SCRS/2022/105

EFFICACY OF A BYCATCH ESTIMATION TOOL

Elizabeth A. Babcock[1] , William J. Harford[2], Todd Gedamke[3], Derek Soto[4], C. Phillip Goodyear[5]

SUMMARY

The bycatch estimation tool developed by Babcock (2022) was subjected to simulation testing using the species distribution model and longline simulator (LLSIM) developed by Goodyear (2021). To evaluate the efficacy of the bycatch estimation tool, generalized representations of ICCAT CPC longline fisheries were created using LLSIM and were coupled with alternative representations of observer programs to produce simulated logbook and observer databases for a range of observer coverage levels and allocation methods. Using a semi-automated model selection process, linear predictors based on negative binomial and delta lognormal models were used to predict total annual bycatch of blue marlin from the simulated datasets. A stratified ratio estimator was also used for comparison. Across representations of observer programs, bycatch estimates were reasonably unbiased, with diminishing variation in bias estimates as observer coverage increased. The use of simulated data sets provides a demonstration of the utility of the bycatch estimation tool as well as evaluation of its reliability.

KEYWORDS

Bycatch, catch statistics, simulation, model testing, LLSIM, blue marlin, swordfish

1. Introduction

A statistical framework for bycatch estimation has been developed by Babcock (2022) and Babcock and Goodyear (2021). This bycatch estimation tool utilizes a model-based procedure to estimate total annual bycatch by expanding a sample, such as an observer database, in relation to total effort from logbooks or landings records. This framework can also be used to estimate an annual index of abundance, calculated only from the observer data. Using this tool, bycatch estimation is carried out by fitting a generalized linear model based on user-defined statistical distributions for the observation error models (e.g. delta-lognormal, and negative binomial) and predictor variables (e.g., year, season, depth). The complexity of the task of identifying a best approximating model is addressed through a semi-automated model selection process based on the user's choice of information criteria (AICc, AIC or BIC). Once a best approximating model is identified, the GLM is used to predict total bycatch in all logbook trips (or only unsampled trips, if desired) and total bycatch is estimated summing across trips. For comparison, the code also calculates simple design-based ratio estimators. This bycatch estimation tool has been developed as an accessible R package (Babcock 2022).

To evaluate the efficacy of the bycatch estimation tool for bycatch species such as billfish (Blue Marlin, White Marlin and Round-scale Spearfish), simulation testing was carried out by coupling the longline simulator LLSIM (Goodyear 2021) with an additional simulation model that produces observer databases. Statistical vetting of the bycatch estimation tool was carried out by creating different scenarios of observer program designs and varying spatiotemporal allocation, allocation based on historical catches, and coverage levels. These simulated observer databases, along with simulated logbooks, were used to calculate bias in annual bycatch estimates relative to 'known true' bycatch from LLSIM. Generalized representations of ICCAT CPC longline fisheries were created in a three-step process. First, the LLSIM platform was used to simulate representations of ICCAT CPC longline fisheries (Goodyear 2021). Second, simulated observer programs were specified in a manner consistent with current, realistic scenarios of CPC data collection programs. Finally, bycatch estimation was conducted

---

[1]Department of Marine Biology and Ecology, Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149. USA. ebabcock@rsmas.miami.edu
[2] Nature Analytics, 551 Lakeshore Rd E, Suite 105, Mississauga, Ontario, Canada, L5G 0A8. bill@natureanalytics.ca
[3]MER Consultants, 5521 SE Nassau Terrace, Stuart, Florida, 34997 USA, todd@merconsultants.org
[4] Department of Science and Technology, Inter American University of Puerto Rico, San Germán Campus, PO Box 5100, San Germán, PR 00683, USA. derek_soto@intersg.edu
[5]686 Hickory LN, Havana, Florida 32333, USA. phil_goodyear@msn.com

according to design-based estimators and model-based estimators and compared against ‘known true’ bycatch. Bycatch estimators included statistical methods used by ICCAT CPCs for bycatch estimation and CPUE standardization (Brown 2011, Porter et al. 1999, Forrestal et al. 2019).

2.Methods

2.1 Longline fishery simulation (LLSIM)

LLSIM was used to simulate three idealized fleets as described by Goodyear (2021). These fleets are a USA-like fleet (fleet 1), Japan-like fleet (fleet 2) and a Brazil-like fleet (fleet 3), with data spanning from 1990 to 2018 to reflect the approximate period for which observer coverage has been established (e.g. Diaz et al. (2009)). The species distribution model (SDM) generates a 3-dimensional distribution of blue marlin and swordfish throughout the Atlantic Ocean based on the habitat preferences of the species (Goodyear 2016; Forrestal and Schrippa 2019). LLSIM then simulates longline sets by distributing hooks throughout the habitat of the species, consistent with the distribution, gear, hooks between floats, use of light sticks and other characteristics of historical longline fishing fleets. The probability of each hook capturing a blue marlin or swordfish is then determined by the location of the hook and the probability of fish presence from the SDM (Goodyear 2021).

While LLSIM initially produces set-level catches, sets were allocated to simulated trips to more accurately reflect the fact that observers are allocated randomly by trips rather than sets. The method of Grüss et al. (2019) was used to allocate sets to the same trip if they were in the same gear, month and spatial area (5 x 5 squares). Trips with more than 100 sets were randomly allocated to different trips so that the median trips had about 20 sets. This algorithm allowed for correlation among sets in the same trip to introduce potential clustering bias into the simulated observer data. These data were aggregated to the trip level before randomly assigning trips to be observed or not, and the trip identifier was retained so that the trips that were observed could be matched to the corresponding logbook trips.

Additionally, LLSIM data aggregated to the trip-level are used as the ‘known true’ blue marlin bycatch in subsequent calculation of estimation bias. **Figure 1** shows the annual trends in blue marlin catch in numbers (**Figure 1A**), swordfish catch in numbers (**Figure 1B**), and fishing effort (**Figure 1C**), as well as the spatial distribution of fishing effort (**Figure 1D**).

2.2 Observer program simulation

The datasets simulated using LLSIM were then provided to the observer program sub-model. Four basic allocation schemes were used to generate observer data sets. The probability of an individual trip being sampled was based on the following:

- Random allocation (5 runs at 19 levels: levels: 5% to 95% by 5%)
- Proportional allocation based on historical swordfish catch (5 runs at 19 levels: 5% to 95% by 5%)
- Proportional allocation based on historical blue marlin catch (5 runs at 19 levels: 5% to 95% by 5%)
- Spatial-temporal allocation based on approximate annual coverage of the three idealized fleets (100 runs)

In the random allocation model, all fleets and all trips were assigned a probability of being sampled from 5% to 95%, by 5% intervals, and trips were randomly selected until the target simulate observer coverage was reached. In the proportional allocation to historical swordfish, and the proportional allocation to blue marlin catch scenarios, the model set the overall coverage at each 5% level as in the random allocation, but trips were selected as a function of swordfish or blue marlin catch. For the swordfish scenario, the logit of the probability of a trip being observed increased linearly with swordfish catch, scaled so that the overall coverage was approximately at the specified level (5% to 95%). This scenario is meant to represent the case where vessels that are larger or take longer trips are more likely to be observed due to logistical constraints. For the blue marlin scenario, the logit of the probability of a trip being observed increased with blue marlin catch, to represent an observer program in which observers are preferentially allocated to sectors with more bycatch to improve bycatch estimates. Together, the two proportional allocations scenarios represent biased (i.e., nonrandom) but potentially realistic allocations that CPC data collection programs may deploy.

Finally, the spatial-temporal allocation coverage is the most “realistic” scenario we explored with values for 3 fleets based on those reported by Diaz et al. (2009), Anonymous (2012 and 2014), and the 2017 -2019 annual reports to ICCAT by USA and Japan. A linear interpolation was used to fill in unknown years for the two fleets which showed increases from 1% to approximately 9% and 13% for each. For the Brazil-like fleet, no values were available, so we included a linear increase from 1% in 1990 to 5% in 2019. See **Figure 2** for examples of allocations from each scenario. This scenario was intended to capture the general trend of increasing observer coverage in the fleets, but it does not capture nuances such as the variability in trip duration between fleets, and the fact that coverage levels may vary within fleets depending on target species and other factors.

2.3 Bycatch estimation tool

The bycatch estimation tool estimates total bycatch as follows. First, mean catch per unit effort (CPUE) of observed sample units (trips in this example) is estimated from a linear model with predictor variables in R (R Core Team 2020). The observation error models that can be used include delta-lognormal, delta-gamma, negative binomial (from either glm.nb in the MASS library or glmmTMB, nbinom1 and nbinom2) and Tweedie (from cpglm or glmmTMB) (Brooks et al. 2017, Dunn and Smyth, 2005, Venables and Ripley 2002, Zhang 2013). Within each observation error model group, potential predictor variables are chosen based on the user's choice of information criteria (AICc, AIC or BIC). The user specifies a most complex and simplest model, and all intermediate models are considered using the information criterion with the dredge function in the MuMIn library (Barton, 2020). The best candidate models in each observation error group may then be compared using 10-fold cross-validation to see which observation error model best predicts CPUE. The best model according to cross-validation is the one with the lowest root mean square error (RMSE) in the predicted CPUE and mean error (ME) closest to zero. Cross validation, rather than information criteria, is needed to compare between observation error models because information criteria may not be used to compare models with different structures in the likelihood (e.g. delta models vs. models that fit all zero and non-data simultaneously).

For the best model in each observation error model group, the total bycatch may be estimated by predicting the catch in all logbook trips or sets (i.e., the whole fishery) from the fitted model and summing across all effort in each year. Alternatively, the model can be used to predict bycatch in only the unobserved effort, which can be added to the observed bycatch as a known constant. The bycatch in each sample unit is predicted directly by the negative binomial models. Tweedie models predict CPUE, which is then multiplied by effort. Delta-lognormal and delta-gamma models have separate components for the probability of a positive CPUE and the CPUE, which must be multiplied together (with appropriate bias corrections) and multiplied by effort to get the total catch. Catch is predicted in each trip in the logbook data using the model fitted to the observer data, and catch is summed across trips to get the total catch in each year. Because the model-predicted catches are not independent between trips, the total variance of the bycatch must be estimated with a formula that includes the covariance among trips. This is done using either a Monte Carlo simulation method or a delta method. See the User's Guide (Babcock 2022) for details.

The software also calculates an annual abundance index from the same models that were selected for bycatch estimation. An annual abundance index is calculated by setting all variables other than year, and any variables required by the user to be included in the index (e.g. region or fleet) to a reference level, which is the mean for numerical variables or the most common value for categorical variables. The index and its standard error are then predicted at these reference levels.

In addition to the total bycatch estimates, the model outputs diagnostics including plots of the residuals calculated using the DHARMa R library (Hartig 2020). The DHARMa library uses simulation to generate scaled residuals based on the specified observation error model so that the results are more clearly interpretable than ordinary residuals. DHARMa draws random predicted values from the fitted model to generate an empirical predictive density for each data point and then calculates the fraction of the empirical density that is greater than the true data point. Values of 0.5 are expected, and values near 0 or 1 indicate a mismatch between the data and the model. A model that does not appear to fit well according to these diagnostics should not be used for bycatch prediction.

For cross-validation, the observer data are randomly divided into 10 folds. Each fold is left out one at a time and the models are fit to the other 9 folds. The same procedure described above is used to find the best model within each observation error group using information criteria and the MuMIn library. The fitted model is used to predict the CPUE for the left-out fold, and the model with the lowest mean RMSE across the 10 folds is selected as the best model. Mean error is also calculated as an indicator of whether the model has any systematic bias.

Finally, for comparison to the model-based total bycatch estimate, the software calculates total bycatch using a simple ratio estimator, stratified by variables input by the user. The software does not include any method to impute the ratio in unsampled strata, so the method should not be used with many stratification variables at low observer coverage levels. The ratio estimator calculates the ratio $\hat{R}_i$ in each stratum, *i,* of mean observed bycatch ($\bar{c}_i$) to mean observed effort ($\bar{e}_i$, in number of hooks) in the observed data for each strata as $\hat{R}_i = {}^{\bar{c}_i}/_{\bar{e}_i}$ so that the total bycatch $\hat{C}$is calculated as the sum of the ratio times total logbook effort E_i , so that $\hat{C} = \sum_{i=1}^{n} \hat{R}_i E_i$, and the variance of the total is (Rao 2000):

$$V(\hat{C}) = \sum_{i=1}^{n} E_i^2 \frac{(1-f)}{n} \left(s_{c+}^2 \hat{R}_i^{\,2} s_e^2 - 2\hat{R}_i s_{ce}\right)$$

were *f* is the fraction of the effort observed, n is the number of sample units observed, and s refers to the standard deviations and covariances between effort and catch in the observed data.

2.4 Scenario exploration

2.4.1 Observer coverage levels

Coarse patterns of bias in bycatch estimation were examined using coverage steps of 5% from 5% to 95% for observer program models of random allocation, swordfish catch allocation, and blue marlin catch allocation. In this step, 5 simulation runs were performed for each observer program model and level of observer coverage. This evaluation was carried out to ensure estimation models were producing, on average, unbiased bycatch estimates, with diminishing variation in bias estimates as observer coverage increased, and to explore the role that observer allocation to trips may have in bycatch estimation reliability.

