no catch-at-length.
- Category C: Stocks without abundance indices but with catch-at-length.
- Category D: Stocks with only landings data.

The sGSL sea scallop stock falls into category B. It has a time series of catches (1923 – 2021) and fisheries-based abundance indices (commercial CPUE), occasionally supplemented with fishery-independent abundance indices. However, there is no catch at length data for most of the fishery.

4. POPULATIONS MODELS

This research document will investigate models with limited data requirements, particularly those designed for category B stocks, as proposed by Boudreau and Duplisea (2022), along with additional models (summarized in appendix 2). The predicted biomass and parameter estimates generated by these models will be compared to select a single model and its corresponding LRP calculated as of 40% of the predicted B_{MSY}. It is important to note that all of these data-limited models assume catch records from a fishery are representative of stock abundance to some extent. However, in the sGSL, the sea scallop fishery is complementary to the herring and lobster fisheries (DFO 1996) and scallop boats tend to be multi-purpose vessels. This raises questions about the assumption's applicability. Nevertheless, the catch and effort time series remains the longest available, and despite potential influences, it is expected to capture major underlying patterns in sea scallop stock abundance trends in the sGSL.

4.1. BOOSTED REGRESSION TREE (ZBRT)

The boosted regression tree (zBRT) method, a catch-only approach, utilizes a machine learning algorithm to predict the underlying biomass trends over time by analyzing patterns in the catch time series. Generally, boosted regression trees employ a two-part model composed of a recursive partitioning tree and a sequential boosting procedure which gradually emphasises observations modelled poorly by the initial fit (Elith et al. 2008). Initially applied to fisheries data by Zhou et al. (2017), the zBRT method uses the RAM Legacy database as its learning dataset, extracting predictor variables from trends in the catch time series. The model output includes estimates of saturation (S) and relative biomass (B/B_{MSY}).

When applied to the core landings data using the zBRT function in the datalimited2 R package (Free 2018), the model suggested that scallops in the core area have oscillated around B_{MSY} for approximately 65 years (1932-1997). However, it indicated a dramatic decrease to a temporarily stable position near the LRP ($0.4B_{MSY}$) between 1995 and 2005. Subsequently, the model results suggested a collapse in 2015, maintaining a relative biomass level of 0.02 from 2015 to 2021, implying that the stock has been in the critical zone of the PA for the last 6 years.

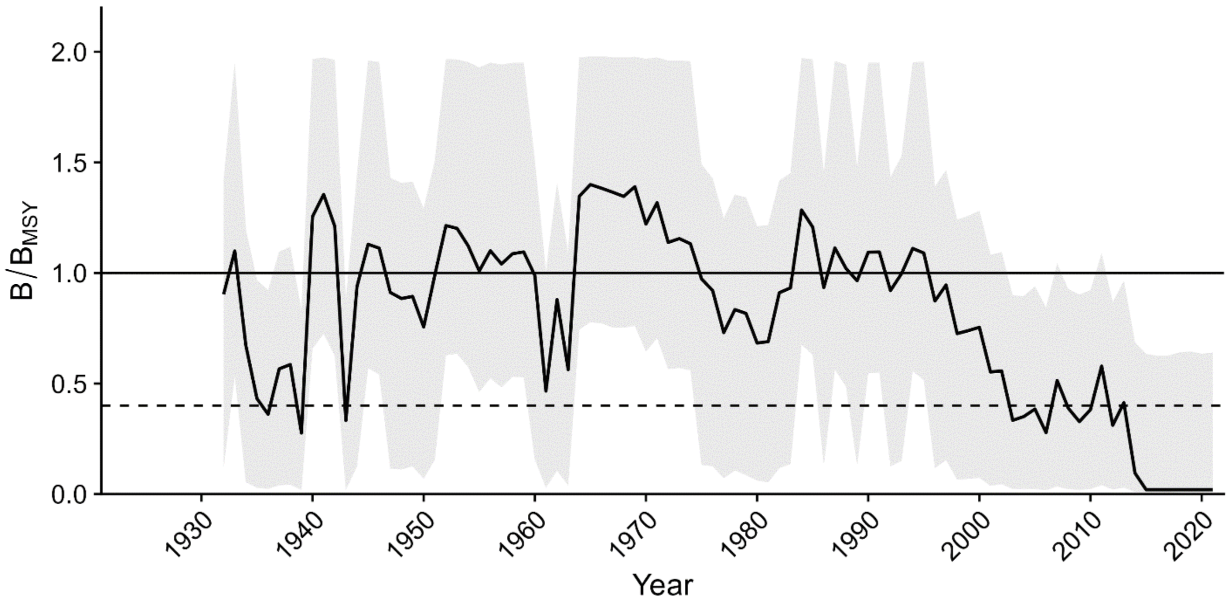

Figure 11. Estimated relative biomass levels for sea scallops in the core area of the sGSL using the boosted regression tree method of Zhou et al. (2017).

4.2. OPTIMIZED CATCH ONLY METHOD (OCOM)

The optimized catch-only method (OCOM) developed by Zhou et al. (2018), incorporates a stock reduction analysis (SRA) based on a Graham-Schaefer surplus production model (Graham 1935; Schaefer 1954b). Utilizing prior information on intrinsic population growth rate (r) and current stock depletion estimated from natural mortality and saturation data derived from the zBRT method, OCOM employs an optimization algorithm to identify suitable parameter combinations. This approach aims to estimate time series data for biomass, fishing mortality, stock status (specifically, B/B_{MSY} and F/F_{MSY}), and relevant biological and management quantities (such as r, K, MSY, B_{MSY}, and F_{MSY}) from catch data and an estimate of natural mortality (M). To account for uncertainty in the input parameter M, a range of values (0.08-0.39) were tested.

Implemented via the OCOM() function in the datalimited2 package in R (Free 2018), the OCOM method, at the median M (0.26), predicts the core area to have K and B_{MSY} estimates of 4235 t and 2117 t respectively. At this level of M, an LRP corresponding to $0.4B_{MSY}$ (847 t) exceeds the estimated biomass in 2021 (454 t, Figure 12) by nearly 83%, implying the stock is currently in the critical zone. According to the OCOM model, the core area should support a maximum harvest rate of 264 t per year at B_{MSY}, roughly twice the current removal rate.

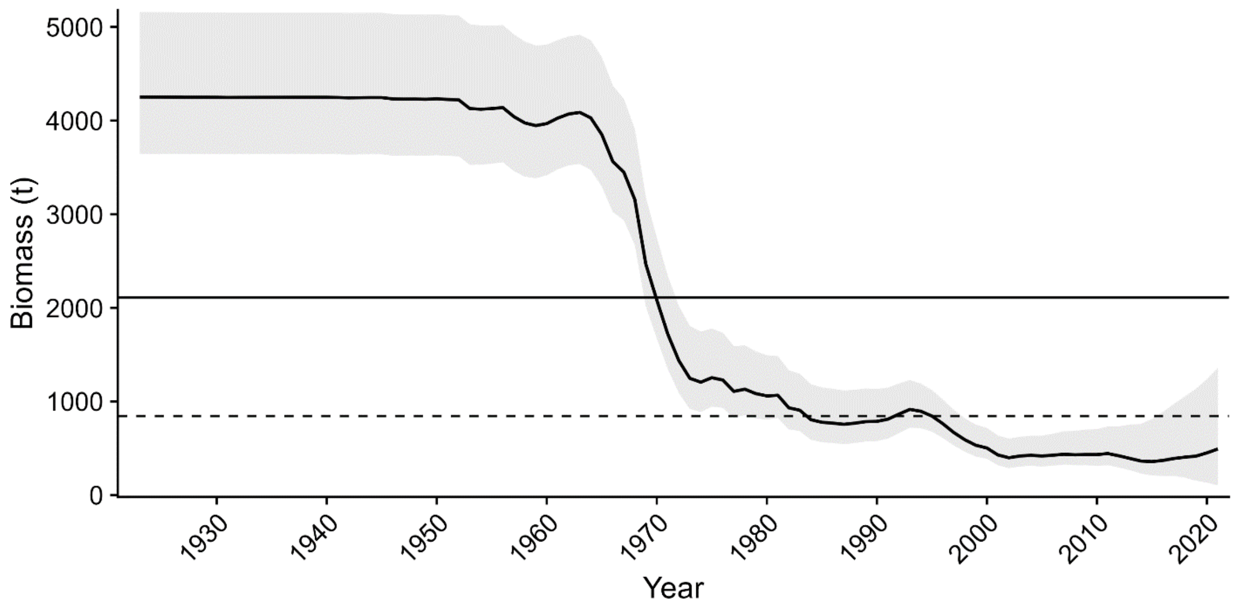

Figure 12. Biomass estimates from the OCOM method applied to sea scallops in the core area of the sGSL along with B_{MSY} and LRP estimates as solid and dashed lines respectively.

The predicted results from the OCOM method exhibited variability across the tested range of M values. Lower mortality rates led to larger estimates of K, B_{MSY}, and B_{2021}, while resulting in smaller estimates of MSY, r, and F_{MSY} (Figure A3.1). Decreasing M had a more pronounced effect on parameter estimates compared to increasing M. Most parameter estimates were within the expected range with the exception of the estimate of r (0.25), which was below the expected range of 0.3 – 0.8 (Figure A3.2).

4.3. DEPLETION-BASED STOCK REDUCTION ANALYSIS (DB-SRA)

The Depletion-Based Stock Reduction Analysis (DB-SRA) extends the DCAC method to estimate reference points and biomass for commercially fished stocks utilizing a delay difference production model (Dick and MacCall 2011). This method requires a comprehensive history of removals from the inception of the fishery, and for the sGSL sea scallop, the assumption that the catch time series represents the entire fishery is reasonable as the commercial fishery only began developing in the 1940s and 50s. Implemented using the dbsra() function in the R package fishmethods (Nelson 2023), the model employs a hybrid Pella-Tomlinson-Fletcher production model (PTF, Pella and Tomlinson 1969; Fletcher 1978). User-provided parameter inputs include the age at maturity (5 for sea scallops in the sGSL, DFO 2019), and distributions for K, B_0/K, S, F_{MSY}/M, B_{MSY}/K, and M. The model was run using a total of 10000 Monte Carlo simulations.

Results for the core area within the sGSL indicate K and B_{MSY} estimates of 5878 t and 876 t respectively (Figure 13). The estimated stock biomass in 2021 is 911 t, surpassing the model's proposed LRP of 350 t and even exceeding the default USR proposed by the PA of $0.8B_{MSY}$, or 426 t.

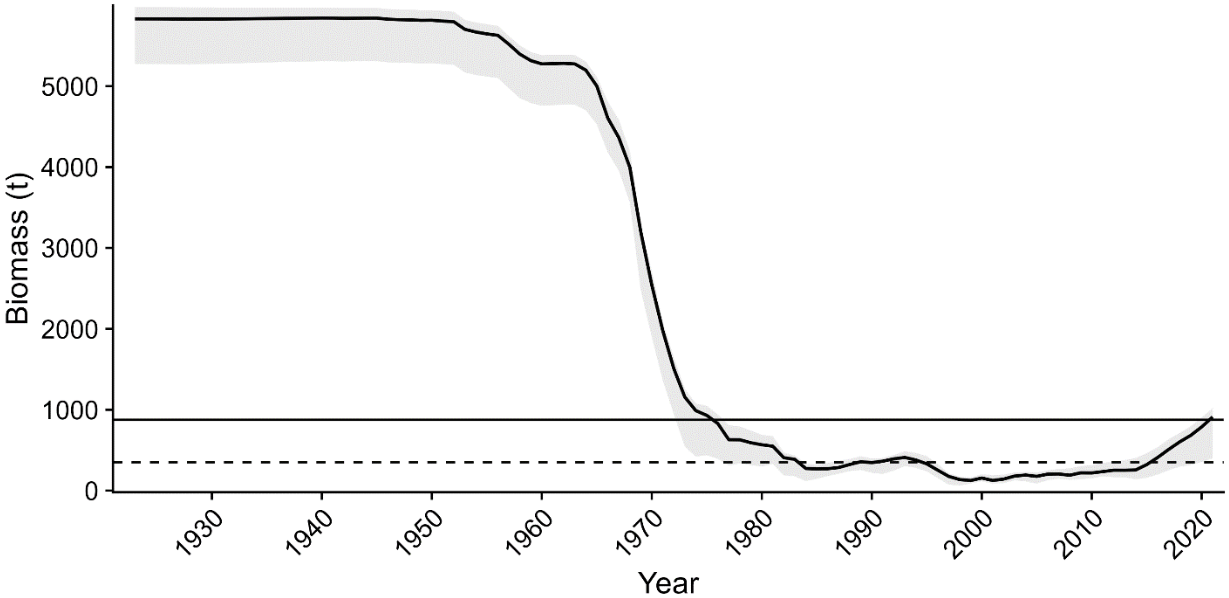

Figure 13. Biomass estimates for sea scallops in the core area of the sGSL, produced by a DB-SRA model. Solid and dashed lines correspond to B_{MSY} and the LRP.

Most parameter estimates align with the defined priors, with the notable exception of B_{MSY}/K (Figures A3.4 and A3.5). Among the 10000 simulations run, only 10 produced acceptable catch and biomass trajectories, leading to wide confidence intervals around some parameter estimates (e.g. B_{MSY}: 660 – 1070 t, or B_{2021}: 399 – 1017 t). Despite the uncertainty, the model fit to the catch data demonstrated small residuals evenly spaced around zero throughout the time series, indicative of a good fit (Figure A3.6).

The model's high relative and absolute biomass estimates result from the selected production curve characterized by an unusually low B_{MSY}/K of 0.16, which seems improbable for sea scallop. The DB-SRA model uses a hybrid Schaefer and Pella-Tomlinson-Fletcher production function in its delay-difference equation that allows for peak latent productivity (B_{MSY}/K) to be anywhere between 0 and 1, but which resembles a Schaefer model below a join-point and a PTF model above the join-point. Given that the result in B_{MSY}/K is unusually low, this type of production model may not be well suited to invertebrate bivalves like scallops.

4.4. CATCH ONLY MODEL WITH SAMPLING-IMPORTANCE-RESAMPLING (COM-SIR)

The COM-SIR model is a Bayesian method proposed by Vasconcellos and Cochrane (in Kruse et al. 2005), utilizing a coupled harvest-biomass dynamics model that combines a Schaefer biomass dynamics model and a logistic harvest dynamics model. Inputs parameters include r and K, along with two variables that shape the harvest rate: *a* (the bioeconomic equilibrium as a proportion of K), and *x* (a multiplier expressing the increase in the harvest rate over time). Employing Monte Carlo Markov Chain (MCMC) sampling methods with a sampling-importance-resampling algorithm, the model assigns additional weight to more probable parameter values while maintaining a diverse sampling distribution to account for uncertainty.

Executed in R using the comsir() function in the datalimited package (Anderson et al. 2016), the model utilized the informative prior ranges for r and K described above, and uninformative prior ranges for *a* (0 – 1) and *x* (0.0001 – 1). After a burn-in period of 5 x 10^7 MCMC simulations, 1 x 10^4 posterior draws were collected.

For the core population of sea scallops in the sGSL, the model predicts a carrying capacity of 2251 t with a biomass level at MSY of 1126 t. An LRP of $0.4B_{MSY}$ corresponds to 447 t, exceeding the current biomass levels of 335 t, placing the stock in the critical zone. The biomass timeseries estimates (Figure 14) show the scallop stock dropped below B_{MSY} in 1980 and below the LRP in 1990. While the model predicts an increasing trend in biomass in more recent years, it is insufficient to bring the stock above the LRP.

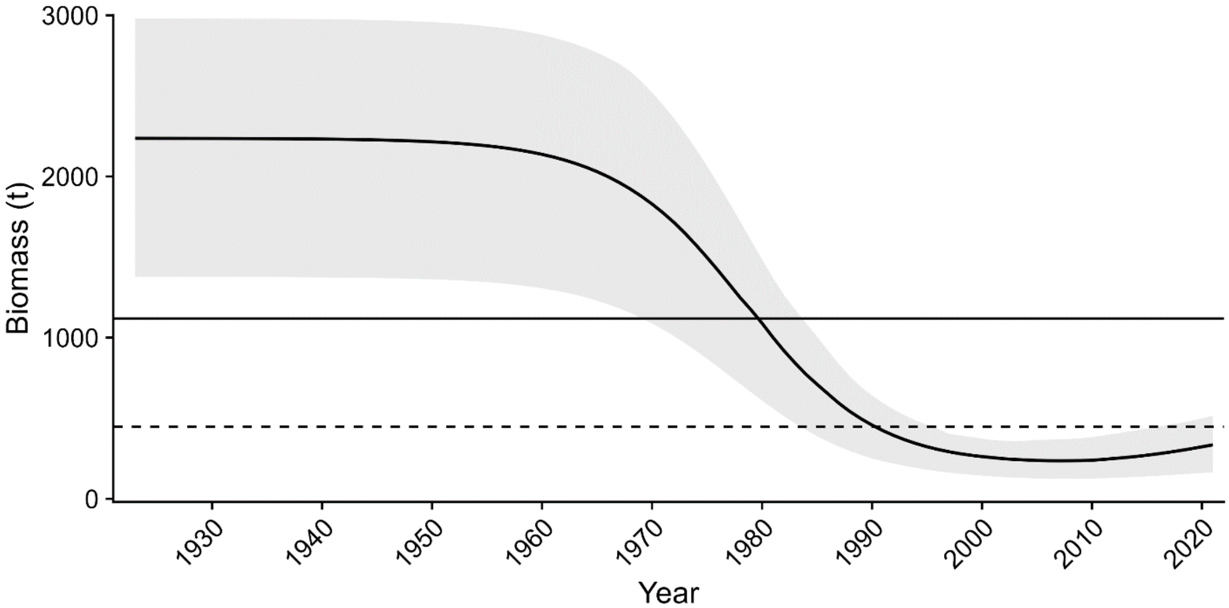

Figure 14. Time series of biomass estimates for the sea scallop stock in the core area within the sGSL produced by a COM-SIR model.

Model parameter estimates fall within the expected range for both r (0.41) and K (2236, Figure A3.7). However, in terms of model fit, catch residuals indicate that the model tends to underestimate catch and is slow to respond to steep increases in catch rates (Figure A3.8). This discrepancy may stem from the logistic harvest model's assumption that effort, determined in part by a constant bioeconomic equilibrium value, remains steady over time. Given the scallop fishery's complementary nature and its sensitivity to the price of alternative fisheries, this assumption may not be entirely reasonable.

4.5. CATCH-MAXIMUM SUSTAINABLE YIELD (CMSY++)

The CMSY++ model is a Monte Carlo method developed by Martell and Froese (2013) and further refined in Froese et al. (2017). CMSY++ estimates biomass and fisheries reference points for data-limited stocks using catch data, resilience information, and qualitative stock status assessments. This method employs Monte Carlo simulations to assess the plausible range of intrinsic growth rate (r) and carrying capacity (K) values. Population trajectories under fishing pressure are modeled, and the most likely r/K combination is identified by comparing these trajectories with actual catch trajectories. CMSY++ requires estimates of r (which can be derived from reported resilience levels, e.g., from SeaLifeBase) and saturation levels for the start, end, and an intermediate year within the catch time series (1975).

Implemented in R using the author's R scripts (Froese et al. 2017) and the parameter ranges described earlier, the CMSY++ model for scallops in the core area estimates a carrying capacity of 4037 t and a B_{MSY} of 2018. However, the predicted biomass in 2021 is 355 t, below the model proposed LRP of 807 t, placing the stock in the critical zone (Figure 15).

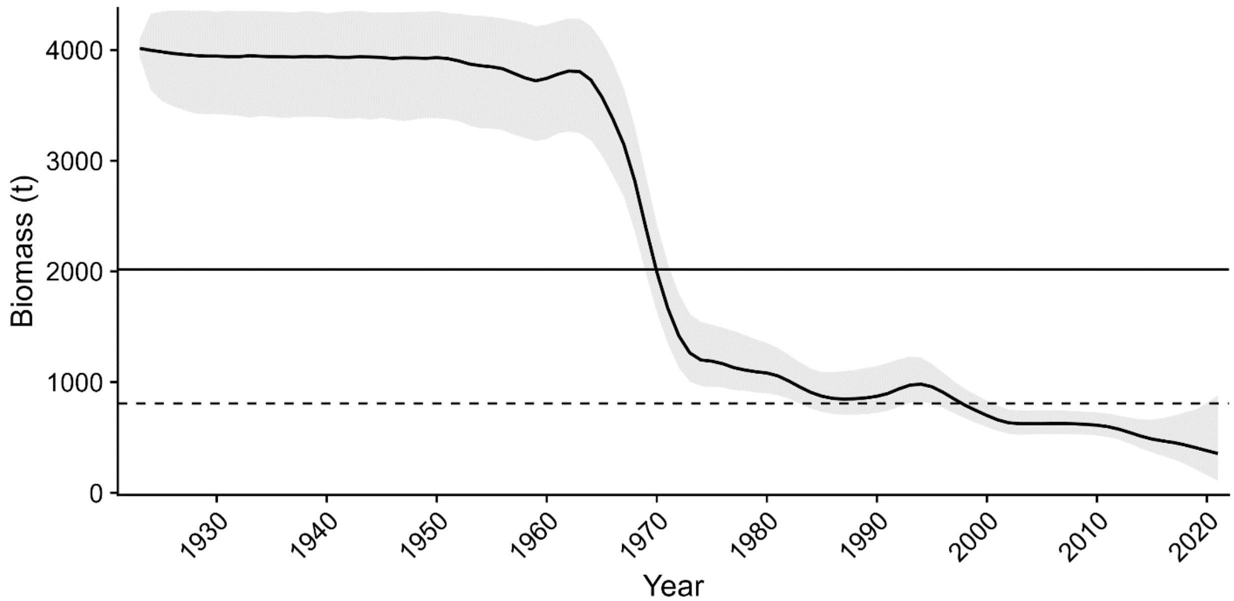

Figure 15. Biomass estimates from the CMSY++ model for sea scallops in the core area within the sGSL. The solid and dashed lines represent B_{MSY} and the LRP respectively.

Model parameter estimates for r and K fall within the ranges defined by the priors and are reasonable for this stock (Figure A3.9). Posterior distributions suggest that the model is data-driven rather than influenced by the priors. However, catch estimates by the model consistently exceed the actual landings records from 1984 onwards (Figure A3.10), indicating potential difficulties by the model in accurately tracking recent catch levels.

4.6. BAYESIAN STATE SPACE MODEL (BSM)

The Bayesian State Space Model (BSM) is implemented alongside the CMSY++ algorithm when CPUE data are available. The BSM incorporates the Schaefer model into a Bayesian framework, considering both process error in population dynamics and observation error in measurement/sampling. Implemented in R using the same input data as CMSY++, the BSM also includes fishery-dependent CPUE data, specifically the total landings per active fishing boat for scallops in this case.

The results from the BSM were similar to those from the CMSY method and produced estimates of K and B_{MSY} of 4135 t and 2067 t respectively. Biomass in 2021 was estimated to be 314 t, or less than half of the LRP of 827 t, placing this stock in the critical zone.

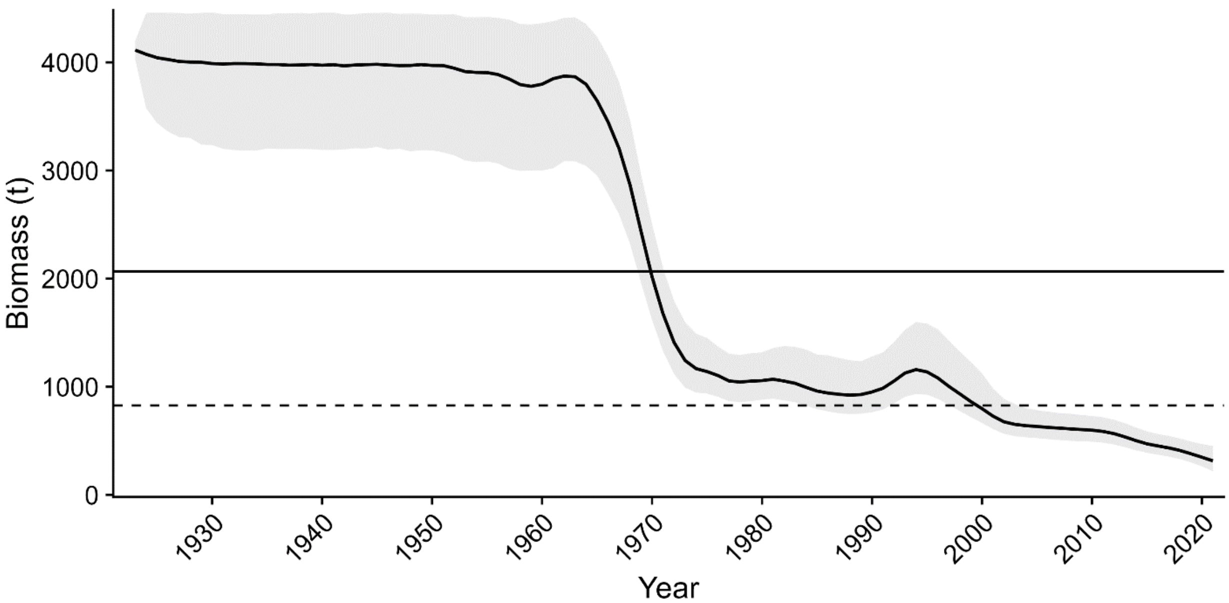

Figure 16. Biomass levels for sea scallops in the core area of the sGSL as estimated by a Bayesian State Space Model implemented in CMSY++. The solid and dashed lines represent the estimate of B_{MSY} and an LRP of $0.4B_{MSY}$ respectively.

Similar to the CMSY++ method, the BSM exhibits a tendency to overestimate catches in recent years (Figure A3.12), consistently predicting catches ~100 t higher than the actual records in recent years.

4.7. STOCHASTIC SURPLUS PRODUCTION MODEL IN CONTINUOUS TIME (SPICT)

The SPiCT model, developed by Pedersen and Berg (2017), estimates stock size and catch potential by incorporating both deterministic and stochastic components. Relying on key parameters like the intrinsic growth rate, the carrying capacity, and the stochasticity in the population dynamics, SPiCT utilizes historical catch data as well as an index of abundance (like commercial CPUE). The model addresses natural system uncertainties through a random walk process, capturing stochastic fluctuations in populations over time and projecting future trends.

Implemented in R, SPiCT requires a historical catch time series and an index of abundance or effort. In the model described below, effort was provided along with priors for initial saturation levels (logbkfrac), r, K, the Pella-Tomlinson parameter (logn), and biomass in a given year. As the model does not accept a prior for saturation in the final year, a prior biomass level was provided with a normal distribution, a mean of 150 t, and a standard deviation of 0.2, to be consistent with the saturation estimates described above.

When executed, the SPiCT model produced K and B_{MSY} estimates of 3626 t and 1123 t respectively (Figure 17). As this model uses a Pella-Tomlinson production model, B_{MSY} does not equal K/2; instead, the model estimated a shape parameter of 1.45 which corresponds to a B_{MSY}/K of 0.31, which is consistent with the prior provided (1.5, see appendix 3 Figure 46). Unlike the previous models, SPiCT predicts a significant drop in abundance in the late 1950s, coinciding with the offshore fleet from the Maritimes briefly fishing the Northumberland straight. The population recovers before declining again in the late 1960s and early 70s. SPiCT ultimately predicts a 2021 biomass level of 234 t , below the corresponding LRP of 449 t, placing this stock in the critical zone.

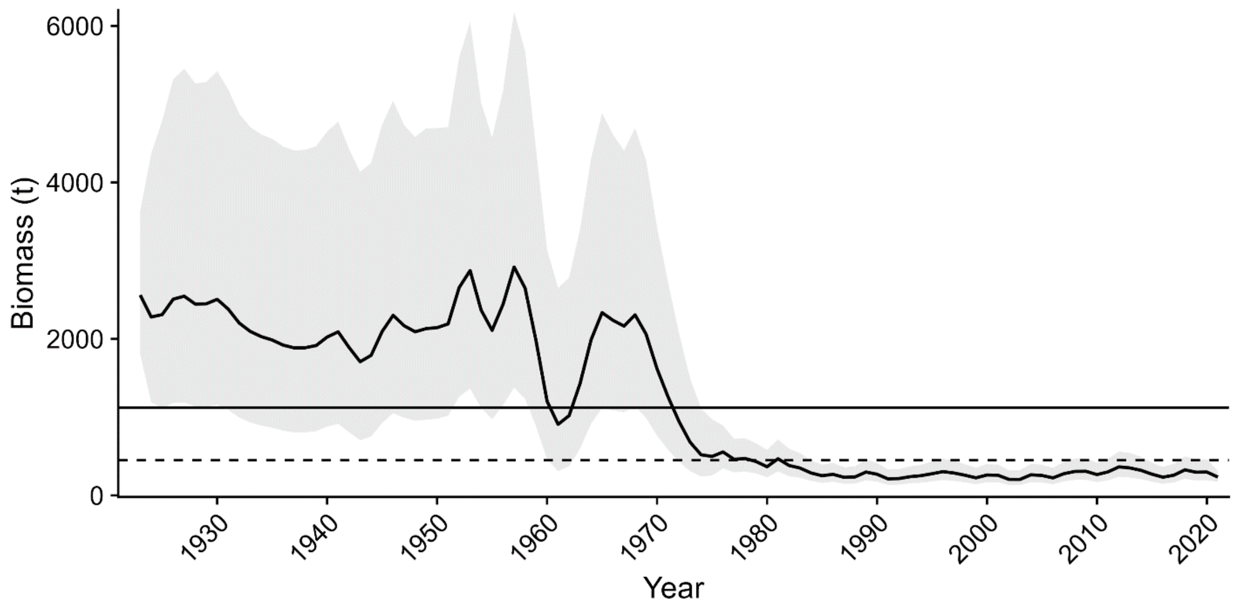

Figure 17. Biomass estimates from a SPiCT model for sea scallops in the core area within the sGSL along with the model estimates for B_{MSY} and the LRP as solid and dashed lines respectively.

Parameter estimates align with expected ranges, indicated by prior and posterior distributions (Figure A3.13). True catch levels fall within model confidence intervals, reflecting a good fit with the data (Figure A3.14).

4.8. JUST ANOTHER BAYESIAN BIOMASS ASSESSMENT (JABBA)

JABBA is a Bayesian state-space estimation framework, building upon previous optimization procedures in Bayesian modeling approaches (Winker et al. 2018, 2020). Formulated based on surplus production models, JABBA incorporates multiple surplus production functions, such as Schaefer, Fox, and Pella-Tomlinson. What distinguishes JABBA is its ability to accommodate multiple CPUE series and consider both process and observation errors, enabling parameter estimation and uncertainty assessment for each time series.

Implemented in R using the JABBA package (Winker et al. 2018), the model requires, at a minimum, a time series of catch history and an index of abundance (commercial CPUE). Priors for r and K, along with estimates of S in 1923 and 2021, were provided. A Pella-Tomlinson model with an intermediate shape parameter (B_{MSY}/K = 0.35) with a CV of 0.1 was used. The model also estimated the observation error, with a minimum allowable estimate of 0.01 for the abundance index. Procedural deviations were only estimated for the time period with an available abundance index (1976 – 2021).

The JABBA model predicts K and B_{MSY} levels of 3658 t and 1377 t, respectively, for core area scallops. The current predicted biomass level 249 t (Figure 18), below the 0.4B_{MSY} LRP of 551 t and placing the stock in the critical zone. The model anticipates a maximum sustainable yield of 298 t for this stock which occurs at B_{MSY} /K = 0.38. The current saturation level is predicted to be 0.07.

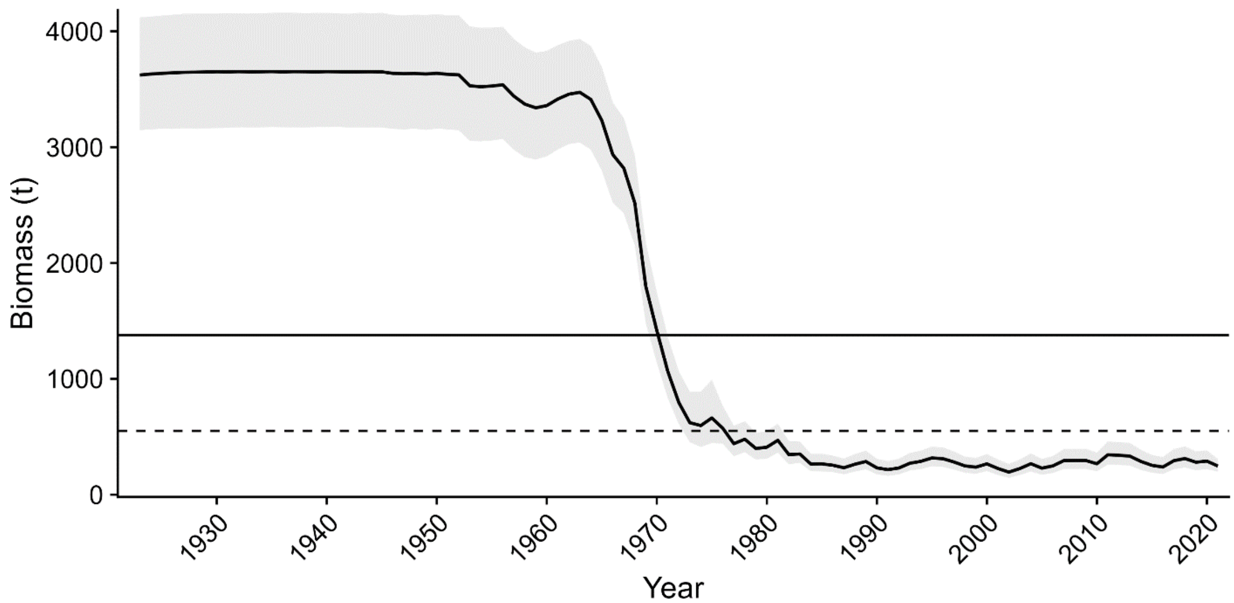

Figure 18. Estimated biomass levels over the time series of catch and effort records for the core area of the sGSL scallop stock as estimated by the JABBA model. The solid and dashed lines represent model predictions for B_{MSY} and the LRP, respectively.

