Approximating uncertainty around indices from stratified-random trawl surveys using the Gamma distribution

Paul M. Regular, Mariano Koen-Alonso, Semra Yalcin, Andrea M.J. Perreault, Laura J. Wheeland

Northwest Atlantic Fisheries Center, Fisheries and Oceans Canada, P.O.Box 5667, St. John’s, NL, A1C 5X1, Canada   
  
2022-06-20

# Introduction

A primary objective of fisheries-independent trawl surveys is to obtain indices of stock size and quantify the uncertainty around these indices. Such information plays a critical role in the assessment and management of fish stocks around the world as they often serve as a leading indicator of change ([Kimura and Somerton, 2006](#ref-kimura2006); [Pennington and Strømme, 1998](#ref-pennington1998)). Surveys indices also influence risk-based decision making; however, such information is typically provided indirectly via estimates of uncertainty from stock assessment models that are calibrated using point estimates of trawlable abundance or biomass. Data limitations often preclude the use of complex assessment models and, as such, many stocks are assessed using survey indices. While model-based indices produced using geostatistical approaches (e.g., [Anderson et al., 2022](#ref-anderson2022); [Thorson et al., 2015](#ref-thorson2015)) are growing in popularity, design-based estimators continue to be widely used. In the Northwest Atlantic, surveys typically follow a stratified-random sampling design with proportional allocation (e.g., [González-Troncoso et al., 2022](#ref-gonzalez2022); [Rideout et al., 2022](#ref-rideout2022)) and indices are obtained using stratified analyses (e.g., [S. Smith and Somerton, 1981](#ref-smith1981)). Unfortunately, the quantification of uncertainty around these estimates remains a challenge. Quantiles from a Student’s t distribution are often used to approximate the uncertainty around stratified estimates; however, the lower limits of this approximation can result in negative values, which is unrealistic ([Cadigan, 2011](#ref-cadigan2011)). We propose an alternate approximation of uncertainty using the Gamma distribution which accounts for the positive and skewed nature of survey indices.

# Methods

Provided data from a stratified-random survey, average trawlable abundance or biomass () and sampling variance () can be estimated using standard design-based formula ([Cochran, 1977](#ref-cochran1977); [S. J. Smith, 1990](#ref-smith1990); [S. Smith and Somerton, 1981](#ref-smith1981)). Instead of using a Student’s t distribution to describe uncertainty and allow negative values, we apply the Gamma distribution by translating and to scale () and shape () parameters as follows:

Provided these values, density, quantile, and random functions for the Gamma distribution can be used to calculate probabilities. For instance, the probability that the index increased from one year to the next can be quantified. For some cases there might also be a need to calculate the probability that the current index is above or below an average level from a reference period, . If the reference period is based on the index, then the level cannot be perfectly known. To account for uncertainty around this reference, , it is necessary to combine the variances across the indices. This is accomplished by averaging the means and summing equally weighted variances across a reference set of years , where is the size of the set,

This assumes that the estimates being averaged are independent. As above, and can be converted to and parameters to approximate uncertainty.

## Simulation

We simulated a redfish-like population using the R package SimSurvey ([Regular et al., 2020](#ref-regular2020)). The simulated population was based on the exponential decay cohort model where parameter settings for mortality, recruitment, and growth were based on assessments of redfish on the Grand Bank. The simulated population were distributed through an area according to the age-year-space covariance with a parabolic relationship with depth. This survey area was 300 x 300 km with 10 km2 cell size and had 30 depth-based strata. We simulated stratified random sampling with a 2 m wide trawl hauled for a distance of 1.5 km. The population and survey were simulated over 20 years. The number of sets in a stratum was proportional to its area (approximately 1 set per 1000 km2) and the minimum set per stratum was 2. The survey simulation was replicated five times over the same population.