In this evaluation, bycatch estimates were made using the stratified ratio estimator, delta-lognormal, and negative binomial observation error models (negative binomial 2 from the glmmTMB library, Brooks et al. 2017). The stratified ratio estimator was specified using the variables included in the simplest linear predictor model. For the delta-lognormal and negative binomial observation models, the simplest linear predictor included year, fleet, and area (North vs. South Atlantic). The most complex model additionally included hooks between floats (i.e., trip-level median hooks between floats, an indicator of depth of fishing) and season. When predicting bycatch using delta-lognormal and negative binomial observation error models, variables used in prediction were those identified as the best approximating model based on BIC for the respective observation error model in each simulation run. Two forms of bycatch estimation are shown: those that predict total bycatch in all trips (not just unobserved trips) and those where observed catches were included in the estimates as a known constant.

2.4.2 Random allocation scheme

Building upon the previous section, a more in-depth exploration of bias in bycatch estimation was conducted using random allocation to trips with coverage levels of 5% and 10%. This evaluation was carried out to quantify bias that could be expected at reasonably realistic levels of observer coverage. In this evaluation, 100 simulation runs were performed, and total bycatch was predicted in all trips (not just unobserved trips). The selection of linear predictors was also examined by exploring the frequency with which the variables of hooks between floats (i.e., trip-level median hooks between floats) and season were included in best approximating models for each observation error type.

2.4.3 Spatial-temporal allocation scheme

To extent possible, the spatial-temporal allocation scheme was specified to simulated observer programs that encompassed true historical coverage levels of CPC data collection programs. This allocation scheme was used to examine the bycatch estimation tool's ability to choose the best observation error model by cross-validation, and also the accuracy of the variance estimates. One hundred runs were performed and included a more diverse set of observation error model types. Further, two forms of bycatch estimation are reported: those that predict total bycatch in all trips (not just unobserved trips) and those where observed catches were included in the estimates as a known constant.

For each of 100 runs, cross-validation was used to select the best model among the options of delta-lognormal, delta-gamma, and the glmmTMB versions of negative binomial 1, negative binomial 2, and Tweedie (Babcock

2022). Variances were calculated by the Monte Carlo simulation method, which involves drawing 1000 values of each of the model coefficients and then drawing simulated values of the predicted catches in each unobserved trip. The accuracy of the variance estimate was evaluated by calculating the coverage for each observation error and model year, where coverage is defined as the fraction of the 100 draws in which the estimated 95% confidence interval contained the true value. This should be around 95% for a correctly specified model.

3. Results

3.1 Bycatch estimation tool

Example outputs produced by the bycatch estimation tool for one draw of the spatiotemporal allocation scenario are provided to illustrate use of all the features of the tool (**Tables 1 & 2; Figures 3, 4, 5**). In this example, cross-validation plots demonstrate that the delta-gamma model best predicts CPUE according to both RMSE and ME (**Table 2, Figure 3**). The other models also performed well. All models predicted very similar trends in the total estimated catch of blue marlin, very close to the correct values, with the width of the 95% confidence interval decreasing as observer coverage increased over time (**Figure 4**). The DHARMa residuals showed adequate model specification in the binomial and gamma models as indicated by the linear qq plot of the scaled residuals (**Figure 5**).

3.2 Scenario exploration

3.2.1 Observer coverage levels

Using 5 simulation runs across observer coverage levels, from 5% to 95%, resulted in reasonably unbiased bycatch estimates, with diminishing variation in bias as observer coverage increased (**Figures 6 through 9**; predictions made for all trips). For comparison, the results are also presented for catch prediction where observed catches were included in the estimates as a known constant (**Figures 10 through 12**). Note that in these figures, year-specific plots are shown for years 2000 and 2010, which were chosen arbitrarily to be representative of bycatch estimation bias.

Each estimation model (i.e., stratified ratio, negative binomial, and delta lognormal) reproduced temporal trends in bycatch, including at 5% observer coverage (**Figures 6B, 7B, 8B**). Among alternative models of observer allocation to trips that were used in this comparison (i.e., random allocation, swordfish catch allocation, and blue marlin catch allocation), minor differences in bycatch estimation bias were observed (**Figures 6A, 7A, 8A, & 9**). While random observer allocation demonstrated unbiased bycatch estimates, especially when observed catch was included in bycatch estimation (**Figures 10 & 11**), systematic bias of bycatch occurred for the swordfish catch-based observer allocation and blue marlin catch-based observer allocation.

3.2.2 Random allocation scheme

The more in-depth exploration of random observer allocation to trips (i.e., random allocation with 5% coverage, random allocation with 10% coverage) reinforced the finding that the bycatch estimation tool produced reasonably unbiased bycatch estimates. Results are presented where catch predictions were made for all trips. The stratified ratio estimator and negative binomial estimation model produced, on average, unbiased bycatch estimates across all years (**Table 3; Figure 13**). The delta lognormal estimation model produced, on average, a positive bias of approximately 7%.

Figure 14 highlights the distribution of bias as well as its central tendency (across 100 simulation runs) for each year from 1990 to 2015. Visual examination of these trends suggests that the distribution of bias produced by the stratified ratio estimator was approximately centered at 0% in all years and across models of observer allocation. Conversely, negative binomial and delta lognormal estimation models consistently produced underestimates of bycatch in some years and overestimates in others.

In examining the linear predictors included in best approximating model of each observation error structure, the negative binomial estimation model frequently included both of the predictors included in the most complex formulation (i.e., season and hooks between floats) (**Table 4**). Likewise, the delta lognormal estimation model frequently included hooks between floats and season in the binomial component, with hooks between floats also commonly included in the lognormal component of this observation error model.

3.2.3 Spatial-temporal allocation scheme

Like the random allocation scheme, spatial-temporal allocation produced reasonably unbiased bycatch estimates for the stratified ratio estimator and negative binomial estimation model, with the delta lognormal model having a small positive bias (**Figure 15; Table 5**). Temporal trends in bycatch estimates tended to be centered at zero percent for the stratified ratio estimator, while negative binomial and delta lognormal estimation models consistently produced underestimates of bycatch in some years and overestimates in others (**Figure 16**). Further, results were consistent between forms of bycatch estimation (i.e., prediction of total bycatch from all trips versus inclusion of observed catches as a known constant). This result is unsurprising for the range of percent coverage included in the spatial-temporal allocation model; however, inclusion of observed catches as a known constant remains an important consideration as percent coverage increases (e.g., compare **Figure 8A** and **Figure 11A**).

The five tested model types produced very similar distributions of RMSE across the 100 draws, but ME was more variable showing a slight consistent positive bias in negative binomial 1 and delta-lognormal, and a consistent negative bias in negative binomial 2 (**Figure 17**). Both RMSE and ME consistently preferred the delta-gamma model, although all of the other models were also preferred occasionally. Looking at individual draws, RMSE was much more consistent between models in the same draw, while the opposite was true for ME (**Figure 18**). The quality of the variance estimates varied by models, with coverage levels close to the correct value of 95% in most years for the delta models and negative binomial 2, but not Tweedie or negative binomial 1 (**Figure 19**). In general, the bycatch estimates were reasonably good even at realistically low coverage levels.

4. Discussion

The use of simulated data sets provided a demonstration of the utility of the bycatch estimation tool as well as evaluation of its reliability. Overall, the semi-automated model selection process produced reasonably unbiased estimates of bycatch, highlighted by diminishing variation in bias as observer coverage increased. Furthermore, the stratified ratio estimator and negative binomial estimation model more routinely produced unbiased bycatch estimates than did the delta lognormal estimation model across years.

The tool provided reasonably good estimates of variance in the total bycatch estimates, particularly for the models that were preferred by cross-validation, such as the delta-gamma. Also, the total bycatch estimates were fairly consistent across estimation methods (e.g. **Figure 4**), implying that the choice of error model may be less important than the specifics of the observer program design (coverage level, allocation over space and time). As expected, the precision of the estimates was strongly influenced by observer coverage levels. At low levels of around 5%, sampling error caused substantial variability between draws in the total bycatch estimates. In the realistic spatiotemporal allocation scenario, in which coverage increased from around 1% in 1990 to up to 13% in some fleets in 2018, the estimates became more precise over time. This indicates that the increase in coverage levels over time in many fleets is likely to improve bycatch estimates substantially.

Recent improvements to the bycatch estimation tool, including code refactoring and parallelization, have been integrated into an R library to make this framework more easily accessible (Babcock 2022). Additional improvements could be added that would be useful for abundance index estimation, such as random effects, GAMS, and spatial correlation. If other methods are being used or considered by ICCAT members, they could be added to the tool.

5. Acknowledgements

This work was developed with financial support from the ICCAT Science Budget and the European Union Grant Agreement SI2.839159 - Strengthening the scientific basis for decision-making in ICCAT. E. A. Babcock's work was supported by NOAA, via the Cooperative Institute for Marine and Atmospheric Science and Goodyear's work was supported by The Billfish Foundation, Ft. Lauderdale, FL, USA.

6. References

Anonymous. 2012. Report of Japan's scientific observer program for the tuna longline fishery in the Atlantic ocean in the fishing years of 2008 to 2010. Collect. Vol. Sci. Pap. ICCAT, 68(5): 2025-2039 (SCRS/2011/200)

Anonymous. 2014. Report of Japan's scientific observer program for the tuna longline fishery in the Atlantic ocean in the fishing years of 2011 and 2012. Collect. Vol. Sci. Pap. ICCAT,70(6): 2845-2855 (SCRS/2013/186)

Babcock, E.A. 2022. Bycatch Estimator R library. https://ebabcock.github.io/BycatchEstimator/

Babcock, E.A. and C. P. Goodyear. 2021. Testing a bycatch estimation tool using simulated blue marlin longline data. ICCAT Collect. Vol. Sci. Pap. ICCAT, 78(5): 179-189

Barton, K. 2020. MuMIn: Multi-Model Inference. https://CRAN.R-project.org/package=MuMIn.

Brooks, M. E., K. Kristensen, K. J. van Benthem, A. Magnusson, C. W. Berg, A. Nielsen, H. J. Skaug, M. Maechler, and B. M. Bolker. 2017. glmmTMB Balances Speed and Flexibility Among Packages for Zero-Inflated Generalized Linear Mixed Modeling. The R Journal 9 (2): 378–400.

Brown, C.A. 2001. Revised estimates of bluefin tuna dead discards by the U.S. Atlantic pelagic longline fleet, 1992-1999. Collect. Vol. Sci. Pap. ICCAT 52(3):1007-1021.

Diaz, G.A., L. R. Beerkircher & V. R. Restrepo. 2009. Description of the U.S. Pelagic Observer Program (POP). Collect. Vol. Sci. Pap. ICCAT 64(7): 2415-2426

Dunn, P. K., and G. K. Smyth. 2005. Series Evaluation of Tweedie Exponential Dispersion Models. Statistics and Computing 15: 267–80.

Forrestal, F.C., and Schirripa, M.J. 2020. Addition of Swordfish distribution model to longline simulator study. Collect. Vol. Sci. Pap. ICCAT, 77(5): 37-66.

Forrestal, F. C., M. Schirripa, C. P. Goodyear, H. Arrizabalaga, E. A. Babcock, R. Coelho, W. Ingram, M. Lauretta, M. Ortiz, R. Sharma, and J. Walter. 2019. Testing robustness of CPUE standardization and inclusion of environmental variables with simulated longline catch datasets. Fisheries Research 210:1-13.

Goodyear, C.P. 2021. Development of new model fisheries for simulating longline catch data with LLSIM. Collect. Vol. Sci. Pap. ICCAT, 78(5): 53-62.

Goodyear, C. P., M. Schirripa, and F. Forrestal. 2017. Longline data simulation: a paradigm for improving CPUE standardization. Collect. Vol. Sci. Pap. ICCAT 74:379-390.

Goodyear, C.P., 2016. Modeling the time-varying density distribution of highly migratory species: Atlantic blue marlin as an example. Fish. Res. 183, 469–481.

Grüss, A., J. F. Walter, E. A. Babcock, F. C. Forrestal, J. T. Thorson, M. V. Lauretta, and M. J. Schirripa. 2019. Evaluation of the impacts of different treatments of spatio-temporal variation in catch-per-unit-effort standardization models. Fisheries Research 213:75-93.

Hartig, F. 2020. DHARMa: Residual Diagnostics for Hierarchical (multi-Level / Mixed) Regression Models. https://CRAN.R-project.org/package=DHARMa.

Porter, J.M., Wood, B.M., Stone, H.H. 1999. Preliminary estimation of the tonnage of dead swordfish and bluefin tuna discards from the 1998 Canadian swordfish longline fishery. Collect. Vol. Sci. Pap. ICCAT 51(5):1460-1468.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rao, P. S. R. S. 2000. Sampling Methodologies with Applications. Texts in Statistical Science. Chapman and Hall/ CRC.