Model estimates for catch closely match the actual catch records (Figure A3.15). The range of residuals is small and centered around zero. Model estimates align with previous models, except for r (0.23), which was lower than the prior range (Figure A3.16).

4.9. A STOCK-PRODUCTION MODEL INCORPORATING COVARIATES (ASPIC)

ASPIC, a surplus production model developed by Michael Prager (1992, Prager 1994), is the final model under consideration. Similar to JABBA, ASPIC can fit a range of model types (Schafer, Fox, Pella-Tomlinson) to catch and effort data. Inputs requirements include priors for K, MSY, q (catchability), and initial S. Unlike the other catch and effort models, ASPIC does not use input for the current saturation levels. Output variables consist of biomass and fishing mortality estimates through time, along with reference points (B_{MSY} and F_{MSY}). The ASPIC model is executed through a standalone program available from the NOAA Fisheries Integrated Toolbox.

When applied to the core area data for the sGSL, a Pella-Tomlinson model estimated the shape parameter to be initially very high ($B_{MSY}/K > 0.7$). Due to convergence issues with this model, the q parameter was fixed at 0.002 in the final run of the model, chosen iteratively and guided by examination of the log residuals from the abundance index. For robust parameter estimation, 1000 bootstrapped runs were conducted to calculate 95% confidence intervals. A bounds multiple of 4 was also applied, limiting the search algorithm to a range for MSY of 0.4 and 4 times the MSY estimate produced by a standard Schaefer model fit.

The ASPIC model generated K and B_{MSY} estimates of 3390 t and 1390 t for the core area scallops. The biomass estimate in the final year was 398 t (Figure 19), falling below the 0.4 B_{MSY} threshold predicted by the model (556 t), and placing the stock in the Critical zone. The predicted biomass trajectory remains near K for most of the time series before dropping below B_{MSY} in the 1960s, and the model proposed LRP in the 1980s. A slight increase in predicted biomass levels in recent years is insufficient to move the stock out of the Critical zone.

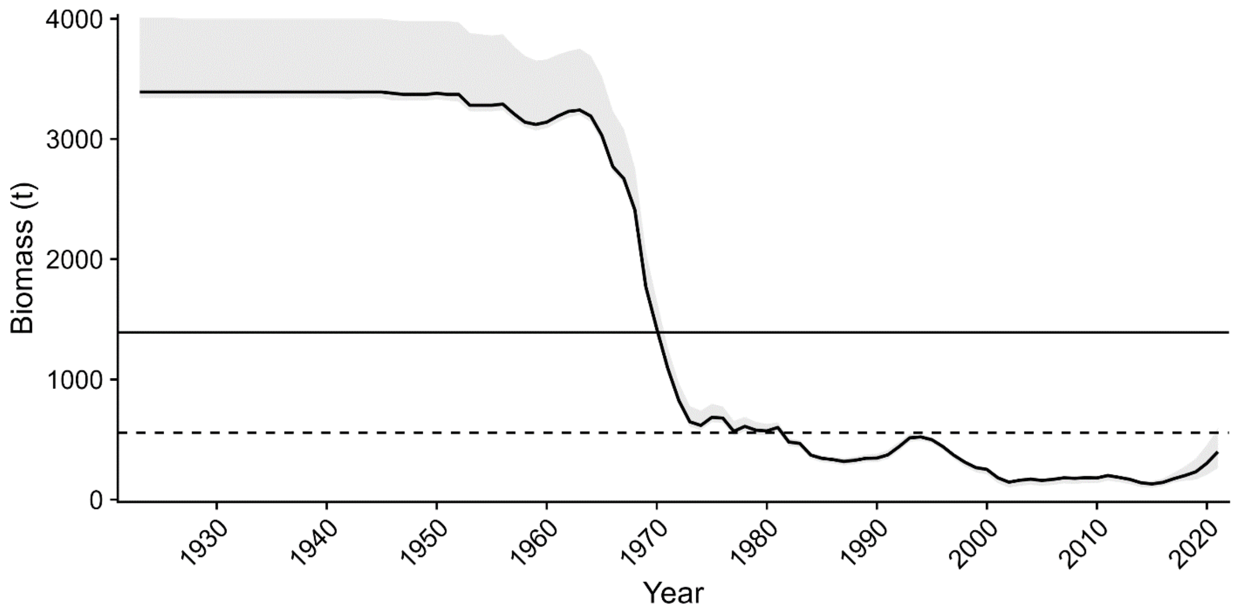

Figure 19. Estimated biomass levels for sea scallops in the core area of the sGSL as estimated by the ASPIC model. Solid and dashed lines correspond to the model estimates for B_{MSY} and the LRP respectively.

The model fit produced relatively small residuals against actual landings data for much of the time series (Figure A3.17). However, the model encountered difficulty with the largest catch data from the 1960s and 70s, overestimating catches at times by up to 150 t, resulting in a skewed residual distribution towards negative values. Parameter estimates for both K and B_{MSY}/K fell within the expected prior range, although the model does not provide an estimate of r when using the Pella-Tomlinson method.

5. POPULATION MODEL SELECTION

This section critically evaluates various data-limited population models for sea scallops in the core area of the sGSL with the goal of establishing an LRP following DFO's Precautionary Approach (DFO 2009), specifically the 0.4B_{MSY} rule. The primary criteria for model selection includes consistency in biomass estimates over time, avoiding overly broad confidence intervals, and considering user knowledge required for implementation.

In general, while the predicted biomass trajectories of these models differ (Figure 20), examining the relative biomass trajectories (B/B_{MSY} and S) reveals commonalities. Most models indicate a stock near its carrying capacity until the late 1960s and early 1970s when a significant drop occurs, often to below 0.4$B_{MSY,}$ following which, the stock biomass levels remain relatively constant. This consistency across models suggests the observed trend is a product of the data rather than variations in model inputs.

The primary differences among the models comes down to whether they use a Schaefer or a Pella-Tomlinson production model. The former predicts relative biomass levels that do not exceed 2B_{MSY}, or B_{MSY}/K = 0.5, while the latter group of models had a tendency to produce B_{MSY}/K is between 0.38 and 0.44 (Table 1), resulting in B_{1923}/B_{MSY} levels closer to 3 and B_{MSY} levels which tend to be smaller. This is not unrealistic for scallops which are a highly resilient species with a generally high intrinsic rate of population growth at lower population densities.

Three models deviate considerably from the overall trend however: the zBRT, DB-SRA, and SPiCT models.

The zBRT model, which predicts relative biomass (S and B/B_{MSY}) directly from the catch timeseries, without taking into consideration biomass in the previous timestep produces trajectories which can be improbable or even impossible. For example, a stock which recovers from a complete collapse despite ongoing fishing, or, as in this case, a stock which supports stable landings for 20 years despite being in a collapsed state (S = 0.02) for the last 6 years. This feature, combined with the zBRT model's inability to produce B_{MSY} estimates make this model unsuited to the task of setting and LRP for scallops.

Biomass and relative biomass trajectories predicted by the DB-SRA model differ considerably from those of the other models due to the model's optimization algorithm settling on a mean B_{MSY}/K ratio of 0.16. This relatively low value produced large B/B_{MSY} estimates as high as 6.6 as well as the largest estimates of K among the models (Figure 20). Despite this difference, the model producing similar saturation estimates to other models (Figures 20 and 21). Another area in which the DB-SRA model stand out is with the improbably steep rise in biomass predicted by the DB-SRA model from 2010 onwards. This rise in abundance is not predicted by any of the other models and is not justified by the data as catch and commercial CPUE have remained relatively stable over the last 20 years. This unlikely trend in more recent years may be due to the very steep production curve predicted by the model, an unlikely result which precludes this model from further consideration.

The SPiCT model produces time series of biomass estimates that are unlike the other remaining models and that is unlikely given the history of the fishery in the area. According to the SPiCT model, the core area scallop stock experienced a considerable depletion of approximately 35% in the 1940s, before the commercial fishery had begun to develop in the area. The abundance levels then rise to approximately 2900, or roughly 16% over the predicted carrying capacity. Despite this unlikely biomass trajectory early in the time series, this model produces parameter estimates that fall within the expected range (Figure 21), including a final biomass estimate which closely aligns to those of the fishery dependent and independent sources. This suggests that while this model may be able to model the population accurately when effort data is provided, it fails to provide realistic estimates for periods prior to the availability of effort indices. As a result, this model is not considered further.

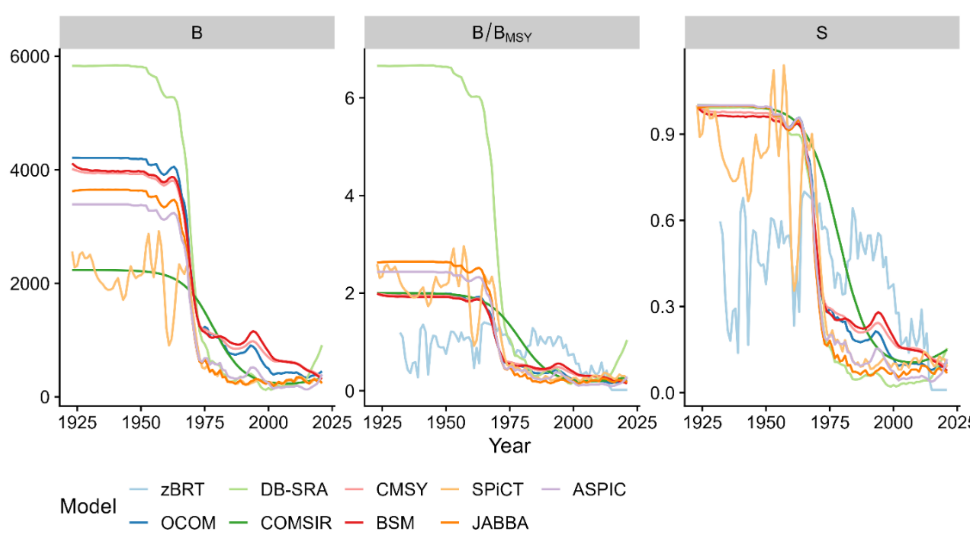

Figure 20. Biomass (B), relative biomass (B/B_{MSY}), and saturation (S) estimates from multiple models applied to catch and effort data for sea scallops in the core area of the sGSL.

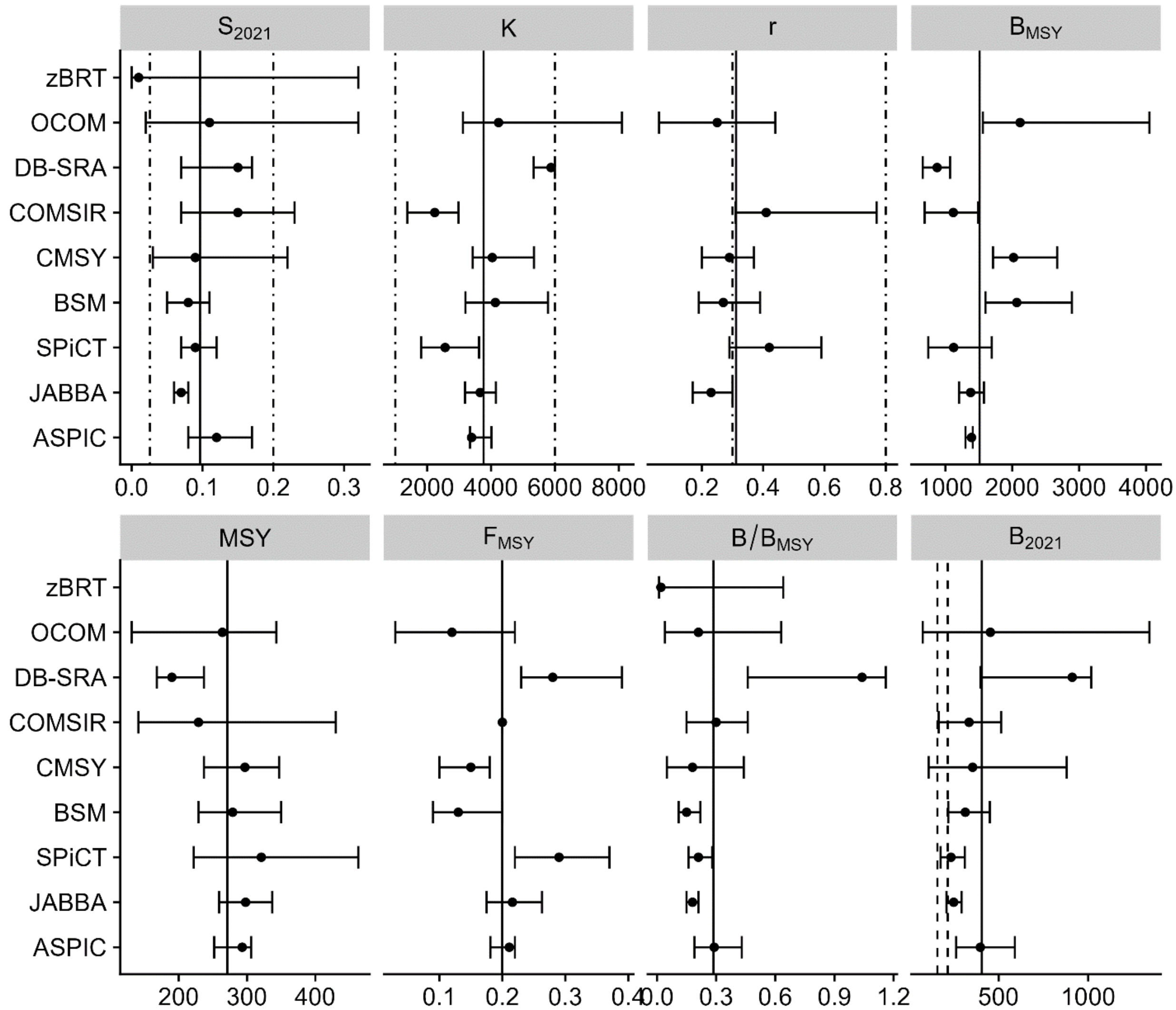

Figure 21. Parameter estimates with their 0.025 and 0.975 percentiles from population models applied to sea scallops in the core area of the sGSL. Dot-dashed lines represent the extent of prior ranges for S, K, and r, while the two dashed lines (B_{2021}) represent the biomass estimates in 2021 based on a fishery-dependent depletion model and a fishery-independent research survey. Solid lines represent means parameter estimates across models.

Table 1. Summary of population parameters estimated from models applied to the catch and effort data from the sGSL core area sea scallop stock.

Method	K	r	B_{MSY}	MSY	B_{2021}	F_{MSY}	B_{2021}/B_{MSY}	B_{MSY}/K	B_1/K	S_{2021}	$0.4B_{MSY}$	Status
zBRT	-	-	-	-	-	-	0.02	0.5	-	0.01	-	Critical
OCOM	4235	0.25	2117	264	454	0.12	0.21	0.5	0.99	0.11	847	Critical
DB_SRA	5878	-	876	190	911	0.28	1.04	0.16	0.99	0.15	350	Healthy
COM-SIR	2236	0.41	1118	229	335	0.2	0.3	0.5	1	0.15	447	Critical
CMSY	4037	0.29	2018	297	355	0.15	0.18	0.5	0.99	0.09	807	Critical
BSM	4135	0.27	2067	279	314	0.13	0.15	0.5	0.99	0.08	827	Critical
Spict	2561	0.42	1123	321	234	0.29	0.21	0.44	1	0.09	449	Critical
Jabba	3658	0.23	1377	298	249	0.216	0.18	0.38	0.99	0.07	551	Critical
ASPIC	3390	-	1390	293	398	0.211	0.29	0.41	1	0.12	556	Critical

The remaining group of models (OCOM, ASPIC, COM-SIR, CMSY++, BSM, and JABBA), includes both catch-only model and catch and effort models, as well as models employing Schaefer and Pella-Tomlinson (P-T) production equations. Despite these differences, though, there is substantial consistency in terms of relative biomass and saturation estimates. Generally, the models predict a rapid decrease in stock abundance in the late 1960s and early 1970s. An exception to this trend is with the COM-SIR model which exhibits a much more gradual 30-year decrease (Figure 20). Given the intensity of fishing that would have occurred to produce the record high catches of ~900 t, this gradual decrease in abundance seems unlikely and precludes the COM-SIR from further consideration as a suitable model from which to select an LRP.

One area where the Schaefer and Pella-Tomlinson models differ is in their predictions of K, B_{MSY}, and B_{2021}. Models employing the Schaefer production curve predict values of K that are, on average, over 600 t greater than those estimated by with B_{MSY}/K ratios less than 0.5. The same can be said for the estimates of B_{MSY}. While this difference isn't necessarily surprising and cannot be used to single out which models to reject, it does have an impact on the resulting LRP as larger B_{MSY} estimates translate to a higher LRP. While there is a tendency for models using a logistic function to produce higher B_{2021} estimates, the disparity between B_{2021} and B_{MSY} (or $0.4B_{MSY}$) estimates from models fit using a Schaefer equation, 367 t on average, tends to be larger than those using the Pella-Tomlinson equation, 225 t on average.

When model fits are compared among the remaining models, both the CMSY++ and BSM models show greater, and more consistent departures from the true catch levels relative to the JABBA, ASPIC, or OCOM models. With both the CMSY++ and BSM models, trends in the residuals suggest these models have difficulties tracking the actual catch histories, particularly following a period of intense fishing (like in the 1960s for sea scallops). This is possibly due, in part, to both models employing a smoother to the catch time series in order to reduce the effect of outlier values. This step supposedly allows the models to better capture the underlying trends, but as a result, they are incapable of accurately predicting the highest or lowest catches and it may in part be responsible for the model over estimating catches following this period of high catch rates. This is not the case with the OCOM, JABBA, or ASPIC models, which each produced catch predictions that closely matched the actual catch values. Among these three models, JABBA produced the most evenly distributed residuals around zero. Both the ASPIC and OCOM models had a tendency to overestimate catch levels, resulting in negatively skewed residuals, though to a much lesser extent than CMSY++ and BSM.

In terms of parameter estimates, the OCOM, ASPIC, and JABBA models all produced parameter estimates of S_{2021} and K that fell within the expected ranges, while both OCOM and JABBA settled on r estimates that were below the expected range of 0.3 to 0.8. In both cases, however, the 95% parameter distributions did overlap this range. When run using a Pella-Tomlinson model, ASPIC does not return an r estimate. Another way in which these models differed from is in their estimation of the biological reference points. Both JABBA and ASPIC predicted a lower B_{MSY} than OCOM, and lower biomass levels in the final year of the time series. With respect to this latter value, JABBA (249 t) was closest to the biomass levels predicted by the fisheries-independent (survey, 216 t) and fisheries-dependent (depletion model, 158 t) estimates. (dashed lines in B_{2021}, Figure 21). The OCOM estimate (454 t) was more than twice the fall survey estimate and exceeded even the larger spring survey estimate (356 t). Similarly, ASPIC's estimate did not include the two alternative biomass estimates for 2021 within it's 95% parameter distribution. Considering the similarity between the JABBA biomass estimates for 2021 and those from alternative sources, and taking into account JABBA's flexibility and ability to incorporate multiple standardized abundance indices, we recommend using the JABBA model for modeling the core sea scallop stock within the sGSL.

6. CONCLUSIONS AND ADVICE

The application of data-limited population models to assess the sea scallop stock in the core area of the sGSL has provided valuable insights. Despite variations in estimates, particularly for key parameters like carrying capacity (K), these models consistently highlight significant trends, especially in biomass.

Both simple interpretations of catch data (the scalar approach, depletion-corrected average catch, and catch ratio) and more advanced analyses of the catch only (OCOM, DB-SRA, COM-SIR, CMSY++) and analysis incorporating catch and effort time series (BSM, SPiCT, JABBA, ASPIC) converge to indicate an approximate Maximum Sustainable Yield (MSY) of around 270 t. Moreover, the majority of models suggest that current biomass levels likely fall below the recommended limit reference point of $0.4B_{MSY}$.

Given the limited availability of fishery-independent indices of scallop abundance in the sGSL and the absence of catch-at-age/size data, we opted for data-limited models to define biological reverence points for this stock and to arrive at an LRP. To address data and model limitations, we assessed and compared multiple models for robustness. Among the nine models investigated, the JABBA model stands out, providing the most reliable estimates based on catch and effort data, offering a robust option for future stock assessments. According to this model, the scallop stock in the core sGSL area is estimated to be at an abundance of 0.18 relative to the biomass at MSY, placing this stock below the model's suggested LRP of 551 t.

It's crucial to note that as more years of data are incorporated, the JABBA model's estimates may be refined. Additionally, future implementation could benefit from incorporating survey estimates as a secondary index of abundance, thereby enhancing overall model confidence.

7. ACKNOWLEDGMENT

We would like to thank J. McDermid for the opportunity to work on this project, the discussion and the comments on the paper as well as R. Sonier and F. Akaishi for the comments on the paper.

8. REFERENCES CITED

Alverson, D.L., and Pereyra, W.T. 1969. Demersal fish explorations in the northeastern Pacific Ocean — an evaluation of exploratory fishing methods and analytical approaches to stock size and yield forecasts. J. Fish. Res. Board Canada 26(8): 1985–2001.

Anderson, S.C., Afflerback, J., Cooper, A.B., Dickey-Collas, M., Jensen, O.P., Kleisner, K.M., Longo, C., Chato Osio, G., Ovando, D., Minte- Vera, C., Minto, C., Mosqueira, I., Rosenberg, A.A., Selig, E.R., Thorson, J.T., and Walsh, J.C. 2016. datalimited: Stock Assessment Methods for Data-limited Fisheries. R package version 0.1.0.

Anderson, S.C., Branch, T.A., Ricard, D., and Lotze, H.K. 2012. Assessing global marine fishery status with a revised dynamic catch-based method and stock-assessment reference points. ICES J. Mar. Sci. 69(8): 1491–1500. SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States.

Beninger, P.G. 1987. A qualitative and quantitative study of the reproductive cycle of the giant scallop, *Placopecten magellanicus* , in the Bay of Fundy (New Brunswick, Canada). Can. J. Zool. 65(3): 495–498.

Bonardelli, J.C., and Himmelman, J.H. 1995. Examination of assumptions critical to body component indices: Application to the giant scallop *Placopecten magellanicus*. Can. J. Fish. Aquat. Sci. 52: 2457–2469.

Boudreau, M., and Duplisea, D. 2022. A decision tool for the selection of methods to obtain indicators and reference points for data-limited stocks. Can. Manuscr. Rep. Fish. Aquat. Sci. 3237: vi + 61 p.

Bourne, N. 1964. Scallops and the offshore fishery of the Maritimes. *In* DFO Bullitin.

Caddy, J.F. 1973. Underwater observations on tracks of dredges and trawls and some effects of dredging on a scallop ground. J. Fish. Res. Board Canada 30(2): 173–180.

Cardin, P.J.A. 1924. Fifty-Seventh Annual Report of the Fisheries Branch for the year 1923-24. Dep. Mar. Fish. 57 (Sessional Paper No.29 A. 1925): 103 p.

Chiasson, L.P. 1951. Scallop investigations and explorations in the southern Gulf of St. Lawrence - 1950. Fish. Res. Board Canada 423: 1–29.

Chiasson, L.P. 1952. Scallop investigations in the Gulf of St. Lawrence and off eastern Cape Breton Island 1951. Fish. Res. Board Canada 449: 1–58.

Cope, J.M., and Hamel, O.S. 2022. Upgrading from M version 0.2: An application-based method for practical estimation, evaluation and uncertainty characterization of natural mortality. Fish. Res. 256: 106493.

Davidson, L.-A., Niles, M., and Légère, L. 2007. Proceedings of the southern Gulf scallop fishery workshop: Moncton, New Brunswick, March 30-31, 2006. Can. Tech. Rep. Fish. Aquat. Sci. 2785: vii + 87 p.

DFO. 1996. Southern Gulf of St. Lawrence sea scallop. Can. Atl. Fish. Sci. Advis. Committee, Stock Status Rep. 96/104E: 1–3.

DFO. 1998. Integrated Fisheries Managment Plan: Sea scallop in the southern Gulf, Scallop Fishing Areas 21, 22, 23, 24 1998. Fish. Manag. Plan G98/2.

DFO. 2006. A harvest strategy compliant with the precautionary approach. Can. Sci. Advis. Secr. Sci. Advis. Rep. 2006/023: 1–7.

DFO. 2009. A fishery decision-making framework incorporating the Precautionary Approach.

DFO. 2019. Fishery and Stock Status of the Sea Scallop (*Placopecten magellanicus*) from the Southern Gulf of St. Lawrence to 2016. Can. Sci. Advis. Secr. Sci. Advis. Rep. 2019/006: 1–24.

DFO. 2021. Science advice for precautionary approach harvest strategies under the fish stocks provisions. Can. Atl. Fish. Sci. Advis. Committee, Sci. Advis. Rep. 2021/004: 1–41.

DFO. 2022. Guidelines for writing rebuilding plans per the Fish Stocks Provisions and A Fishery Decision-making Framework Incorporating the Precautionary Approach. Fish. Ocean. Canada (June): 42 p.

Dick, E.J., and MacCall, A.D. 2011. Depletion-Based Stock Reduction Analysis: A catch-based method for determining sustainable yields for data-poor fish stocks. Fish. Res. 110(2): 331–341.

Dickie, L.M. 1951. Boughton Island, P.E.I., Scallop Investigations, 1946. Fish. Res. Board Canada 415: 1–29.

Dickie, L.M., and MacInnes, C.D. 1958. Gulf of St. Lawrence Scallop Explorations - 1957. Fish. Res. Board Canada 650: 1–60.

Dickie, L.M., and Medcof, J.C. 1963. Causes of mass nortalities of scallops (*Placopecten magellanicus*) in the southwestern Gulf of St. Lawrence. J. Fish. Res. Board Canada 20(2): 451–482.

Elith, J., Leathwick, J.R., and Hastie, T. 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77(4): 802–813.

Fletcher, R.I. 1978. On the restructuring of the Pella-Tomlinson system. Fish Bull. 76(3): 515–521.

Fox, W.W. 1970. An Exponential Surplus-Yield Model for Optimizing Exploited Fish Populations. Trans. Am. Fish. Soc. 99(1): 80–88.

Free, C.M. 2018. datalimited2: More stock assessment methods for data-limited fisheries.

Froese, R., Demirel, N., Coro, G., Kleisner, K.M., and Winker, H. 2017. Estimating fisheries reference points from catch and resilience. Fish Fish. 18(3): 506–526.

Froese, R., and Kesner-Reyes, K. 2002. Impact of fishing on the abundance of marine species. ICES Counc. Meet. Rep. C.: 1–12.

Graham, M. 1935. Modern theory of exploiting a fishery, and application to North Sea trawling. ICES J. Mar. Sci. 10(3): 264–274.

Gulland, J.A. 1971. The fish resources of the ocean. Preface. FAO Tech. Pap. 97: ix–xi.

Hart, D.R. 2006. When do marine reserves increase fishery yield? Can. J. Fish. Aquat. Sci. 63(7): 1445–1449.

Hart, D.R. 2013. Quantifying the tradeoff between precaution and yield in fishery reference points. ICES J. Mar. Sci. 70(3): 591–603. Human Kinetics.

Hart, D.R., and Chang, J.-H.H. 2022. Estimating natural mortality for Atlantic Sea scallops (Placopecten magellenicus) using a size-based stock assessment model. Fish. Res. 254(June): 106423. Elsevier B.V.

ICES. 2012. ICES Implementation of advice for data-limited stocks in 2012 in its 2012 Advice. ICES DLS Guid. Rep. CM 2012(ACOM 68): 42.

ICES. 2021. Tenth workshop on the development of quantitative assessment methodologies based on life-history traits, exploitation characteristics, and other relevant parameters for data-limited stocks (WKLIFE X). ICES Sci. Reports 2(98): 22.

Jamieson, G.S. 1978. Status and assessment of Northumberland Strait scallop stocks. Can. Atl. Fish. Sci. Advis. Committee, Res. Doc. 78(42): 45.

Kruse, G.H., Gallucci, V.F., Hay, D.E., Perry, R.I., Peterman, R.M., Shirley, T.C., Spencer, P.D., Wilson, B., and Woodby, D. 2005. Fisheries Assessment and Management in Data-Limited Situations. Alaska Sea Grant Coll. Progr. Univ. Alaska Fairbanks.: 958 pp.

Lanteigne, M., and Davidson, L.-A. 1991. Catch and effort statistics for the giant scallop (*Placopecten magellanicus*) fishery in the southern Gulf of St. Lawrence - historical review from 1923 to 1989. Can. Tech. Rep. Fish. Aquat. Sci. 1804: viii + 59 p.

Lanteigne, M., and Davidson, L. 1992. Status of the giant scallop (*Placopecten magellanicus*) fishery in the southern Gulf of St. Lawrence (DFO, Gulf Region) - 1990 update. Can. Manuscr. Rep. Fish. Aquat. Sci. 2148: iv + 21 p.

Lidgard, D.C., and Norden, W. 2011. Sea Scallop *Placopecten magellanicus*. *In* Seafood Watch.

MacCall, A.D. 2009. Depletion-corrected average catch: a simple formula for estimating sustainable yields in data-poor situations. ICES J. Mar. Sci. 66(10): 2267–2271.

Mallet, M. 2010. Commercial scallop (*Placopecten magellanicus*) fishery profile in the Gulf Region. Stat. Econ. Anal. Ser. 1–5(1): v–25.

Martell, S., and Froese, R. 2013. A simple method for estimating <scp>MSY</scp> from catch and resiliencev. Fish Fish. 14(4): 504–514.

McDonald, R.R., Keith, D.M., Sameoto, J.A., Hutchings, J.A., and Flemming, J.M. 2021. Explicit incorporation of spatial variability in a biomass dynamics assessment model. ICES J. Mar. Sci. 78(9): 3265–3280.

McDonald, R.R., Keith, D.M., Sameoto, J.A., Hutchings, J.A., and Flemming, J.M. 2022. Incorporating intra-annual variability in fisheries abundance data to better capture population dynamics. Fish. Res. 246: 106152. Elsevier B.V.

Merrill, A.S., and Posgay, J.A. 1964. Estimating the Natural Mortality Rate of the Sea Scallop (*Placopecten magellanicus*). Int. Comm. Northwest Atl. Fish. (NAFO SCS Doc. 14/023): 88–106.

Naidu, K.S., and Meron, S. 1986. Predation of scallops by american plaice and yellowtail flounder. Can. Atl. Fish. Sci. Advis. Committee, Res. Doc. 86/62: 1–25.

Nelson, G. 2023. fishmethods: Fishery Science Methods and Models.

Newman, D., Berkson, J., and Suatoni, L. 2015. Current methods for setting catch limits for data-limited fish stocks in the United States. Fish. Res. 164: 86–93.

Niles, M., Barrell, J., Sameoto, J., Keith, D., and Sonier, R. 2021. Scallop fishery assessment of the southern Gulf of St. Lawrence in 2018: Commercial fishery and survey data. Can. Sci. Advis. Secr. Res. Doc. 2021/038: xii + 178 p.

Pedersen, M.W., and Berg, C.W. 2017. A stochastic surplus production model in continuous time. Fish Fish. 18(2): 226–243.

Pella, J.J., and Tomlinson, P.K. 1969. A generalized stock production model. Inter-American Trop. Tuna Comm. Bull. 13(3): 416–497.

Posgay, J.A. 1957. The range of the sea scallop. Nautilus (Philadelphia). 70(2): 55–56.

Prager, M.H. 1994. A suite of extensions to a nonequilibrium surplus-production model. Fish. Bull. 92(2): 374–389.

Punt, A.E., Dunn, A., Elvarsson, B.Þ., Hampton, J., Hoyle, S.D., Maunder, M.N., Methot, R.D., and Nielsen, A. 2020. Essential features of the next-generation integrated fisheries stock assessment package: A perspective. Fish. Res. 229(May): 105617. Elsevier.