Average trawlable abundance () and sampling variance () was calculated by year and replicate (20 years across 5 surveys) using standard design-based estimators ([S. J. Smith, 1990](#ref-smith1990); [S. Smith and Somerton, 1981](#ref-smith1981)) and these estimates were translated to scale () and shape () parameters for the Gamma distribution as described above. To compare densities obtained from the Gamma distribution with densities based on an empirical approach, we applied a non-parametric bootstrap to resample the observations (sets) independently within each stratum with replacement. The resampling and calculation of the mean bootstrap estimator were repeated 1000 times with the R package boot ([Canty and Ripley, 2021](#ref-canty2021)). Densities from these boostrap samples were computed for each year and survey replicate for comparison to the Gamma approximation.

# Application

# Results and Discussion

The Gamma probability density distribution showed high variability among survey simulations at Year 20 (Figure 1). The bootstrapped estimates of each survey simulation also showed a similar pattern with the gamma probability distribution at Year 20 (Figure 2). When looking at the distribution of individual survey simulations, the gamma distribution showed a wider but very close approximation to the bootstrapped estimates distribution (Figure 3). Further quantitative analysis is required to assess the performance of these methods for calculating the confidence intervals.

# References

Anderson, S. C., Ward, E. J., English, P. A., and Barnett, L. A. (2022). sdmTMB: An R package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields. *bioRxiv*.

Cadigan, N. G. (2011). Confidence intervals for trawlable abundance from stratified-random bottom trawl surveys. *Canadian Journal of Fisheries and Aquatic Sciences*, *68*(5), 781–794.

Canty, A., and Ripley, B. D. (2021). *Boot: Bootstrap r (s-plus) functions*.

Cochran, W. G. (1977). *Sampling techniques*. John Wiley & Sons.

González-Troncoso, D., Garrido, I., Rábade, S., Fabeiro, M., Román, E., Tarrío, C., Sánchez, J. M. C., and Alpoim, R. (2022). Results from Bottom Trawl Survey on Flemish Cap of June-July 2021. *NAFO SCR Doc*, *22/004*.

Kimura, D. K., and Somerton, D. A. (2006). Review of statistical aspects of survey sampling for marine fisheries. *Reviews in Fisheries Science*, *14*(3), 245–283.

Pennington, M., and Strømme, T. (1998). Surveys as a research tool for managing dynamic stocks. *Fisheries Research*, *37*(1-3), 97–106.

Regular, P. M., Robertson, G. J., Lewis, K. P., Babyn, J., Healey, B., and Mowbray, F. (2020). SimSurvey: An R package for comparing the design and analysis of surveys by simulating spatially-correlated populations [Journal]. *PLOS ONE*, *15*(5), 1–28. <https://doi.org/10.1371/journal.pone.0232822>

Rideout, R. M., Rogers, B., Wheeland, L., and Koen-Alonso, M. (2022). Temporal And Spatial Coverage Of Canadian (Newfoundland And Labrador Region) Spring And Autumn Multi-Species RV Bottom Trawl Surveys, With An Emphasis On Surveys Conducted In 2021. *NAFO SCR Doc*, *22/007*.

Smith, S. J. (1990). Use of statistical models for the estimation of abundance from groundfish trawl survey data. *Canadian Journal of Fisheries and Aquatic Sciences*, *47*(5), 894–903.

Smith, S., and Somerton, G. (1981). *STRAP: A User-Oriented Computer Analysis System for Groundfish Research Trawl Survey Data* (p. 66). Canadian Technical Report of Fisheries; Aquatic Sciences No. 1030.

Thorson, J. T., Shelton, A. O., Ward, E. J., and Skaug, H. J. (2015). Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. *ICES Journal of Marine Science*, *72*(5), 1297–1310.

