Summing the parts: Improving population estimates using a state-space multispecies production model

Paul M. Regular1\*, Mariano Koen-Alonso1, M. Joanne Morgan1, Pierre Pepin1, Rick M. Rideout1

1 Fisheries and Oceans Canada, Northwest Atlantic Fisheries Center, 80 East White Hills, St. John’s, Newfoundland and Labrador, A1C 5X1, Canada  
  
\*Corresponding author  
E-mail: [Paul.Regular@dfo-mpo.gc.ca](mailto:Paul.Regular@dfo-mpo.gc.ca)  
  
2022-09-22

**Abstract:** …  
  
**Keywords:** …

# Introduction

# Methods

## Model formulation

Trends in fish populations have frequently been described using state-space production models of the form

where is biomass at the start of year , is the catch through year , is production as a function of biomass, is an index of relative abundance in year from survey , is the time-invariant catchability coefficient for survey index , is process error, and is observation error. Statistical challenges aside, the most difficult aspect of this model to parameterize is the production function as it needs to capture changes caused by growth, recruitment, and natural mortality. Schaefer ([1](#ref-schaefer1954)) proposed a solution by applying the logistic equation to describe self-limiting growth,

where is the maximum per-capita rate of change and is the carrying capacity. That is, a populations’ intrinsic ability to grow () is limited by the size of the current population relative to the maximum biomass the system can support (). While this formulation offers an elegant description of single-species population dynamics, it assumes that density-dependent effects are solely caused by intraspecific competition and ignores the potential effects of other species inhabiting the same ecological area. We present an extension of equation (3) that attempts to account for intra and interspecific competition by assuming that density-dependent effects are incurred when the total biomass of multiple species, represented by , exceeds the capacity of the system,

While intrinsic rates of growth may vary across species, this formulation implies that the growth of all species is ultimately limited by the finite amount of energy in a region (i.e., as the population in the system increases towards , year-over-year growth of all species slows). Combining equations (1), (2), and (4), our model becomes

The inclusion of multiple species in the model permits the estimation of covariance. While covarying changes in observed populations may be described using observation errors, we assume that most covariance stems from population processes. We therefore assume that observation errors are independent and normal distribution such that , where the standard deviation parameter represents species and survey specific levels of observation error. A more flexible error structure is used to describe the process errors as ecological processes may contribute to species or temporal dependencies. Beyond the global limiting effect of , species interactions may elicit positive or negative population responses resulting from direct or indirect associations. We therefore apply the multivariate normal distribution to account for the possibility that process errors are not independent across species. Deviations from expected production may also display temporal dependence if the factors contributing to the process errors change slowly through time. Such inertia may cause positive or negative process errors to persist across several years. A first-order autoregressive (AR1) process was therefore applied to account for temporal dependence. Both sources of dependence are modeled using a multivariate AR1 process where

and

The degree of temporal correlation is controlled by , where values between 0 to 1 represent low to high correlation, and species-to-species correlations are described by , where values between -1 to 1 represent negative to positive correlation. This is a flexible structure that allows for the testing of alternate hypotheses that process errors are independent through time or across species (i.e, or ). The possibility that process errors are similarly correlated across all species may also be tested by estimating only one parameter. Finally, the magnitude of the process error deviations are controlled by the species-specific standard deviation parameters, .

Minor extensions of the formulation also permits the fitting of covariates which may describe an underlying linear effect. Two options were implemented, one that affects the process errors by substituting in equation (5) with , and another that affects the carrying capacity by substituting in equation (5) with . The idea is that some factors may affect positive or negative changes in the populations while others may affect change in the total carrying capacity of the system. A covariate option for intrinsic rates of increase was not implemented as one goal of this model is to obtain estimates of this vital rate, which is not expected to change rapidly as it is shaped by natural selection ([2](#ref-hutchings2021)). The formulation was also modified to fit the single-species Schaefer production function by dropping the summation of biomass in equation (5) and estimating species-specific carrying capacities, (i.e., apply equation (3) indexed by species).