Restrepo, V.R., Thompson, G.G., Mace, P.M., Gabriel, W.L., Low, L.L., MacCall, A.D., Methot, R.D., Powers, J.E., Taylor, B.L., Wade, P.R., and Witzig, J.F. 1998. Technical guidance on the use of precautionary approaches to implementing National Standard 1 of the Magnuson-Stevens Fishery Conservation and Management Act. NOAA Tech. Memo. NMFS–F/SPO: 56 p.

Reuter, R.F., Conners, M.E., Dicosimo, J., Gaichas, S., Ormseth, O., and Tenbrink, T.T. 2010. Managing non-target, data-poor species using catch limits: Lessons from the Alaskan groundfish fishery. Fish. Manag. Ecol. 17(4): 323–335.

Ricard, D., Minto, C., Jensen, O.P., and Baum, J.K. 2012. Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. Fish Fish. 13(4): 380–398.

Schaefer, M.B. 1954a. Fisheries dynamics and the concept of maximum equilibrium catch. Proc. Gulf Caribb. Fish. Inst. 6: 1–11.

Schaefer, M.B. 1954b. Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. Bull. Inter-American Trop. Tuna Comm. 1(1–2): 27–56.

Shumway, S.E., Selvin, R., and Schick, D.F. 1987. Food resources related to habitat in the scallop *Placopecten magellanicus* (Gmelin, 1797): a qualitative study. J. Shellfish Res. 6(2): 89–95.

Smith, S.J., and Lundy, M.J. 2002. Scallop Production Area 4 in the Bay of Fundy: Stock status and forecast. Can. Sci. Advis. Secr. Res. Doc. 2002/018: 1–89.

Smith, S.J., Nasmith, L., Glass, A., Hubley, B., and Sameoto, J.A. 2015. Framework assessment for SFA 29 West scallop fishery. Can. Sci. Advis. Secr. Res. Doc. 2014/110: v + 69 p.

Smith, S.J., and Rago, P. 2004. Biological reference points for sea scallops (*Placopecten magellanicus*): The benefits and costs of being nearly sessile. Can. J. Fish. Aquat. Sci. 61(8): 1338–1354.

Smith, S.J., Sameoto, J.A., and Brown, C.J. 2017. Setting biological reference points for sea scallops (*Placopecten magellanicus*) allowing for the spatial distribution of productivity and fishing effort. Can. J. Fish. Aquat. Sci. 74(5): 650–667.

Wilson, B.R., Brown, C.J., Sameoto, J.A., Lacharité, M., Redden, A.M., and Gazzola, V. 2021. Mapping seafloor habitats in the Bay of Fundy to assess megafaunal assemblages associated with *Modiolus modiolus* beds. Estuar. Coast. Shelf Sci. 252: 107294.

Winker, H., Carvalho, F., and Kapur, M. 2018. JABBA: Just Another Bayesian Biomass Assessment. Fish. Res. 204(November 2017): 275–288. Elsevier.

Winker, H., Carvalho, F., Thorson, J.T., Kell, L.T., Parker, D., Kapur, M., Sharma, R., Booth, A.J., and Kerwath, S.E. 2020. JABBA-Select: Incorporating life history and fisheries' selectivity into surplus production models. Fish. Res. 222(February 2019): 105355. Elsevier.

Worms, J., Lanteigne, M., and Davidson, L.-A. 1986. Status of the southern Gulf of St. Lawrence scallop stocks - 1985. Can. Atl. Fish. Sci. Advis. Committee, Res. Doc. 86/55: 1–47.

Zhou, S., Punt, A.E., Smith, A.D.M., Ye, Y., Haddon, M., DIchmont, C.M., and Smith, D.C. 2018. An optimized catch-only assessment method for data poor fisheries. ICES J. Mar. Sci. 75(3): 964–976.

Zhou, S., Punt, A.E., Ye, Y., Ellis, N., Dichmont, C.M., Haddon, M., Smith, D.C., and Smith, A.D.M. 2017. Estimating stock depletion level from patterns of catch history. Fish Fish. 18(4): 742–751.

Zhou, S., Yin, S., Thorson, J.T., Smith, A.D.M., and Fuller, M. 2012. Linking fishing mortality reference points to life history traits: An empirical study. Can. J. Fish. Aquat. Sci. 69(8): 1292–1301.

APPENDIX 1

Under the Fisheries Act (L.R.C. (1985), ch. F-14) (Government of Canada Fisheries Act 1985), as amended in 2019, the Department of Fisheries and Oceans (DFO) applies the Precautionary Approach (PA) to fisheries management, incorporating Limit Reference Point (LRPs) linked to ecosystem indicators (DFO 2021). Evaluating data availability is crucial to establishing viable methods for determining LRPs, encompassing effort, catch, biomass estimates, biological parameters, and spatial locations, alongside their temporal coverage (Boudreau and Duplisea 2022).

The sea scallop (*Placopecten magellanicus*) fishery in the southern Gulf of St. Lawrence (sGSL) has historically interacted economically with the lobster fishery (MacPhail 1954; Jamieson 1978; Lanteigne and Davidson 1991), impacting scallop fishing effort. Various factors, including politics, technological advancements, environmental conditions, and habitat quality concerns, have also influenced activity levels of the scallop fishery. These factors, in turn, have shaped the evolution of management measures, the frequency of scallop surveys and assessments, and the collection of both fishery-independent and fishery-dependent data (Worms and Chouinard 1983; Lanteigne and Davidson 1992; Mallet 2010; DFO 2013, 2020).

FISHERY DEPENDENT DATA AND REGULATIONS

The earliest records of scallop landings, dating back to 1923, were documented in the Dominion's Annual Report (Cardin 1924; Chiasson 1949). During this period, landings were reported on a county and provincial basis, and licensed fish harvesters were permitted to fish throughout the entire Gulf. Between 1947 and 1966, a growing commitment from communities and political circles aimed at enhancing and diversifying fishing income sources led to advancements in boat technology and fishing gear. Exploratory reports during this time sought to understand the scallop fishing potential in areas of the sGSL that were already being fished (Dickie 1951; Rowell and Lord 1965).

These efforts resulted in a significant increase in scallop landings, reaching a historical peak in 1968 (Chandler 1973; Amaratunga et al. 1976; Lanteigne and Davidson 1991), followed by a rapid decline. The decline prompted concerns about the necessity of obtaining reliable catch and effort data to improve stock assessments, deepen our understanding of scallop abundance, biology, ecology, and inform regulatory decisions (Amaratunga et al. 1976; Caddy and Chandler 1976; Jamieson 1978).

Subsequent decades saw the introduction of regulations governing boat size, gear specifications, season length, and ongoing efforts to enhance data collection from each purchase slip. This data included information about registered boats, licensed fish harvesters assigned to well-defined Scallop Fishing Areas (SFAs), and the introduction of logbooks to improve the understanding of fishing effort.

As of 2021, the management of the sGSL sea scallop fishery is organized into SFAs, including 21A (Chaleur Bay), 21B (Miscou), 21C (Escuminac), 22 (West Point and Cape Tormentine), 23 (North Prince Edward Island), and 24 (Pictou, Boughton Island, and George’s Bay beds), each with one or more known scallop beds. Each fishing area is independently managed, with its unique season, gear restrictions, meat count, and fish harvester regulations (Niles et al. 2021). Refer to Table 1 for a detailed history of data acquisition and quality improvement.

Table A1.1. General description of the data and regulations introduced to the Gulf Region scallop fisheries from 1923 to 2020 (Modified from Lanteigne and Davidson 1991;Table 10).

Years	Descriptions
1923 - 1946	Data are compiled annually (*year*) by county and *province*. Landings of live scallops are reported in barrels. One barrel produces two gallons of meat. Landings of meat (*landings*) are reported in gallons or cases. A cases consists of 48 cans of one pound each. The number of licences issued (*licences*) each year refers to a boat or a fish harvester (not mentioned). Fish harvesters are allowed to fish the entire southern Gulf. Not all landings are reported, especially prior to 1945 when catch was for local and/or personal use. (PARL n.d.; Cardin 1926; Duranleau 1933; Dickie 1951; Caddy and Chandler 1976; Lanteigne and Davidson 1991)
1947 - 1966	Data are compiled annually (*year*) by statistical district (*district*) and province (*province*). Statistical districts were introduced in 1947. Landings are reported in pounds of scallop meat (*landings*). (Caddy and Chandler 1976; Lanteigne and Davidson 1988, 1991)
1962 - 1978	An annual census provides a number of self-identified active licensed fish harvesters. There are usually 2 crew members (including a captain) per scallop vessel. The actual number of active licences in a season depends on the value and catches of other fisheries by these fish harvester. First landings of meat with attached roe in 1963. (Bourne 1965; Fisheries and Marine Service Canada 1977; Jamieson 1978; Jamieson et al. 1981; Lanteigne and Davidson 1991)
1967-1976	Data are compiled monthly (*year, month*) by landing port (*port.landed)*; this variable is composed of the variables *community*, *district* and *province*. Landings (*landings*) are recorded in a computerized databases in pounds and kilograms. The different landing types (meat, roe or live) are identified. (Amaratunga et al. 1976; Jamieson 1978; Jamieson et al. 1981; Lanteigne and Davidson 1991)
1978 - 1984	First regulations on licences, allocation of a licence to a Lobster Fishing Areas (LFA), gear width, boat size, season, meat count and voluntary logbook.

Years	Descriptions
	Creation of the DFO Gulf Region (1981). (Jamieson 1978; Jamieson et al. 1980; Robert and Jamieson 1983; Worms and Chouinard 1984; Lanteigne et al. 1987; DFO 1991)
1977 - 1981	Each sales transaction (*purchase slip*) for each landing is compiled by port of landing (*date.landed*, *port.landed*). Some landings are weekly landings and are identified by *fish.grade* == 'W'. Landings (*landings*) are recorded in pounds and kilograms and the different landing types (*type*; 'meat', 'roe' or 'live' or the corresponding *type.id*; 612, 623, 929 sometimes coded in *target.sp*) are identified. Fishery officers estimate the number of active fish harvester (*active*). (Jamieson 1978; Lanteigne and Davidson 1991)
1982 - 1983	Each sales transaction (purchase slip; mainly: *slip.id*, *date.landed*, *port.landed*) is compiled. Landings (*landings*) are recorded in pounds and kilograms per landing type and per purchase slip type (standard, supplementary B; local sales estimates are assigned to *buyer.code* 9000, or supplementary A; irregular sales estimates). (Lanteigne and Davidson 1991)
1985 -	Each sales transaction (purchase slip) is compiled. Landings are recorded in pounds and kilograms per landing type and per purchase slip type. The Commercial Fishing Vessel identification (*cfv*) is recorded for each purchase slip. The count of all *CFV*s in a *year* (defined as *boats*) could be used to provide information on the number of *active* fishing vessels. Scallop Fishing Areas (SFA; *sfa* assessed with *port.landed*) have been created and replaced LFAs as the area assigned to a fish harvester's licence. (Worms et al. 1986; Lanteigne et al. 1987; Lanteigne and Davidson 1988)
1996 -	SFA 21 is divided into 3 sub-areas: 21A, 21B and 21C to facilitate management and becomes official in 2008. (DFO 1996; Mallet 2010)
1998 -	Mandatory logbooks were included as a condition of licence, but data have only been available since 2001 and used since 2003. The logbook data used include:

Years	Descriptions
	Landings: kilograms for each fishing day (*date.caught*), sales to buyers or locals (via *slip.id*), *date.landed*, *port.landed* and *cfv*), effort: *date.caught* and location in *latitude* and *longitude*, number of tows (*drag.no*), average duration of a tow in minutes (*ave.tow* since 2003, *hours.fished* since 2001) for each fishing day)). Implementation of scallop buffer zones to protect the American lobster nursery habitats (SFA 23 (Gentlemen's Agreement), 24, 21A and 21B). (DFO 1999, 2011; DAFA 2007; Davidson et al. 2007, 2012)
2005 -	Implementation of the SFA 22 scallop buffer zone to protect the American lobster nursery habitats. (Davidson et al. 2007; UN 2009; DFO 2018)
2014 -	Vessel Monitoring Systems (VMS) became mandatory for the SFA 21A. (Bayer et al. 2016)

FISHERY-INDEPENDENT AND DEPENDENT SURVEY DATA

Smith et al. (2009) reported an association between scallop abundance and, substrate type and depth, suggesting that scallop bed locations should remain constant unless the bottom depth and substrate type change (Davidson et al. 2012). This association enables the use of previous surveys that focus on known beds rather than relying on random stratified survey over all area. Surveys conducted between 1946 and 1986 provide fishery-independent indices of abundance, biomass, biological characteristics (i.e. shell height, whole and/or meat weight, age, clapper rate), along with environmental data (i.e. depth, substrate type, associated species/bycatch) within the extent of the surveyed area or delineated beds.

In 1997, Hanson (1998) conducted the first random stratified survey covering all fished and unfished areas to assess SFA 22. This survey included comparisons of estimates per square meter with earlier works from 1967, 1979, 1980, 1981, 1983, and 1986. Unfortunately, these data sets are not yet fully digitized and available.

Recent data sources include a sea sampling project from 2001 to 2005 aimed at enhancing the understanding of scallop stocks and the fishery while providing cost-effective science advice (Davidson et al. 2007). The second project, spanning 2012 to 2016, constitutes an annual, rotational (i.e. different SFA each year), multispecies research survey program for scallops in the sGSL. This initiative aimed to obtain fishery-independent indices of abundance, biomass, and biological characteristics (i.e. shell height, meat weight, age) for all SFAs except 23 (Niles et al. 2021). A stratified random design was applied to each survey (Smith and Gavaris 1993; Smith and Addison 2003). Since 2019, annual surveys have focused on the three main scallop beds in the Gulf Region (i.e. West Point, Cape Tormentine and Pictou), accounting for 80% of the Gulf landings (Niles, pers. comm.).

Table A1.2. General description of the latest scallop surveys and related data availability for the Gulf Region from 2001 to 2023.

Years	Descriptions
2001-2005	All SFAs, sea-sampling program on fishing boats. One fishing day per year, from 2001 to 2005, by main bed in each SFA (One bed in 21A, One bed in 21B, 2 beds in 21C, 5 beds in 22, One bed in 23, 5 beds in 24). Information on boats (see also, Poirier et al. 2021), gears, length of tows (at every second tow) are recorded with GPS (at the start and end of the tow) and time length and speed of the tow. A fishing day consisted of between 15 and 70 tows lasting from 10 to 30 minutes each at a speed of between 2 and 2.7 knots. Bottom type Scallops and clappers are counted and measured, and a subsample of scallops is aged, sexed, sized for growth and weighed (live, meat, gonad and viscera) (Davidson et al. 2007, 2012)
2012-2016	All SFA (except 23 because of low fishing effort, fishery independent survey) Rotational as each year a different area is surveyed 22 south: 2012 21A: 2013 22 north: 2014 24 2015 21BC: 2016 Stratified random design; 3 to 4 effort-based strata, plus one including buffer areas, following Smith and Gavaris 1993 for area deeper than 5.5 m, number of tow is proportional to the strata area and weighted by fish effort, the tows are randomly allocated Eight-gang toothed Digby scallop drag, all buckets are lined with 14 mm Vexar® mesh. The length of a tow is measured with GPS and the time length and speed of the tow are fixed. The catch of each tow was sorted, counted and weighed by species. All scallops and clappers are measured. A subsample of scallops is weighted, aged sexed. (Niles et al. 2021)

Years	Descriptions
2019 - 2023	SFA 22 (beds: West Point, Cape Tormentine) and SFA 24 (bed: Pictou) and Pictou (SFA 24). SFA beds were surveyed each October between 2019 and 2023, except for SFA 24 bed in 2020. SFA 22 beds were also surveyed in April of 2022 and 2023 to obtain pre fishing season estimates. Stratified random design; effort-based strata (in logbooks for years 2001 to 2016), for area deeper than 5.5 m, number of tow is proportional to the strata area and weighted by fish effort, the tows are randomly allocated. Eight-gang toothed Digby scallop drag, all buckets are lined with 18 mm Vexar® mesh. The length of a tow is measured with GPS and the time length and speed of the tow are fixed. The catch of each tow was sorted, counted and weighed by species. All scallops and clappers are measured. A subsample of scallops is weighed, aged, and sexed. (Niles et al. unpublished)

Within this historical context, this appendix attempts to :

- Collect historical and currently used sGSL scallop fishery dependent data (data collection process) in order to obtain the longest possible time series of effort and landing data.
- Develop a structured workflow scheme to improve the understanding of the data source, its constraints and the transformation developed in the data cleansing process (data workflow).
- Initiate the development of a tool (set of scripts) to systematically and automatically extract, standardise, clean and collate data sets (data cleansing process).
- Briefly present, the fishery independent data used to estimate the von Bertalanffy parameters.

METHODOLOGY

The methods of fishery dependent data acquisition, cleansing, collating and aggregation presented in this appendix differ from those previously used in the sGSL scallop assessment. This is mainly because for the LRP analysis, an attempt was made to produce the longest timeseries of landings and effort possible, mainly through the use of historical documents.

To acquire the historical document data we used the websites nanonets and automeris to convert the .pdf or .png files into .csv files. The digitized datasets were corrected manually when necessary, particularly when the quality of images and figures was poor. Additional datasets were obtained from the archives of the Gulf region scallop fishery biologists and the Statistic Division of DFO.

The methodology for achieving workflow and script development objectives is implemented using the R language; R Foundation for Statistical Computing Platform, version 4.2.2 (R Core Team 2016). Scripts related to the models are stored in RStudio projects (Posit team 2022), which are archived and available on GC/Code, a shared service of the Department of Fisheries and Oceans Canada (DFO) hosted by GitLab services. Input and output documents for each project are stored in their respective subfolders.

This work is subdivided into three projects:

- ***ss_fishery_data***, which document the acquisition, cleansing and transformation of fishery dependent data;
- ***ss_survey_data***, which document the fishery independent data used to estimate the von Bertalanffy parameters; and,
- ***ss_lrp***, which document all modeling work done in the research document (not presented in this appendix).

The purpose of this appendix is to introduce the ***ss_fishery_data*** R project which automates the execution of a series of fishery data cleansing and aggregation scripts. These scripts are classified by the data sources and tasks performed and then, produces the ***scallop.RData*** and the ***scallop_sfa.RData*** files containing the requested data and variables used in the ***ss_lrp*** R project. Also, this appendix briefly presents a section of the ***ss_survey_data*** R project that processes fishery independent survey data to produce the von Bertalanffy parameters.

In this regard, the results section is divided into five main parts:

in the ***ss_fishery_data*** project;

1. Fishery dependent data acquisition process,
2. Scripts workflow, 00. numbering,
3. Data cleansing process, 01. numbering,
4. Data transformation for the purpose of fishery analysis, 02. numbering, and

in the ***ss_survey_data*** project;

1. Fishery independent data and estimates of von Bertalanffy parameters.

Results will be presented for the whole sGSL (all the SFAs) and separately for the SFAs 22 and 24.

RESULTS AND DISCUSSION

Fishery Dependent: Data acquisition process

The landing and effort datasets for the sGSL scallop fishery used in the research document come from different sources and are stored by source in their respective data folders within the project.

The first source is historical data which includes published documents such as these, government reports, research documents and peer review documents. The data were digitized from various locations in the reports, including results, tables, appendices and figures. However, some data were obtained from raw and transformed, but already published, data from the archive datafiles of the fishery biologists. The datasets are:

- landings_1947-73_multisp_caddy_chandler76.csv, located in Table 1 (Caddy and Chandler 1976);
- cpue_landings_jamieson80.csv, located in Tables 4 and 7 (in Jamieson et al. 1980);
- cpue_landings_jamieson81.csv, located in Tables 4 and 15 (in Jamieson et al. 1981);

- cpue_landings_worms_chouinard84.csv, located in Appendix 4, Tables 3, 6 and 7 (in Worms and Chouinard 1984);
- cpue_landings_worms86.csv, located in text, Tables 1 and 8 (in Worms et al. 1986);
- landings_1967-87_lanteigne_davidson88.csv, located in Figure 4, Tables 3 and 4 (in Lanteigne and Davidson 1988);
- cpue_landings_lanteigne_davidson89.csv, located in Figure 1 and Table 3 (in Lanteigne and Davidson 1989);
- landings_1968-89_lanteigne_davidson91.csv, located in Appendix 1 (in Lanteigne and Davidson 1991);
- landings_1982-89_lanteigne_davidson91.csv, located in Appendix 2 (in Lanteigne and Davidson 1991);
- landings_1923-89_lanteigne_davidson91.csv, located in Figure 2 (in Lanteigne and Davidson 1991);
- landings_1968-09_mallet10.csv, located in Appendix 3 (in Mallet 2010); and
- landings_1968_16_MN.csv, located in Tables 6 and 7 (in Niles et al. 2021).

The second source of data is from Statistic Division. The datasets are from purchase slips (1984 to 2021):

- sgsl_scallop_slip_data_1984-2002.xlsx,
- 1990-2000 scallop_original_MN.xls,
- sgsl_scallop_slip_data_2003-2020.csv,
- qc_slips_2003_2020.csv, and
- GLF2021_165_Scallop 2021 slipmaster YYYYMMDD.xlsx (YYYYMMDD equals to year, month and day),

and from logbooks (2003 to 2021):

- sgsl_scallop_log_data_2003-2020.csv,
- qc_logs_2003_2020.csv, and
- GLF2021_165_Scallop 2021 Logmaster YYYYMMDD.xlsx (YYYYMMDD equals to year, month and day).

The third source of data corresponds to Regulations and Fleet, which documents the opening, length and closing of seasons (2008 to 2021):

- scallop_seasons.csv;

and the number of annual licences, boats and active boats (1960 to 2023):

- licence_records_1960-2023.csv, and
- licence_records_1962-1984_sfa.csv.

Scripts workflow

The script workflow consists of two main scripts : *0_ss_fishery_data_master.R* and *00.0_update_data.R*. The Figure A1.1 illustrates the general operations of the semi-automated workflow implemented in these scripts.

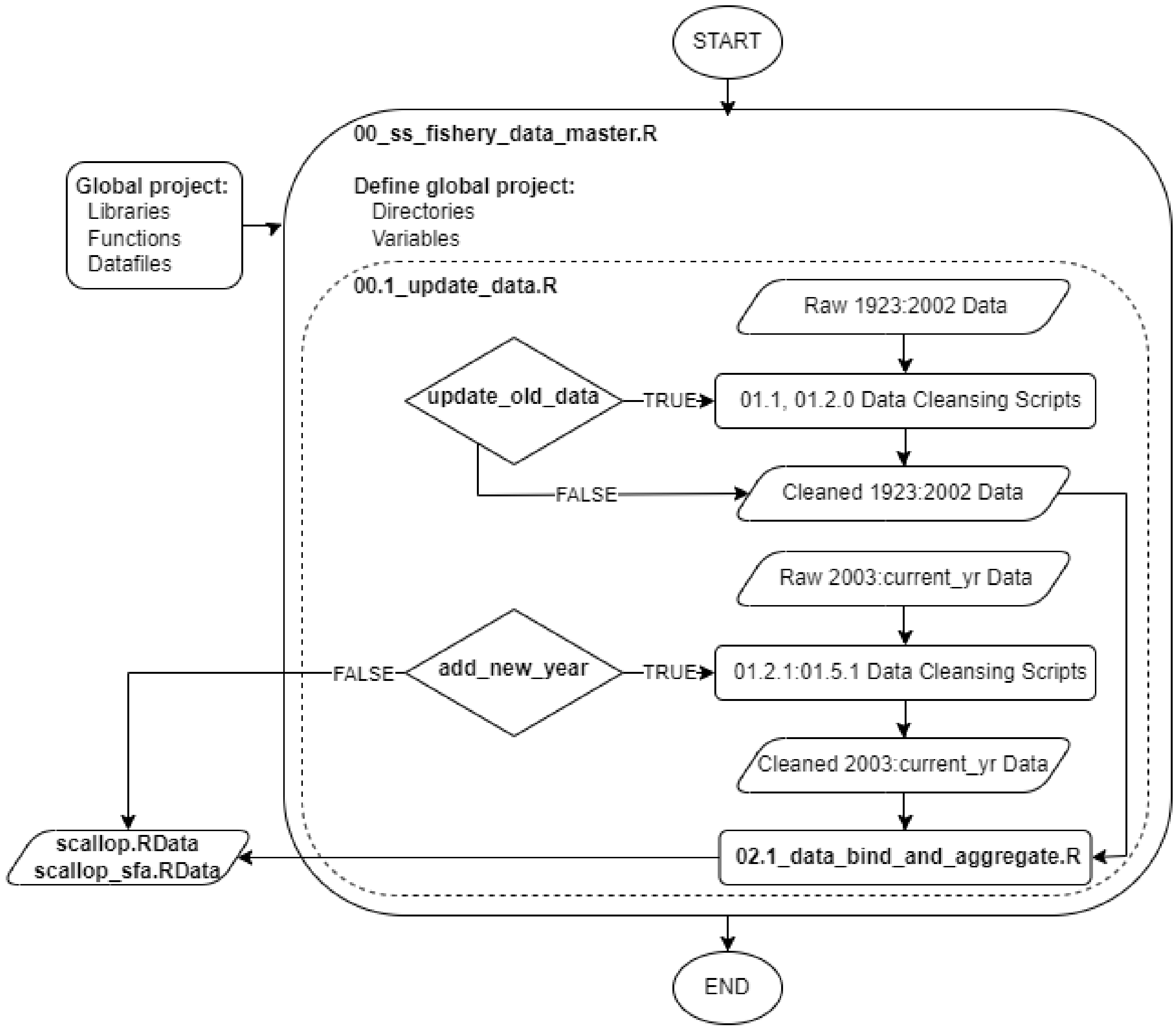

Figure A1.1. Diagram to summarise the semi-automated cleansing and aggregation steps performed in the 00_ss_fishery_data_master.R to produce the final data file required for LRP analysis. See Figure A1.2 for details of the steps within the dashed rectangle 00.1_update_data.R.

0_ss_fishery_data_master.R

The objective of 0_ss_fishery_data_master.R is to define and load shared directories, libraries, variables, datafiles and functions requested to call and execute the data cleansing and aggregation scripts that will end up producing **scallop.RData** and **scallop_sfa.RData** datafiles used in the research document.

Directories

This directories are the different paths used to assess folders containing the datasets from the different sources. They are defined and stored in the following variables:

- ***main_dir***, is the location of the project on the user computer and is automatically defined;
- ***historic_dir1*** (year < 1990), ***historic_dir2*** (year < 2010) and ***historic_dir3*** (year = 2016) are the locations of historical data;
- ***slip_dir***; is the location of the raw slips data files;

- ***log_dir***; is the location of the raw logbooks data files;
- ***fleet_dir***; is the location for fleet and regulatory related data files; and
- ***output_dir***; is the location of the data outputs of all the scripts.

Libraries

The required libraries are downloads is this script. They are presented below in alphabetical order with their main purpose in this project:

- *base*, *base* R libraries including the *stats* library of functions (R Core Team 2016).
- *caret*, tools for Classification And REgression Training library is used for resampling (Kuhn 2022).
- *data.table*, an *improved version of the default; base used to work with tabular data (Dowle and Srinivasan 2021)*
- *ggplot2*, functions to produce graphs (Wickham 2016).
- *gulf*, functions shared by the DFO Gulf Region to support fisheries stock assessment (Surette et al. 2021).
- *librarian*, functions to install and attach all required packages in one step (Quintans 2021).
- *lubridate*, functions for handling with data and time (Grolemund and Wickham 2011).
- *missRanger*; functions to impute missing values based on Random Forest (Stekhoven and Bühlmann 2012; Wright and Ziegler 2017).
- *patchwork*, provides functions to combine multiple plots (Pedersen 2020)
- *readxl*, functions for reading and loading MS-Excel spreadsheets (Wickham and Bryan 2022).
- rlang, collection of frameworks and APIs for programming with R, such as the dynamic dot (Wickham 2022).
- sf, functions to a standardized and encode spatial vector data (Pebesma 2018; Pebesma and Bivand 2023)
- *splitstackshape*, functions to reshape a column, here the *cSplit* function is used to split a list into modalities of a variable in many rows (Mahto 2019)
- *stringdist*, function *amatch* used to match a variable name to a line number where it is (van der Loo 2014)

Variables

The variables presented below are predefined to avoid duplication of information or multiple definitions of them, which could occur because they are used by many scripts. The script defines variables for the species being worked on:

- ***target_sp*** stored the *"sea scallop"* character value relative to the studied species, and
- ***stacac_code*** stored the *612* integer value retrieved with the function *gulf::species.code(**target_sp**, output.coding = "STACAC").* STACAC means Statistical Co—Ordinating Committee for the Atlantic Coast.

It defines a variable for current year the user is working on:

- ***current_yr*** stored the *2021* integer value that could also be retrieved by the function *base::as.integer(format(Sys.Date(), "%Y")).*

It assigns conversion factors to different global variables:

- ***bushel_to_lb*** stored the numeric value *4.98*, 1 bushel equals 4.98 lbs of meat (Chiasson 1952);
- ***lb_to_kg*** stored the numeric value *2.20462*, 1 lb equals 2.20462 kg;
- ***ft_to_m*** stored the numeric value *3.28084*, 1 ft equals 3.28084 m; and
- ***live_to_muscle*** stored the numeric value *8.3*, 1 kg of live scallop equals 8.3 kg of meat (Lanteigne and Davidson 1991).

Datafiles

The script download the files stored in the ***main_dir*** *data/* folder that will be used by different scripts:

- ***port_list***, a list of Gulf ports used to link port number to SFA in pre-2003 slip records. This file is in the subfolder and file: *stats_dep/ports/**Gulf Region Community Ports.csv***;
- ***area_link***, a list of area variables to establish the link between their definitions for use in historic datasets. This file is located in the subfolder and file: *stats_dep/ports/**area_relations.csv***;
- ***var_key***, a variable dictionary for slips and logs data (***year*** *>= 2003*), this file identifies which variables apply to which species and provides the standardised variable names and types. This file is located in the subfolder and file: *stats_dep/**var_key.csv***; and
- ***scallop_poly***, a vector of links towards spatial polygons:
  - *maps/beds/**beds_contour_20.shp***, where the element name of the vector is "bed.20"*, and*
  - *maps/beds/**scallop_sfa.shp***, where the element name of the vector is "sfa.map"*.*

Functions

It downloads functions stored in the ***main_dir*** *r_scripts/functions* folder*.* They are presented below.

CleanNames.R

The objective of *CleanNames* function (Doane 2019) is to take the column names from a data file or vector and to standardise the column/variable names. It is not explained on the web site, so here is a summary of what it does.

The function requires the *rlang and gulf* libraries and has 2 attributes:

- ***.data*** corresponds to a dataframe or a vector of column names, and
- ***unique***, corresponds to a logical variable that allows or not to correct duplicate variable names by setting unique to *TRUE* or *FALSE* (default). For example, if the raw vector of variable names is equal to *c("a", "a")*, the vector of variable names at the end will be equal to *c("a", "a.1")*.

Main steps of this function:

- Replaces symbols with text equivalent plus surrounded by underscores,

- Replaces a single non-alphanumeric/underscore/whitespace(s) character with an underscore,
- Places an underscore before an uppercase letter followed by a lowercase letter (changes camel case to snake case),
- Converts all to lower case,
- Removes leading and trailing underscore(s),
- Replace double underscores with one,
- Makes the variable names unique if the attribute ***unique*** *equals TRUE* by adding an underscore followed by a unique number, and
- Replaces underscores with periods.

MapBedSfa.R

The objective of the *MapBedSfa* function is to assign a sample to a polygon: a delineated bed or an SFA, whether its latitude and longitude fall within the polygon.

The function requires the *sf* and *ggplot2* libraries and has ten attributes: ***input_data***, which is the dataset to modify; ***poly_paths***, a vector of the paths where to find the shapefiles which the (default is ***scallop_poly***); ***poly_var_name*** a vector of the variable containing the polygon names in the shape file (default is ***scallop_poly_var_name***); ***lat_name*** and ***lon_name*** character attributes containing names of the variable in ***input_data*** associated respectively with latitude and the longitude; ***final_crs*** numeric attribute for the coordinate reference system (CRS) of your project (default is 4326), and; the ***look*** logical attribute which prints a map of the samples that fall inside or outside (NA) the polygons. If the ***look*** is *TRUE* the attributes ***date_var*** character attribute related to the a date variable in the ***input_data*** file; and, ***cooord_sf_ylim*** (default is c(-67, -60)) and ***coord_sf_xlim*** (default is c(44.5, 49)) associated with the limit of the map figure created must not be assigned to NA.

Main steps of this function:

- Transforms the latitude (***lat_name***) and longitude (***lon_name***) of each observation in the input dataset (***input_data***) into a spatial vector and assigns the coordinate reference system (***crs***).
- Inside a loop, the same steps are performed for each shapefile present in the ***poly_paths*** attribute:
    - It transforms the shapefile into a spatial vector;
    - It transforms the coordinate reference system to correspond to **crs**;
    - It associates each sampling point with a polygon,
    - It merges the new variable, named after the name of the polygon in the **poly_path** vector, with the **input_data** dataset, and;
    - if **look** is TRUE, it generates a map that displays whether the points are located inside or outside each polygon.

The function will returns the ***input_data*** file.