# Figures

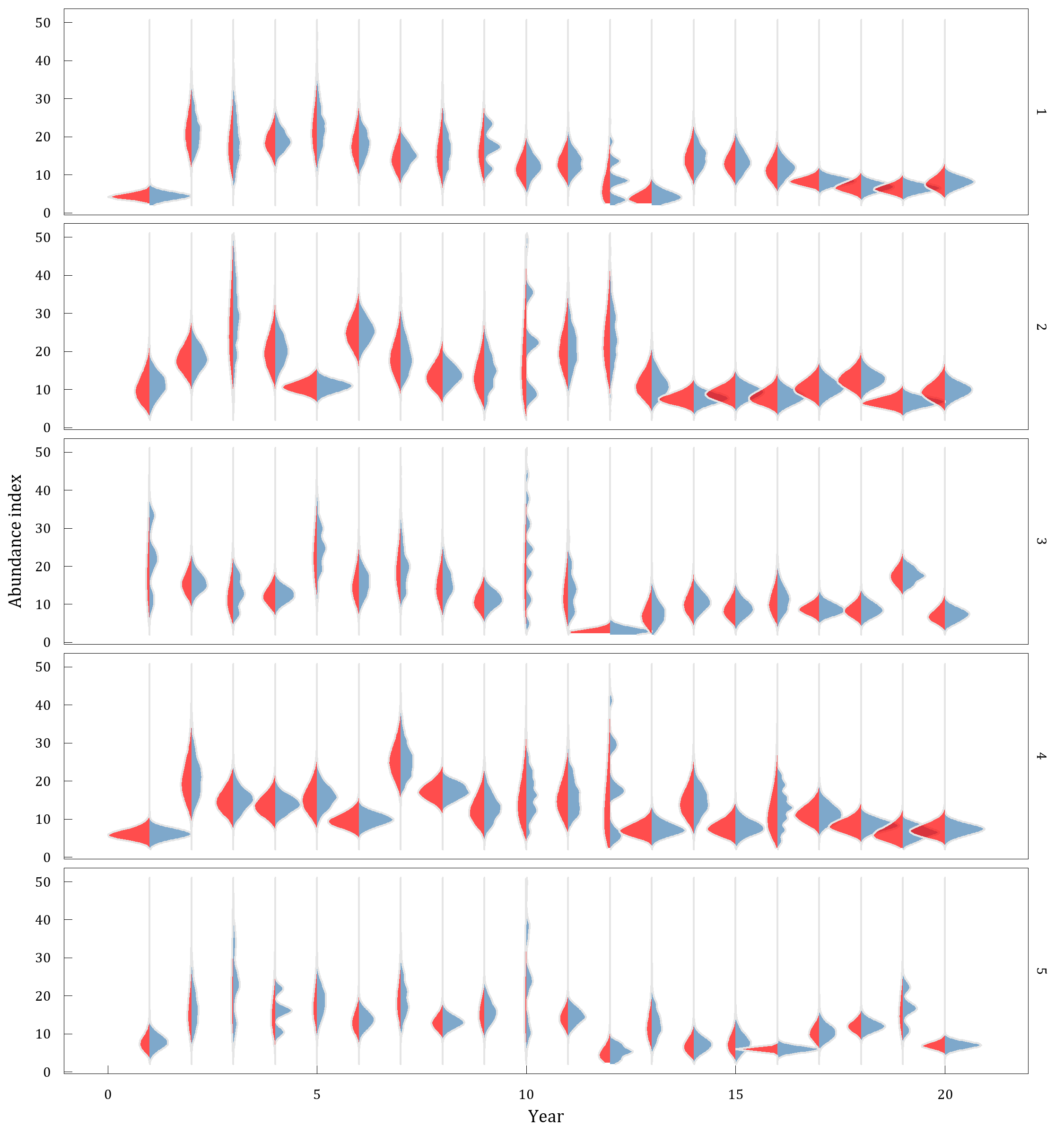


Fig 1: The bootstrap and gamma distributions estimated using simulated data from five independent surveys conducted over the same population across 20 years. The red area shows the density distribution from 1000 bootstrapped samples from each year and survey replicate. The blue area shows the gamma probability distribution from each year and survey replicate based on the mean and standard deviation of the design-based index.