## Statistical framework

This model was implemented using template model builder (TMB; [3](#ref-kristensen2015)), which is package for R ([4](#ref-R)) that enables the fitting of complex nonlinear random effects such as the latent variable in state-space production models (equation (1)). Such variables are not directly measured but are inferred indirectly via observed values. Data fitting is accomplished using a combination of Laplace approximation and automatic differentiation to evaluate the joint likelihood ([3](#ref-kristensen2015)). Like the production model descried by Pedersen et al. ([5](#ref-pedersen2017)), both frequentist and Bayesian inference of model parameters are possible. In development, we found that estimation was generally more successful when vaguely informative priors are specified as parameters were, in some cases, not identifiable when unconstrained.

## Case study

### Data

The multispecies production model described above requires two basic inputs for one or more species in a region: 1) a time-series of catch ( in equation (5)), and 2) an index of population size ( in equation (6)). The Northwest Atlantic Fisheries Organization (NAFO) and Fisheries and Oceans Canada (DFO) have been collecting and curating such information for multiple fish populations along the shelves of Newfoundland and Labrador (NL) since the 1970s. The communities inhabiting these shelves can be divided into several regions with distinct productivity [i.e. ecosystem production units; ([6](#ref-pepin2014))]. For our case study, we tallied catch data and calculated survey indices of multiple demersal fish populations from three regions: 1) the Northeast NL Shelf (NAFO divisions 2J3K), 2) the Grand Bank (NAFO divisions 3LNO), and 3) Southern NL (NAFO sub-division 3Ps; Figure 1).

Catch data were extracted from the STATLANT 21A database maintained by NAFO (<https://www.nafo.int/Data/STATLANT-21A>, accessed 2022-01-21) and aggregated by region, species, and year. Survey indices were derived from the standardized, stratified random bottom-trawl surveys conducted each spring and fall by DFO. Since the inception of this program in 1971, survey protocol have undergone a series of changes that affect the continuity of the data collected in each region and season. A Yankee then Engel otter trawl, with nets designed to catch large demersal fish, were used between 1971 to 1994. Since 1995, a Campelen shrimp trawl has been used and, because of its smaller mesh size, a broader range of species and size groups have been captured ([7](#ref-chadwick2007)). Within each era of the survey (Yankee, Engel, or Campelen) and for each season and region, samples were limited to strata that were covered most years (> 80%) and to species found across more than 10% of these core strata. Stratified analyses ([8](#ref-smith1981)) were then conducted on the remaining species to obtain indices of total biomass. To minimize bias introduced by inconsistent survey coverage, indices from years where more than 20% of the biomass was likely missed, inferred from time averaged percent occupancy within strata, were excluded from our analysis [*sensu* NAFO guidelines; page 10, ([9](#ref-nafo2019))]. Finally, species were ranked by cumulative catch and limited to the the seven most commonly caught species, or species group when catch was not consistently distinguished, within each region. On the Northeast NL Shelf, species include Redfish spp. (*Sebastes fasciatus* or *S. mentella*), Wolffish spp. (*Anarhichas lupus* or *A. minor*), Witch Flounder (*Glyptocephalus cynoglossus*), American Plaice (*Hippoglossoides platessoides*), Greenland Halibut (*Reinhardtius hippoglossoides*), Atlantic Cod (*Gadus morhua*), and Skate spp. (*Amblyraja radiata* or *Malacoraja senta*). On the Grand Bank, species include Redfish spp., Yellowtail Flounder (*Limanda ferruginea*), American Plaice, Greenland Halibut, Haddock (*Melanogrammus aeglefinus*), and Atlantic Cod. Finally, along Sourthern NL, species include Redfish spp., Witch Flounder, American Plaice, White hake (*Urophycis tenuis*), Haddock, Atlantic Cod, and Skate spp.