ReadSlipLog.R

The objective of the *ReadSlipLog* function is to load .csv annual master file of slips and logs provided by the DFO's Statistic Division.

The function requires the *readxl* library and has as attributes ***dir_path*** which corresponds to the path where the slips or log files are located, ***yr_fct*** which corresponds to the year to be download and ***region*** which could be "GLF" for the DFO Gulf region or "QC" for the DFO Quebec region.

The function will return a list of files whose names contain a pattern equals to *paste0(**region**, **yr_fct**)*.

RenameSubareas.R

The objective of the *RenameSubareas* function is to associate the different types of subarea (***subarea***) with a bed (***bed***) or a SFA (***sfa***).

The function has only the attribute ***input_data***, which is the dataset to be modified.

Mainly, the ***area_link*** file is used to convert the ***subarea*** variable in the proper modality of the ***bed*** and ***sfa*** character variables. If further requests are made to use this function, il will need to be uptdated.

TblStdze.R

The objective of the ***TblStdze*** function is to take a raw dataset provided by the DFO Statistics Division and to convert it to a standardised format in terms of column names and types according to the information in the ***var_key*** helper file presented earlier.

The function requires the *lubridate*, *stringdist* and *splitstackshape* libraries and has 4 attributes:

- ***input_data***, the data file to be standardised,
- ***species***, defined in the ***target_sp*** variable,
- ***var_key***, is the ***var_key*** file, and
- ***date_formats***, to specify valid date formats, see the lubridate package "cheatsheets" on https://rawgit.com/rstudio/cheatsheets/main/lubridate.pdf.

Main steps of this function:

- Creates a vector of variable names from the ***input_data*** file;
- Separates multiple aliases of each ***var_key$alias*** record into separate rows and creates a ***var_key2*** file;
- For each variable in the vector of former variable names, conducts a while-loop that checks if the variable exists in the ***var_key2*** file and, if so, renames and assigns it to the appropriate class. In detail, it will:
    - identify the line number of the **var_key2** file corresponding to the former variable name if it exists in **var_key2$alias**;
    - If no line number has been identified, then the former name is kept in **input_data**;
    - *If a line number has been identified, it verifies whether this variable is applicable to the species:*
        - if the variable is not applicable, the variable is removed; or
        - if the variable is applicable, the loop assigns the **var_key2$var.name** of the previously identified row to the variable name in the former file;
    - It identifies the appropriate class for the variable.

- It then checks that each column name is unique for the updated former file, if not, the file is run through the ***CleanNames*** function again with the ***unique*** equals to *TRUE* attribute to make each column name unique.

Loading and Cleansing

To load and execute the cleansing of the raw data files, the script sources the *00.0_update_data.R* script. In order to update the data from historical sources from 1923 to 2002, the ***update_old_data*** variable must be set to *TRUE*. To add a new year of data from logbooks and purchase slips, ***add_a_new_year*** must also be set to *TRUE*. If these variables are set to *FALSE*, then the previously cleaned files will be downloaded (see decision symbol *update* in Figure A1.2). Note that, in the research document ***impute_missing_data*** for the effort variable ***hours.towed*** is always set to *TRUE* (see the section *01.4.1_impute_missing_data.R*).

00.1_update_data.R

The objective of this script is mechanical and is to execute or not the data cleansing scripts according to the instruction *TRUE* or *FALSE* stored in the logical variables ***update_old_data*** and ***add_a_new_year***. The action in this script is illustrated in Figure A1.1 inside the dashed line rectangle. This script only requires the *base* R library.

But mainly, if ***update_old_data*** equals to *FALSE* and all the Cleaned Data files (see Figure A1.2) are in ***output_dir*** folder, then it loads all the following standardised files:

- *historic_raw_dat_list.RData*, historic datasets from 1923 to 1989 documents;
- *slip_historic_list.RData*, annual purchase slip datasets from 1984 to 2002;

The same is done for the following files if **add_a_new_year** is FALSE:

- ***full_dat.RData***, annual purchase slip datasets merged with logs from 2003 to 2021;
- ***licences.RData***, annual number of licences issued and estimated active boats, and; and
- ***seasons.RData****,* annual number of fishing days allowed.

But it includes also the loads of the final files:

- ***scallop.RData***, and
- ***scallop_sfa.RData***.

Otherwise, if the files don't exist, it sources the scripts as if **update_old_data** (or **add_a_new_year**) were set to TRUE. Thus, if **update_old_data** equals to TRUE, it sources the scripts to do the Data Cleansing which cleans the Raw Datasets (see Figure A1.2), stored in the **main_dir** r_scripts folder:

- 01.1_reports_historic.R which produces historic_raw_dat_list.RData,
- 01.2.0_slips_historic.R which produces slip_historic_list.RData,

and the same is done for the following scripts if ***add_a_new_year*** is *TRUE*:

- 01.4.0_merge_logs_slips.R which produces **full_dat.RData**
- 01.5.0_fishing_seasons.R which produces **seasons.RData**
- 01.5.1_fleets.R which produces **licences.RData**

Creating the data files for analysis

Then if ***update_old_data*** or ***add_a_new_year*** were set to *TRUE* or if all files don't exist it sources the script *02.1.0_bind_and_aggregate.R*. This script will aggregate the data and produce of the final ***scallop.RData*** and ***scallop_sfa.RData*** files.

Data cleansing process

The data cleansing process has the objective to load, standardize, validate and store data for assessment. The Figure A1.2 shows detailed of the steps to import, clean and export datasets from the different sources. It corresponds to steps where:

- Raw data are loaded;
- Variable names, definitions and units are standardised;
- Data structure is transformed in order that each row corresponds to one observation;
- All data available or required variables are retained;
- True duplicates are removed;
- Missing data are handled; and
- Outliers and irrelevant observations are removed or corrected by a semi-automated Quality Assurance and Quality Control (QAQC) steps (currently under development and are not included in this appendix);
- Data are stored in the ***output_dir*** folder***.***

The script presentation below describes each of these steps for each of the fishery dependent sGSL scallop stocks data sources. The associated scripts are numbered from 01.1 to 01.5, stored in the ***main_dir*** *r_scripts* folder*.*

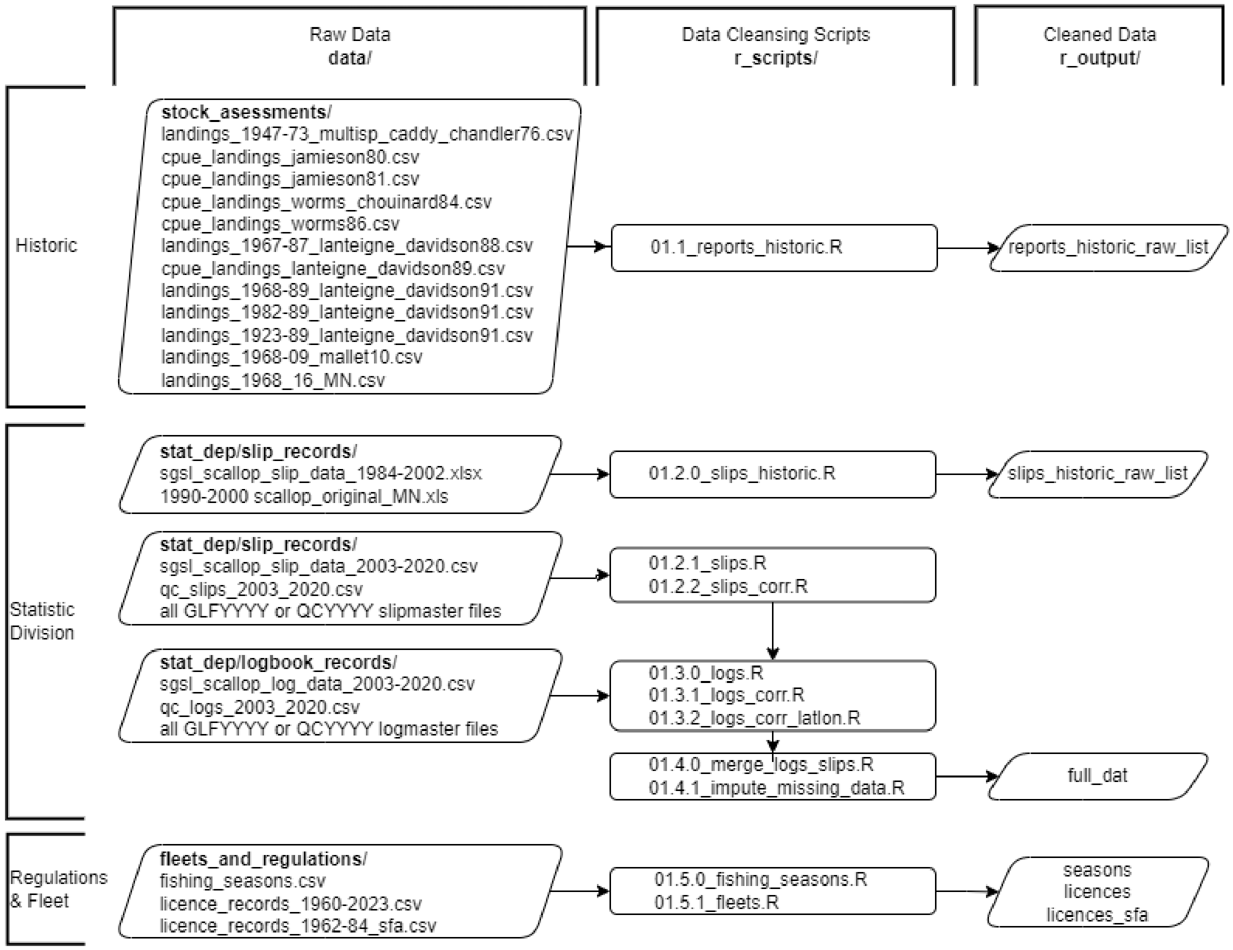

Figure A1.2. Steps to import, clean and export data from different sources; historical documents, DFO's Statistic Division and regulations and fleet related informations. YYYY in GLFYYYY and QCYYYY is equal to the year between 2021 and the value of ***current_yr*** *(see Figure A1.1).*

01.1_reports_historic.R

The objective of the script *01.1_report_historic.R* is to obtain published data on landings, catch per unit effort and effort by year; 1923 to 2016, from historical documents. To achieve this task, the script imports data, then standardizes names and transforms data variable units to the International System.

It requires the ***port_list*** and ***area_link*** files, the *RenameAreas* function, the conversion factor variables: ***bushel_to_lb***, ***lb_to_kg***, ***ft_to_m***, ***live_to_muscle***, and; the historical directory path variables: ***historic_dir1***, ***historic_dir2***, ***historic_dir3***, and their enclosed datasets.

Main steps of this script for each data files, except for the *Lanteigne and Davidson 1991*.csv, are to:

- Load the .csv data;
- Standardise the column names and their modalities:
    - For effort character modalities of the column **variable**, they are renamed according to their definition:

- "fishing.days", if it corresponds to the number of fishing days of a trip, a logbook record, or a catch date (date.caught), but before 2003 a purchase slip was considered as a fishing day, see notes below.
  - **"**monthly.reports", when unit corresponds to many fishing days in a month.
  - The numerical modalities of the landings **value.units** column are converted to kg.
  - The RenameAreas function is used to add the **bed** and **sfa** variables corresponding to the modalities of the variable **subarea**.

- Remove the data that are not in the Gulf region (or Quebec region which are part of the Gulf region assessments), and;
- Remove duplicate information, real duplicates or 2 similar information reported by the same observer in 2 different publications.

Particularities for Lanteigne and Davidson (1991):

- The ***port_list*** file is merged with the file by district to obtain SFAs (***sfa***).
- Also, before 1976 fishing days are monthly reports and therefore the data for this part of the time series for the column *variable* modality "*fishing.days"* are changed to "*monthly.trips"*. In order to have a longer time series, the data from the Appendix 1 covering the years 1968 to 1989 inclusively are supplemented with the data from the Figure 1, covering the years 1923 to 1967 inclusively.

All datasets are stored in a list called ***reports_historic_raw_list***, in the ***output_dir*** folder*.* The information on the historical data set is summarized in Figure A1.3.

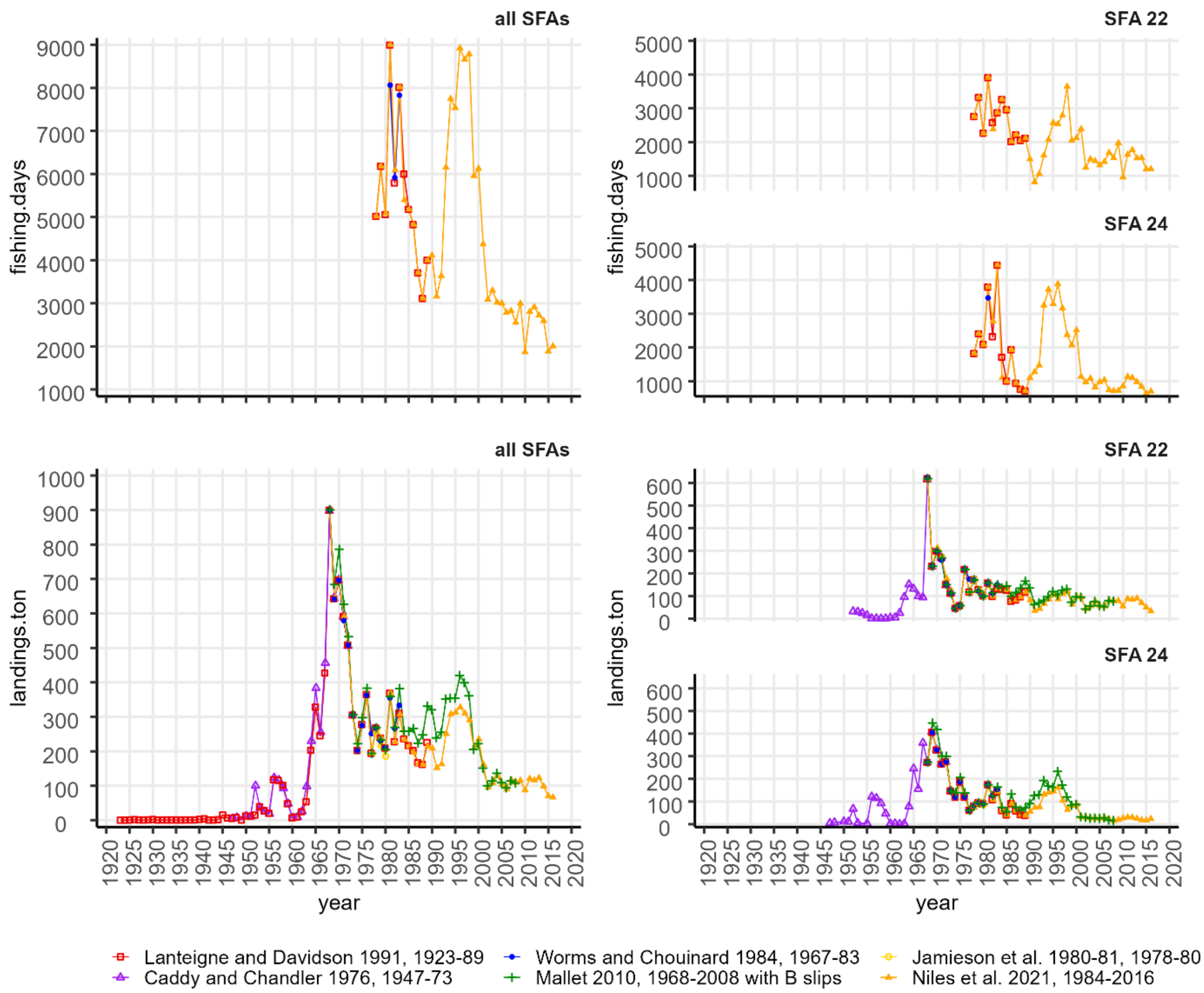

Figure A1.3. Comparison of fishing days from 1923 to 2016 and landings in tons of scallop meat between different historical documents. Left panels shown information about all Scallop Fishing Areas (SFAs: 21A, 21B, 21C, 22, 23, 24) together and; right panels SFA 22 and SFA 24.

0.1.1.1._Notes on "fishing.days" and "trips" definitions

Since 1978, all sales slips have been recorded and a voluntary logbook has been introduced. However, a percentage of catches were not sold to registered buyers (Jamieson 1978), so no purchase slips were produced. It was then assumed that it is possible to get an idea in the variation of effort in terms of number of days fished (i.e. "*fishing.days*") from year to year (Worms and Chouinard 1984) by recording the number of days on which scallops were sold (Jamieson et al. 1980; Worms and Chouinard 1983), even if this underestimates effort.

Later, in 1985, the CFV (Canadian Fishing Vessel Number) was systematically recorded on each purchase slip. This improvement helped to identify many problems in purchase slips data (Worms et al. 1986):

- Many duplicate purchase slips underlies double entry or the splitting of the daily catch between several slips for one vessel;

- Some fish harvester wait two or three days before selling their catch, resulting in a single sales slip for several fishing days;
- A number of landings is recorded without vessel number and could be the result of combination of landings by many vessels;
- Some of the above are estimates of unsold to registered buyers (sold to individuals or locals) catches in Supplementary B slips.

However, this improvement makes it possible to assume that landings recorded on a sales slip represent the catch of a fishing day ("*fishing.day*") by a vessel and can provide a rough estimate of the number of fishing days and the number of active fishing vessels.

Retrospectively, the definition of *fishing.days* was approximately equal to *trips* for all authors and was imperfect until logbook data became available in 2003 (Davidson et al. 2012). The logbook introduces a new variable, *date.caught*, which together with *date.landed* allows the calculation of real trips stored in the *trips* variable, equal to one *date.landed* per vessel, and real fishing days stored in the *fishing.days* variable, equal to one *date.caught* per vessel.

In this appendix we have therefore changed the name of the "*trips*" variable to "*fishing.days*" (i.e. the modality of ***variable*** in the datasets included in the ***reports_historic_raw_list***) for data before 2002 inclusively to distinguish between the two variable definitions. However, see also, the section *0.2.0_slips_historic* for the computation of these variables between 1984 and 2002 and all sections related to *01.4.0_merge_logs_slips.R.* for the computation of 2003 and onward.

01.2.0_slips_historic.R

The objective of the script *01.2.0_slips_historic.R* is to import and reformat the archival DFO's Statistic Division slips data on scallop landings from 1984 to 2002.

To execute the script, it requires the *readxl* library, the ***port_list*** file, the *TblStdze* function and slip files corresponding to years before 2003.

Main steps of this script:

Before the execution of the loop, each data file is loaded and some column/variable names and their modalities are corrected to allow the loop to recognize them. Then the loop takes each file one by one:

- Standardises variables using the *TblStdze* function;
- Merges ***sfa***, ***nafo.div*** and ***nafo.subdiv*** from the ***port_list*** file per ***port_landed***. NAFO stands for Northwest Atlantic Fisheries Organization;
- Converts landings units to kg;
- Removes records outside the Gulf region (not in NAFO Division 4T) or with missing SFAs;
- Removes catches of species other than scallops. ***target.sp*** must be 612, 623 or 929;
- Removes duplicate lines, true duplicates;
- Removes supplementary B slip, where ***buyer.code*** equals to 9000;
- Removes records where ***cfv*** is unknown;
- Corrects ***fishing.days*** if ***fish.grade*** equals to "W" related to weekly completion of landings for a ***cfv***:
    - creates a list of weekly fish harvesters;

- selects data of weekly fish harvesters for weeks where they fish daily (***fish.grade*** is not equal to “W”);
- counts ***fishing.days*** (number of rows) per ***cfv***, ***week*** and ***year***, then calculates an average per year (***fishing.days.ave***);
- assigns to the whole dataset ***fishing.days*** these ***fishing.days.ave*** if fish.grade == “W” and 1 to the others.

* Pastes the two files into the list named ***slips_historic_raw_list.RData*** and save it in the *output_dir* folder.

Figure A1.4 illustrates the data included in the historical purchase slip datasets and the differences between archival data and data from recent requests to the DFO Statistic Division. This figure also shows that there is a difference between these two datasets which could be corrected by adding missing observations from the archives to the recent request to the DFO Statistic Division.

Figure A1.5, which includes data from 2003 to 2021, shows that observations/purchase slips were missing but only prior to 2002 inclusively and mainly affected SFAs 21. It also seems to have unresolved issues for landings and fishing days in SFA 21C between 1991 and 1999 inclusively.

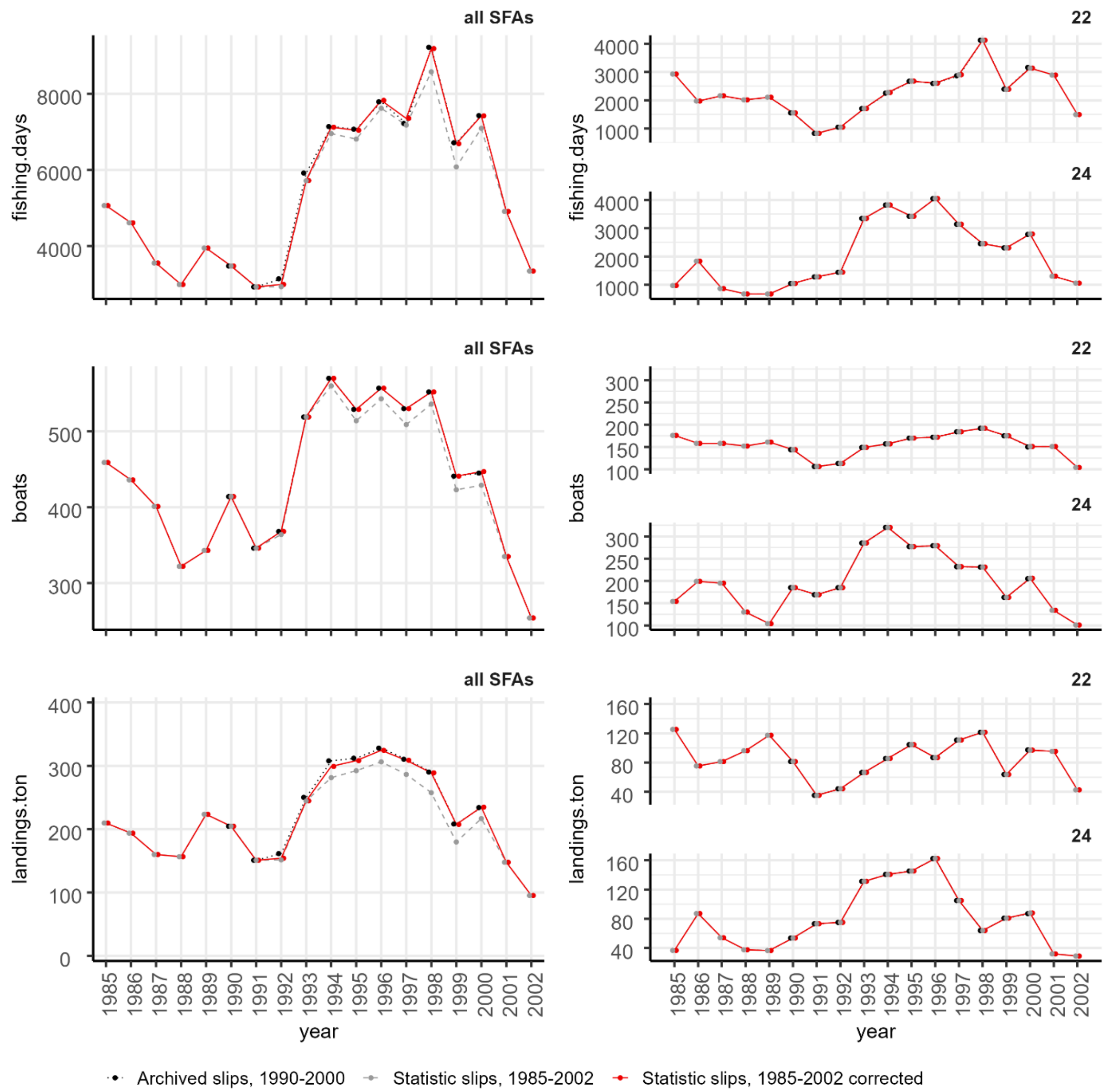

Figure A1.4. Information on number of fishing days, number of boats and landings in tons of scallop meat included in the historical purchase slips dataset and comparisons between 3 different purchase slips datasets from 1985 to 2002: archived purchase slips (black dotted), recent requests from DFO Statistic Division (gray dashed), and DFO Statistic Division where we included the missing observations (red full). Left panels show SFAs together: 21A, 21B, 21C, 22, 23 and 24. Right panels show SFA 22 and SFA 24.

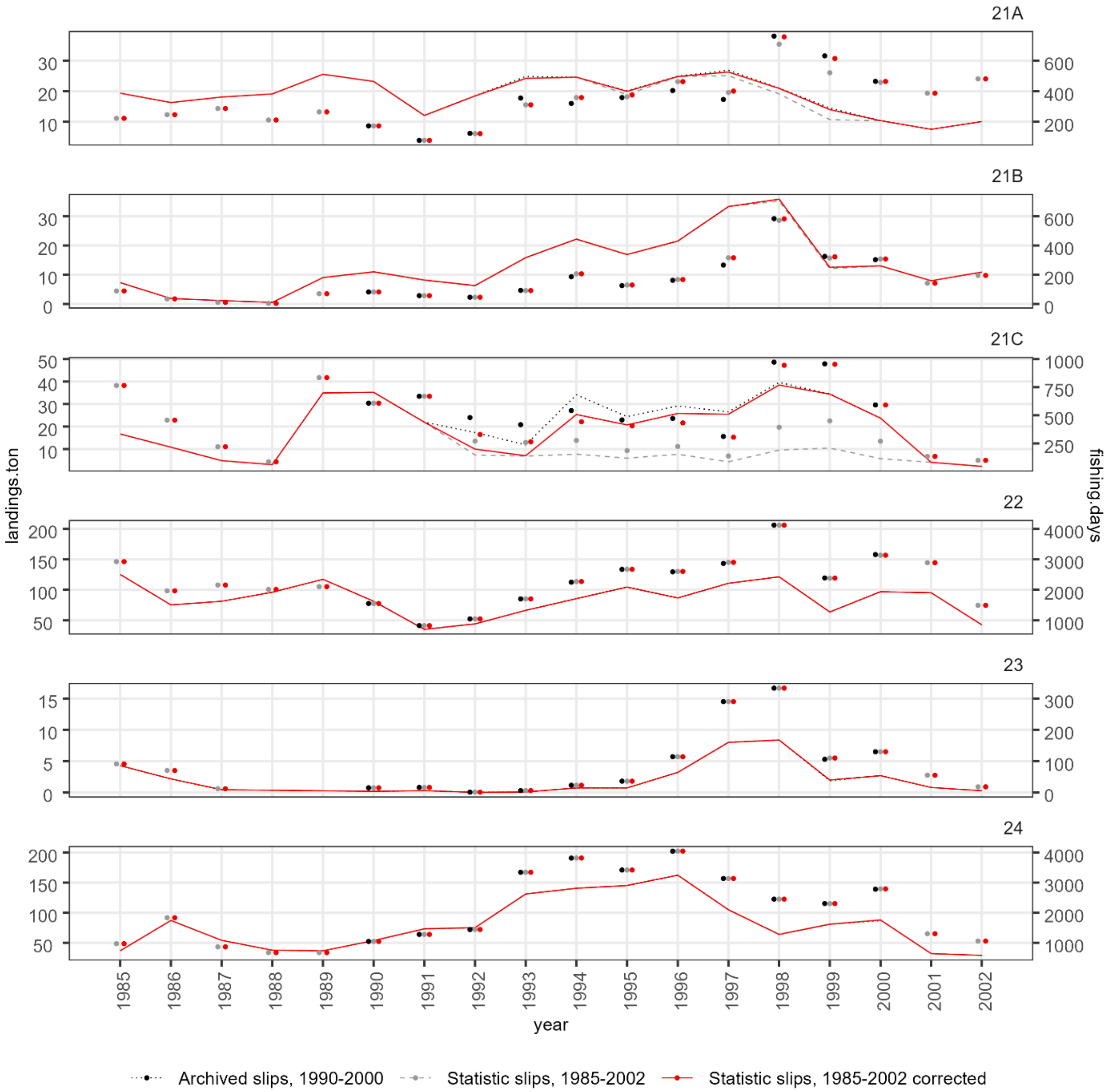

Figure A1.5. Comparison of landings in tons (lines) and number of fishing days (dots) between 3 different purchase slips datasets from 1984 and 2002: archived purchase slips (black dotted), recent requests from DFO Statistic Division (gray dashed), and DFO Statistic Division where we included the missing observations (red full). Detailed by Scallop Fishing Area (SFAs: 21A, 21B, 21C, 22, 23, 24).

01.2.1_slips.R

The objective of the *01.2.1_slips.R* script is to import, format and standardize the raw purchase slip files from 2003 to 2021 obtained from the DFO Statistic Division. The information about landings comes from these slips.

It requires the *ReadSlipLog*, *TblStdze, CleanNames* and *AllLatLon* functions and ***lb_to_kg*** variable along with the raw purchase slip files for years greater of equal to 2003.

Main steps of this script are to:

- Import the files;
- Standardise their column names using the *TblStdze* function;
- Remove duplicated lines, i.e. true duplicated records;
- Assign ***slip.id*** values to the Quebec records:
  - Since the Quebec records are currently missing ***slip.id*** values, the script assigns the ***slip.id*** value of the Gulf records file per ***cfv*** and ***date.caught***. For each record where no ***slip.id*** was found in the Gulf slips, the script creates a new ***slip.id*** value. The new value combines the word "pseudo-", the ***cfv*** and a unique number is assigned to a trip. For example, the unique number is "1" for the first record of a ***cfv*** for a ***date.landed*** and "2" for the second record for that ***cfv*** and the same ***date.landed***;
- Bind the two data frame (Gulf and Québec) together;
- Remove the supplementary B slip records: ***buyer.code*** that are equal to 9000;
- Removes records where ***cfv*** is unknown;
- Convert the landing variables: ***landings.total***, ***landings. local*** and ***landings.com***, from lb to kg using the variable ***lb_to_kg***;
- Recover the small landings weights from the rounding to zero by dividing ***landings.value*** by ***unit.price***;
- Source the *01.2.2_slips_corr.R* script to apply the corrections to slip records, based on QAQC analysis (not shown);
- Separate ***landings.total*** into ***landings.local***: ***buyer.code*** equal to 9997, and ***landings.com***: ***buyer.code*** not equal to 9997;
- The final data file is aggregated by:
  - Identification: cfv, slip.id, summary.id, licence.id,
  - Dates: date.caught, date.landed, date.sailed
  - Location: nafo.subdiv, fish.area.landed, fish.area.licenced, port.home, port.landed

by summing the landing variables: landings.local, landings.com, landings.total.

The ***slips.RData*** files is stored in the ***output_dir*** folder.

01.3.0_logs.R

The objective of the *01.3.0_logs.R* script is to import, format and standardize the commercial fish harvester logbook files from 2003 to 2021 obtained from the DFO Statistics Division. The information about fishing effort comes mainly from these logs.

It requires the *ReadSlipLog* and *TblStdze* functions, ***lb_to_kg*** variable, along with the raw log files.

Main steps of this script are to:

- Import the files;
- Standardise their column names using the *TblStdze* function;
- Assign ***slip.id*** values to the Quebec records (see the previous slips section for explanation);
- Join the Gulf and Quebec region records;

- Convert the weights of ***landings.est*** from pounds to kg *with the **lb_to_kg** variable;*
- Select the variable to be kept in the ***logs*** dataset, this is an important feature for merging this dataset with the ***slips*** dataset;
    - Identification: cfv, slip.id, summary.id, trip.id, licence.id,
    - Dates: date.caught, date.landed, date.sailed
    - Estimated landings: landings.est
    - Effort: drag.no: number of tow, hours.fished, ave.tow: average duration of a tow in minutes
    - Location: nafo.subdiv, fish.area.log, lat, lon
- Then, source two scripts (both of them are in development and will not be presented in this appendix):
    - *01.3.1_logs_corr.R* to apply the corrections based on QAQC analysis, and;
    - *01.3.2_corr_latlon.R* to apply corrections on latitude and longitude on each records, it uses the *MapBedSfa* function to verified the assign SFA to each record based on their latitude and longitude.

The script will then produces the ***logs*** file and stored it in the ***output_dir*** folder.

01.4.0_merge_logs_slips.R

The objective of the *01.4.0_merge_logs_slips.R* script is to combine the slips, ***slips.RData***, with the logs, ***logs.RData***, datasets from 2003 to 2021 into a single dataset, ***full_dat.RData***, and to impute missing data using a random forest algorithm if the variable ***impute_missing_values*** is assigned to *TRUE*.

It requires ***slips.Rdata***, ***logs.RData*** (or associated scripts to produce them) and ***port_list*** files.