Venables, W. N., and B. D. Ripley. 2002. Modern Applied Statistics with S. Fourth Edition. New York: Springer.

Zhang, Y. 2013. Likelihood-Based and Bayesian Methods for Tweedie Compound Poisson Linear Mixed Models. Statistics and Computing 23: 743-757.

Table 1. Example of summary of the simulated observer and logbook data for blue marlin, for one simulation with the spatiotemporal allocation of observer coverage. Cat Est is the blue marlin bycatch estimated by a simple ratio estimator stratified only by year.

Year	Obs Cat	Obs Eff	Obs trips	CPUE	Pos	Pos Frac	Effort	trips	trips Obs Frac	Cat Est	Cat se
1990	258	793	30	0.18	16	0.53	89813	3218	0.01	29210	5721
1991	296	1035	36	0.28	25	0.69	93130	3224	0.01	26643	3246
1992	598	1922	53	0.24	29	0.55	84827	3148	0.02	26386	7150
1993	641	2587	89	0.3	52	0.58	112551	3237	0.03	27892	2813
1994	301	1706	88	0.35	40	0.45	105370	3374	0.03	18587	3564
1995	461	2791	97	0.14	52	0.54	109494	3471	0.03	18085	2584
1996	524	3155	79	0.2	55	0.7	121937	4019	0.02	20251	2370
1997	442	2846	78	0.19	49	0.63	102645	3542	0.02	15944	4004
1998	580	3353	93	0.16	52	0.56	112927	3658	0.03	19531	4286
1999	357	2672	97	0.22	63	0.65	97008	3392	0.03	12959	1729
2000	753	3781	102	0.23	70	0.69	103886	3278	0.03	20688	3291
2001	736	3511	103	0.24	70	0.68	90824	2939	0.04	19039	2016
2002	555	2905	96	0.2	66	0.69	72489	2474	0.04	13849	1513
2003	613	4203	131	0.09	54	0.41	88922	2777	0.05	12970	2969
2004	494	4013	135	0.13	67	0.5	98444	2912	0.05	12117	1901
2005	704	4982	142	0.15	80	0.56	88131	2735	0.05	12453	1338
2006	467	4439	122	0.09	63	0.52	83396	2591	0.05	8773	1805
2007	433	4644	136	0.08	77	0.57	78185	2416	0.06	7289	1233
2008	532	4424	149	0.08	72	0.48	86300	2492	0.06	10377	2107
2009	470	5602	160	0.06	76	0.48	70920	2213	0.07	5950	832
2010	630	7544	204	0.07	103	0.5	72313	2156	0.09	6039	658
2011	152	2480	111	0.04	34	0.31	64880	2073	0.05	3976	760
2012	155	2844	136	0.05	47	0.35	67809	2212	0.06	3696	624
2013	102	2891	140	0.05	46	0.33	58227	1988	0.07	2054	250
2014	130	3323	153	0.06	52	0.34	57166	1971	0.08	2237	317
2015	92	2832	142	0.03	40	0.28	46652	1683	0.08	1515	227
2016	125	2643	70	0.05	44	0.63	43522	1013	0.07	2059	282
2017	124	3672	81	0.04	41	0.51	46431	1041	0.08	1568	306
2018	75	2940	75	0.03	38	0.51	49482	1192	0.06	1262	181

Table 2. Example of best models of each type according to the BIC, along with root mean square error (RMSE) and mean error (ME) from the cross-validation.

Model	Formula	RMSE	ME
Binomial	hbf + season + 1 + area + fleet + Year	NA	NA
Lognormal	1 + area + fleet + Year	0.292	0.001
Gamma	1 + area + fleet + Year	0.292	-0.001
TMBnbinom1	hbf + season + 1 + area + fleet + Year + offset(log(Effort))	0.293	0.010
TMBnbinom2	1 + area + fleet + Year + offset(log(Effort))	0.293	-0.009
TMBtweedie	1 + area + fleet + Year	0.293	0.000

Table 3. Summary of percent bias for random allocation to trips with 5% observer coverage and random allocation to trips with 10% observer coverage - predictions made for all trips. Estimation models are: stratified ratio estimator, negative binomial, and delta-lognormal. Shown are results across all years 'All', and for years 2000 and 2010 across 100 simulation runs.

				Centered 95%	
Year	**Observer allocation**	**Estimation model**	**Median**	**Lower**	**Upper**
All	Random allocation 5%	Stratified ratio	-0.76	-27.79	30.23
All	Random allocation 5%	Negative binomial	-0.12	-26.63	30.37
All	Random allocation 5%	Delta lognormal	7.92	-18.88	40.24
All	Random allocation 10%	Stratified ratio	-0.06	-19.48	21.25
All	Random allocation 10%	Negative binomial	0.46	-22.67	21.77
All	Random allocation 10%	Delta lognormal	7.70	-13.78	30.32
2000	Random allocation 5%	Stratified ratio	-0.03	-29.57	28.35
2000	Random allocation 5%	Negative binomial	-3.10	-24.81	15.02
2000	Random allocation 5%	Delta lognormal	0.61	-18.22	19.92
2000	Random allocation 10%	Stratified ratio	1.10	-17.75	21.15
2000	Random allocation 10%	Negative binomial	-4.06	-17.44	12.15
2000	Random allocation 10%	Delta lognormal	-0.03	-14.92	16.11
2010	Random allocation 5%	Stratified ratio	3.48	-29.95	44.79
2010	Random allocation 5%	Negative binomial	-7.37	-27.08	16.49
2010	Random allocation 5%	Delta lognormal	5.48	-17.90	33.51
2010	Random allocation 10%	Stratified ratio	-0.85	-25.57	30.84
2010	Random allocation 10%	Negative binomial	-6.96	-29.81	12.55
2010	Random allocation 10%	Delta lognormal	4.98	-18.77	27.50

Table 4. Percentage of times that linear predictors Season and hooks between floats (HBF) were included in the best approximating model of each observation error model across n = 100 simulation runs.

Observer allocation	Estimation model	Percent season included	Percent HBF included	n
Random allocation 5%	Negative binomial	80	97	100
Random allocation 5%	Delta lognormal (binomial component)	99	100	100
Random allocation 5%	Delta lognormal	5	89	100
Random allocation 10%	Negative binomial	100	100	100
Random allocation 10%	Delta lognormal (binomial component)	100	100	100
Random allocation 10%	Delta lognormal	27	100	100

Table 5. Summary of percent bias for spatial-temporal allocation of observer coverage. Estimation models are: stratified ratio estimator, negative binomial, and delta-lognormal. Shown are results across all years 'All', and for years 2000 and 2010 across 100 simulation runs.

				Centered 95%	
Year	**Include observed catch in estimate**	**Estimation model**	**Median**	**Lower**	**Upper**
All	TRUE	Stratified ratio	-1.46	-31.19	35.41
All	TRUE	Negative binomial	1.79	-29.05	48.57
All	TRUE	Delta lognormal	6.95	-22.91	50.83
All	FALSE	Stratified ratio	-1.46	-31.19	35.41
All	FALSE	Negative binomial	1.59	-29.18	49.18
All	FALSE	Delta lognormal	7.39	-22.92	51.38
2000	TRUE	Stratified ratio	0.01	-27.65	30.43
2000	TRUE	Negative binomial	-1.43	-27.18	23.40
2000	TRUE	Delta lognormal	1.20	-22.37	25.76
2000	FALSE	Stratified ratio	0.01	-27.65	30.43
2000	FALSE	Negative binomial	-1.52	-26.87	23.96
2000	FALSE	Delta lognormal	1.01	-22.18	26.37
2010	TRUE	Stratified ratio	-2.93	-28.24	34.61
2010	TRUE	Negative binomial	-6.39	-23.79	18.88
2010	TRUE	Delta lognormal	2.95	-15.00	32.33
2010	FALSE	Stratified ratio	-2.93	-28.24	34.61
2010	FALSE	Negative binomial	-7.51	-23.73	18.45
2010	FALSE	Delta lognormal	2.72	-14.94	33.48

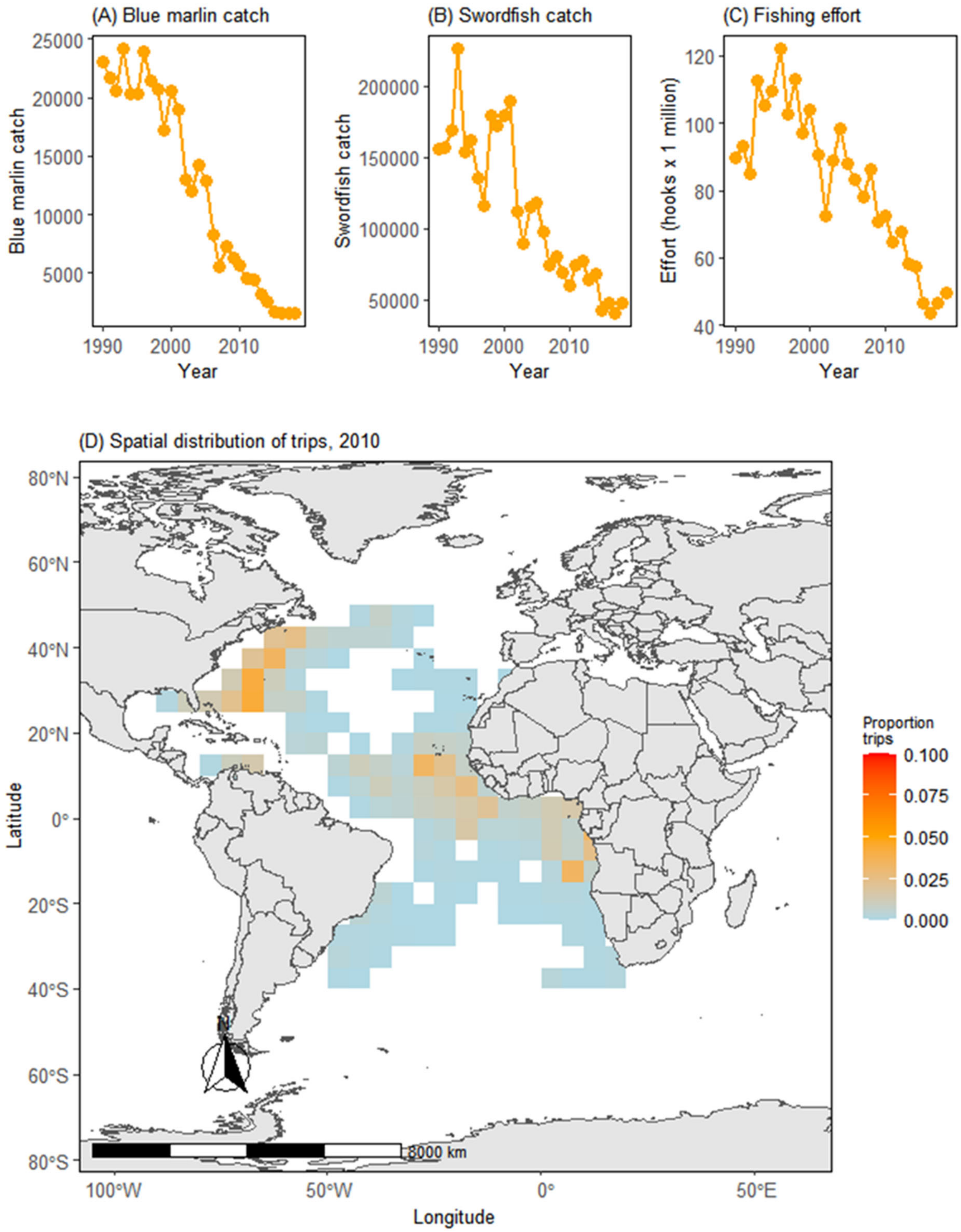


Figure 1. Annual catches, fishing effort and spatial distribution of fishing effort from simulated longline fleets using LLSIM.