### Priors

For simplicity, all priors were normally distributed and upper and lower inflection points (; ~68% of the total area under the curve) of each normal prior were defined using values in log or logit space. Upper and lower values were based on previous research or knowledge to impose fairly generic and vaguely informative priors.

#### *Intrinsic growth rate, , and carrying capacity,*

To capture a broad range of predicted intrinsic growth rates for all fishes worldwide ([10](#ref-thorson2020)), 0.01 and 0.1 were chosen as the lower, , and upper, , values; this translate to a normal prior with a mean of -1.15 and a standard deviation of 1.15 on the log-scale. The log-scale prior was informed by total levels of catch and and ,

where and are the lower and upper values. On the lower end, this prior imposes the assumption that the fishery is unlikely to have caught more fish in a single year than the system is capable of supporting and, on the upper end, it assumes that the maximum observed catch represents a portion of the true carrying capacity of the system (*sensu* [11](#ref-froese2017)). The division by the lower and upper values for also accounts for the potential range of productivity. Note that the maximum time-series catches are species specific when estimating species-specific carrying capacities.

#### *Starting biomass,*

Like , catch was used to constrain the plausible range of the biomass at the beginning of the time series, , which we will denote as to simplify notation. Specifically,

where and are the lower and upper values that are log transformed to define the normal prior for . Again, adjusting for the potential range of productivity from to , this assumes that the fishery did not catch all of the biomass in the first year and it assumes that the catch represents a portion of the biomass in the first year. A flaw with the upper value for this prior is that a lack of market demand or conservation concerns may contradict the assumption that landings are coarsely proportional to stock size. However, the upper range chosen was considered reasonable for this case study as there was an active fishery for each species at the start of the time series.

# Results

# Discussion

# Acknowledgements

We thank the many ship’s crew and research staff involved in leading the surveys and collecting the data used in this study. These surveys were supported by Fisheries and Oceans Canada (DFO). This work has benefited from valuable feedback from numerous colleagues, including …

# References

1. M. B. Schaefer, Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. *Inter-American Tropical Tuna Commission Bulletin* **1**, 23–56 (1954).

2. J. A. Hutchings, *A primer of life histories: Ecology, evolution, and application* (Oxford University Press, 2021).

3. K. Kristensen, A. Nielsen, C. Berg, H. Skaug, B. Bell, [TMB: Automatic differentiation and laplace approximation](https://doi.org/10.18637/jss.v070.i05). *Journal of Statistical Software* **70**, 1–21 (2016).

4. R Core Team, [*R: A language and environment for statistical computing*](https://www.R-project.org/) (R Foundation for Statistical Computing, 2021).

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

6. P. Pepin, J. Higdon, M. Koen-Alonso, M. Fogarty, N. Ollerhead, Application of ecoregion analysis to the identification of Ecosystem Production Units (EPUs) in the NAFO Convention Area. *NAFO Sci. Counc. Res. Doc* **14**, 069 (2014).

7. E. Chadwick, *et al.*, History of annual multi-species trawl surveys on the Atlantic coast of Canada. *Atlantic Zonal Monitoring Program Bulletin* **6**, 25–42 (2007).

8. S. Smith, G. Somerton, *STRAP: A User-Oriented Computer Analysis System for Groundfish Research Trawl Survey Data* (Canadian Technical Report of Fisheries; Aquatic Sciences No. 1030, 1981).

9. NAFO, [Report of the Scientific Council, 31 May – 13 June 2019, Halifax, Canada](https://www.nafo.int/Portals/0/PDFs/sc/2019/scs19-20.pdf). *NAFO SCS Doc* **19/20** (2019).

10. J. T. Thorson, Predicting recruitment density dependence and intrinsic growth rate for all fishes worldwide using a data-integrated life-history model. *Fish and Fisheries* **21**, 237–251 (2020).