Main steps of this script are to:

- Create the ***full_dat*** file by merging ***slips*** and ***logs*** files by all variables, for all rows of both;
    - Before the merge, it adds a variable ***slip*** and its modality 1 and the variable ***log*** and its modality 1 to their respective ***slips*** and ***logs*** files.
    - After the merge, it corrects the modalities for the NAs in the ***slip*** and ***log*** variables, thus if ***slip*** equals 1 and is.na(***log***) then ***log*** equals 0 and vice versa.
- Assign values to ***sfa***:
    - By merging ***full_dat*** with ***port_list*** which contains the ***port.landed*** and ***fish.area.port*** variables by the modalities of the variable ***port.landed***,
    - By filling the ***full_dat$sfa*** variable with ***fish.area.landed*** coming from ***slips*** and then for all ***sfa*** equal to 21 or NA, are filled with ***fish.area.port*** coming from ***port_list***.
    - If there are still NA values, they will be populated with ***fish.area.log*** coming from ***logs***.
- Keep only records that are: slips and logs, or slips only, i.e. records where ***slip*** equals1;
- Impute missing values for ***ave.tow***, ***drag.no*** and ***hours.fished*** logbook variables using the *1.4.1_impute_missing_data.R* script, if the ***impute_missing_values*** variable is *TRUE*.

The final ***full_dat*** files with imputed data is saved in the ***output_dir*** folder.

The Figure A1.6 presents comparison between landings, fishing days and number of boats data found in: Niles et al. (2021), Mallet (2010) and the results from this script. The effort variable related to the number of hour towed is only given as an indication to illustrate the correlation with the number of boats.

01.4.1_impute_missing_data.R

The objective of the *01.4.1_impute_missing_data.R* script is to impute missing data using a random forest algorithm.

It requires the *missRanger* and *caret* libraries and the ***full_dat*** dataset.

Main steps of this script are to:

- Impute missing data for ***ave.tow***, ***drag.no*** and ***hours.fished*** variables. The *missRanger::missRanger* function imputes missing values by conducting a non parametric multivariate imputation using a chained random forest algorithm.
- It creates a file with columns of interest: ***sfa***, ***cfv***, ***year***, ***week***, ***day***, ***landings.total***, ***ave.tow***, ***drag.no***, ***hours.fished***, ***slip.id*** which also contains the effort variables: ***ave.tow***, ***drag.no*** and ***hours.fished***.
- It creates a validation set by duplicating the effort variable columns: ***ave.tow.miss***, ***drag.no.miss***, ***hour.fished.miss***, and adding an extra 10% of NA values to these new duplicated variables.
- It identifies:
  - variables to be imputed; stored in the vector imputee which includes ave.tow.miss, hours.fished.miss, drag.no.miss and;
  - variables that random forest works on to estimate the NA values of imputes. They are stored in the vector imputer which includes sfa, cfv, year, week, day, landings.total, ave.tow.miss, hours.fished.miss, drag.no.miss
- It creates a formula to impute the values with the *missRanger::missRanger* function:

$$\begin{aligned} imp\_formula \leftarrow\ & ave.tow.miss + hours.fished.miss + drag.no.miss \sim \\ & sfa\ +\ cfv\ +\ year\ +\ week\ +\ day\ +\ landings.total\ + \\ & ave.tow.mis\ +\ hours.fished.miss\ +\ drag.no.miss \end{aligned}$$

To describe the attributes used to run the function and why they were chosen, more information on the *missRanger* library algorithm can be found in these references: Carruthers et al. 2011; Li et al. 2015, and cited references and websites mentioned in the script or by typing the command: *help(missRanger)*, in the R console.

- Then, it calculates the error rate using the *caret::postReseample* function and in particular a relationship between the observed and predicted data (RMSE, Rsquared , MAE) is calculated from each imputed variable over the observed variables.

The result without any NA values of each variable is attributed to new variables named as ***ave.tow.imp***, ***drag.no.imp***, ***hour.fished.imp*** in the ***full_dat*** file.

The comparison between raw values and implemented values; and the effect on the number of hours towed in relation to number of NAs are presented at the Figure A1.7.

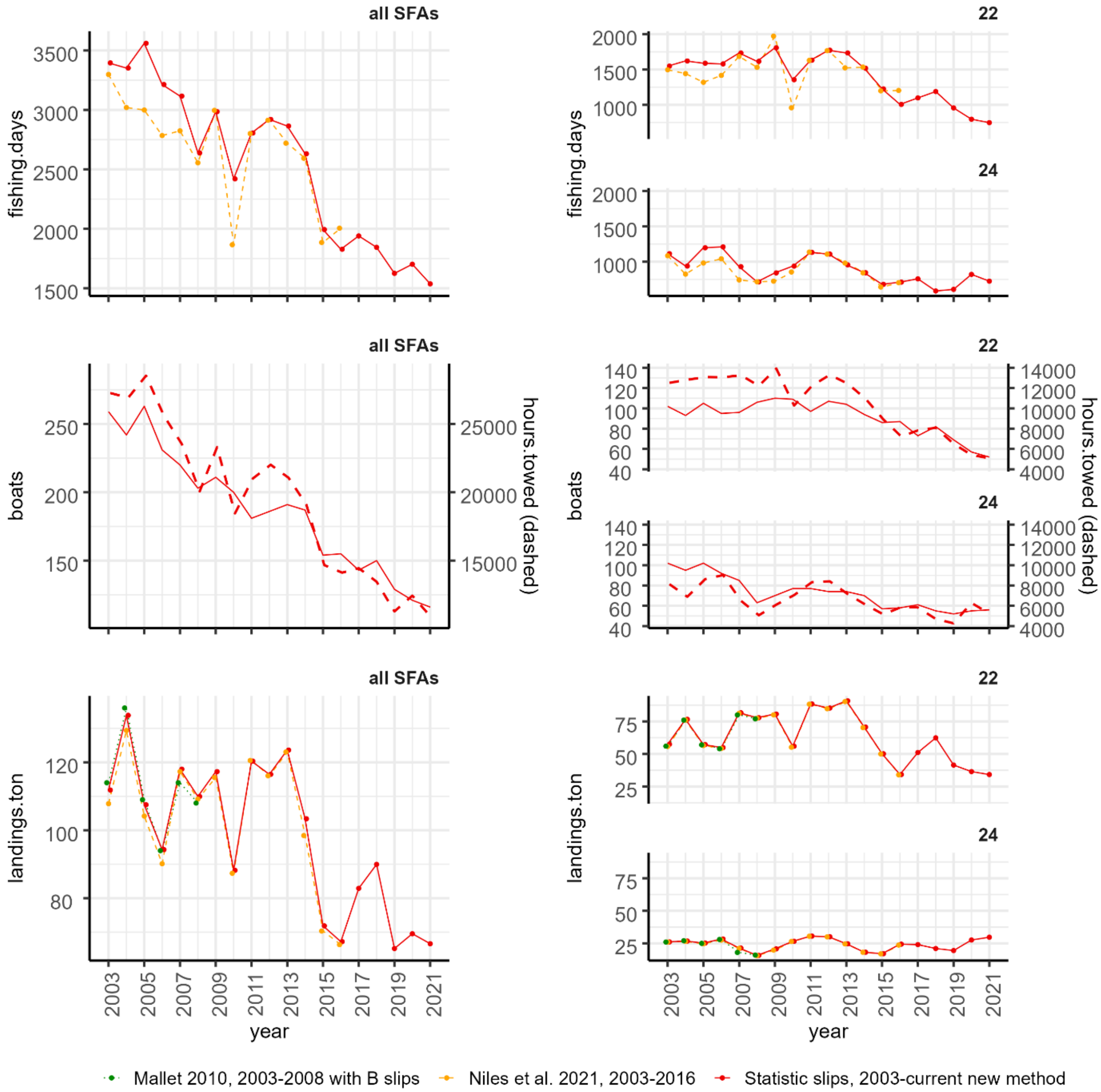

Figure A1.6. Comparison fishing days, number of boats (solid line)/number of hours towed (long dash line) and landings in tons between different historical references (Mallet 2010: green; Niles et al. 2021: orange) and a recent request to the DFO Statistic Division (red) for years between 2003 and 2021. Left panels show SFAs together: 21A, 21B, 21C, 22, 23 and 24. Right panels show SFA 22 and SFA 24.

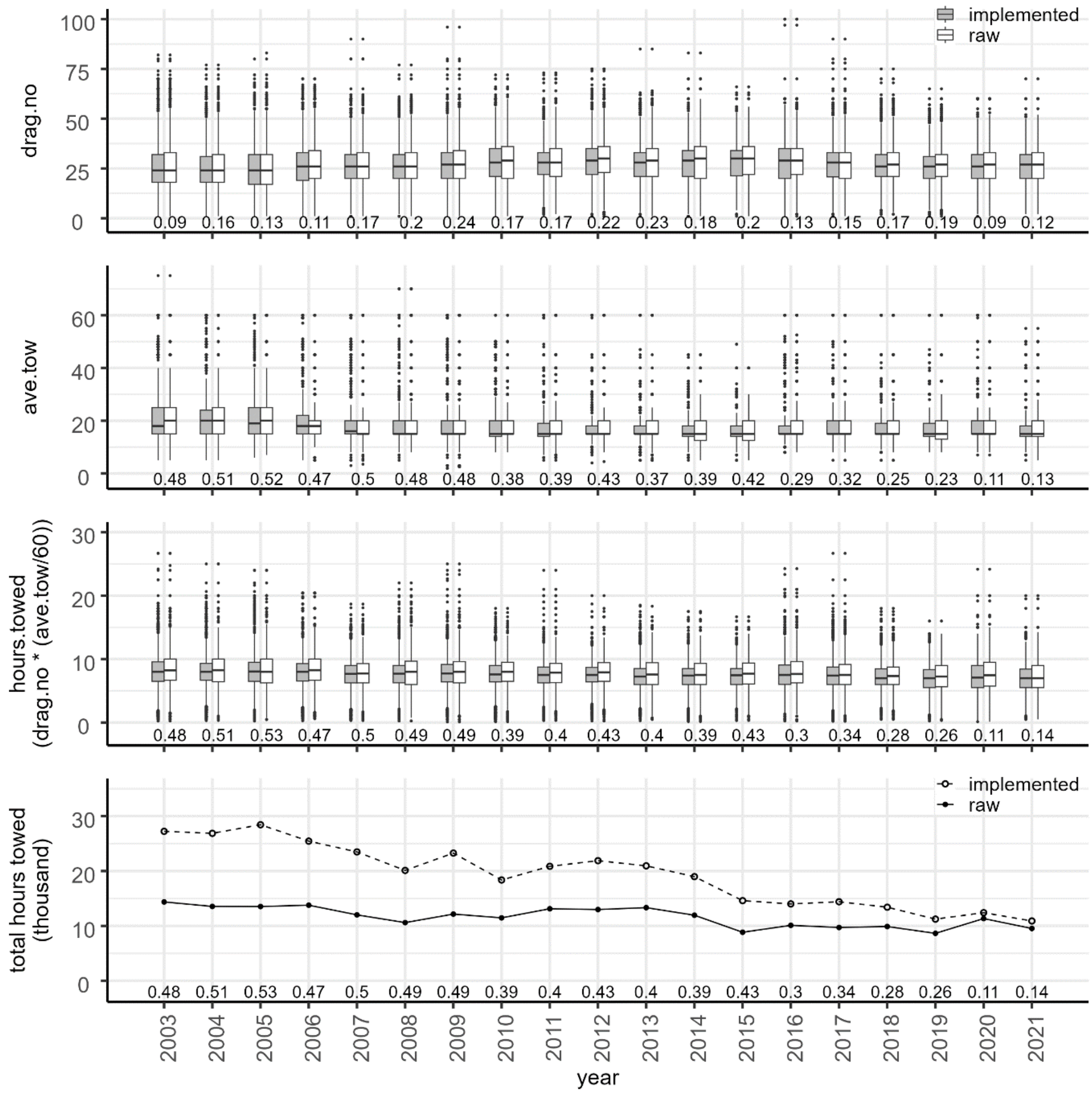

Figure A1.7. Difference between logbook values from DFO Statistics Division (raw, gray boxes and black dots) and logbook values where missing values were implemented (white boxes and dots) for drag numbers, average tow in minutes and hours towed by fishing days; and total hour towed (drag.no(ave.tow/60)) in thousand for the whole Gulf region. Numbers under boxplots and dots stand for proportion of NA values in raw data. Note that fish harvester are not allowed to fish more than 12h a day.*

01.5.0_ fishing_seasons.R

The objective of the ***01.5.0_fishing_seasons.R*** script is to format fishing season variables, but specifically for this work to provide the official annual length of the sGSL fishing season for 2010 to 2021.

It requires the *TblStdze* function. It also imports the season data.

Main steps of this script are to:

- Standardize its column names using the *TblStdze* function;
- Remove the year record when all other variable values are NA;
- Format dates; ***date.season.start***, ***date.season.end***;
- Improve the modalities of the ***weekend.closure*** variable, by adding the following logical class variables:
  - ***season.closed*** is *TRUE* if the whole season is closed for one year in an SFA, otherwise it's *FALSE*
  - ***season.sat.closure*** is *TRUE* if Saturdays are closed in the season, otherwise it's *FALSE*
  - ***season.sun.closure*** is *TRUE* if Sundays are closed in the season, otherwise it's *FALSE*
- Add new variables to store the information about the second season per **year** and **sfa**, if the variable ***notes*** equals *"2nd season"*: ***date.season.start.2***, ***date.season.end.2***.
- Create Julian day variables: ***yday.season.start***, ***yday.season.end***, ***yday.season.start.2 and yday.season.end.2***, to complete missing information for ***yday.season.start***, ***ydays.season.end*** with the mean of every other present value per ***year*** and ***sfa***.
- Complete missing value for logical variables that provide information about closures.
- Find the number of days in a season, ***num.season.days***, by counting the number of days between ***date.season.start*** and ***date.season.end*** for each ***year*** and ***sfa***.
- Removes the number of days closed and,
- Adds the days of the second season if it exists.
- Correct for SFA 21B for the season that have been fixed to a maximum of 42 days in 2013 and to 50 days for the prior years.

The ***seasons.RData*** files are stored in the ***output_dir*** folder. The Figure A1.8 shows the number of allowed fishing days in each season from 2010 to 2021 for each SFA: 21A, 21B, 21C, 22, 23 and 24.

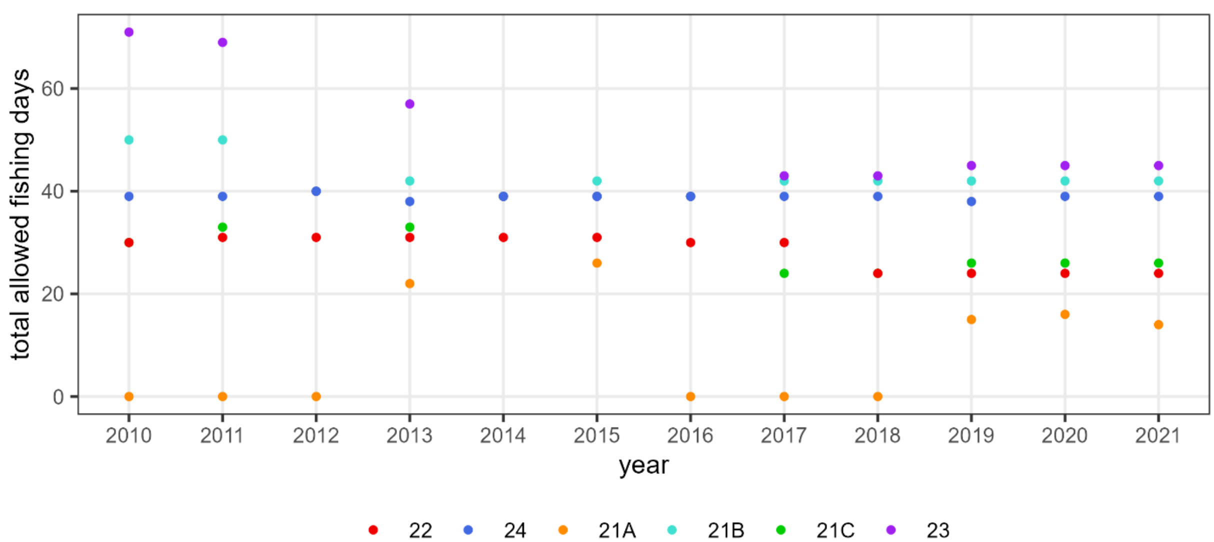

Figure A1.8. Comparison of the number of fishing days allowed per year (2010 to 2021) and SFA.

01.5.1_fleets.R

The objective of the *01.5.1_fleets.R* script is to produce a longer and complete effort time series from 1962 to 2021 using historic active boats data (Fisheries and Marine Service Canada 1977; Jamieson 1979; Jamieson et al. 1981; Worms and Chouinard 1983, 1984; Lanteigne and Davidson 1988, 1991, 1992; DFO 1996; Hanson 1998; Mallet 2010; Davidson et al. 2012).

The year 1962 is the first year we have data on active boats (Fisheries and Marine Service Canada 1977). It uses the current slip dataset ***full_dat*** (2003-2021), the ***licences*** and ***licences_sfa*** datasets and ***slips_historic_raw_list$stat_slips_1984_2002_and_missing***.

Main steps of this script (that is also followed by ***licences_sfa*** (1976:1984) file by keep the ***sfa*** variable) is to:

- Load the ***licences*** file containing the variables:
  - ***year***, ***licences*** (1960:2016)
  - ***licence***, number of licences issued between 1976:1987, 1990, 1995:1997 and 2001:2016
  - ***active.in.questionnaire***, number of active vessels or fish harvesters according to a questionnaire between 1962 and 1978
  - ***active***, number of active vessels as reported by fishery officers between 1976:1990, 1995 and 2001:2016.
- Create the ***slips_yearly_boats*** file, it:
  - Creates a file with the variable ***year*** (1962:1984) and the variable ***boats*** with values equals to NA;
  - For ***stat_slips_1984_2002_and_missing*** (1985:2002) and ***full_dat*** (2003:current_yr) datasets, keep only the ***cfv*** and ***year*** variables and creates the variable ***boats*** by aggregating the number of different vessel numbers: ***cfv*** per ***year*** (between 1962 to 2021), and;
  - Binds the three datasets together.
- Merge slips_yearly_boats with the licences dataset per year.
- Fill the NAs of the variable active which become active.hat, to do that it:

  Calculates the mean ratio active.in.questionnaire per active, for years between 1976 and 1978, fishers_to_active (it equals 2.300904), and if an active.in.questionnaire value is available for a year, then ***active.in.questionnaire*** is divided by ***fishers_to_active***.
- Fill the NAs of the variable ***boats,*** which become ***boats.hat***, with ***active.hat***.

Then the ***licences*** and ***licences_sfa*** files are saved in the ***output_dir*** folder. The Figure A1.9 shows the different variable used to create the variable ***boats.hat***.

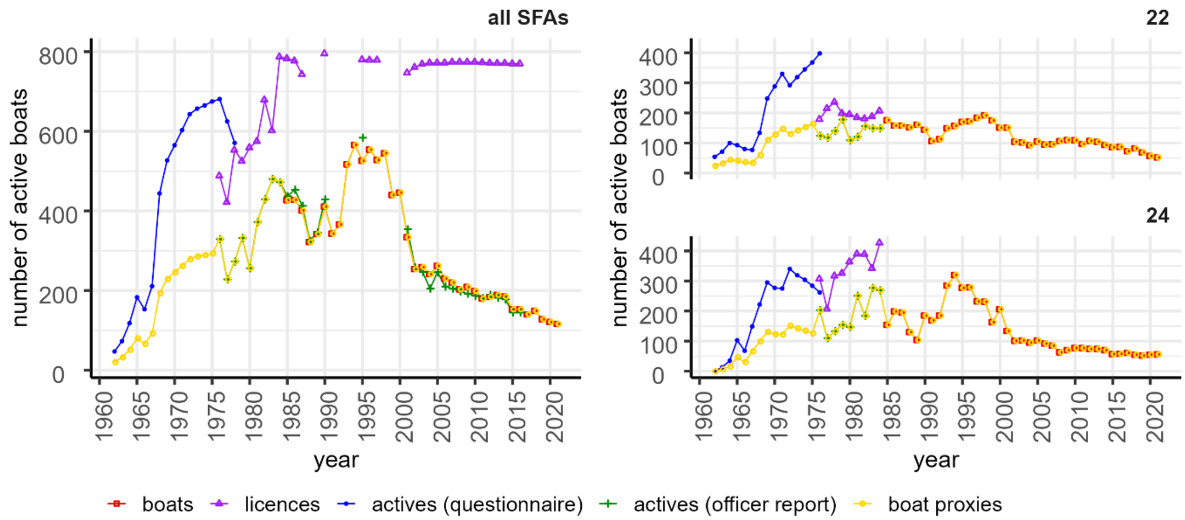

Figure A1.9. Comparison between different raw variables and variables where missing values were estimated relative to number of actively fishing boats per year, from 1962 to 2021. Left panels show SFAs together: 21A, 21B, 21C, 22, 23 and 24. Right panels show SFA 22 and SFA 24.

Data transformation for the purpose of fishery analysis

02.1.0_bind_and_aggregate.R

The objective of the *02.1.0_bind_and_aggregate.R* script is to bind, aggregate per ***year*** (and per ***year*** and ***sfa***) all sGSL scallop datasets and calculate CPUEs (i.e. Catch per Unit Effort for different effort variables).

It uses ***reports_historic_raw_list***, ***slips_historic_raw_list*** lists, ***full_dat***, ***seasons***, ***licences*** *and* ***licences_sfa*** dataset and outputs the ***scallop*** and ***scallop_sfa*** files required for LRP analysis.

Main steps of this script to produce ***scallop***(the following procedure is similar to produce the ***scallop_sfa***) is to:

- Bind *reports_historic*_raw_dat_list, *slips_historic_raw_list* lists of files to form respectively *historic_dat* and *old_slips*
  - For ***historic_dat***, the script:

Selects the modalities “landings” and “fishing.days” from the column named variable;

Removes the records where the modality equals “suppB” for the column named source

Reshapes the file from wide to long to allow each modality of the column ***variable*** to become a distinct column.

  - For ***full_dat***, it calculates the effort variable ***hours.towed***; by the multiplication of the imputed variable ***drag.no.imp*** (number of tow) with ***avg.tow.imp**** 60 (average number of minutes per tow transformed to hours).

- Construct ***scallop_long*** by binding all the three files together and aggregates them by ***year***, ***sfa*** and ***reference*** and then, by ***year*** and ***reference***. The script calculates a sum of effort and landing variables.

The files contains the variables:

- ***landings***, converts landings from kg to ton
- ***fishing.days.2***, .N or number of records for years after 2002
- ***fishing.days***, one ***fishing.day.2*** equals one ***date.caught*** per ***cfv*** for years after 2002 and equals the number of records before or equal to 2002.
- ***boats***, number of unique ***cfv*** per ***year***
- ***hours.towed***, number of towed hours
- ***hours.fished***, number of fished hours

The script removes outliers corresponding to these following notes:

- There is a discrepancy between the recorded landings in 1967 by Lanteigne and Davidson (1991) and those recorded in Lanteigne and Davidson (both 1989 and 1988), and Worms and Chouinard (1984), so we will disregard the records from Lanteigne and Davidson (1991).
- There is probably an error with the number of fishing days published for 1977 since the values here are much lower than what is expected given the catch that year so this will also be removed and the timeseries of effort will start in 1978.
- It the year 1984, many ***cfv*** numbers are missing which results in an obvious underestimate of the number of boats and fishing days. So, we will disregard these variables and will be assigned to NA for this year.

- To construct ***scallop.RData*** (***or scallop_sfa.RData***), it prioritises, in order;
  - data review in long temporal series recorded in historic documents as in the reference of caddy_chandler_1976 (1947 to 1973) and complete with Lanteigne and Davidson (1991) (i.e. years from 1923 to 1989),
  - the raw data files from the actual scallop assessment teams;
  - the raw data files from Statistics Division, and then;
  - the other historical references.
- The script then merges per year the ***scallop*** (or per ***sfa*** and per ***year*** the ***scallop_sfa***) file with the ***licences*** (or ***licences_sfa***) file to paste in the ***boats.hat*** variables.
- It calculate diverse catch per unit effort variables:
  - ***cpue.hours*** equals ***landings*** in kg divided by ***hours.towed***
  - ***cpue.days*** equals ***landings*** in kg divided by ***fishing.days***
  - ***cpue.boats*** equals ***landings*** in tons divided by ***boats***

It saves the resulting ***scallop_sfa_long***, ***scallop_long***, ***scallop_sfa*** and ***scallop*** files in the ***output_dir*** folder. The result of this script is illustrated in the Figure A1.10 for landing and effort variables and in the Figure A1.11 for the CPUE variables.

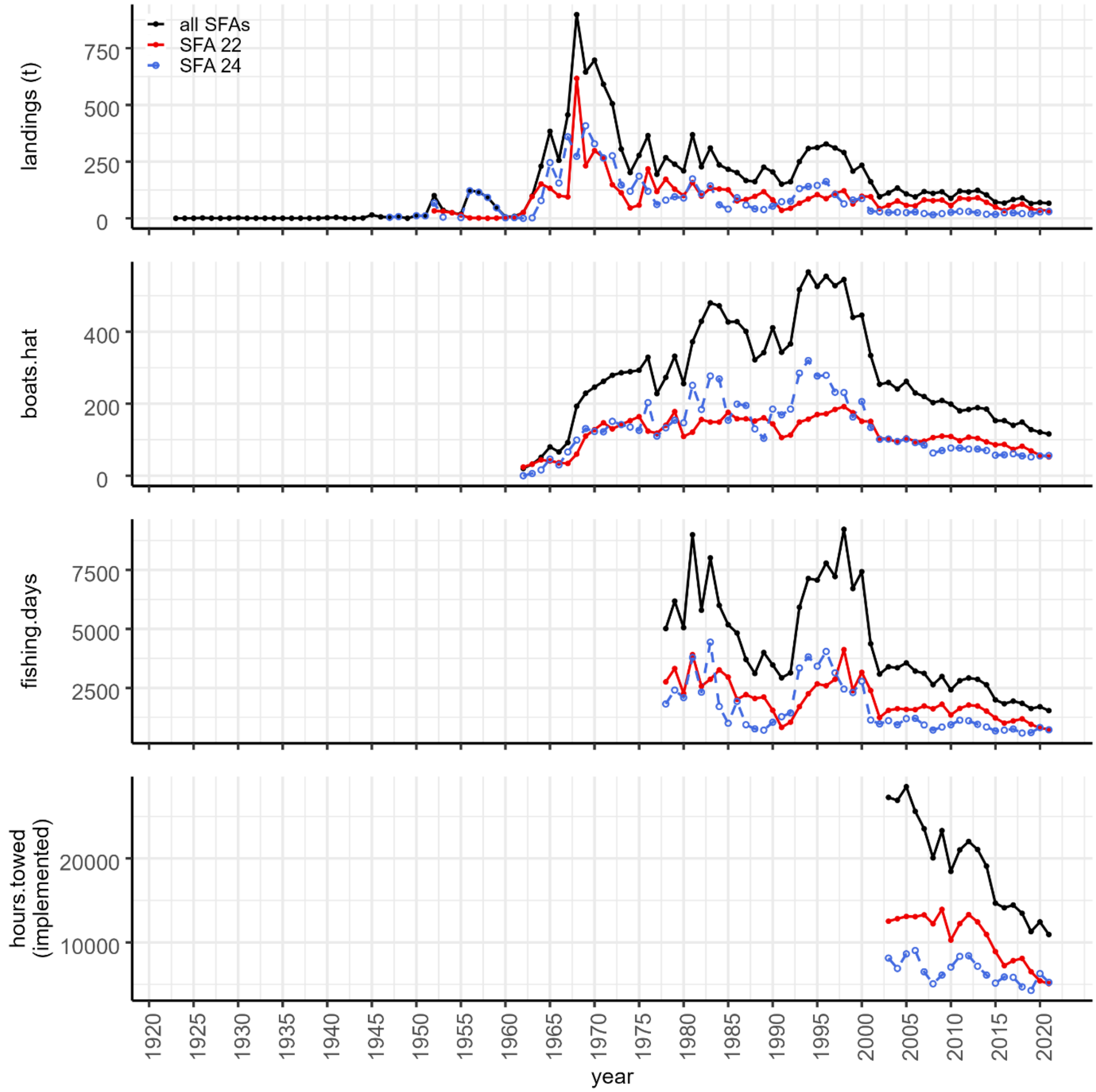

Figure A1.10. Landing and effort variables by year (1923 to 2021) included in the ***scallop_sfa*** *and* ***scallop*** *files. All SFAs: 21A, 21B, 21C, 22, 23, 24 together (black), SFA 22 (red) and SFA 24 (blue).*

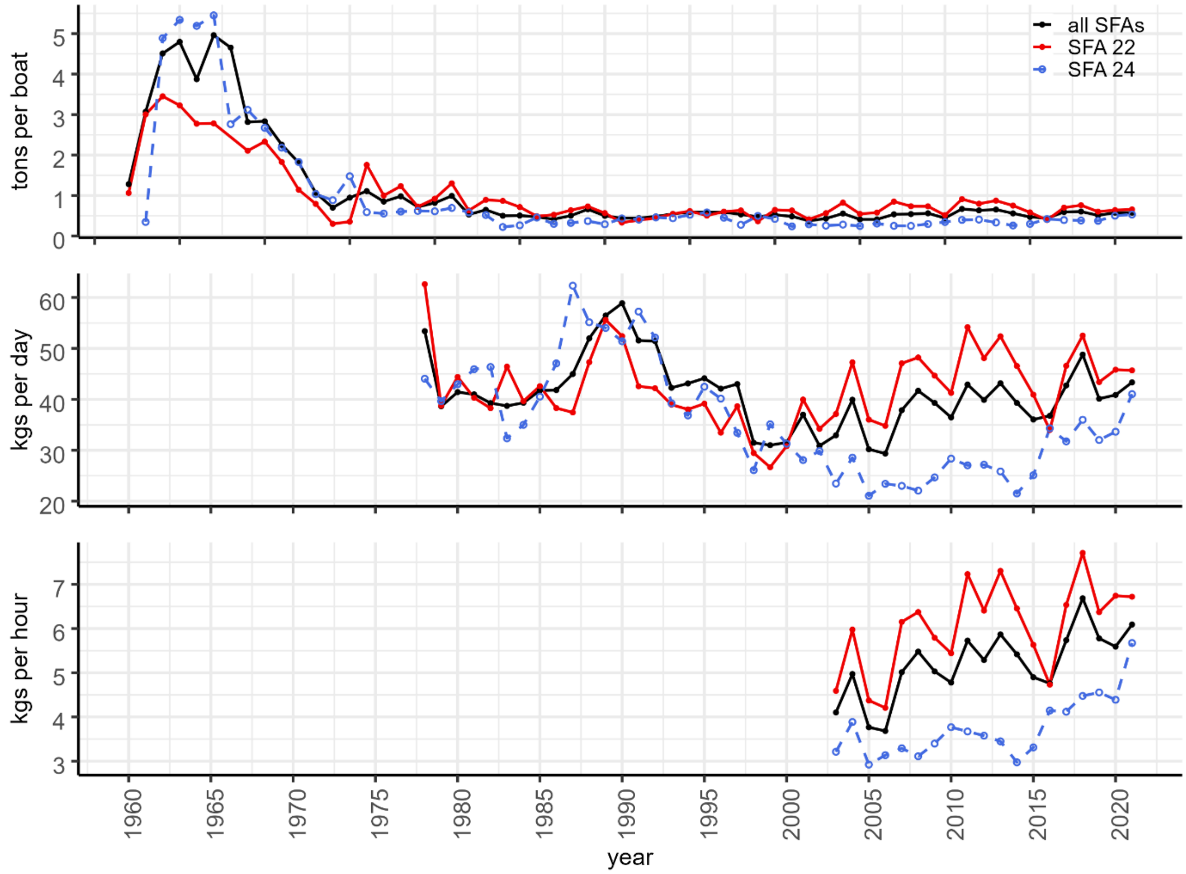

Figure A1.11. CPUE variables by year (1962 to 2021) in the ***scallop_sfa*** *and* ***scallop*** *files. All SFAs: 21A, 21B, 21C, 22, 23, 24 together (black), SFA 22 (red) and SFA 24 (blue).*

Fishery Independent Data: von Bertalanfy parameters estimates

The ***ss_survey_data*** project is used to manipulate all the research survey datasets.

Briefly, for the present research document, the scallop survey datasets from 2001 to 2021 are stored in an Oracle database and schema *PTRAN.glf_groundfish in* the view *v_gscard_type_s_scallop (set)*.

And in datasets from detailed biological analysis of scallops from these surveys stored in the archive data files of the scallop fishery biologists Length at age, shell ht age 2001-2022_MN.csv.

Two scripts allowed to obtain the estimates of von Bernalanffy growth parameters used in the research document *01.1_data_prep_survey.R* and *01.2_growth_params.R.*

01.1_data_prep_surveys.R

The objective of the script *01.1_data_prep_survey.R* is to import and collate the annual research surveys data for more methodological details of the surveys see Niles et al. (2021).

To execute the script, it requires the *gulf* and *lubridate* libraries, the *MapBedSfa* function, and, the scallop research survey database and datafile.