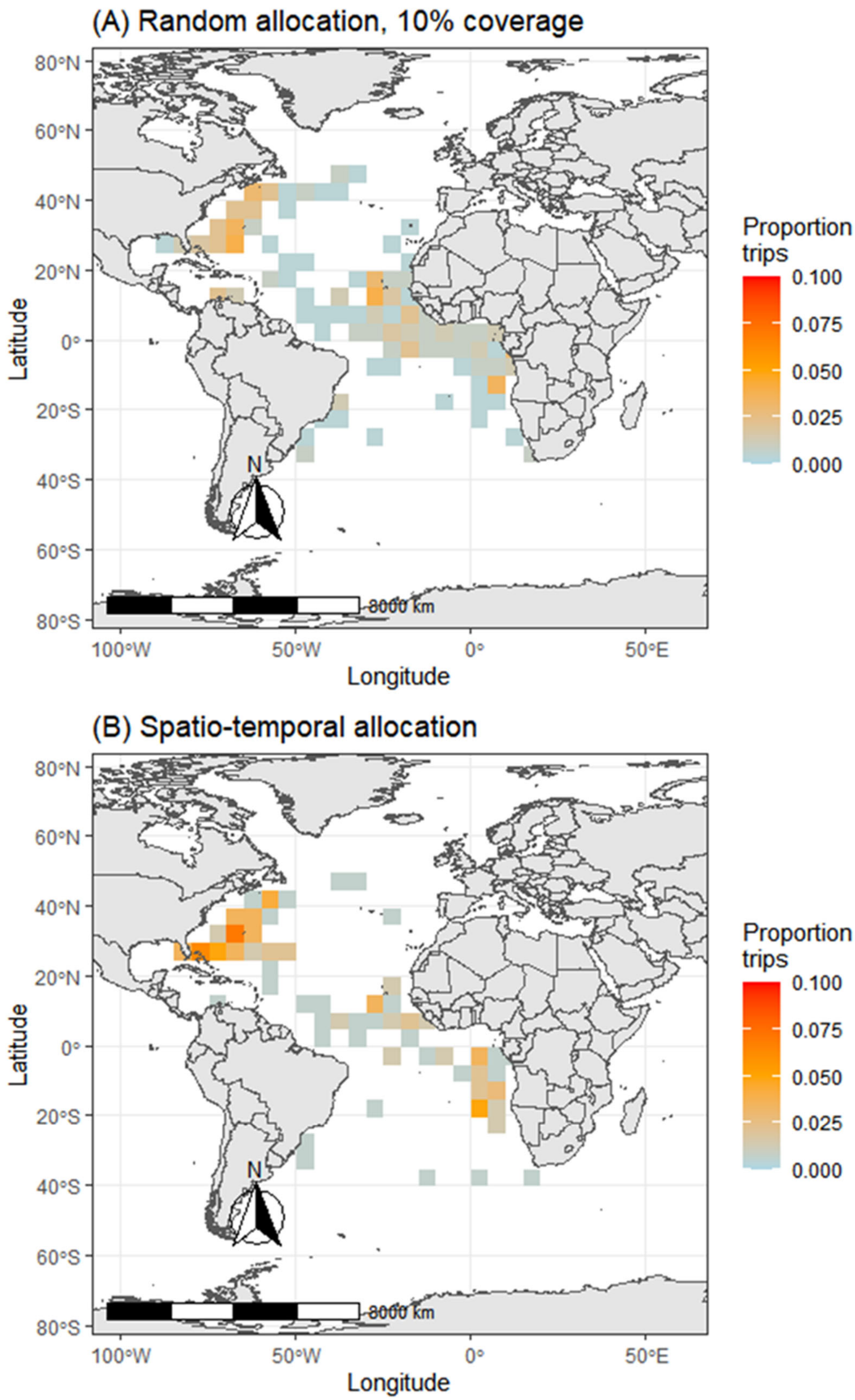
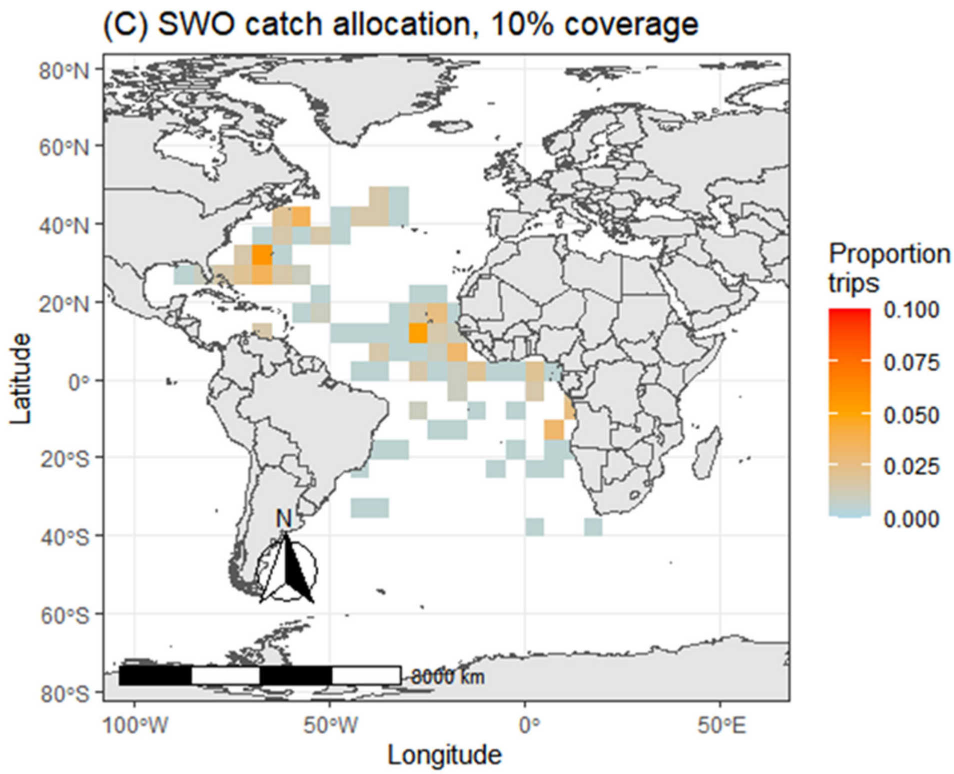


Figure 2. Example simulated observer coverage based on (A) random allocation to fishing trips with 10% observer coverage and (B) spatio-temporal allocation model.



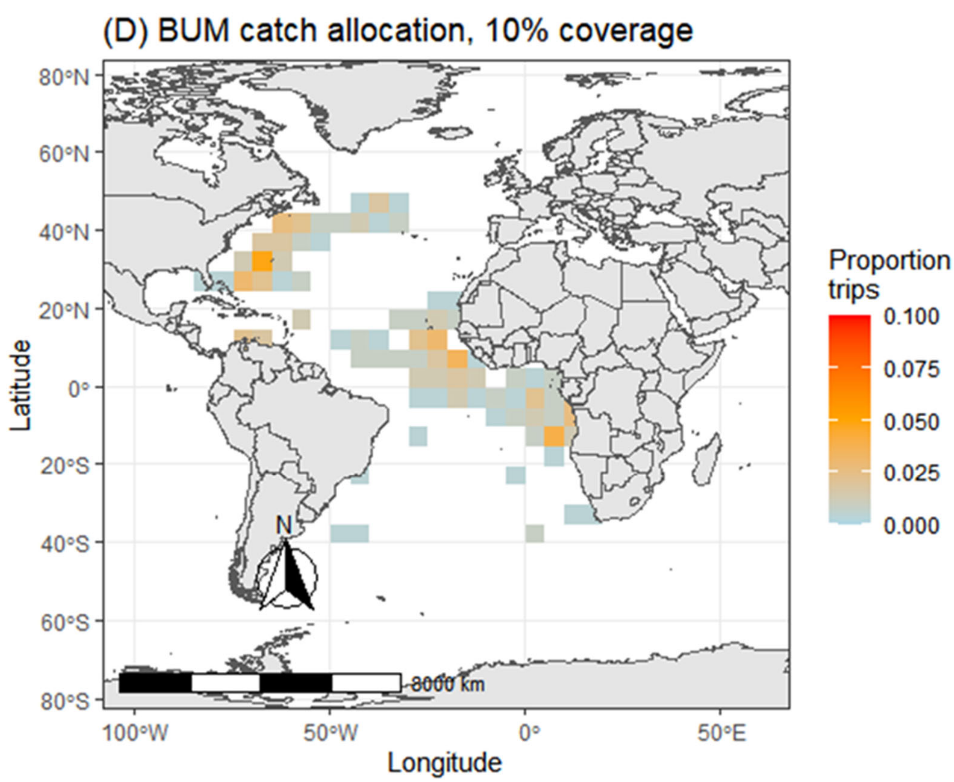


Figure 2 continued. Example simulated observer coverage based on (C) proportional allocation to fishing trips based on historical catch of swordfish (SWO) with 10% observer coverage and (D) proportional allocation to fishing trips based on historical catch of blue marlin (BUM) with 10% observer coverage.

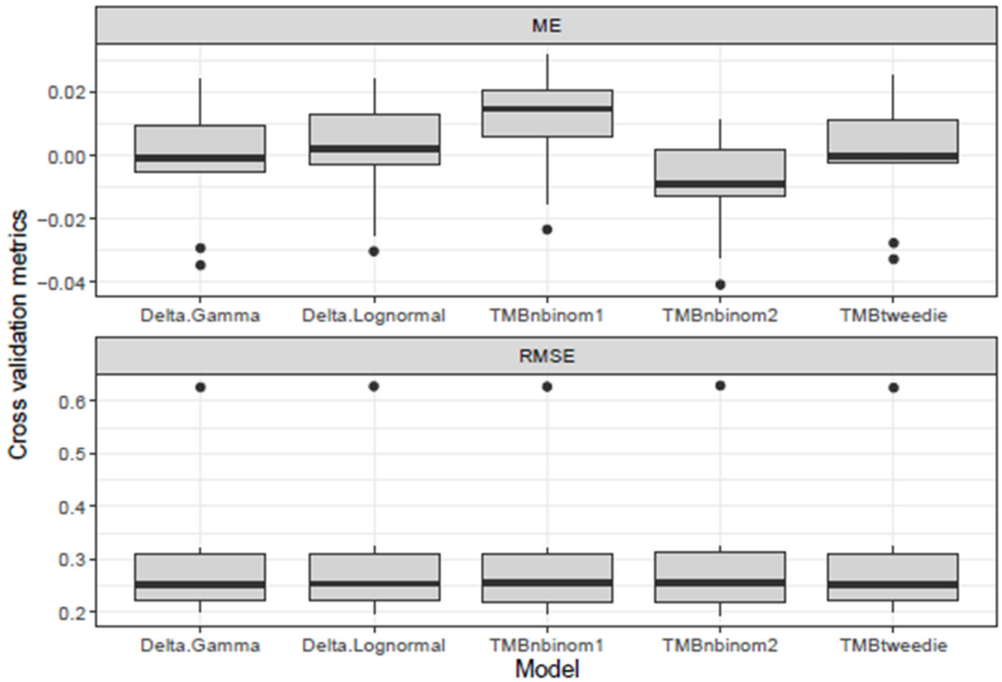


Figure 3. Cross-validation results for each model group by trip for one example run using the spatiotemporal coverage scenario Delta-Gamma performed best in this case.

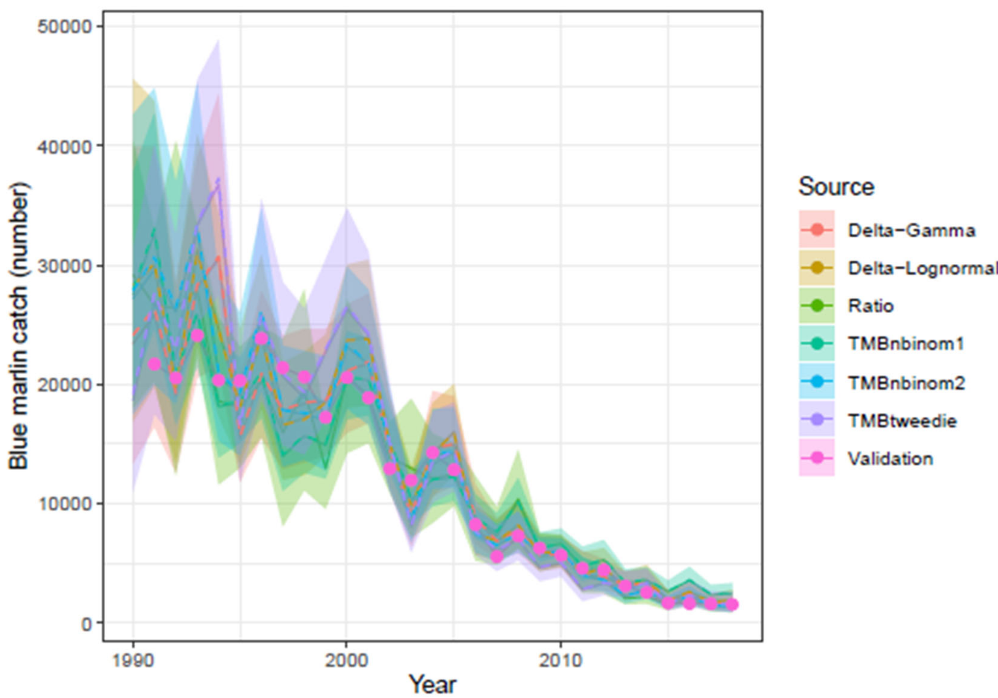


Figure 4. Total bycatch estimate from BIC best model in each error group, along with the correct totals (pink circles) for one example run of the spatiotemporal coverage scenario.