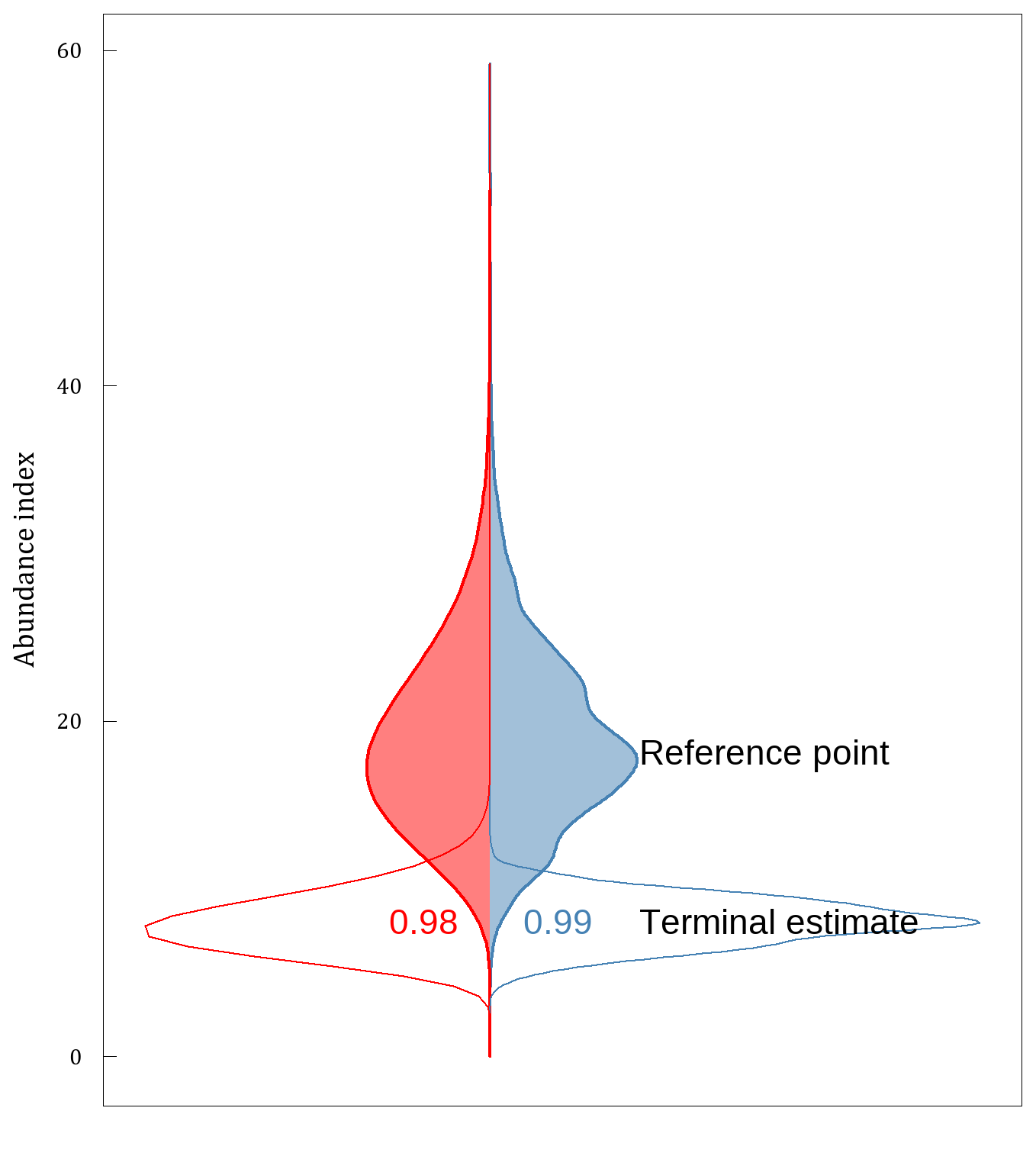


Fig 2: Bootstrap (blue) and gamma (red) distributions estimated from simulation 1 of a redfish-like population, where terminal estimates (year 20; open area) are compared to a reference period (aggregate estimates from years 2-9; shaded area). Densities for the reference period were obtained by combining the bootstrap samples and by aggregate parameters across the reference period (see Methods section). Probability that the terminal value is below the reference point is indicated.