Main steps of this script is to:

- Load data from the Oracle database's set view and the length at age datafile;
- Standardise variables names and modalities;
- Merge the set table to the length at age file per ***year*** and ***set.no***;
- Assign ***sfa.map*** to the survey samples if their latitude and longitude modality fall within the SFA polygon using the *MapBedSfa* function;
    - The ***sfa*** variable in the final file will correspond to the ***sfa.map*** variable and if ***sfa.map*** is equal to NA then the ***sfa*** variable will be equal to the ***sfa.survey*** variable (i.e. SFA found in the length at age file).

The imported and standardised data will then be stored in the ***survey_data_list$surv_age*** in the ***output_dir*** folder.

01.2_growth_params.R

To execute the script, it requires the *FSA* (Ogle et al. 2023), *car* (Fox and Weisberg 2019), and *ggplot2* libraries, and the ***survey_data_list$surv_age*** datafile.

Main steps of this script is to:

- Select data for surv_age where source equals to "survey" for sfa in "21A", "21B", "21C", "22", "23" and "24".
- Estimate starting values for the von Bertalanffy parameters Linf, K and t0
- Select the parametrisation "Typical" (*FSA::vbFuns*) and fit the von Bertalanffy model using nonlinear least squares optimisation:

$$\begin{aligned} scal_{growth} \leftarrow \text{stats} :: nls(formula \;=\; shell.height \sim vbTypical(t \;=\; age,\\ Linf \;=\; Linf,\\ K \;=\; K,\\ t0 \;=\; t0),\\ data \;=\; age\_dat,\\ start \;=\; sv\_typical) \end{aligned}$$

- Calculate the 95% confidence intervals around the parameters:

$$vb\_pars \leftarrow stats::confint(car::Boot(object \;=\; scal\_growth), level \;=\; 0.95, type \;=\; "perc")$$

$$vb\_pars \leftarrow cbind(coef(scal\_growth), vb\_pars)$$

- Construct a table of predicted lengths, where $Linf \,*\, (1 - exp(-K \,*\, (t - t0)))$:

$$vb\_fit \leftarrow data.table(age \;=\; 1{:}max(age\_dat\$age))$$

$$vb\_fit[, lower := vb\_pars[1,2] \,*\, (1 - exp(-vb\_pars[2,2] \,*\, (age - vb\_pars[3,2])))]$$

$$vb\_fit[, fit := vb\_pars[1,1] \,*\, (1 - exp(-vb\_pars[2,1] \,*\, (age - vb\_pars[3,1])))]$$

$$vb\_fit[, upper := vb\_pars[1,3] \,*\, (1 - exp(-vb\_pars[2,3] \,*\, (age - vb\_pars[3,3])))]$$

Then the result is exported and shown at Figure A1.12.

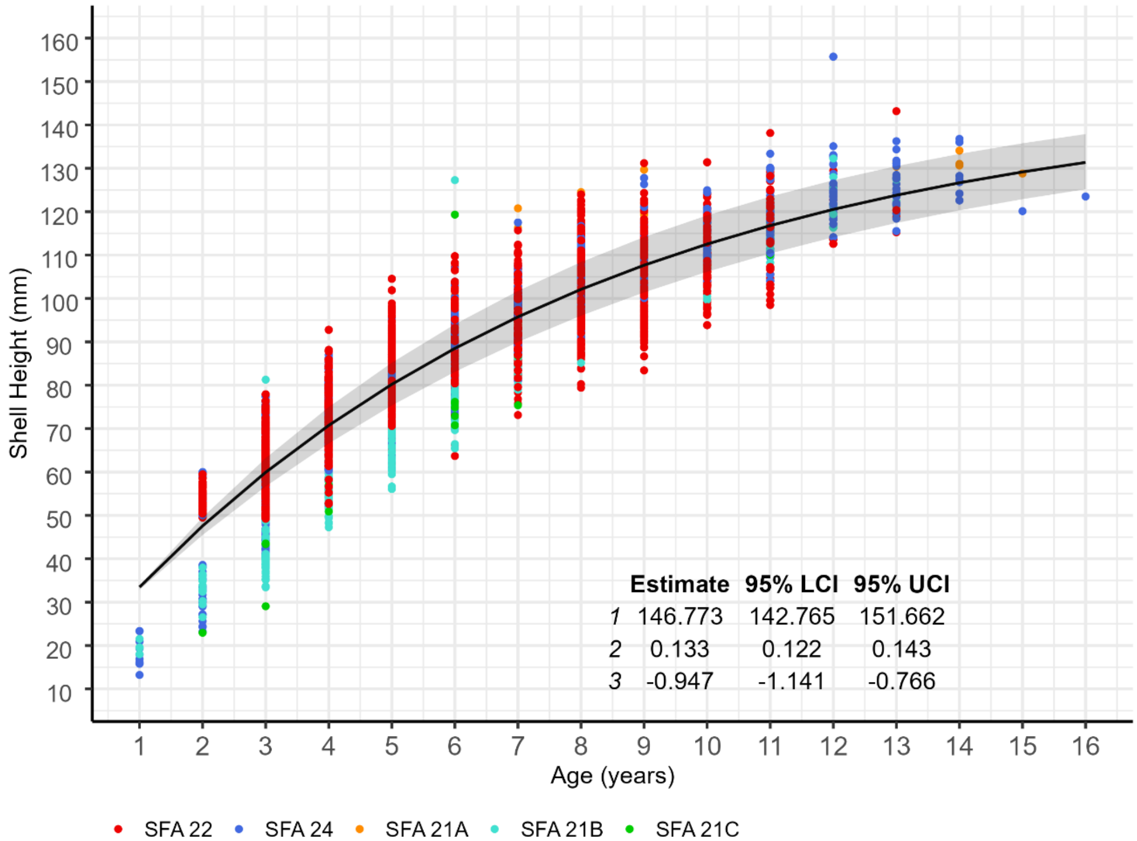

Figure A1.12. Von Bertalanffy growth curve and parameter estimates (Linf, K, t0) and confidence limits for years 2012 to 2021 and all SFAs.

CONCLUSION

The purpose of this work was to gather historical and current sGSL scallop fishery dependent data to create the longest possible time series of effort and landing data. Additionally, we aimed to develop a semi-automated R workflow that could be used to extract, clean, and organize data from databases or recurring sources to enhance understanding of the data source, its constraints, and transformations.

We have successfully extracted the longest possible time series of effort and organized it in our R project. This information is now available for future users and documented in the associated scripts and in this appendix as well as being available through the DFO's open access portal. However, we still need to add to this project some aspect of effort data, such as those related to the type and width of gears and the mesh size. While we have begun applying this approach to fishery-independent data, much of it is still presented only in the form of maps and figures within PDF documents.

REFERENCES CITED

Amaratunga, T., Caddy, J.F., and Stasko, A.B. 1976. Northumberland Strait project: An interdisciplinary study of the declining shellfish resources. ICES CM

Bayer, S., Guenther, C.M., East, P., and Sameoto, J. 2016. Proceedings of the US and Canada Scallop Science Summit: St. Andrews, New Brunswick, May 6–8, 2014. Can. Tech. Rep. Fish. Aquat. Sci. 3151: v + 48 p.

Boudreau, M., and Duplisea, D. 2022. A decision tool for the selection of methods to obtain indicators and reference points for data-limited stocks. Can. Manuscr. Rep. Fish. Aquat. Sci. 3237: vi + 61 p.

Bourne, N. 1965. Annual Report, Sea Scallop Investigations, January 1, 1964 - December 31, 1964. Fish. Res. Board Canada 1007: 1–14.

Caddy, J.F., and Chandler, R.A. 1976. Historical statistics of landings of inshore species in the Maritime Provinces 1947-73. Tech. Report, Fish. Mar. Serv. 639: 1–250.

Cardin, P.J.A. 1924. Fifty-Seventh Annual Report of the Fisheries Branch for the year 1923-24. Dep. Mar. Fish. 57(Sessional Paper No.29 A. 1925): 103 p.

Cardin, P.J.A. 1926. Fifty-Ninth Annual Report of the Fisheries Branch for the year 1925-26. Dep. Mar. Fish. 59: 191 p.

Carruthers, T., Ahrens, R.N.M., McAllister, M.K., and Walters, C.J. 2011. Integrating imputation and standardization of catch rate data in the calculation of relative abundance indices. Fish. Res. 109(1): 157–167.

Chandler, R.A. 1973. Fisheries Fact Sheet: The Sea Scallop. Fish. Mar. Serv. 4: 1–3.

Chiasson, L.P. 1949. Report of scallop investigations and explorations in the southern Gulf of St. Lawrence - 1949. Fish. Res. Board Canada 395: 1–29.

Chiasson, L.P. 1952. Scallop investigations in the Gulf of St. Lawrence and off eastern Cape Breton Island 1951. Fish. Res. Board Canada 449: 1–58.

DAFA. 2007. The Fishing and Aquaculture Industry, Prince Edward Island. PEI Dep. Agric. Fish. Aquac. May: 16 p.

Davidson, L.-A., Biron, M., and Niles, M. 2012. Scallop fishery assessment of the southern Gulf of St. Lawrence in 2010: Commercial fishery data. Can. Sci. Advis. Secr. Res. Doc. 2012/007: vi + 40 p.

Davidson, L.-A., Niles, M., and Légère, L. 2007. Proceedings of the southern Gulf scallop fishery workshop: Moncton, New Brunswick, March 30-31, 2006. Can. Tech. Rep. Fish. Aquat. Sci. 2785: vii + 87 p.

DFO. 1991. A brief history of the Gulf Region. : 14 p.

DFO. 1996. Southern Gulf of St. Lawrence sea scallop. Can. Atl. Fish. Sci. Advis. Committee, Stock Status Rep. 96/104E: 1–3.

DFO. 1999. Atlas of Ecologically and Commercially Important Areas in the Southern Gulf of St. Lawrence. Environ. Stud. Res. Fund: 1–74.

DFO. 2011. Assessment of the scallop fishery (*Placopecten magellanicus*) of the southern Gulf of St. Lawrence. Can. Sci. Advis. Secr. Sci. Advis. Rep. 2011/039: 1–17.

DFO. 2013. American lobster, *Homarus americanus*, stock status in the southern Gulf of St. Lawrence LFA 23, 24, 25, 26A and 26B. Can. Sci. Advis. Secr. Sci. Advis. Rep. 2013/029: 1–21.

DFO. 2018. Annex B: Definitions of the scallop buffer zones. : 5 pp.

DFO. 2020. Lobster Fishing Areas 27 - 38: Integrated Fisheries Management Plan.

DFO. 2021. Science advice for precautionary approach harvest strategies under the fish stocks provisions. Can. Atl. Fish. Sci. Advis. Committee, Sci. Advis. Rep. 2021/004: 1–41.

Dickie, L.M. 1951. Boughton Island, P.E.I., Scallop Investigations, 1946. Fish. Res. Board Canada 415: 1–29.

Doane, W. 2019. clean_names: Clean, Consistent Column Names.

Dowle, M., and Srinivasan, A. 2021. data.table: Extension of `data.frame`. R package version 1.14.2.

Duranleau, A. 1933. Sixty-Sixth Annual Fisheries Report of the Dominion: Third Annual Report of the Department of Fisheries for the year 1932-33. Dep. Fish. 66: 273 pp.

Fisheries and Marine Service Canada. 1977. The Future of the Northumberland Strait Scallop Fishery. Fish. Mar. Serv. Bull. 77–5E: 4 pp.

Fox, J., and Weisberg, S. 2019. An {R} Companion to Applied Regression. Sage, Thousand Oaks {CA}.

Government of Canada Fisheries Act. 1985. Fisheries Act - Loi sur les pêches. R.S.C., 1985, c. F-14, Last Amend. August 28, 2019: ix + 102 p.

Grolemund, G., and Wickham, H. 2011. Dates and times made easy with lubridate. J. Stat. Softw. 40(3): 1–25.

Hanson, J.M. 1998. Survey of sea scallop abundance and distribution in western Northumberland Strait (SFA 22), June 1997. Can. Sci. Advis. Secr. Res. Doc. 98/71: 1–17.

Jamieson, G.S. 1978. Status and assessment of Northumberland Strait scallop stocks. Can. Atl. Fish. Sci. Advis. Commitee, Res. Doc. 78(42): 45.

Jamieson, G.S. 1979. Status and Assessment of Northumberland Strait Scallop Stocks. Fish. Mar. Serv. Tech. Rep. 78/42: 1–45.

Jamieson, G.S., Witherspoon, N.B., and Lundy, M.J. 1980. Assessment of Northumberland Strait scallop stocks - 1979. Can. Atl. Fish. Sci. Advis. Committee, Res. Doc. 80/73: 1–38.

Jamieson, G.S., Witherspoon, N.B., and Lundy, M.J. 1981. Assessment of Northumberland Strait scallop stocks - 1980. Can. Tech. Rep. Fish. Aquat. Sci. 1017: iv + 44 p.

Kuhn, M. 2022. caret: Classification and Regression Training. R package version 6.0-92.

Lanteigne, M., and Davidson, L.-A. 1988. Status of the southern Gulf of St. Lawrence giant scallop (*Placopecten magellanicus*) stock - 1987. Can. Atl. Fish. Sci. Advis. Committee, Res. Doc. 88/55: 1–39.

Lanteigne, M., and Davidson, L.-A. 1989. Évaluation de la pêcherie du pétoncle géant (*Placopecten magelianicus*) dans le sud du golfe du Saint-Laurent - 1988. Can. Sci. Advis. Secr. Res. Doc. 1989/020: 1–22.

Lanteigne, M., and Davidson, L.-A. 1991. Catch and effort statistics for the giant scallop (*Placopecten magellanicus*) fishery in the southern Gulf of St. Lawrence - historical review from 1923 to 1989. Can. Tech. Rep. Fish. Aquat. Sci. 1804: viii + 59 p.

Lanteigne, M., Davidson, L.-A., and Worms, J.M. 1987. Status of the southern Gulf of St. Lawrence scallop stocks - 1986. Can. Atl. Fish. Sci. Advis. Committee, Res. Doc. 87/84: 1–44.

Lanteigne, M., and Davidson, L. 1992. Status of the giant scallop (*Placopecten magellanicus*) fishery in the southern Gulf of St. Lawrence (DFO, Gulf Region) - 1990 update. Can. Manuscr. Rep. Fish. Aquat. Sci. 2148: iv + 21 p.

Li, Z., Ye, Z., Wan, R., and Zhang, C. 2015. Model selection between traditional and popular methods for standardizing catch rates of target species: A case study of Japanese Spanish mackerel in the gillnet fishery. Fish. Res. 161: 312–319.

MacPhail, J. 1954. The Inshore Scallop fishery of the Maritime Provinces. 22(22): 2–4.

Mahto, A. 2019. splitstackshape: Stack and Reshape Datasets After Splitting Concatenated Values. R package version 1.4.8.

Mallet, M. 2010. Commercial scallop (*Placopecten magellanicus*) fishery profile in the Gulf Region. Stat. Econ. Anal. Ser. 1–5(1): v–25.

Niles, M., Barrell, J., Sameoto, J., Keith, D., and Sonier, R. 2021. Scallop fishery assessment of the southern Gulf of St. Lawrence in 2018: Commercial fishery and survey data. Can. Sci. Advis. Secr. Res. Doc. 2021/038: xii + 178 p.

Ogle, D.H., Doll, J.C., Wheeler, A.P., and Dinno, A. 2023. FSA: Simple Fisheries Stock Assessment Methods.

PARL. (n.d.). parl.ns.ca/lobster/history.

Pebesma, E. 2018. Simple Features for R: Standardized Support for Spatial Vector Data. R J. 10(1): 439–446.

Pebesma, E., and Bivand, R. 2023. Spatial Data Science: With Applications in {R}. Chapman and Hall/CRC, Boca Raton.

Pedersen, T.L. 2020. patchwork: The Composer of Plots.

Posit team. 2022. RStudio: Integrated Development Environment for R. Boston, MA.

Quintans, D. 2021. librarian: Install, Update, Load Packages from CRAN, “GitHub”, and “Bioconductor” in One Step.

R Core Team. 2016. R: A language and environment for statistical computing. Vienna, Austria.

Robert, G., and Jamieson, G.S. 1983. Assessment of Northumberland Strait scallop stocks and review, 1978 to 1981. Can. Tech. Rep. Fish. Aquat. Sci. 1150: v + 37 p.

Rowell, T.W., and Lord, E.I. 1965. Gulf of St. Lawrence Scallop Survey - 1965. Fish. Res. Board Canada 842: 1–43.

Smith, M.T., and Addison, J.T. 2003. Methods for stock assessment of crustacean fisheries. Fish. Res. 65(1–3): 231–256.

Smith, S.J., Black, J., Todd, B.J., Kostylev, V.E., and Lundy, M.J. 2009. The impact of commercial fishing on the determination of habitat associations for sea scallops (*Placopecten magellanicus*, Gmelin). ICES J. Mar. Sci. 66: 2043-2051.

Smith, S.J., and Gavaris, S. 1993. Improving the Precision of Abundance Estimates of Eastern Scotian Shelf Atlantic Cod from Bottom Trawl Surveys. North Am. J. Fish. Manag. 13(1): 35–47.

Stekhoven, D.J., and Bühlmann, P. 2012. Missforest--Non-parametric missing value imputation for mixed-type data. Bioinformatics 28(1): 112–118.

Surette, T., Vergara, P., and Ricard, D. 2021. gulf: Access, manipulate, display and analyze southern gulf data.

UN. 2009. Appendix 1: Canada’s legislative and policy framework. : 10–46.

van der Loo, M. 2014. The stringdist package for approximate string matching. R J. 6: 111–122.

Wickham, H. 2016. ggplot2: Elegant graphics for data analysis.

Wickham, H., and Bryan, J. 2022. readxl: Read Excel Files. R package version 1.4.0.

Wickham, H. 2022. rlang: Functions for Base Types and Core R and “Tidyverse” Features. R package version 1.0.4.

Worms, J., Lanteigne, M., and Davidson, L.-A. 1986. Status of the southern Gulf of St. Lawrence scallop stocks - 1985. Can. Atl. Fish. Sci. Advis. Committee, Res. Doc. 86/55: 1–47.

Worms, J.M., and Chouinard, G. 1983. Status of southern Gulf of St. Lawrence scallop stocks - 1982. Can. Atl. Fish. Sci. Advis. Committee, Res. Doc. 83/68: 1–52.

Worms, J.M., and Chouinard, G. 1984. Status of southern Gulf of St. Lawrence scallop stocks - 1983. Can. Atl. Fish. Sci. Advis. Committee, Res. Doc. 84/57: 1–48.

Wright, M.N., and Ziegler, A. 2017. ranger : A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. J. Stat. Softw. 77(1).

APPENDIX 2

Table A2.1. Summary of the models.

Model	Reference	Input	Output	Production Model	Brief Description	Assumption*	Comment	Relevant Example
Scalar Approach	Restrepo et al. 1998	° catch time series ° effort time series ° informed estimate of current B/B_{MSY}	° MSY_{proxy} ° Removal limit	NA	Uses periods in the catch time series with stable effort and landings as a proxy of MSY and applies a scalar to reduce this to a precautious removal level	Assumes that CPUE is linearly related to stock size.	Has limitations, particularly when dealing with variable effort or developing fisheries and may result in overly cautious removal limits.	Weathervane scallop (*Patinopecten caurinus)*, King scallop (*Pecten maximus)*, Softshell clam (*Mya arenaria)*, Common whelk (*Buccinum undatum)* (Miethe et al. 2016; Armstrong et al. 2020; DFO2020, 2022)
DCAC	Alverson and Pereyra 1969; MacCall 2009; Carruthers and Hordyk 2018	° catch time series ° B_{MSY}/B_0 ° F_{MSY}/M ° M ° depletion (1-S)	° Y_{sust}	NA	Provides an estimate of a sustainable yield (Y_{sust}) by incorporating a windfall ratio into the potential-yield equation of Alverson and Pereyra (1969)	Assumes that the stock is maintained near historical abundance levels.	Relies on expert knowledge of approximate depletion, without which the method may be overly cautious.	*Ensis* spp. (Marine Institute and Bord Iascaigh Mhara 2015, 2019)
Catch Ratio	Froese and Kesner-Reyes 2002; Anderson et al. 2012	° catch time series	° Stock status ° Time series of catch ratios	NA	Estimates stock status based on the ratio of catch to maximum historical catch, using relationships established in the RAM Legacy database (RAMLDB, Ricard et al. 2012).	Assumes patterns of fishery development are similar across stocks	Sets threshold C/C_{max} limits based on stock assessments within the RAMLDB wherein invertebrates are underrepresented.	Invertebrates (i.e. Crustaceans, Molluscs, Echinoderms) (Anderson et al. 2011)

Model	Reference	Input	Output	Production Model	Brief Description	Assumption*	Comment	Relevant Example
zBRT	Elith et al. 2008; Zhou et al. 2017; Free 2018	° catch time series	° Relative biomass (B/B_{MSY}) ° saturation (S) trajectories	NA	Predicts the underlying biomass through time based on patterns in the catch time series. Uses a boosted regression tree and the RAMLDB selected stocks as training data.	Assumes observed relationships between reference points are applicable across stocks	Uses the RAMLDB as it's learning dataset, wherein invertebrates are underrepresented.	191 stocks where B/B_{MSY} < 1 or from 135 stocks (10 invertebrates) with various B/B_{MSY} from the RAM Legacy database (Zhou et al. 2017; Free et al. 2020)
OCOM	Graham 1935; Schaefer 1954 ; Free 2018; Zhou et al. 2018	° catch time series ° M	° biomass (B) ° fishing mortality (F) trajectories ° biological and management quantities (r, K, MSY, B_{MSY}, F_{MSY})	Graham-Schaefer	A stock reduction analysis which uses a combination of user provided priors and priors estimated from fitting a zBRT model.	° estimates r from M by assuming r = $2F_{MSY}$ and using the relationship between F_{MSY} and M from teleost and chondrichthyans (F_{MSY}/M < 0) ° MSY = rK/4	Uses assumptions based on the RAMLDB which contains few invertebrates to calculate r and K.	135 stocks (10 invertebrates) with various B/B_{MSY} from the RAMLDB (Free et al. 2020)
DB-SRA	Pella and Tomlinson 1969; Fletcher 1978; Dick and MacCall 2011; Nelson 2023	° catch time series (from beginning of fishery) ° age at maturity ° K, ° B_0/K, ° S, ° F_{MSY}/M, ° B_{MSY}/K, ° M	° biomass (b) and catch ($\hat{C}$) trajectories ° biological and management quantities (K, MSY, B_{MSY}, F_{MSY})	Pella-Tomlinson-Fletcher, delay difference	Uses Monte Carlo simulations to generate sets of plausible b and $\hat{C}$ trajectories from starting points sampled from the supplied parameters.	Assumes that the time series of catch data starts at the conception of the commercial fishery.	Result in B_{MSY}/K is unusually low, this type of production model may not be well suited to invertebrate bivalves like scallops.	Lake sturgeon (*Acipenser fulvescens*) (Sweka et al. 2018)

Model	Reference	Input	Output	Production Model	Brief Description	Assumption*	Comment	Relevant Example
COM-SIR	Vasconcello and Cochrane in Kruse et al. (2005); Anderson et al. 2016	° catch time series ° range of K ° range of r ° effort parameters: x and bioeconomic equilibrium a	° biomass (b) and catch ($\hat{C}$) trajectories ° biological quantities (K, r)	Schaefer, logistic harvest dynamics	Uses a Bayesian method (MCMC sampling with a sampling-importance-resampling algorithm) coupled with a harvest-biomass dynamics model to predict catches.	° Fishing effort follows a logistic growth pattern in response to stock abundance ° q is constant over time ° a bioeconomic equilibrium exists which is constant over time	Estimates of B_{MSY}, MSY, and F_{MSY} are not directly provided, but can be calculated (see Table 8.1 in Hilborn and Walters 2003)	135 stocks (10 invertebrates) with various B/B_{MSY} (Free et al. 2020)
CMSY++	Martell & Froese 2013; Froese et al. 2017	° catch time series ° range of r ° Initial, intermediate, and final range of S	° biomass (b) and fishing mortality (f) trajectories ° biological and management quantities (K, r, MSY, B_{MSY}, F_{MSY})	Schaefer	A Monte Carlo stock reduction analysis for estimating fisheries reference points from catch, resilience and qualitative stock status.	Estimates K based on assumptions about the relationship between K, F_{MSY}, and r.	Includes an improvement over the original for adjusting r when populations are highly depleted.	84 fish and invertebrate data-limited stocks (Schijns 2022)
BSM	Froese et al. 2017	° cpue time series ° catch time series ° range of r ° S at the start, intermediate and last year of the fishery	° biomass (b) and fishing mortality (f) trajectories ° biological and management quantities (K, r, MSY, B_{MSY}, F_{MSY})	Schaefer	A Bayesian state-space implementation of the Schaefer model (Schaefer 1954), implemented by CMSY when catch-per-unit-effort data are available.	See above	Provides a means of comparing CMSY output to full Bayesian Schaefer production model using identical inputs and priors	12 stocks (invertebrates and fish), 5 stocks (4 fish and *Pecten jacobaeus*) (Wang et al. 2020; Armelloni et al. 2021)

Model	Reference	Input	Output	Production Model	Brief Description	Assumption*	Comment	Relevant Example
SPiCT	Pedersen and Berg 2017; Mildenberger et al. 2022	° cpue time series ° catch time series	° biomass (b) and fishing mortality (f) trajectories ° biological and management quantities (K, r, MSY, B_{MSY}, F_{MSY})	Pella-Tomlinson, Schaefer	A Pella-Tomlinson surplus production model in continuous time which addresses natural system uncertainties through a random walk process.	Assumes the stock is closed and that age and size distribution are stable, with a constant catchability of gear used to produce CPUE index	Accepts a multitude of additional starting parameter guesses and prior distributions	Brown crab (*Cancer pagurus*) (Marcussen 2022)
JABBA	Winker et al. 2023	° catch time series (one or more) ° cpue time series (one or more)	° biomass (b) and fishing mortality (f) trajectories ° biological and management quantities (K, r, MSY, B_{MSY}, F_{MSY})	Schaefer, Fox, or Pella-Tomlinson	A Bayesian state-space surplus production models with the ability to accommodate catch and CPUE time series from multiple fleets and considers process and observation errors for each.	Assumes catch observations are error free.	Accepts priors for: K,r, $S_{initial}$, S_{final}, P-T shape parameter	Sea Scallop (*Placopecten magellanicus*), Kona Crab (*Ranina ranina*) (Kapur et al. 2019; DFO 2023)
ASPIC	Prager 1994	° catch time series ° cpue time series (one or more) ° K, ° MSY ° $S_{initial}$	° biomass (b) and fishing mortality (f) trajectories ° biological and management quantities (K, r, MSY, B_{MSY}, F_{MSY})	Schaefer, Fox, Pella-Tomlinson	A continuous time non-equilibrium surplus production model.	Assumes there no process error, provides a deterministic outcome.	Requires little input from user, but output is not easily accessible.	Longfin inshore squid (*Loligo pealeii*); Redfish (Sebastes spp.), Atlantic blue marlin (Makaira nigricans) (Goodyear and Prager 2001; Moustahfid et al. 2009; Ávila de Melo et al. 2014)

* all assume catch records from a fishery are representative of stock abundance to some extent, *** exceptions

REFERENCES CITED

Alverson, D., and Pereyra, W. 1969. Demersal fish explorations in the northeastern Pacitic Ocean – an evaluation of exploratory fishing methods and analytical approaches to stock size and yield forecasts. Journal of the Fisheries Research Board of Canada, 26: 1985 – 2001.

Anderson, S.C., Afflerback, J., Cooper, A.B., Dickey-Collas, M., Jensen, O.P., Kleisner, K.M., Longo, C., Chato Osio, G., Ovando, D., Minte- Vera, C., Minto, C., Mosqueira, I., Rosenberg, A.A., Selig, E.R., Thorson, J.T., and Walsh, J.C. 2016. datalimited: Stock Assessment Methods for Data-limited Fisheries. R package version 0.1.0.

Anderson, S.C., Flemming, J.M., Watson, R., and Lotze, H.K. 2011. Rapid global expansion of invertebrate fisheries: Trends, drivers, and ecosystem effects. PLoS One 6(3): 1–9.

Anderson, S.C., Branch, T.A., Ricard, D., and Lotze, H.K. 2012. Assessing global marine fishery status with a revised dynamic catch-based method and stock-assessment reference points. ICES J. Mar. Sci. 69(8): 1491–1500.

Armelloni, E.N., Scanu, M., Masnadi, F., Coro, G., Angelini, S., and Scarcella, G. 2021. Data Poor Approach for the Assessment of the Main Target Species of Rapido Trawl Fishery in Adriatic Sea. Front. Mar. Sci. 8(June): 1–11.

Armstrong, J., Beder, A., Burt, R., Byerly, M., Henry, K., Jackson, T., Miller, S., Olson, A., Olson, J., Richardson, N., Rumble, J., Russ, E., William, B., and Zheng, J. 2020. Stock assessment and fishery evaluation report for the scallop fishery off Alaska. North Pacific Fish. Manag. Counc.: 1–102.

Ávila de Melo, A.M., Brites, N., Alpoim, R., and Troncoso, D.G. 2014. An ASPIC Based Assessment of Redfish (*S. mentella and S. fasciatus*) in NAFO Divisions 3LN (assuming that the highest apparently sustained historical average level of catch is a sound proxy to MSY). 1999(June): 1–14.

Carruthers, T.R., and Hordyk, A.R. 2018. The Data-Limited Methods Toolkit (DLMtool): An R package for informing management of data-limited populations. Methods Ecol. Evol. 9(12): 2388–2395.

DFO. 2020. Assessment of softshell clam stocks in Québec coastal waters. Can. Sci. Advis. Secr. Sci. Advis. Rep. 2020/032(June): 1–14.

DFO. 2022. Assessment of 4Vs And 4W offshore whelk (*Buccinum undatum*) stock status. Can. Sci. Advis. Secr. Sci. Advis. Rep. 2022/043: 1–14.

DFO. 2023. Scallop stocks assessment in Quebec coastal waters in 2022. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2023/034.

Dick, E.J., and MacCall, A.D. 2011. Depletion-Based Stock Reduction Analysis: A catch-based method for determining sustainable yields for data-poor fish stocks. Fish. Res. 110(2): 331–341.

Elith, J., Leathwick, J.R., and Hastie, T. 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77(4): 802–813.

Fletcher, R.I. 1978. On the restructuring of the Pella-Tomlinson system. Fish Bull. 76(3): 515–521.

Free, C.M. 2018. datalimited2: More stock assessment methods for data-limited fisheries.

Free, C.M., Jensen, O.P., Anderson, S.C., Gutierrez, N.L., Kleisner, K.M., Longo, C., Minto, C., Osio, G.C., and Walsh, J.C. 2020. Blood from a stone: Performance of catch-only methods in estimating stock biomass status. Fish. Res. 223: 105452.

Froese, R., Demirel, N., Coro, G., Kleisner, K.M., and Winker, H. 2017. Estimating fisheries reference points from catch and resilience. Fish Fish. 18(3): 506–526.

Froese, R., and Kesner-Reyes, K. 2002. Impact of fishing on the abundance of marine species. ICES Counc. Meet. Rep. C.: 1–12.

Goodyear, C.P., and Prager, M.H. 2001. Fitting Surplus-Production Models With Missing Catch Data Using Aspic: Evaluation With Simulated Data on Atlantic Blue Marlin. Col. Vol. Sci. Pap. ICCAT 53(January 2001): 146–163.

Graham, M. 1935. Modern theory of exploiting a fishery, and application to North Sea trawling. ICES J. Mar. Sci. 10(3): 264–274.

Hilborn, R., & Walters, C. J. (2003). Quantitative Fisheries Stock Assessment, Choice, Dynamics, And Uncertainty (Vol. 2003).

Kapur, M.R., Fitchett, M.D., Yau, A.J., and Carvalho, F. 2019. 2018 Benchmark Stock Assessment of Main Hawaiian Islands Kona Crab. NOAA Tech. Memo. NMFS-PIFSC-77: 1–9.

Kruse, G.H., Gallucci, V.F., Hay, D.E., Perry, R.I., Peterman, R.M., Shirley, T.C., Spencer, P.D., Wilson, B., and Woodby, D. 2005. Fisheries Assessment and Management in Data-Limited Situations. Alaska Sea Grant Coll. Progr. Univ. Alaska Fairbanks.: 958 pp.

MacCall, A.D. 2009. Depletion-corrected average catch: a simple formula for estimating sustainable yields in data-poor situations. ICES J. Mar. Sci. 66(10): 2267–2271.

Marcussen, J.B. 2022. Fishing for more data: Exploratory stock assessment of the data-limited brown crab (Cancer pagurus) stock in Norway. Thesis of the University of Agder. : 75 pp.

Marine Institute, and Bord Iascaigh Mhara. 2019. Shellfish Stocks and Fisheries Review 2019: An assessment of selected stocks. Irish Sea Fish. Board: 200 p.

Marine Institute, and Bord Iascaigh Mhara. 2015. Shellfish Stocks and Fisheries. Review 2015: An assessment of selected stocks. Irish Sea Fish. Board: 63 pp.