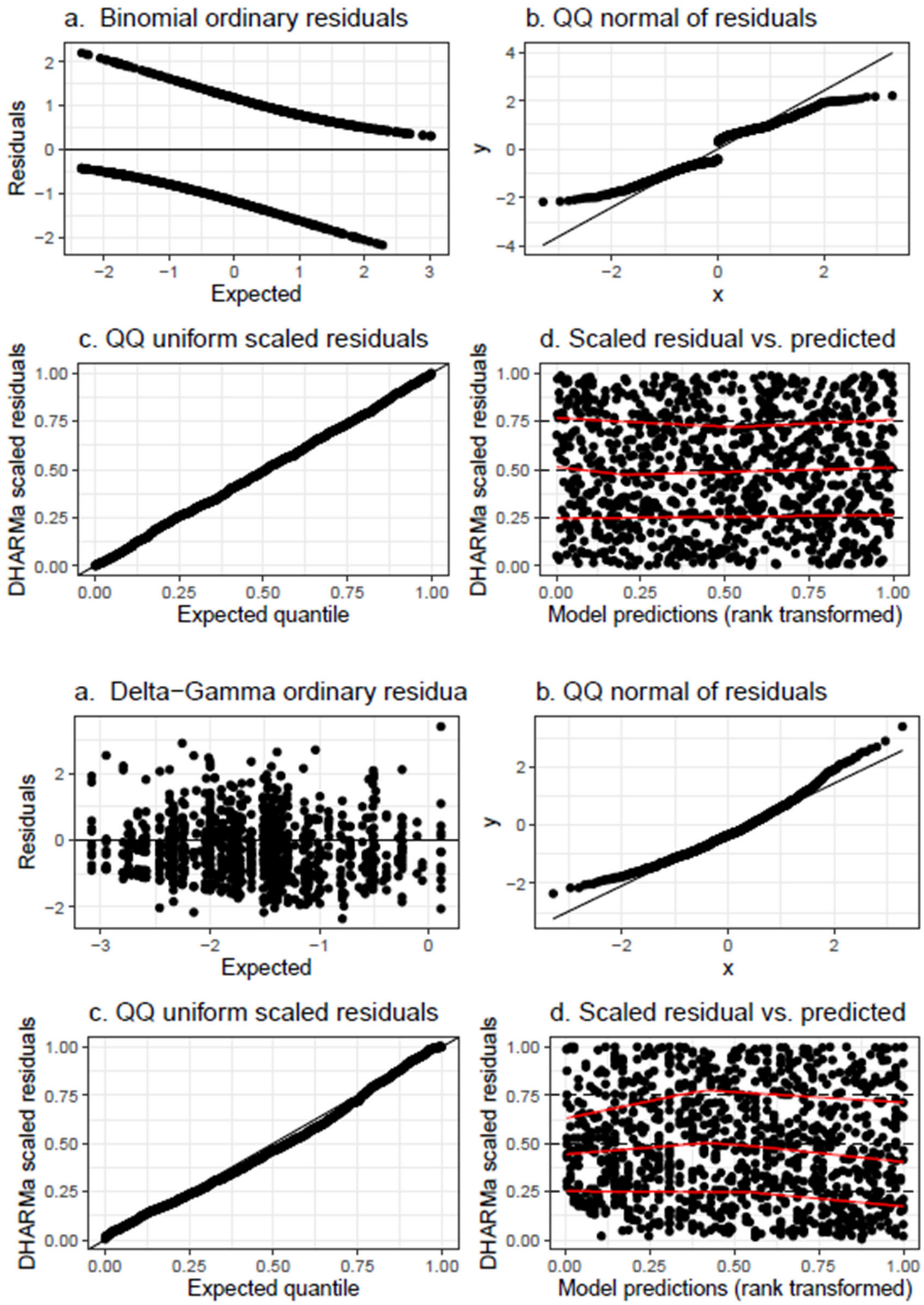


Figure 5. Diagnostics for the best model for the example run for the realistic spatiotemporal scenario, which was a delta model with both binomial and gamma components.

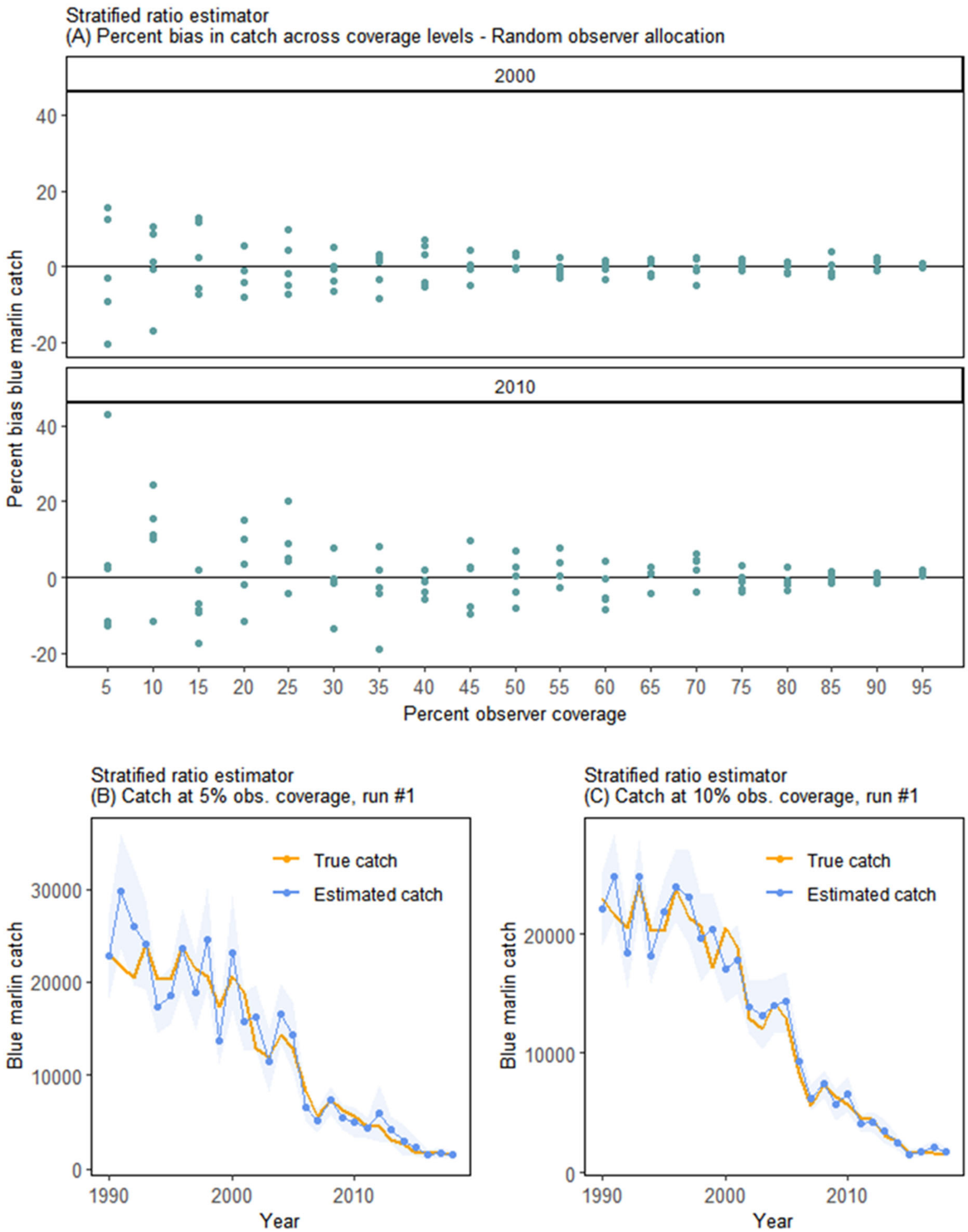


Figure 6. Stratified ratio estimator of blue marlin bycatch. (A) Percent bias resulting from random observer allocation to fishing trips across coverage levels of 5% to 95% for years 2000 and 2010, (B) example of annual bycatch estimates (blue shaded area is 95% confidence interval) at 5% observer coverage, and (C) example of annual bycatch estimates (blue shaded area is 95% confidence interval) at 10% observer coverage.

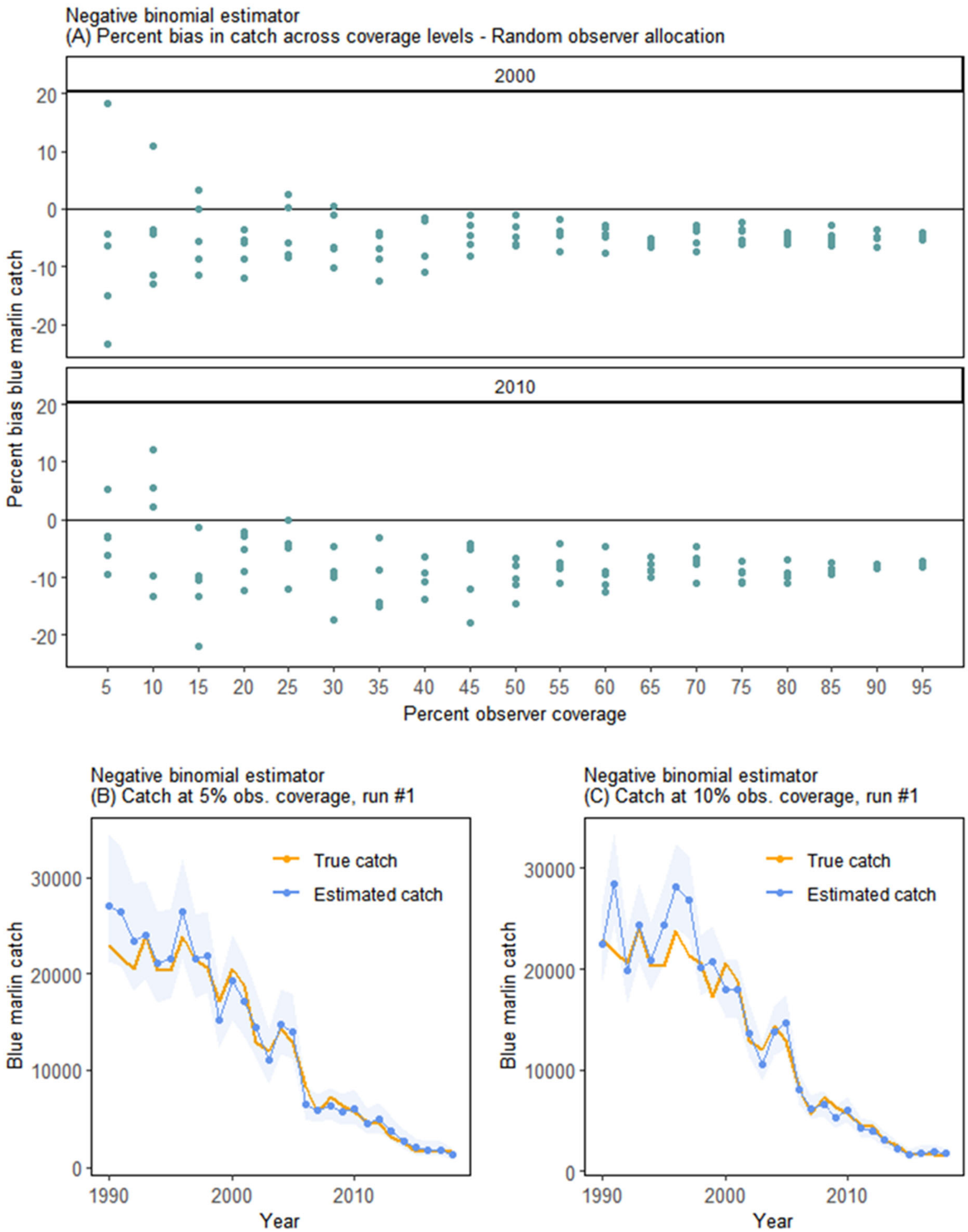


Figure 7. Negative binomial estimator of blue marlin bycatch - predictions made for all trips. (A) Percent bias resulting from random observer allocation to fishing trips across coverage levels of 5% to 95% for years 2000 and 2010, (B) example of annual bycatch estimates (blue shaded area is 95% confidence interval) at 5% observer coverage, and (C) example of annual bycatch estimates (blue shaded area is 95% confidence interval) at 10% observer coverage.