# Appendix A

TODO

# Colophon

This version of the document was generated on 2022-06-20 12:51:33 using the R markdown template for SCR documents from [NAFOdown](https://github.com/nafc-assess/NAFOdown).

The computational environment that was used to generate this version is as follows:

#> - Session info ---------------------------------------------------------------  
#> setting value  
#> version R version 4.1.2 (2021-11-01)  
#> os Windows 10 x64 (build 19042)  
#> system x86\_64, mingw32  
#> ui RTerm  
#> language (EN)  
#> collate English\_United States.1252  
#> ctype English\_United States.1252  
#> tz America/St\_Johns  
#> date 2022-06-20  
#> pandoc 2.14.0.3 @ C:/Program Files/RStudio/bin/pandoc/ (via rmarkdown)  
#>   
#> - Packages -------------------------------------------------------------------  
#> package \* version date (UTC) lib source  
#> assertthat 0.2.1 2019-03-21 [1] CRAN (R 4.1.2)  
#> base64enc 0.1-3 2015-07-28 [1] CRAN (R 4.1.0)  
#> bookdown 0.24 2021-09-02 [1] CRAN (R 4.1.1)  
#> cachem 1.0.6 2021-08-19 [1] CRAN (R 4.1.1)  
#> callr 3.7.0 2021-04-20 [1] CRAN (R 4.1.1)  
#> cli 3.1.0 2021-10-27 [1] CRAN (R 4.1.2)  
#> colorspace 2.0-2 2021-06-24 [1] CRAN (R 4.1.1)  
#> crayon 1.4.1 2021-02-08 [1] CRAN (R 4.1.1)  
#> data.table 1.14.0 2021-02-21 [1] CRAN (R 4.1.1)  
#> DBI 1.1.1 2021-01-15 [1] CRAN (R 4.1.1)  
#> desc 1.3.0 2021-03-05 [1] CRAN (R 4.1.1)  
#> devtools 2.4.3 2021-11-30 [1] CRAN (R 4.1.2)  
#> digest 0.6.29 2021-12-01 [1] CRAN (R 4.1.2)  
#> dplyr 1.0.7 2021-06-18 [1] CRAN (R 4.1.1)  
#> ellipsis 0.3.2 2021-04-29 [1] CRAN (R 4.1.1)  
#> evaluate 0.14 2019-05-28 [1] CRAN (R 4.1.1)  
#> fansi 0.5.0 2021-05-25 [1] CRAN (R 4.1.1)  
#> farver 2.1.0 2021-02-28 [1] CRAN (R 4.1.1)  
#> fastmap 1.1.0 2021-01-25 [1] CRAN (R 4.1.1)  
#> flextable \* 0.6.9 2021-10-07 [1] CRAN (R 4.1.0)  
#> fs 1.5.2 2021-12-08 [1] CRAN (R 4.1.2)  
#> gdtools 0.2.3 2021-01-06 [1] CRAN (R 4.1.1)  
#> generics 0.1.1 2021-10-25 [1] CRAN (R 4.1.2)  
#> ggplot2 \* 3.3.5 2021-06-25 [1] CRAN (R 4.1.1)  
#> ggridges 0.5.3 2021-01-08 [1] CRAN (R 4.1.3)  
#> ggthemes 4.2.4 2021-01-20 [1] CRAN (R 4.1.3)  
#> glue 1.4.2 2020-08-27 [1] CRAN (R 4.1.1)  
#> gtable 0.3.0 2019-03-25 [1] CRAN (R 4.1.1)  
#> here \* 1.0.1 2020-12-13 [1] CRAN (R 4.1.1)  
#> highr 0.9 2021-04-16 [1] CRAN (R 4.1.1)  