Martell, S., & Froese, R. (2013). A simple method for estimating <scp>MSY</scp> from catch and resilience. Fish and Fisheries, 14(4), 504–514.

Miethe, T., Dobby, H., and McLay, A. 2016. The Use of Indicators for Shellfish Stocks and Fisheries: a Literature Review. Scottish Marine and Freshwater Science. Scottish Mar. Freshw. Sci. 7 (16)(16): 78.

Mildenberger, T.K., Kokkalis, A., and Berg, C.W. 2022. Guidelines for the stochastic production model in continuous time (SPiCT). : 6–10.

Moustahfid, H., Tyrrell, M.C., and Link, J.S. 2009. Accounting Explicitly for Predation Mortality in Surplus Production Models: An Application to Longfin Inshore Squid. North Am. J. Fish. Manag. 29(6): 1555–1566.

Nelson, G. 2023. fishmethods: Fishery Science Methods and Models.

Pedersen, M.W., and Berg, C.W. 2017. A stochastic surplus production model in continuous time. Fish Fish. 18(2): 226–243.

Pella, J.J., and Tomlinson, P.K. 1969. A generalized stock production model. Inter-American Trop. Tuna Comm. Bull. 13(3): 416–497.

Prager, M.H. 1994. A suite of extensions to a nonequilibrium surplus-production model. Fish. Bull. 92(2): 374–389.

Restrepo, V.R., Thompson, G.G., Mace, P.M., Gabriel, W.L., Low, L.L., MacCall, A.D., Methot, R.D., Powers, J.E., Taylor, B.L., Wade, P.R., and Witzig, J.F. 1998. Technical guidance on the use of precautionary approaches to implementing National Standard 1 of the Magnuson-Stevens Fishery Conservation and Management Act. NOAA Tech. Memo. NMFS–F/SPO: 56 p.

Ricard, D., Minto, C., Jensen, O.P., and Baum, J.K. 2012. Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. Fish Fish. 13(4): 380–398.

Schaefer, M.B. 1954. Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. Bull. Inter-American Trop. Tuna Comm. 1(1–2): 27–56.

Schijns, R. 2022. A fuller picture of the state of Canada's Fisheries: Assessments for data-limited stocks. Oceana: 1–49.

Sweka, J.A., Neuenhoff, R., Withers, J., and Davis, L. 2018. Application of a Depletion-Based Stock Reduction Analysis (DB-SRA) to lake sturgeon in Lake Erie. J. Great Lakes Res. 44(2): 311–318.

Wang, Y., Liang, C., Wang, Y., Xian, W., and Palomares, M.L. 2020. Stock Status Assessments for 12 Exploited Fishery Species in the Tsushima Warm Current Region, Southwest Japan and East China, Using the CMSY and BSM Methods. Front. Mar. Sci. 7(640).

Winker, H., Carvalho, F., and Kapur, M. 2023. JABBA: Just Another Bayesian Biomass Assessment. Fish. Res. 204: 275–288.

Zhou, S., Punt, A.E., Ye, Y., Ellis, N., Dichmont, C.M., Haddon, M., Smith, D.C., and Smith, A.D.M. 2017. Estimating stock depletion level from patterns of catch history. Fish Fish. 18(4): 742–751.

Zhou, S., Punt, A.E., Smith, A.D.M., Ye, Y., Haddon, M., DIchmont, C.M., and Smith, D.C. 2018. An optimized catch-only assessment method for data poor fisheries. ICES J. Mar. Sci. 75(3): 964–976.

APPENDIX 3

OCOM MODEL FIT ASSESSMENT

The optimized catch only method relies on an estimate of the natural mortality rate (M) as well as a time series of catches as its input values. To evaluate the model's sensitivity to varying M inputs, iterative runs were conducted with the same catch tame series while varying this parameter. The outcomes reveal that an increase in M led to the carrying capacity (K), biomass at maximum sustainable yield (B_{MSY}), and biomass in the most recent year (B_{2021}), while decreasing estimates of MSY, the intrinsic rate of population growth (r), and fishing mortality at MSY (F_{MSY}). Notably, the parameter estimates of r and F_{MSY} exhibited considerable variations, increasing by 123% and 133%, respectively, between the lowest and highest estimates of M (Figure A3.1).

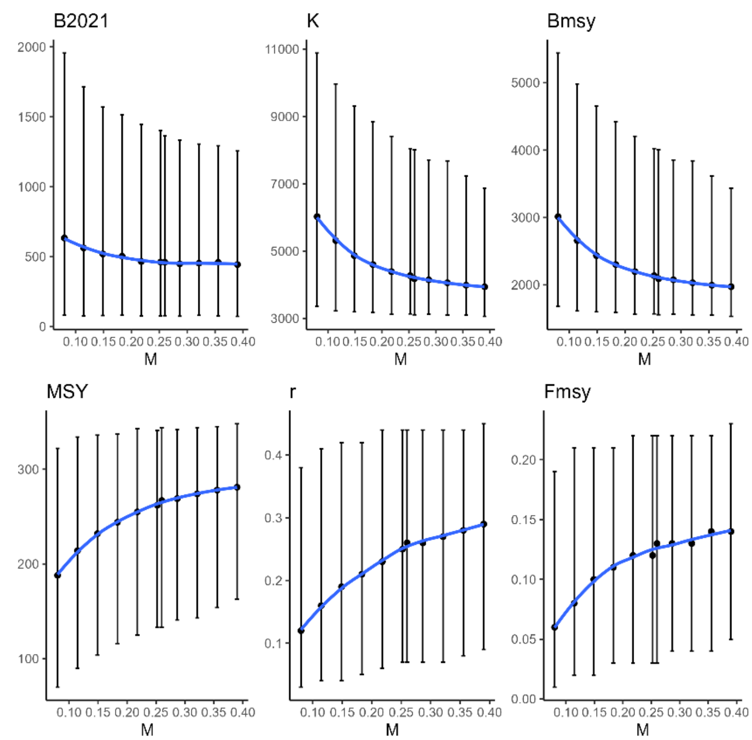

Figure A3.1. Parameter responses to a range of natural mortality rate (M) inputs using the OCOM method of Zhou et al. 2018 applied to sea scallop landings in the core area of the sGSL.

The model's evaluation of parameter estimates aligns reasonably well with our understanding of the system. Notably, the estimated range of the K for the core habitat, based on available knowledge, spans from 1000 to 6000. Interestingly, the OCOM model, without predefined input regarding the acceptable range of K values, generated a skewed distribution of K estimates. The 95% distribution fell between 3120 and 8100, as illustrated in Figure A3.2.

Similarly, the estimated range of population saturation in the final year (S) falls comfortably within the anticipated range of 0.026 – 0.20. In contrast, the model's selection of r values appears lower than expected, exhibiting a mean of 0.25. This discrepancy may be attributed to the model's utilization of the F_{MSY}/M relationship to estimate r, after making the assumption that r = $2F_{MSY}$ and that F_{MSY}/M is < 0 – an assumption based on observations in teleosts and chondrichthyans.

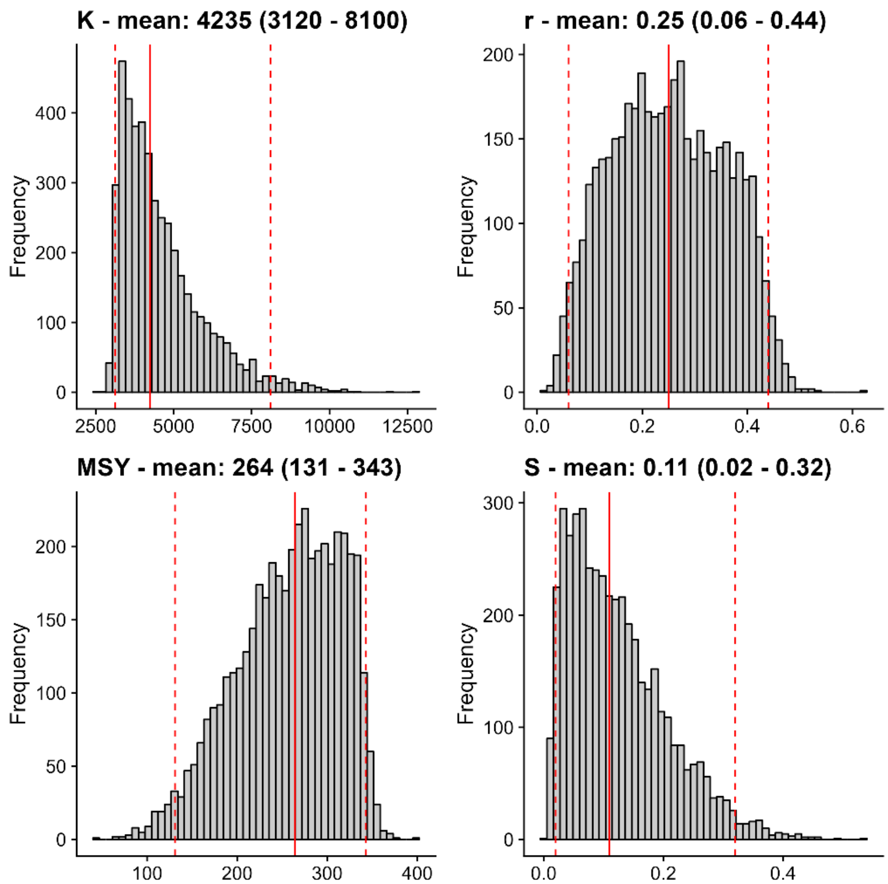

Figure A3.2. Parameter estimate distributions from an OCOM model fit to the core area landings of sea scallops from the sGSL using an estimate of natural mortality (M) of 0.26. Solid red lines represent median values while dashed lines represent the 95% parameter distribution.

Predicted catch can be calculated from the model output for the timeseries of biomass (b) and exploitation rate (F) estimates. The resulting modelled catch rates follow the true catch rates extremely closely, with a mean residual value very near to zero (Figure A3.3). The model consistently overestimated the catch, however, resulting in all residuals being negative.

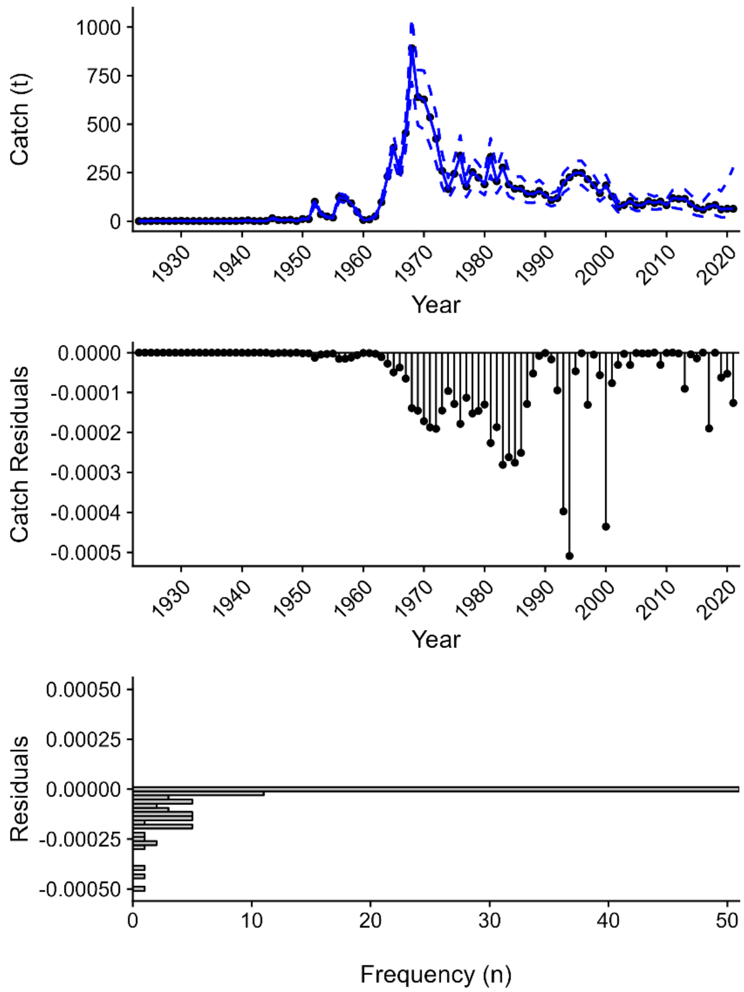

Figure A3.3. Model residual plots for predicted catch values from an OCOM model fit to the catch data for sea scallop in the core area of the sGSL. Predicted catch was calculated as the product of predicted biomass and exploitation estimates. Residuals are calculated as catch minus predicted catch.

DB-SRA MODEL FIT ASSESSMENT

From the prior and posterior parameter estimate distributions produced by the DB-SRA model applied to the core area sea scallop landings in the sGSL we can see that the wide range of MCMC starting parameters (10000 simulations, Figure A3.4) produced only a small number of combinations resulting in an acceptable biomass trajectory (Figure A3.5). All successful runs used K estimates near the upper limit of the estimated prior range.

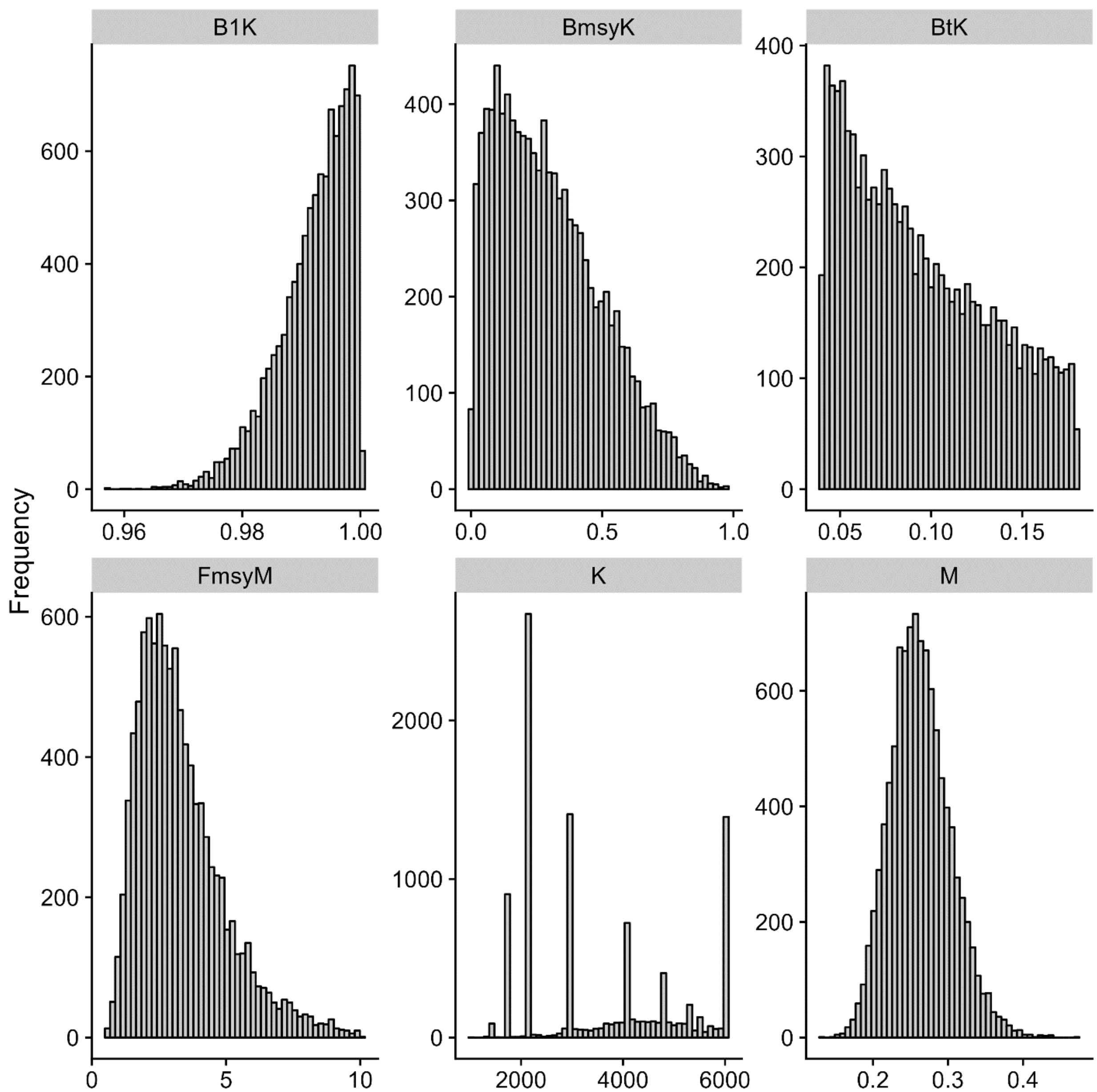

Figure A3.4. Prior parameter distributions for MCMC sampling process in a DB-SRA model of sea scallops in the core area of the sGSL.

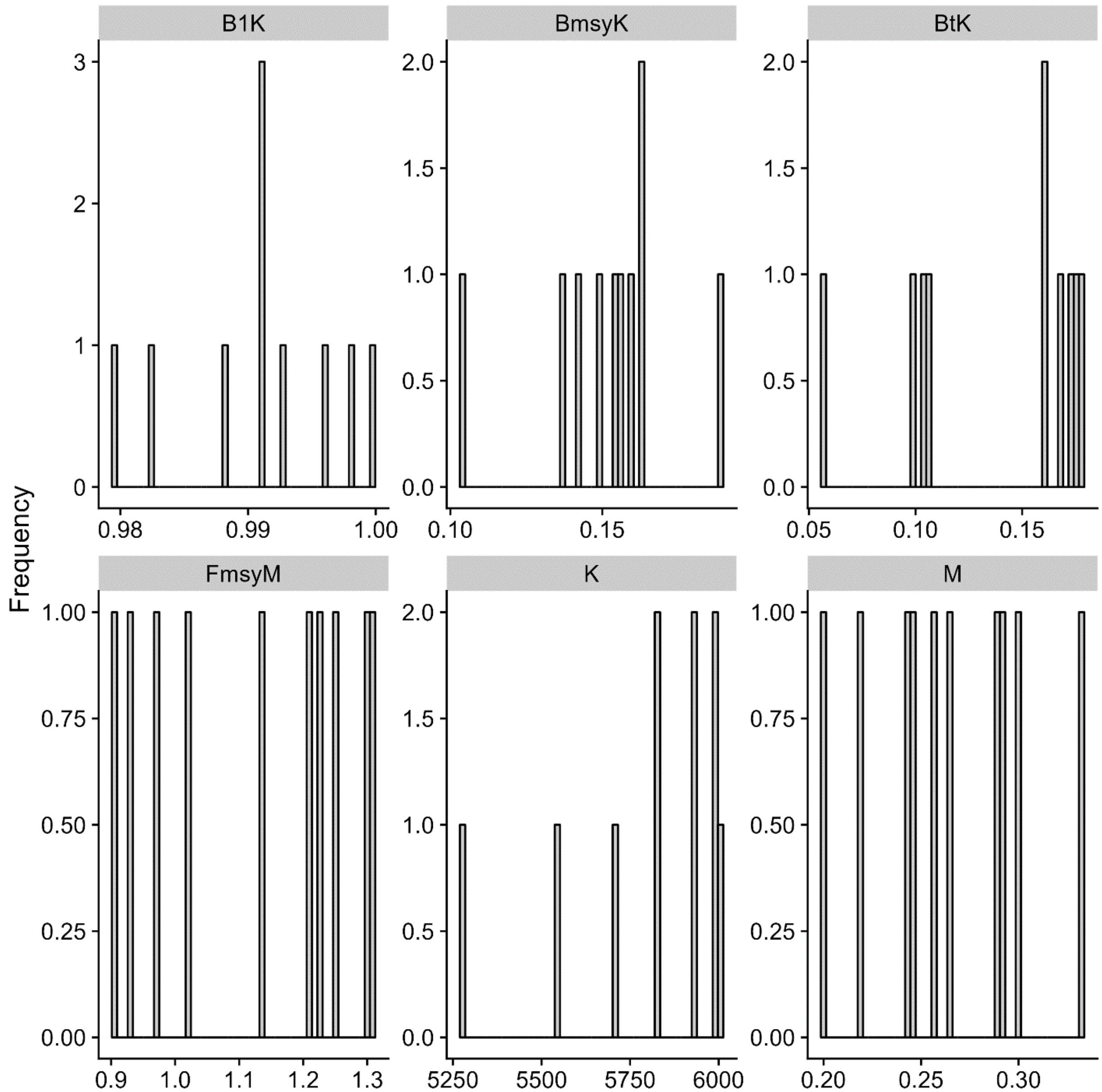

Figure A3.5. Posterior parameter distributions for successful runs of a DF-SRA model of sea scallops in the core area of the sGSL.

The residual plots (Figure A3.6) indicate that the model, despite using parameter estimates outside of the expected range (see B_{MSY}/K) was able to produce catch estimates which closely matched the true values. The resulting residuals were small and equally centered around zero, with the largest residuals produced during the period when catches were at their greatest, between the mid 1960s to the mid 1980s.

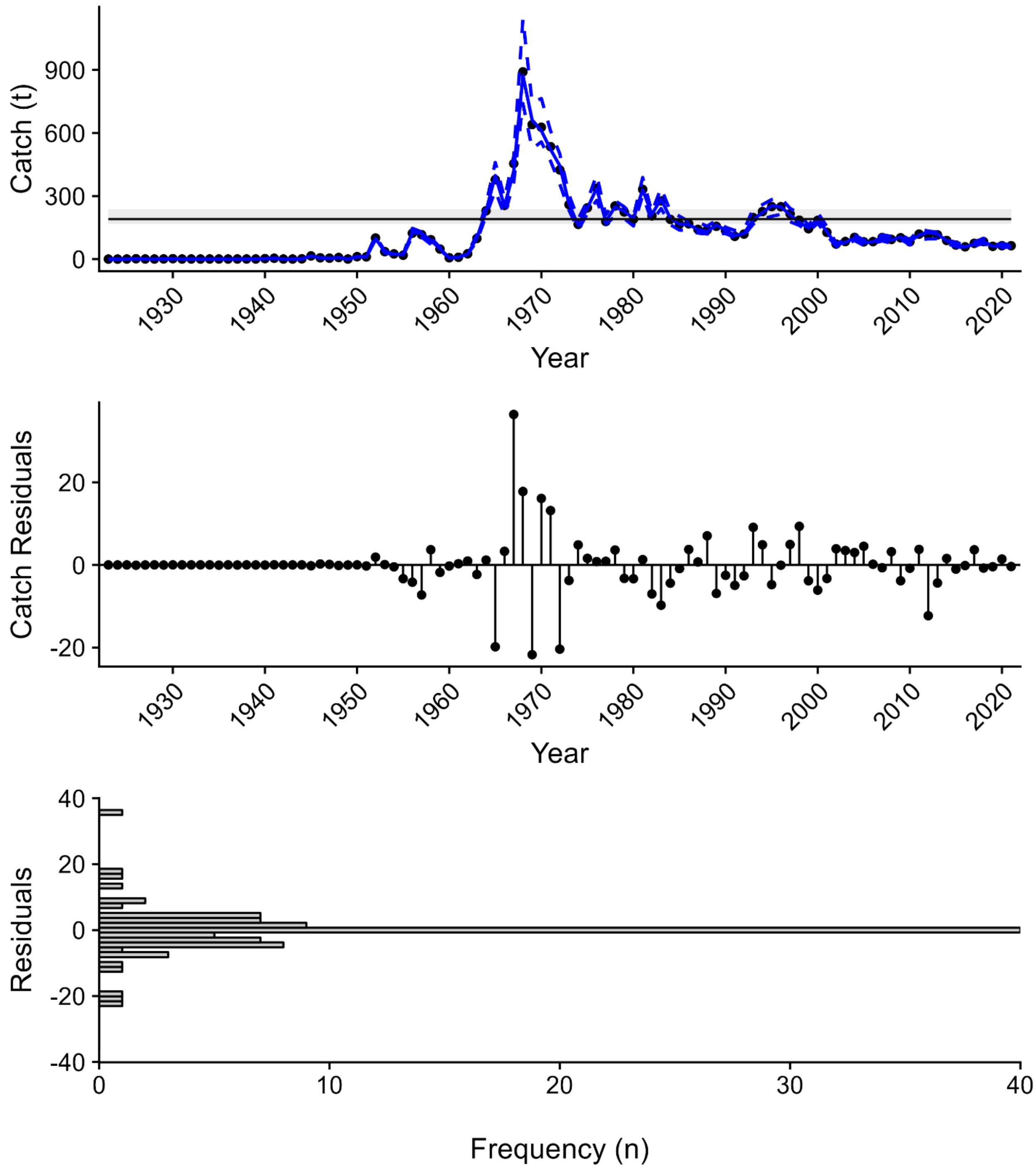

Figure A3.6. Model residual plots for a DB-SRA model fit to the sea scallop catch data in the core area of the sGSL.

COM-SIR MODEL FIT ASSESSMENT

For the COM-SIR model, posterior parameter distributions are much tighter than the those of the uniform distributions priors formed from the ranges provided (Figure A3.7). A notable exception to this is with the r, which had the greatest density around it’s median of 0.41, but exhibited noticeable density throughout it’s range.

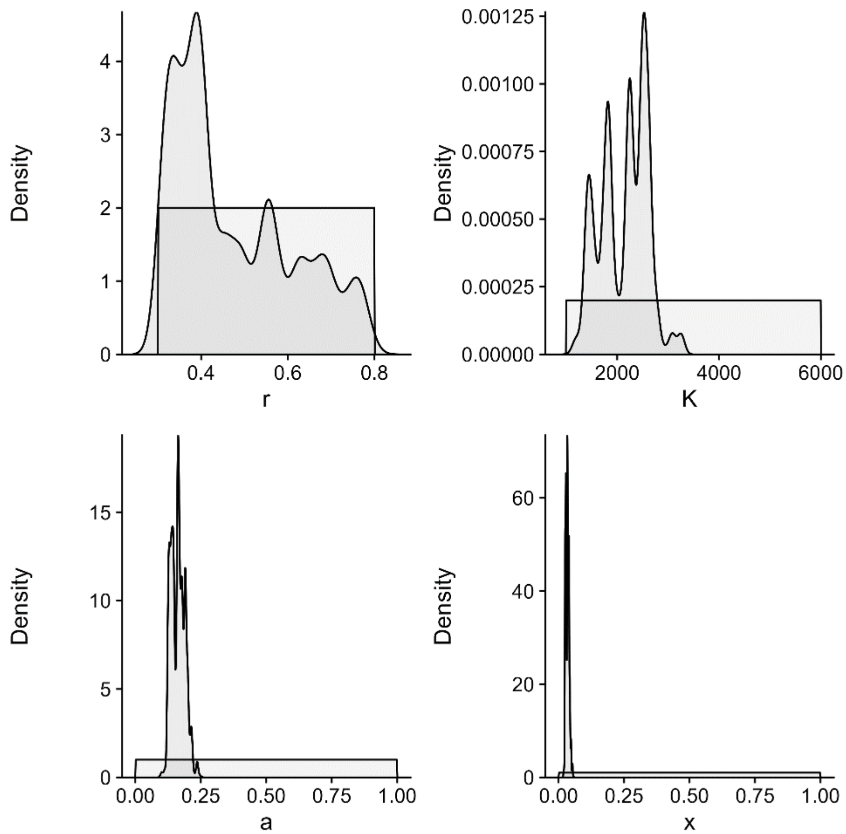

Figure A3.7. Prior and posterior distributions for parameters used by the COM-SIR model for sea scallops in the core area within the sGSL. Priors followed a uniform distribution within user defined ranges.

The catch levels predicted by the COM-SIR model did not match the true values particularly well (Figure A3.8), resulting in residuals which: a) were excessively large during periods when catches here high, and b) oscillate around the mid point due to an inability of the model to react quick enough to steep changes in the catch trajectory, resulting in underestimates followed by overestimates.

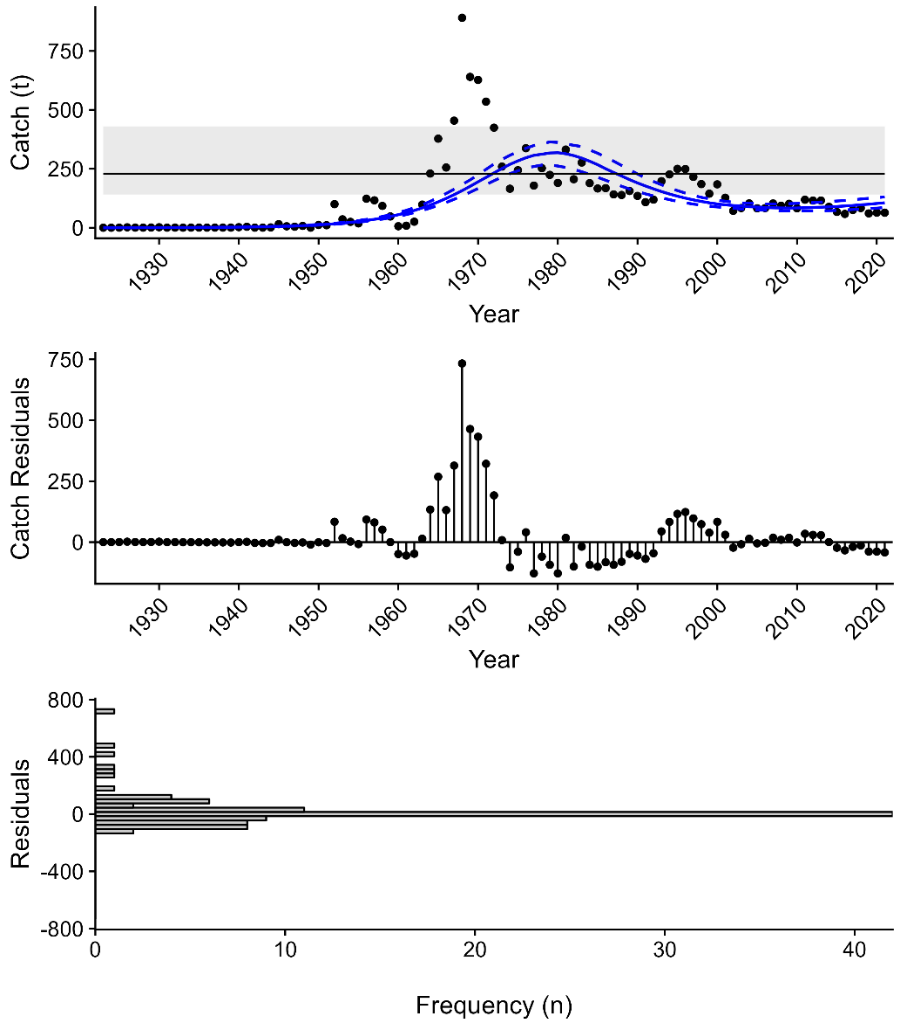

Figure A3.8. Model residuals for catch estimates by the COM-SIR model fit to the landings data for the core area within the sGSL.

CMSY++ MODEL FIT ASSESSMENT

Posterior distributions for parameter estimates for the CMSY model (Figure A3.9) were within the expected prior ranges for this stock. A notable result for r, however, indicates that the model preferred values at the lower limit of the provided prior range. In general, the posterior distributions showed a substantial reduction compared to the priors, with mean values differing from those of the priors in most instances. This suggests that the model is predominantly influenced by the available data rather than being strongly guided by the specified prior ranges.