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

# Figures

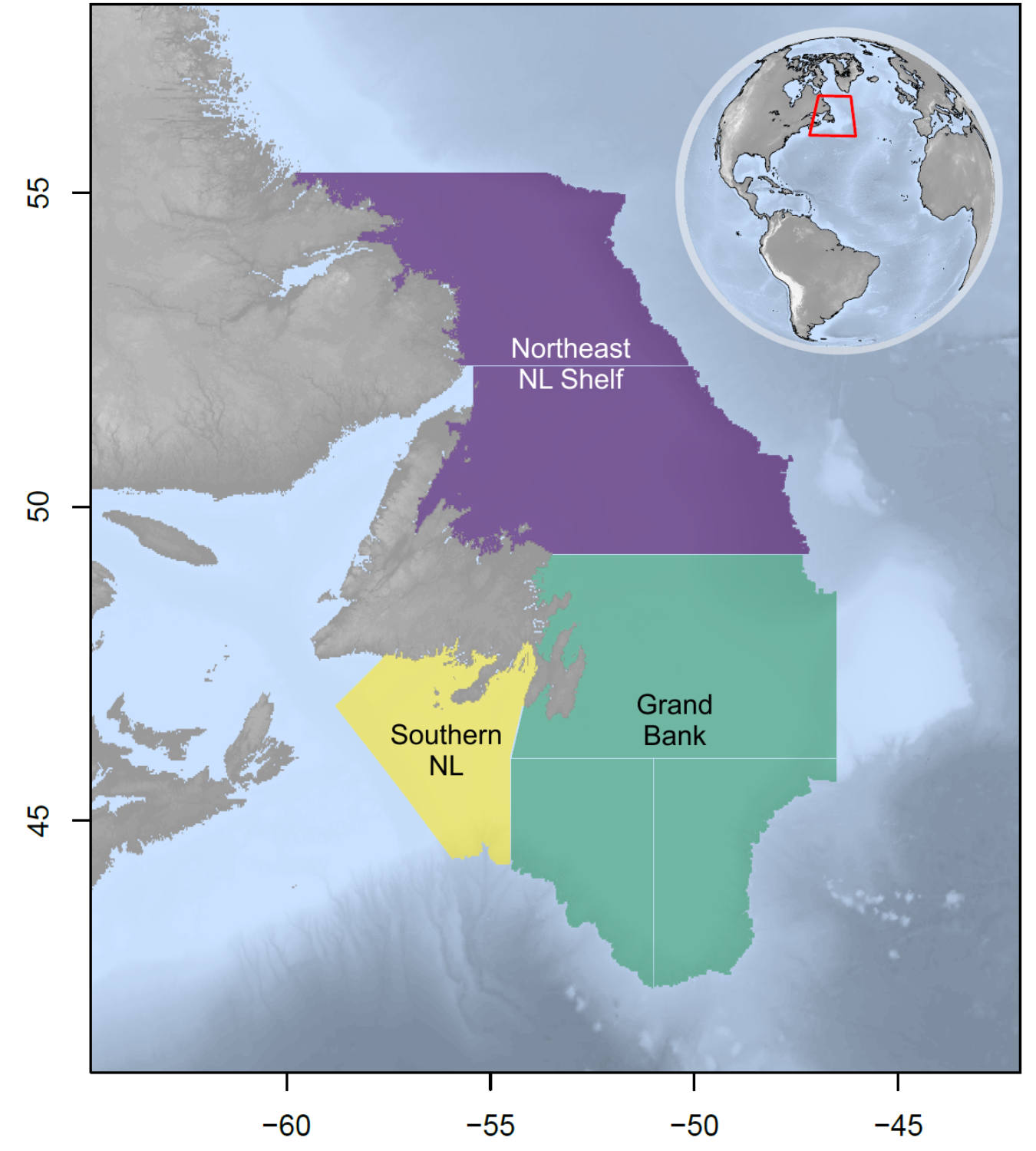


Fig 1: Map of Newfoundland and Labrador (NL) case study area showing the Northeast NL Shelf (purple), Grand Bank (green), and Southern NL (yellow) ecosystem production units.