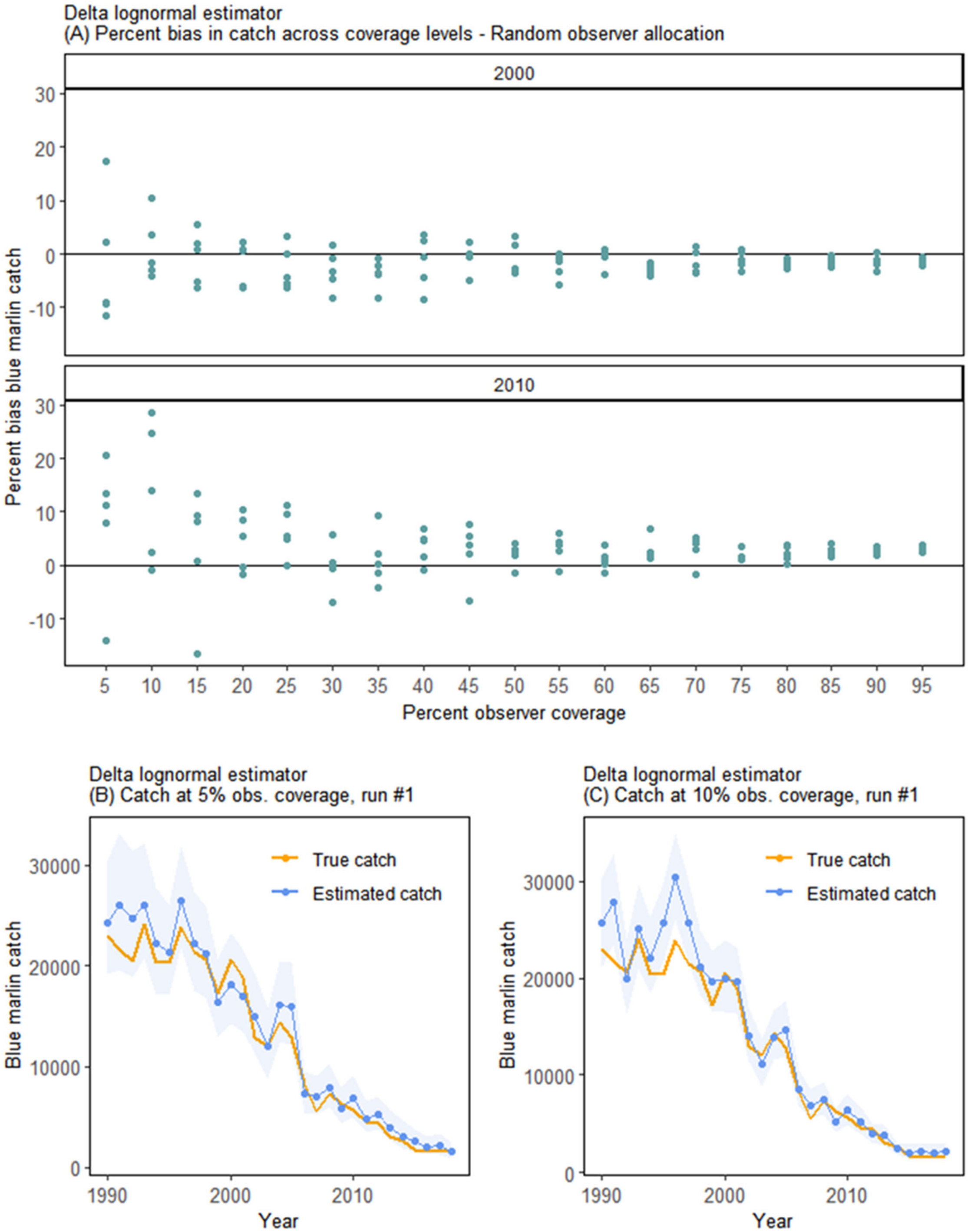


Figure 8. Delta lognormal estimator of blue marlin bycatch - predictions made for all trips. (A) Percent bias resulting from random observer allocation to fishing trips across coverage levels of 5% to 95% for years 2000 and 2010, (B) example of annual bycatch estimates (blue shaded area is 95% confidence interval) at 5% observer coverage, and (C) example of annual bycatch estimates (blue shaded area is 95% confidence interval) at 10% observer coverage.

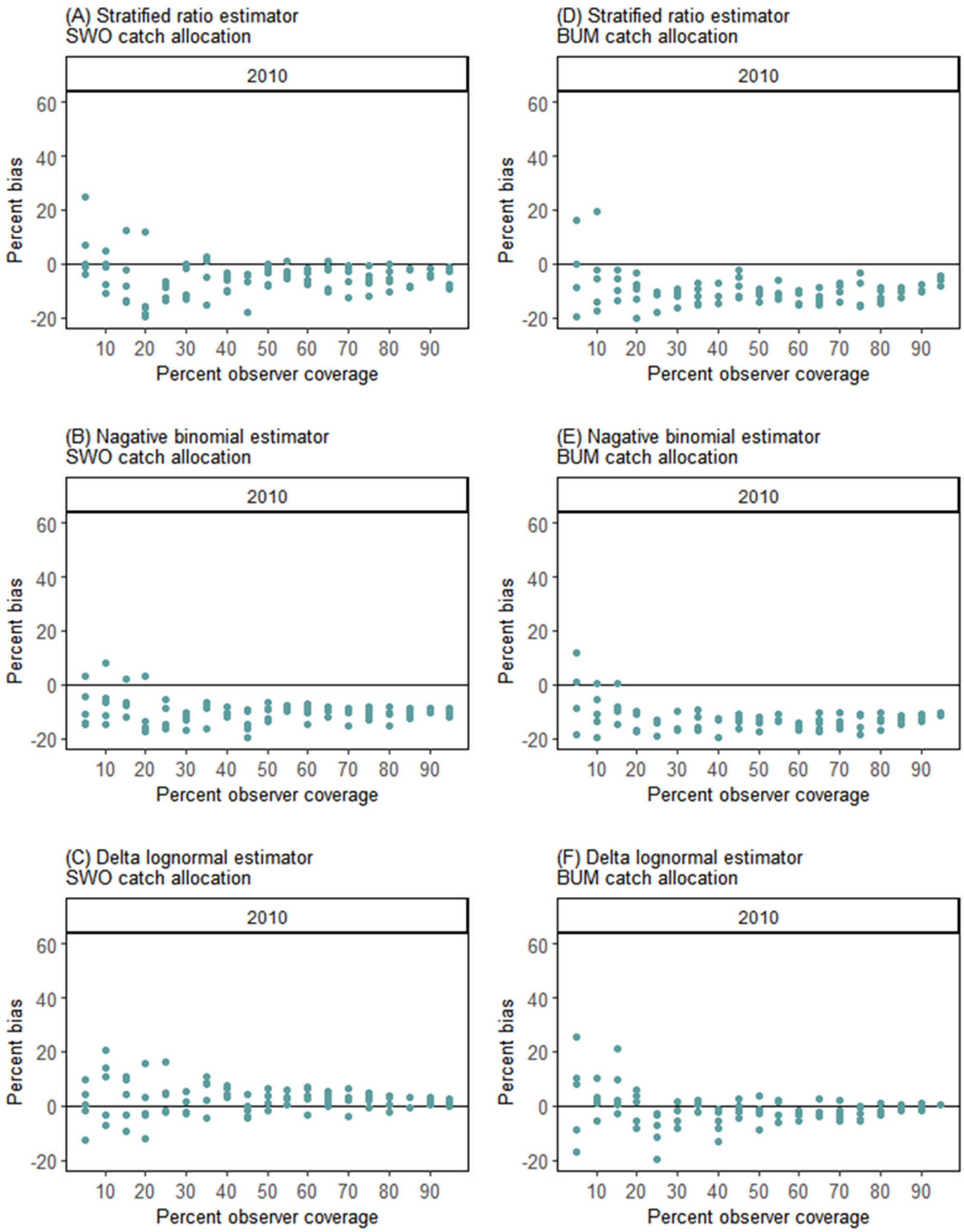


Figure 9. Percent bias resulting from swordfish catch allocation (A, B, C) and blue marlin catch allocation (D, E, F) to fishing trips across coverage levels of 5% to 95%. Predictions made for all trips. Plot titles indicate estimation models used to predict bycatch. Shown are examples for years 2000 and 2010.

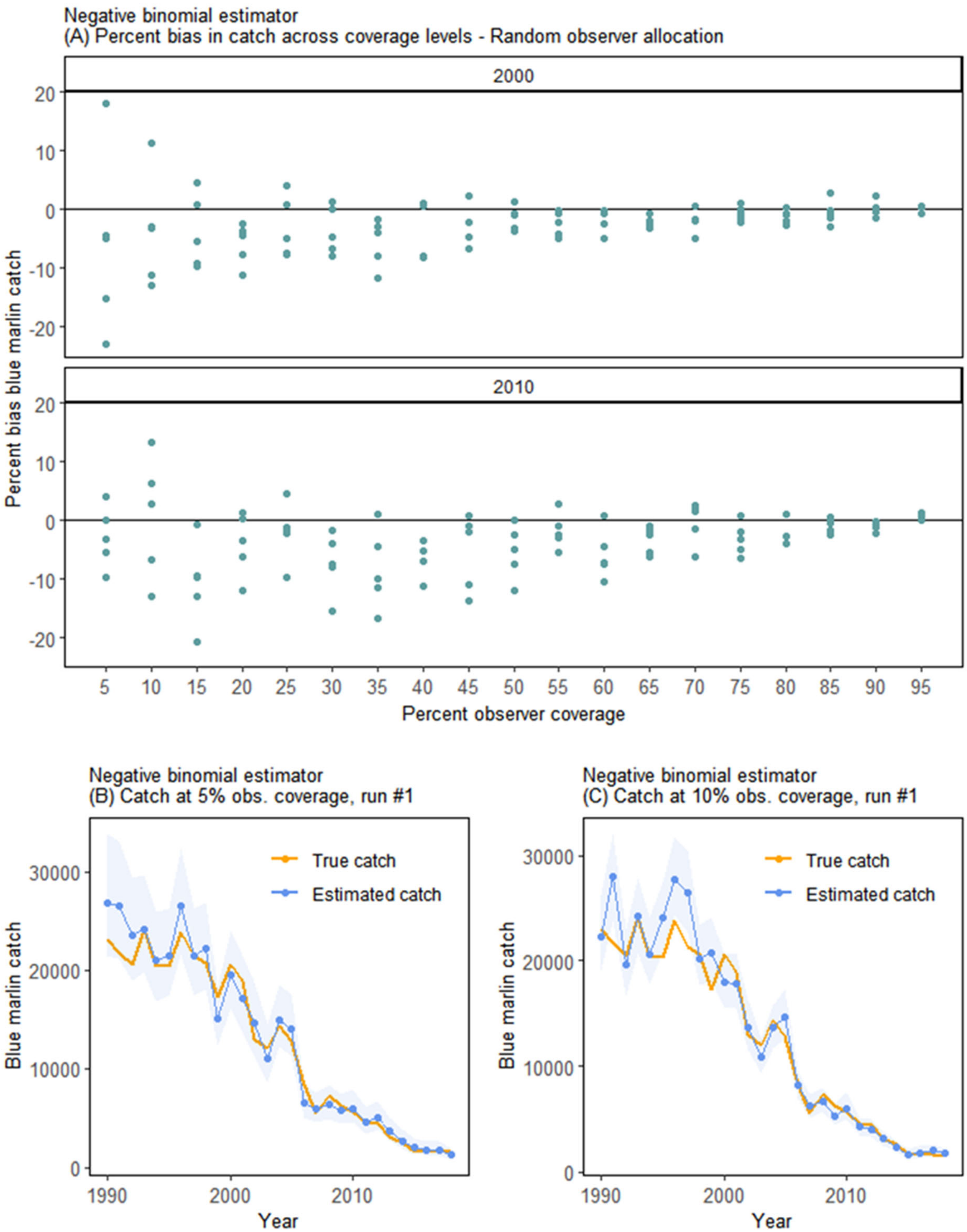


Figure 10. Negative binomial estimator of blue marlin bycatch - observed catches included as a known constant. (A) Percent bias resulting from random observer allocation to fishing trips across coverage levels of 5% to 95% for years 2000 and 2010, (B) example of annual bycatch estimates (blue shaded area is 95% confidence interval) at 5% observer coverage, and (C) example of annual bycatch estimates (blue shaded area is 95% confidence interval) at 10% observer coverage.