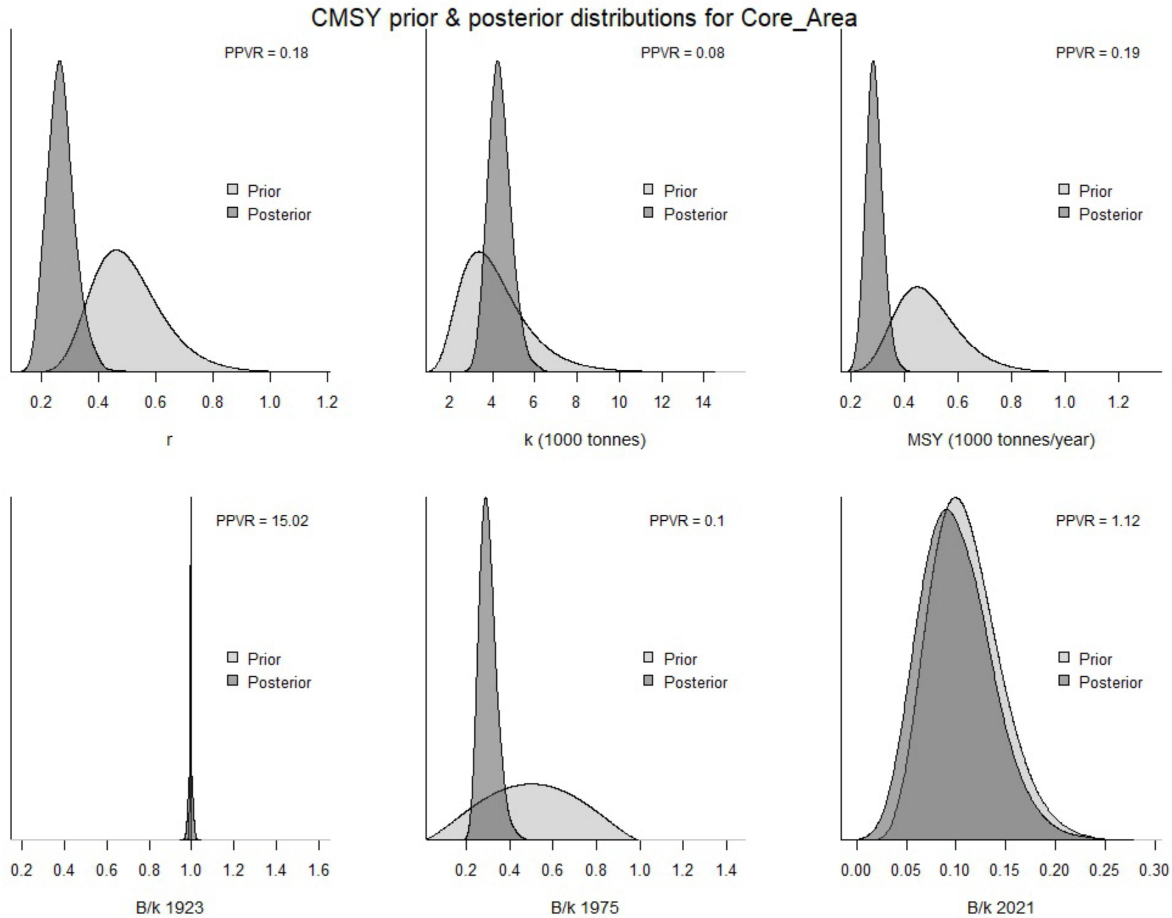

Figure A3.9. Prior and posterior parameter distributions for the CMSY model run on the sea scallop catch history in the core area within the sGSL.

The model's fit to the catch data, however, indicates a departure by the model from the true catch levels in the mid 1980s (Figure A3.10). Until this point, the residuals are minimal and evenly distributed around zero. After this point, however, the residuals are consistently negative, implying the model is overestimating catch. This result may be due to the model forcing parameter estimates that return final saturation levels near the mean of the distribution provided. In order to do this, the model may be favoring parameter estimates that fit initial catches well, but overestimate in more recent years in order to reach the desired output for S_{2021}.

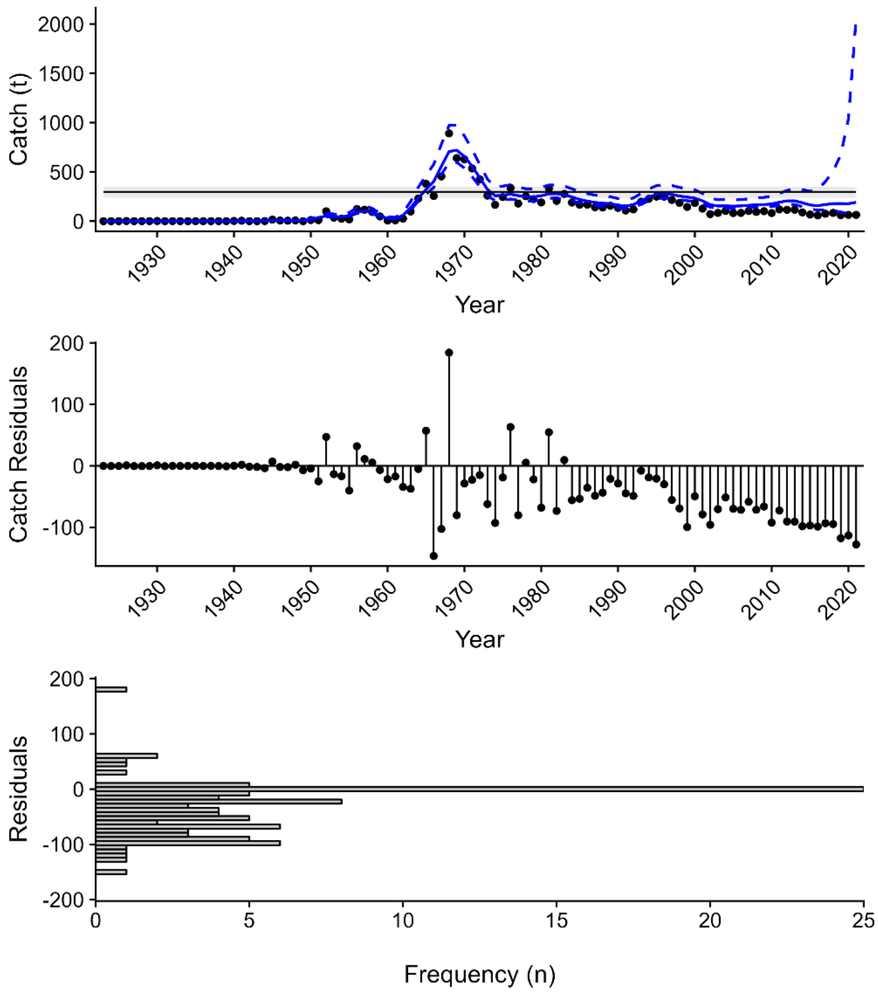

Figure A3.10. Model fit plots of the catch time series from a CMSY++ model fit to scallop landings data from the core area within the sGSL.

BSM MODEL FIT ASSESSMENT

As with the CMSY method above, the BSM model produced posterior distributions which were narrower than the original priors and whose means differed, indicating the model was drawing conclusions from the data rather than simply returning the priors (Figure A3.11). This is perhaps most evident in the prior and posterior for saturation in the final year (B/K 2021) where we see a considerably larger shift towards lower saturation levels relative to the CMSY++ model above. This is likely the result of the additional inclusion of CPUE data in the BSM model compared to the CMSY++ model.

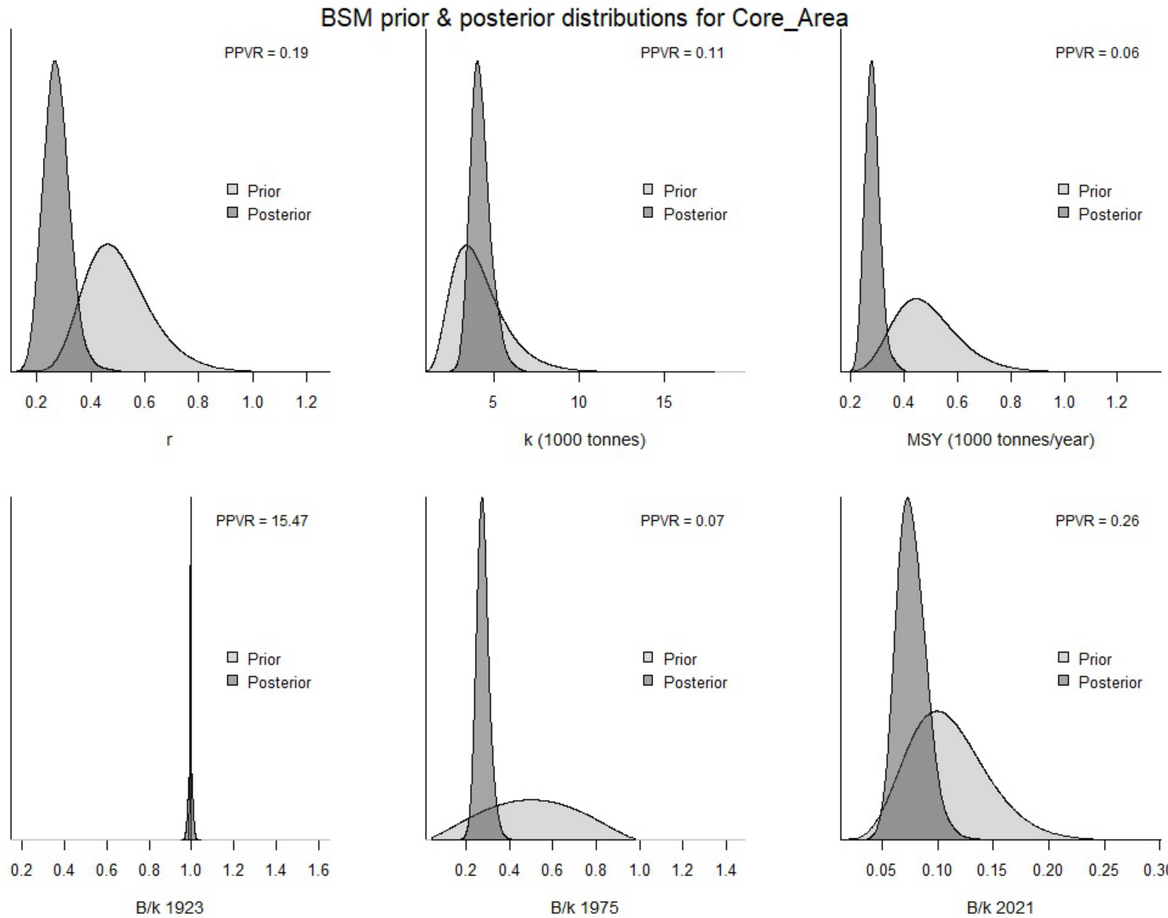

Figure A3.11. Prior and posterior parameter distributions for the BSM model fit to sea scallop data in the core area of the sGSL.

Despite this availability of additional information, the BSM model residuals indicate a similar pattern to the CMSY++ method, as the model initially predicts catch levels quite accurately, but overestimates the catch in more recent years (Figure A3.12). Unlike the CMSY++ method, however, the divergence between the predicted and actual catch values occurred in the mid 1990s rather than the mid 1980s. The extent to which the model overestimated catch is quite similar among the two models, with the model overestimating catch by more than 100 t.

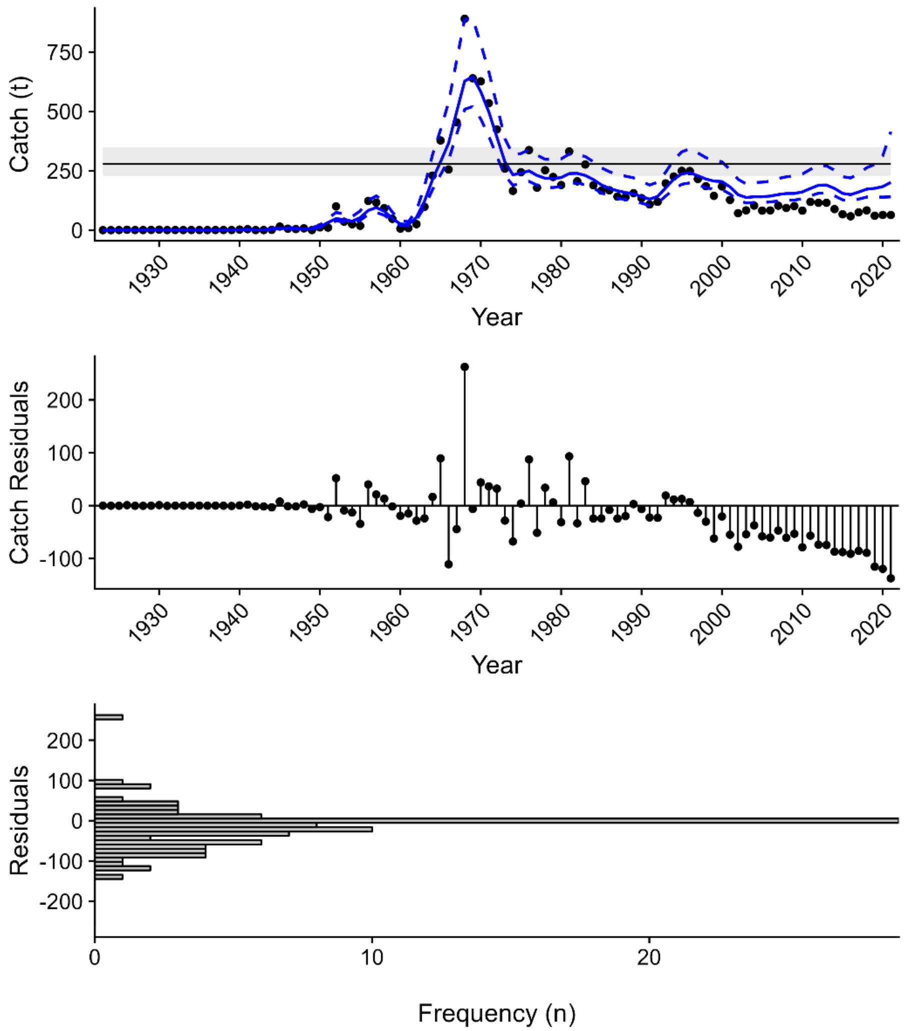

Figure A3.12. Model residual plots of the catch time series from a BSM model fit to scallop landings data from the core area within the sGSL.

SPICT MODEL FIT ASSESSMENT

Parameter estimates from the SPiCT model indicate the input data were somewhat informative as the mean of posterior distributions differ considerably from those of the priors with the exception of the initial biomass estimate (bkfrac) and the Pella-Tomlinson shape parameter (Figure A3.13). With bkfrac, this is unsurprising as the parameter was essentially fixed by the use of a highly informative prior. The prior for the shape parameter (n), however, was deliberately left uninformative. Despite this, the model returned essentially the same distribution. Without this prior, however, the model often settled on an estimate of n which exceeded 2, indicating that B_{MSY}/K was near 1 which is unrealistic.

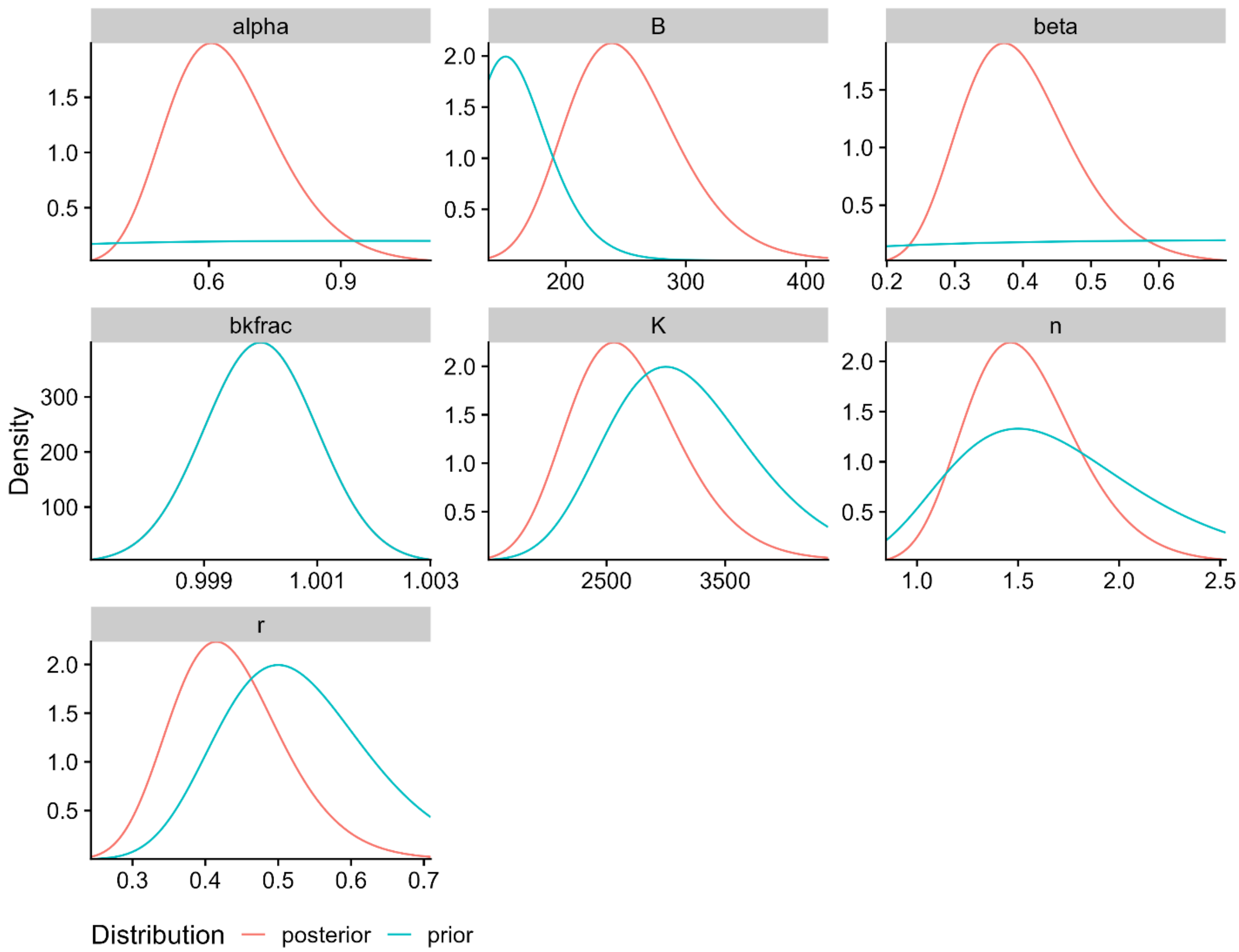

Figure A3.13. Prior and posterior parameter distributions for a SPiCT model fit to sea scallop landings data from the core area within the sGSL. Priors bkfrac and B correspond to the relative initial biomass and the absolute final biomass respectively.

Model fit was generally quite good, with the largest difference between estimated and actual catch levels occurring in 1968, the year with the largest landings of the entire time series. Apart from this the majority catch estimates differed only by 50 t or less from the true landings and virtually all fell within the models catch confidence intervals (Figure A3.14).

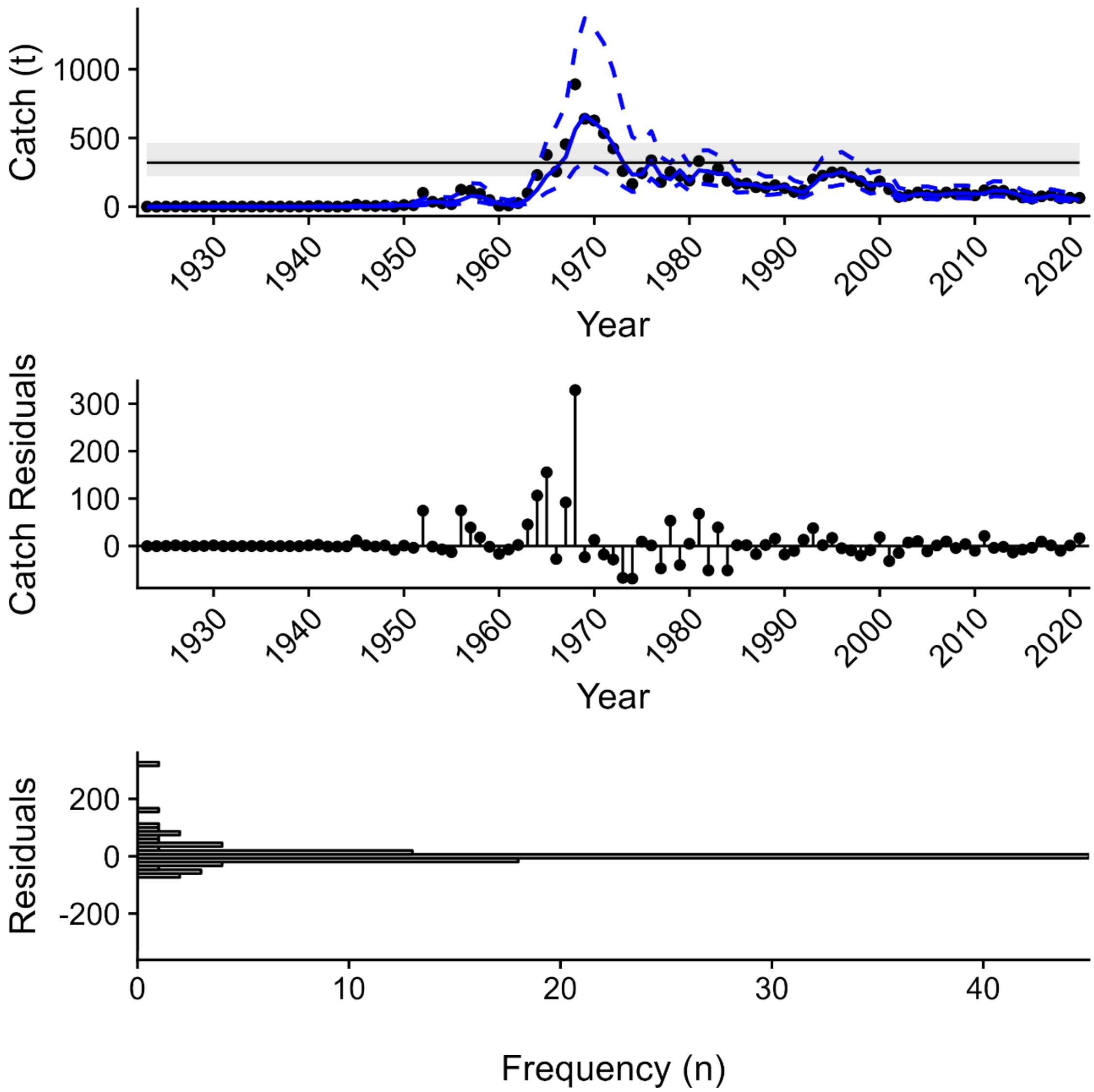

Figure A3.14. Model residual plots from a SPiCT model of sea scallops in the core area of the sGSL.

JABBA MODEL FIT ASSESSMENT

Posterior distributions from the JABBA model demonstrate that the model drew much from the data provided, as means shift considerably relative to the prior distributions. This is most notable with the estimate of r which falls outside of the expected range set by the prior. An exception to this shift is with the initial biomass estimates (psi) which were highly informative (i.e. fixed).

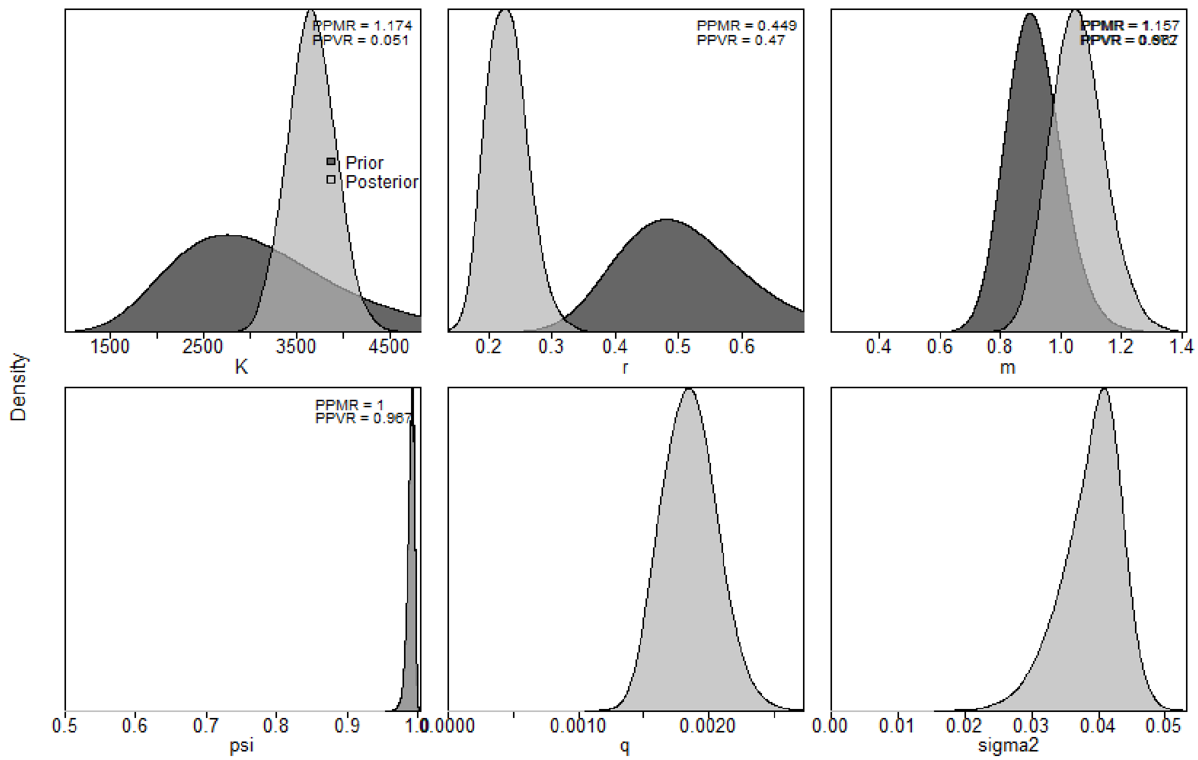

Figure A3.15. Posterior and prior parameter distributions from a JABBA model fit to sea scallop from the core area of the sGSL.

The model residuals plots highlight how this model produced catch estimates which closely matched the true catch levels, resulting in small residuals which were evenly spaced around the mean of zero. While there is some indication of an oscillating pattern in the catches, this was found to be insignificant by a runs test performed by the model.

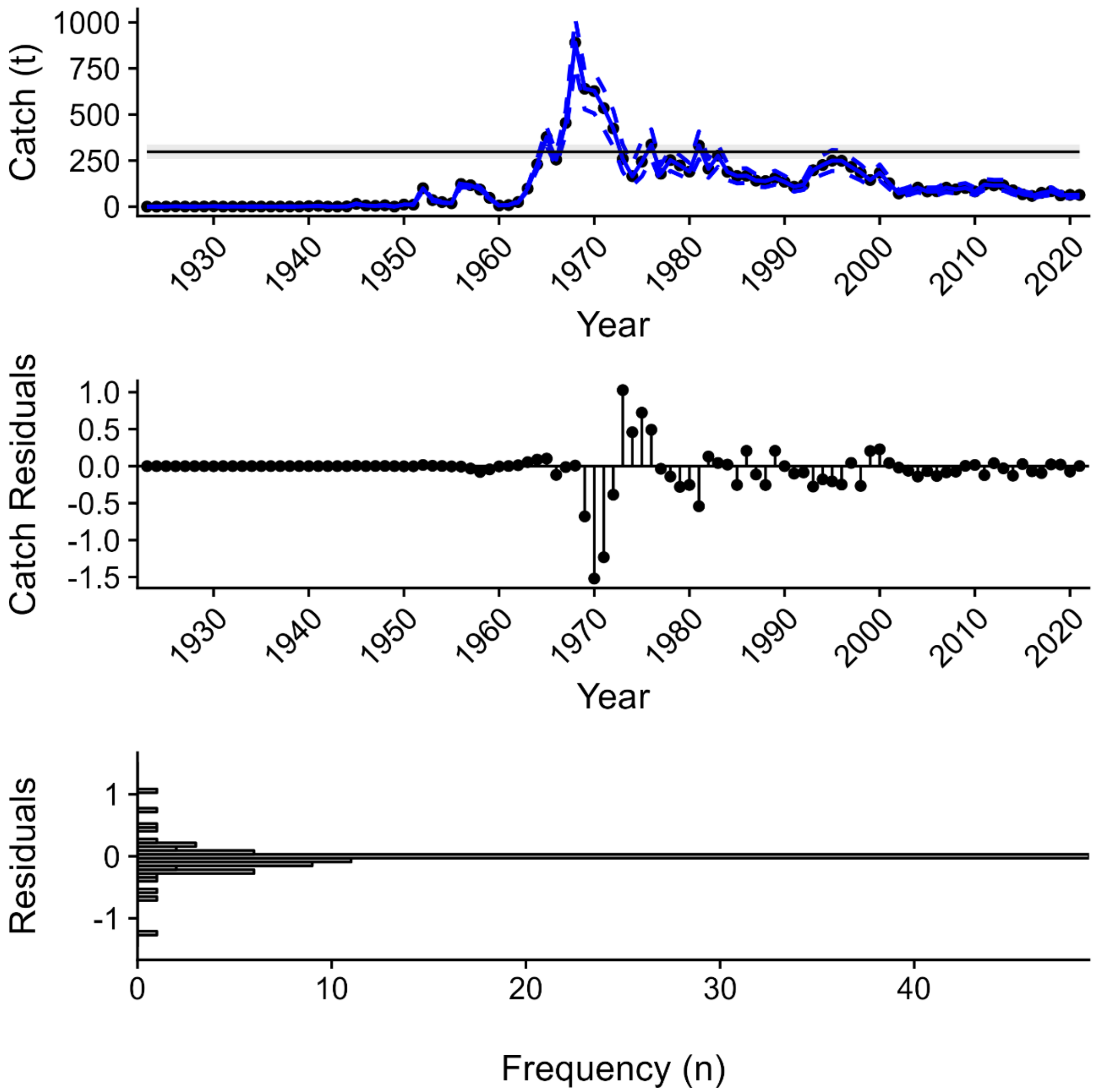

Figure A3.16. Model fit plots for a JABBA model fit to the catch and effort data for scallops within the core area of the sGSL.

ASPIC MODEL FIT ASSESSMENT

The residual plots for the ASPIC model show a pattern of generally small residuals equally spaced around zero with a few major exceptions. First, the model overestimated the catch by 150 t during the period when catches were their highest. Second, there are two periods when the model estimates were consistently off for multiple consecutive years. The first period is in the mid 1990s when there was a tendency by the model to overestimate catches, followed by a more recent period when the model underestimated the catch by 5 – 15 t each year for 7 consecutive years.

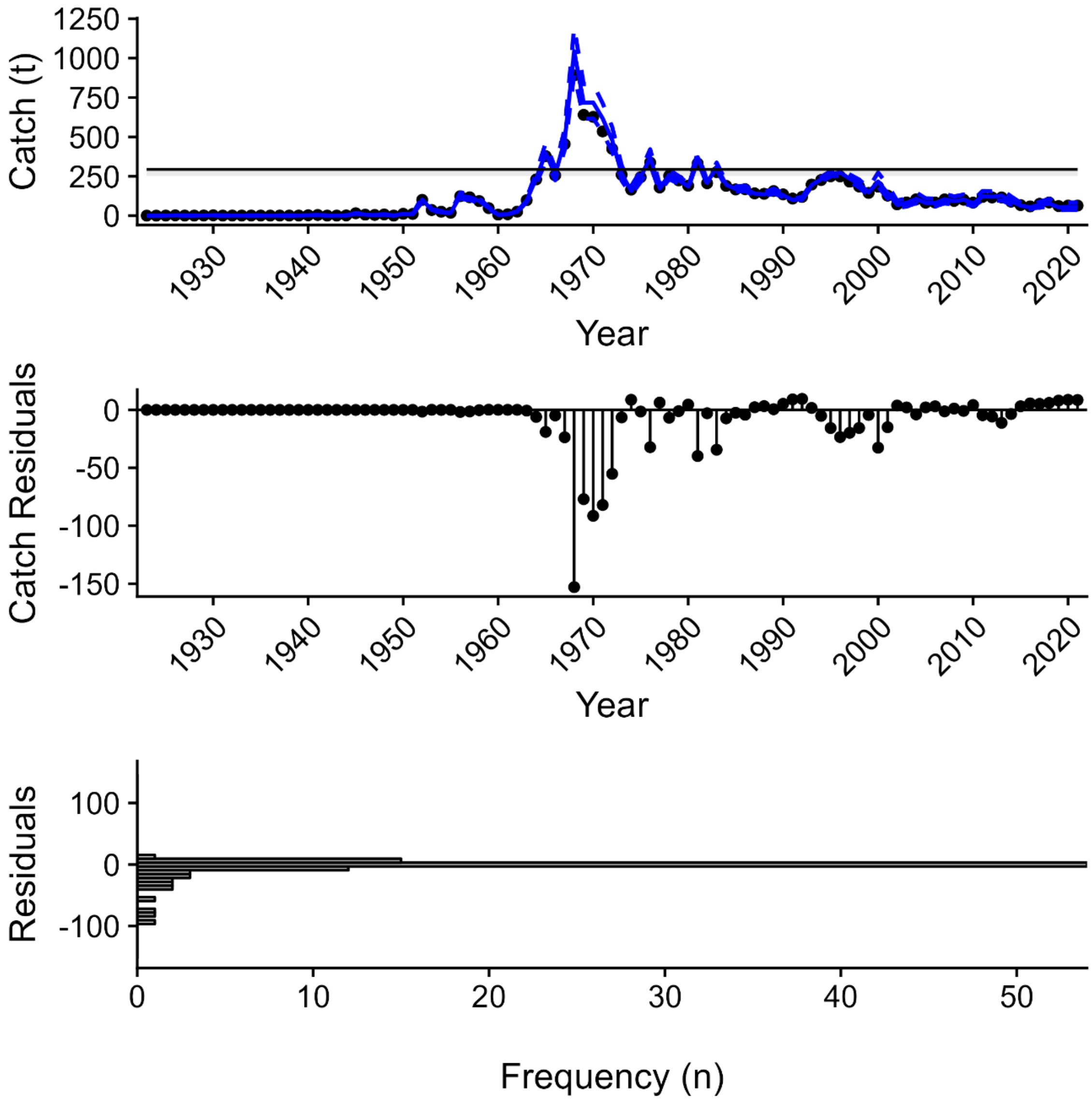

Figure A3.17. Model residual plots for an ASPIC model fit to sea scallop data for the core area of the sGSL.