#> htmltools 0.5.2 2021-08-25 [1] CRAN (R 4.1.1)  
#> knitr 1.34 2021-09-09 [1] CRAN (R 4.1.1)  
#> labeling 0.4.2 2020-10-20 [1] CRAN (R 4.1.0)  
#> lifecycle 1.0.1 2021-09-24 [1] CRAN (R 4.1.3)  
#> magrittr 2.0.1 2020-11-17 [1] CRAN (R 4.1.1)  
#> memoise 2.0.1 2021-11-26 [1] CRAN (R 4.1.2)  
#> munsell 0.5.0 2018-06-12 [1] CRAN (R 4.1.1)  
#> NAFOdown \* 0.0.1.9000 2022-06-13 [1] local  
#> officer 0.4.0 2021-09-06 [1] CRAN (R 4.1.1)  
#> pillar 1.6.2 2021-07-29 [1] CRAN (R 4.1.1)  
#> pkgbuild 1.2.0 2020-12-15 [1] CRAN (R 4.1.1)  
#> pkgconfig 2.0.3 2019-09-22 [1] CRAN (R 4.1.1)  
#> pkgload 1.2.2 2021-09-11 [1] CRAN (R 4.1.0)  
#> plyr 1.8.6 2020-03-03 [1] CRAN (R 4.1.1)  
#> prettyunits 1.1.1 2020-01-24 [1] CRAN (R 4.1.1)  
#> processx 3.5.2 2021-04-30 [1] CRAN (R 4.1.1)  
#> ps 1.6.0 2021-02-28 [1] CRAN (R 4.1.1)  
#> purrr 0.3.4 2020-04-17 [1] CRAN (R 4.1.1)  
#> R6 2.5.1 2021-08-19 [1] CRAN (R 4.1.1)  
#> Rcpp 1.0.7 2021-07-07 [1] CRAN (R 4.1.1)  
#> remotes 2.4.0 2021-06-02 [1] CRAN (R 4.1.1)  
#> rlang 1.0.2 2022-03-04 [1] CRAN (R 4.1.3)  
#> rmarkdown 2.11 2021-09-14 [1] CRAN (R 4.1.1)  
#> rprojroot 2.0.2 2020-11-15 [1] CRAN (R 4.1.1)  
#> rstudioapi 0.13 2020-11-12 [1] CRAN (R 4.1.1)  
#> scales 1.1.1 2020-05-11 [1] CRAN (R 4.1.1)  
#> sessioninfo 1.2.2 2021-12-06 [1] CRAN (R 4.1.2)  
#> showtext 0.9-4 2021-08-14 [1] CRAN (R 4.1.1)  
#> showtextdb 3.0 2020-06-04 [1] CRAN (R 4.1.1)  
#> stringi 1.7.4 2021-08-25 [1] CRAN (R 4.1.1)  
#> stringr 1.4.0 2019-02-10 [1] CRAN (R 4.1.1)  
#> sysfonts 0.8.5 2021-08-09 [1] CRAN (R 4.1.1)  
#> systemfonts 1.0.3 2021-10-13 [1] CRAN (R 4.1.2)  
#> testthat 3.1.1 2021-12-03 [1] CRAN (R 4.1.2)  
#> tibble 3.1.4 2021-08-25 [1] CRAN (R 4.1.1)  
#> tidyselect 1.1.1 2021-04-30 [1] CRAN (R 4.1.1)  
#> usethis 2.0.1 2021-02-10 [1] CRAN (R 4.1.1)  
#> utf8 1.2.2 2021-07-24 [1] CRAN (R 4.1.1)  
#> uuid 0.1-4 2020-02-26 [1] CRAN (R 4.1.1)  
#> vctrs 0.3.8 2021-04-29 [1] CRAN (R 4.1.1)  
#> withr 2.4.3 2021-11-30 [1] CRAN (R 4.1.2)  
#> xfun 0.26 2021-09-14 [1] CRAN (R 4.1.0)  
#> xml2 1.3.2 2020-04-23 [1] CRAN (R 4.1.1)  
#> yaml 2.2.1 2020-02-01 [1] CRAN (R 4.1.0)  
#> zip 2.2.0 2021-05-31 [1] CRAN (R 4.1.1)  
#>   
#> [1] C:/Users/RegularP/Documents/R/win-library/4.1  
#> [2] C:/Program Files/R/R-4.1.2/library  
#>   
#> ------------------------------------------------------------------------------