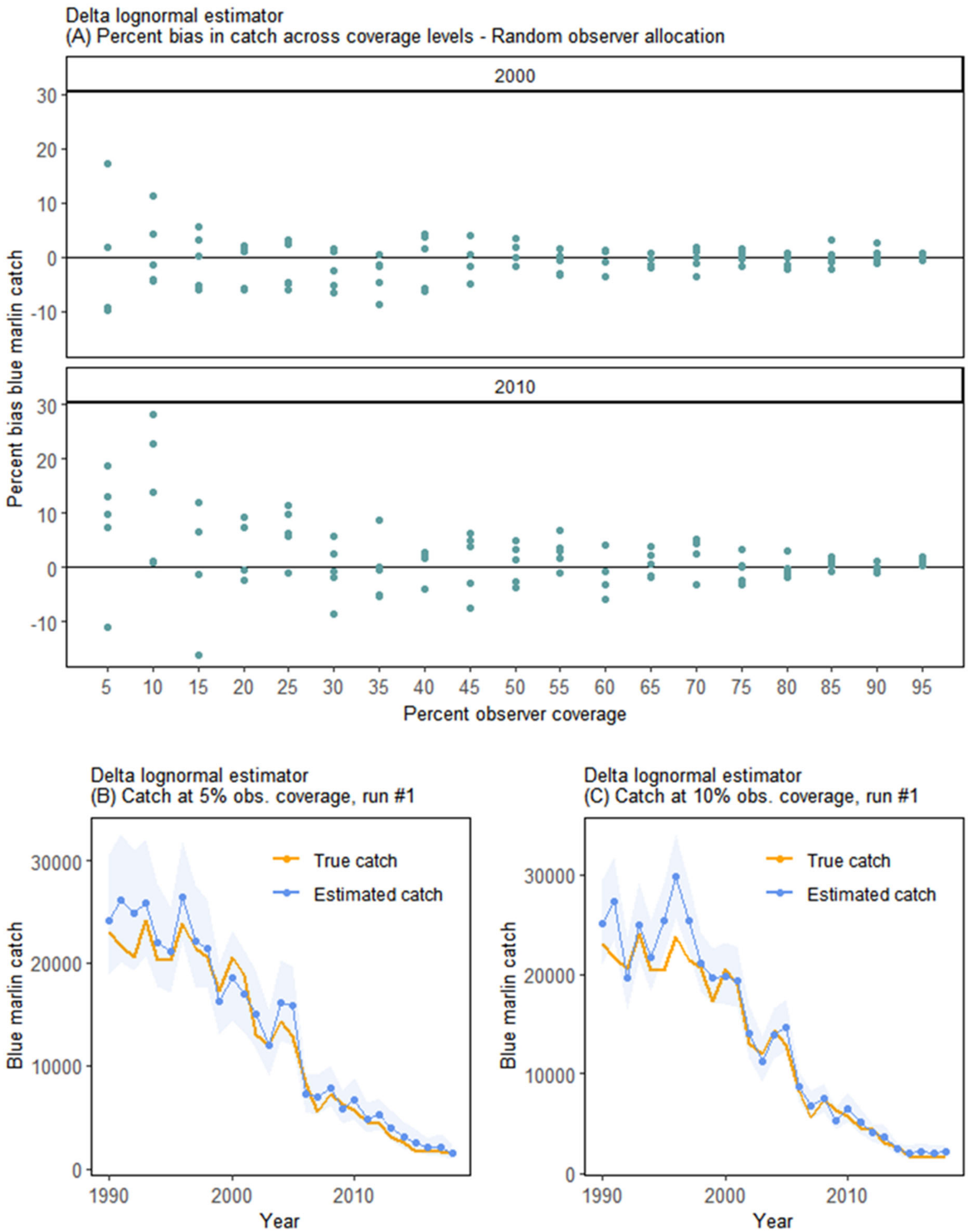


Figure 11. Delta lognormal estimator of blue marlin bycatch - observed catches included as a known constant. (A) Percent bias resulting from random observer allocation to fishing trips across coverage levels of 5% to 95% for years 2000 and 2010, (B) example of annual bycatch estimates (blue shaded area is 95% confidence interval) at 5% observer coverage, and (C) example of annual bycatch estimates (blue shaded area is 95% confidence interval) at 10% observer coverage.

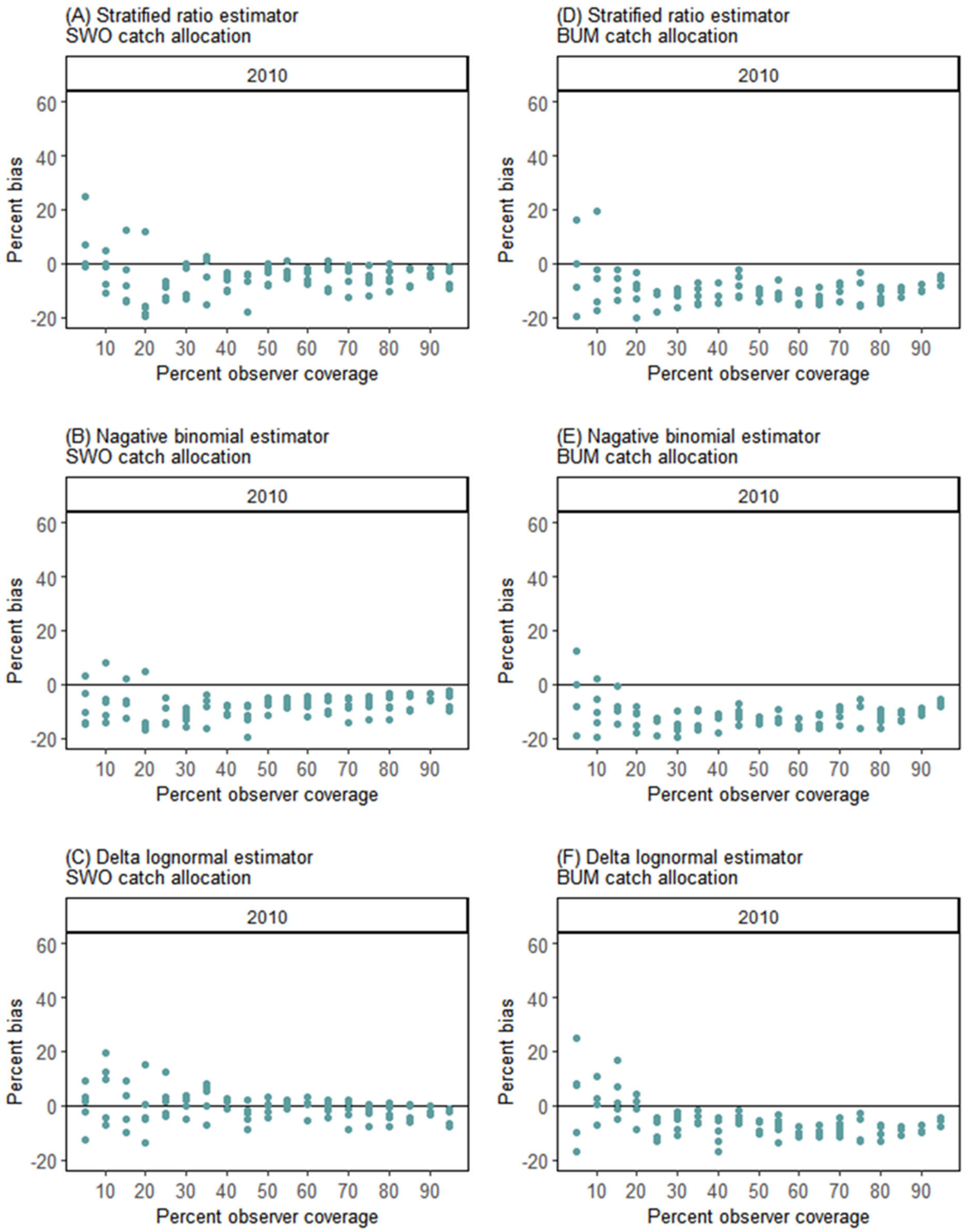


Figure 12. Percent bias resulting from swordfish catch allocation (A, B, C) and blue marlin catch allocation (D, E, F) to fishing trips across coverage levels of 5% to 95%. Observed catches included as a known constant. Plot titles indicate estimation models used to predict bycatch. Shown are examples for years 2000 and 2010.

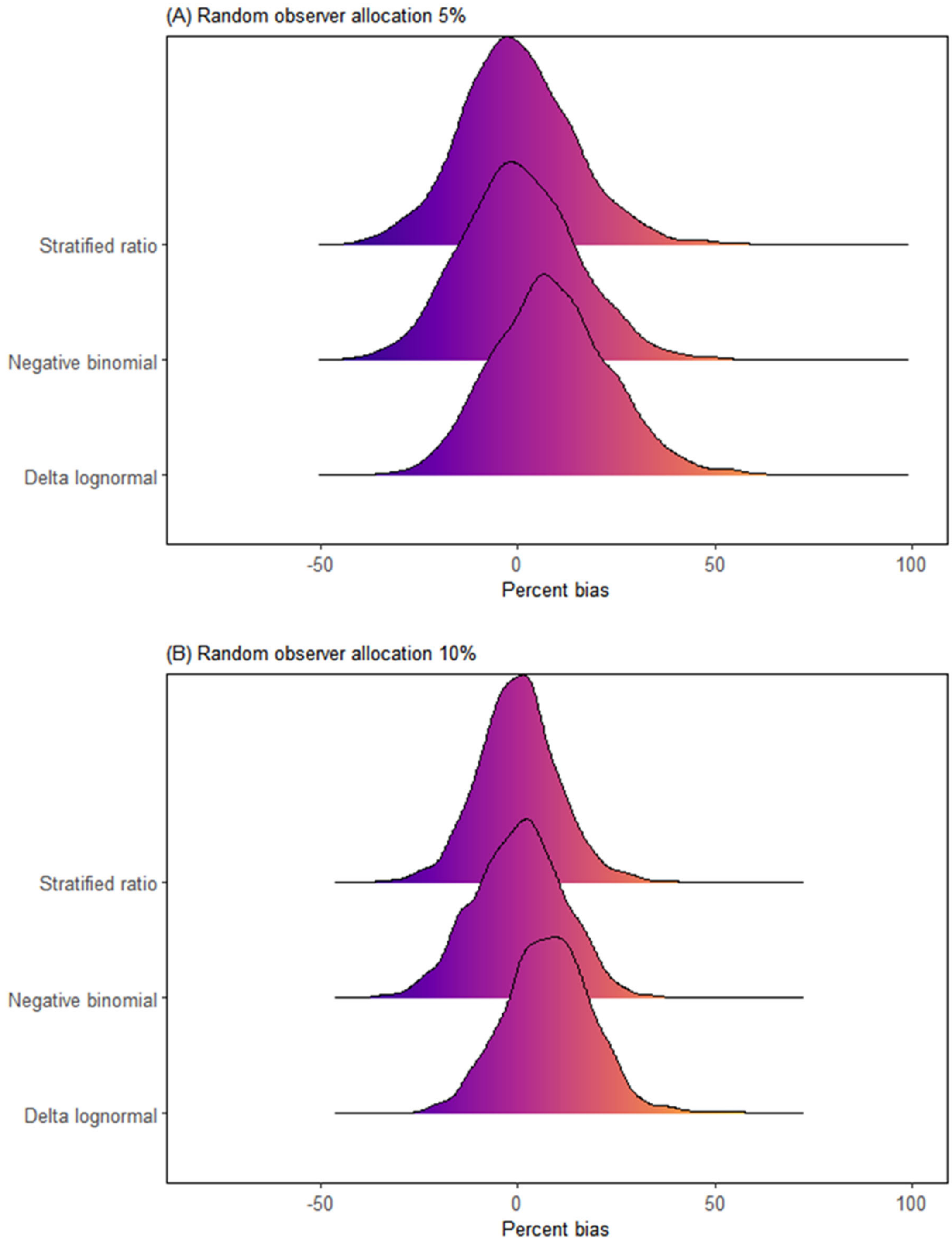


Figure 13. Percent bias across all years for (A) random allocation to trips with 5% observer coverage, and (B) random allocation to trips with 10% observer coverage. Predictions made for all trips. Percent bias across 100 simulation runs from estimation models: stratified ratio estimator, negative binomial, and delta-lognormal.

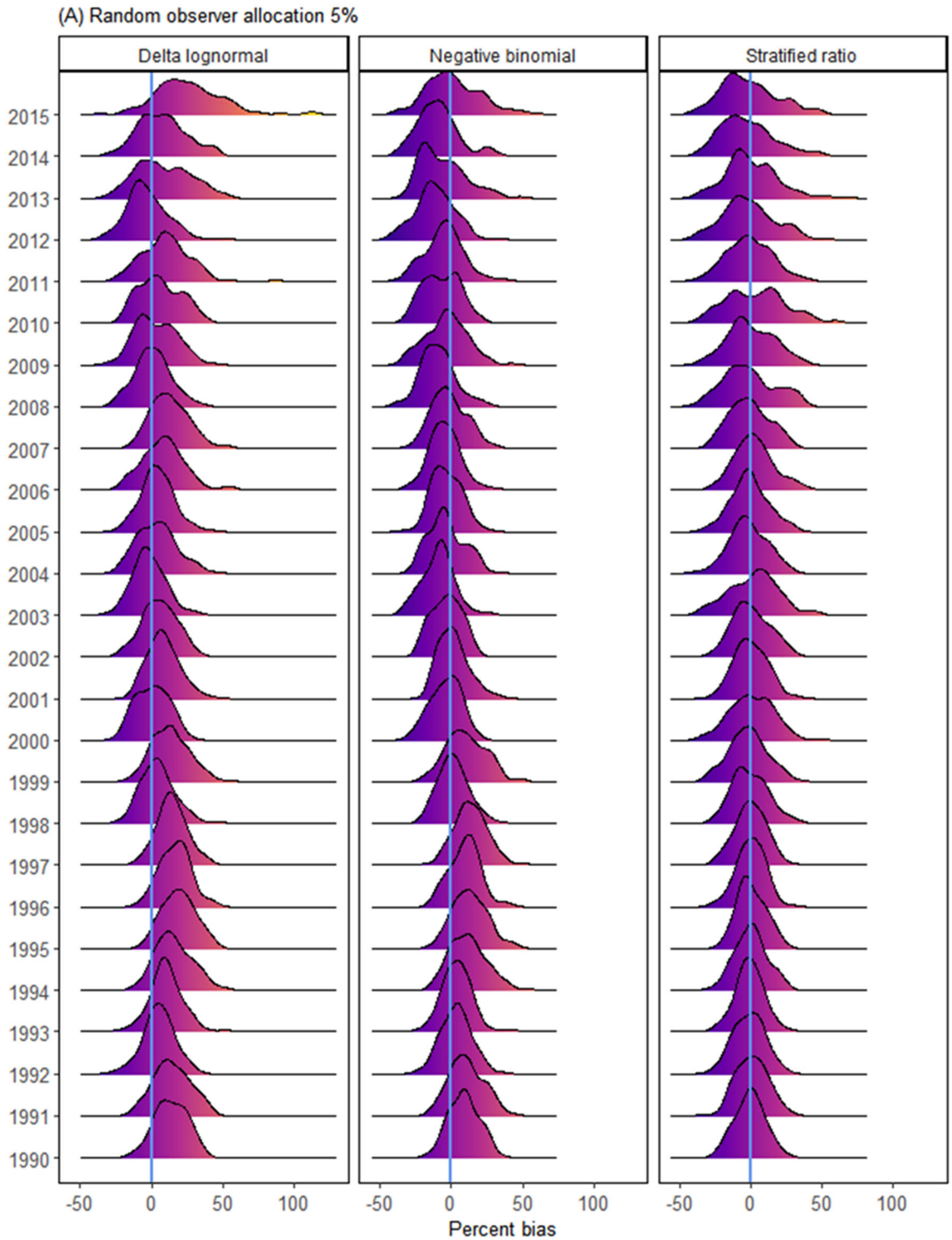


Figure 14. Annual distribution of percent bias across (A) random allocation to trips with 5% observer coverage and (B) random allocation to trips with 10% observer coverage. Shown are percent bias across 100 simulation runs from estimation models: stratified ratio estimator, negative binomial, and delta-lognormal.

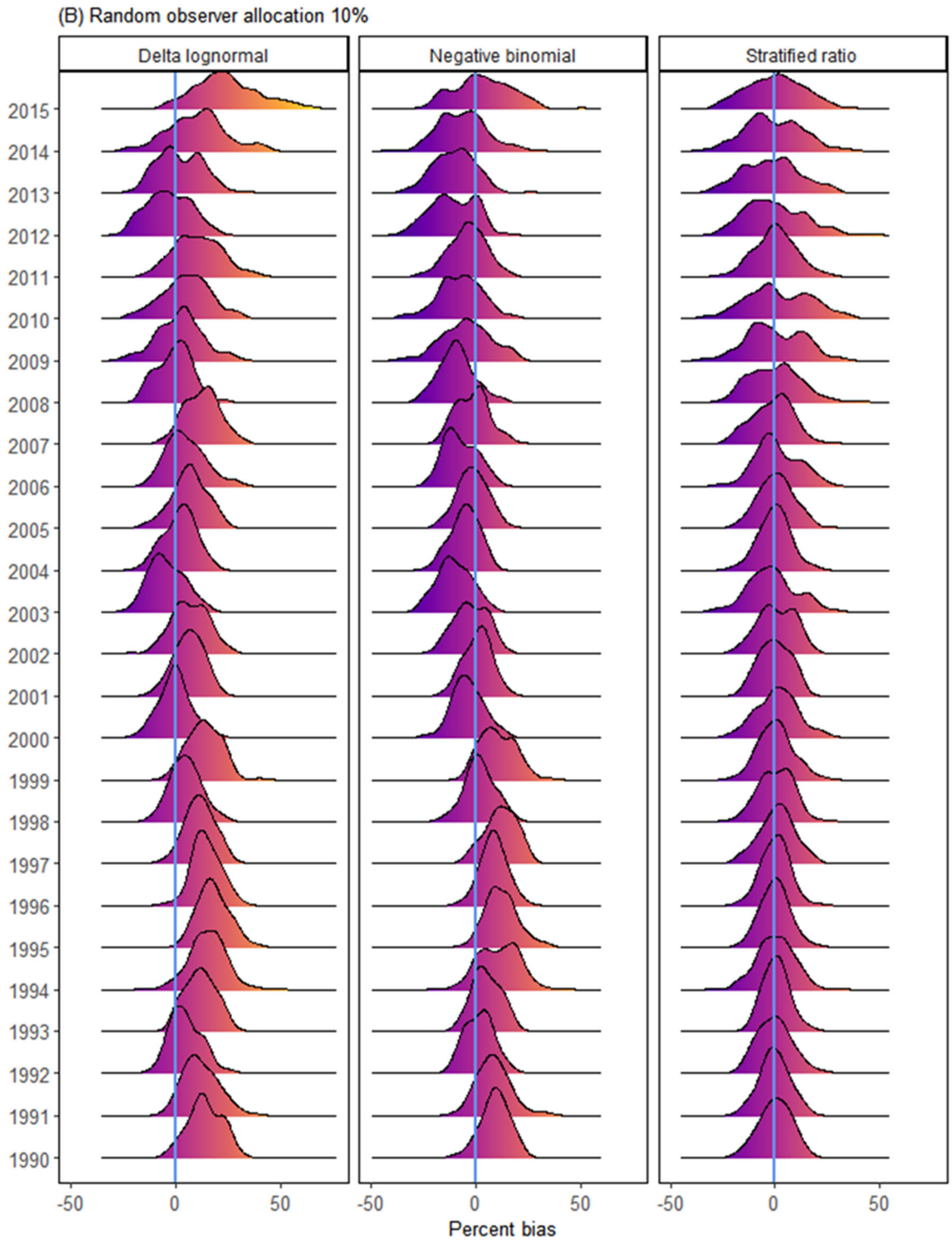


Figure 14. Continued.

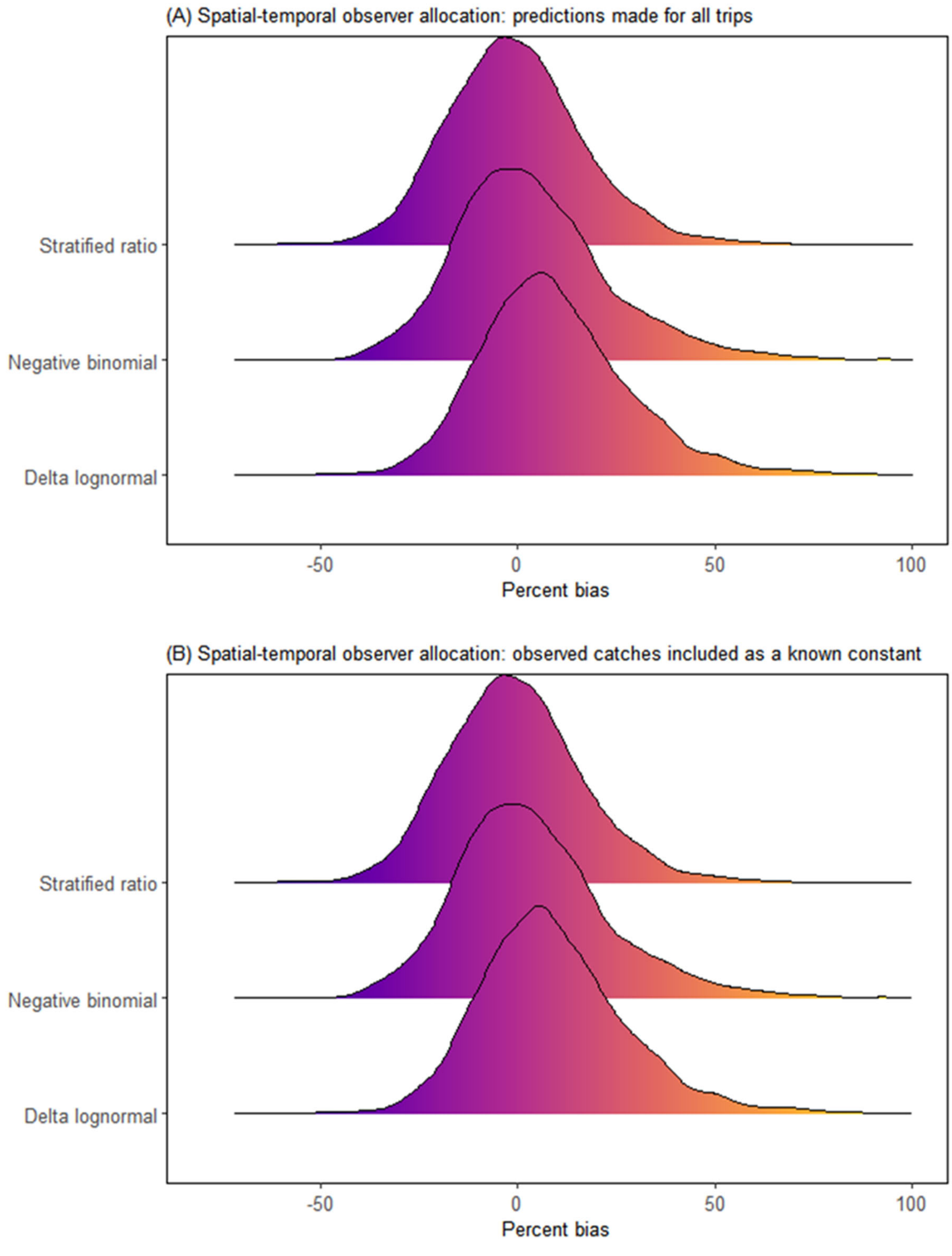


Figure 15. Percent bias across all years for spatial-temporal observer allocation for (A) predictions made for all trips, and (B) observed catches included as known constant. Percent bias across 100 simulation runs from estimation models: stratified ratio estimator, negative binomial, and delta-lognormal.

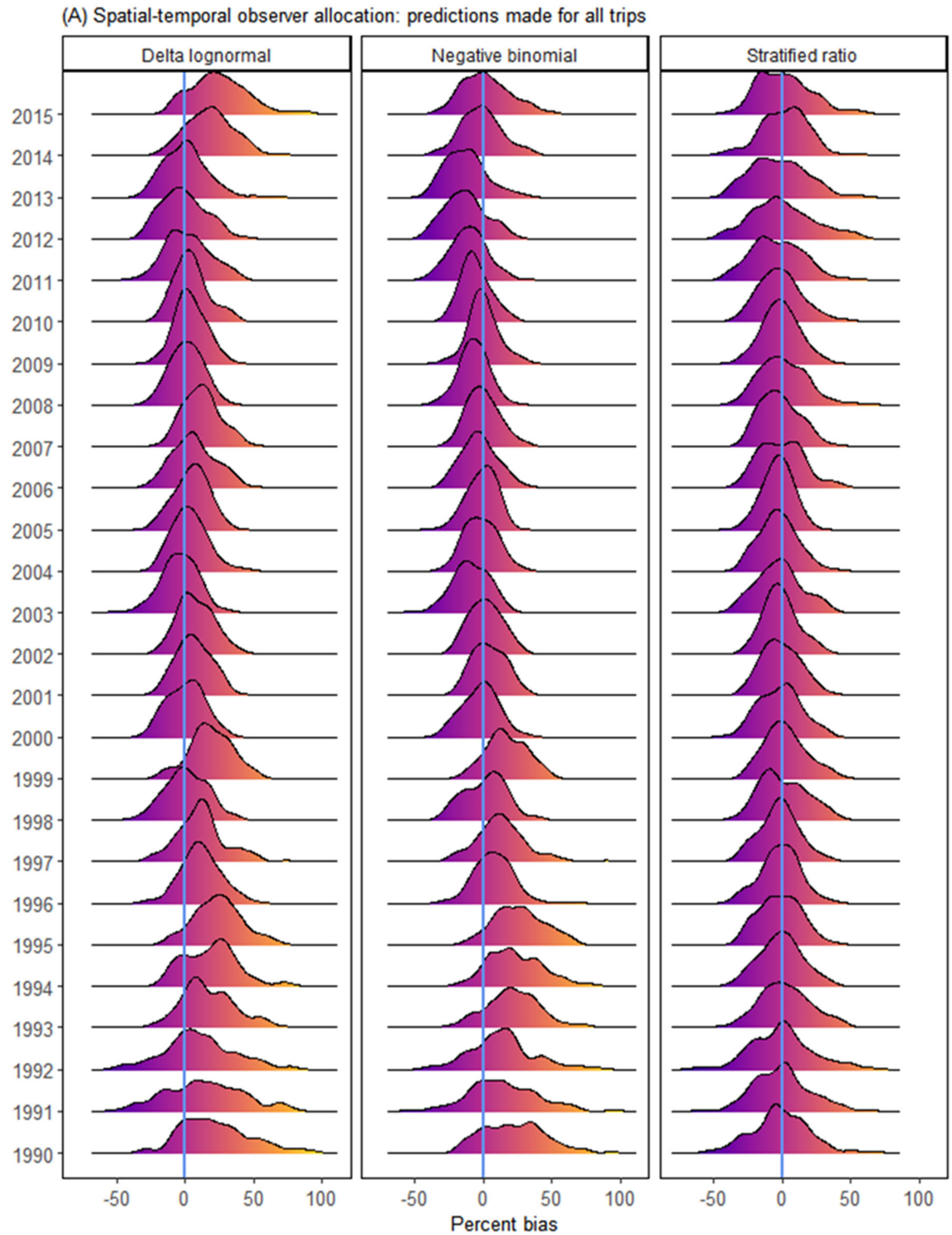


Figure 16. Annual distribution of percent bias for spatial-temporal allocation of observer coverage. Shown are (A) predictions made for all trips, and (B) observed catches included as known constant across 100 simulation runs from estimation models: stratified ratio estimator, negative binomial, and delta-lognormal.

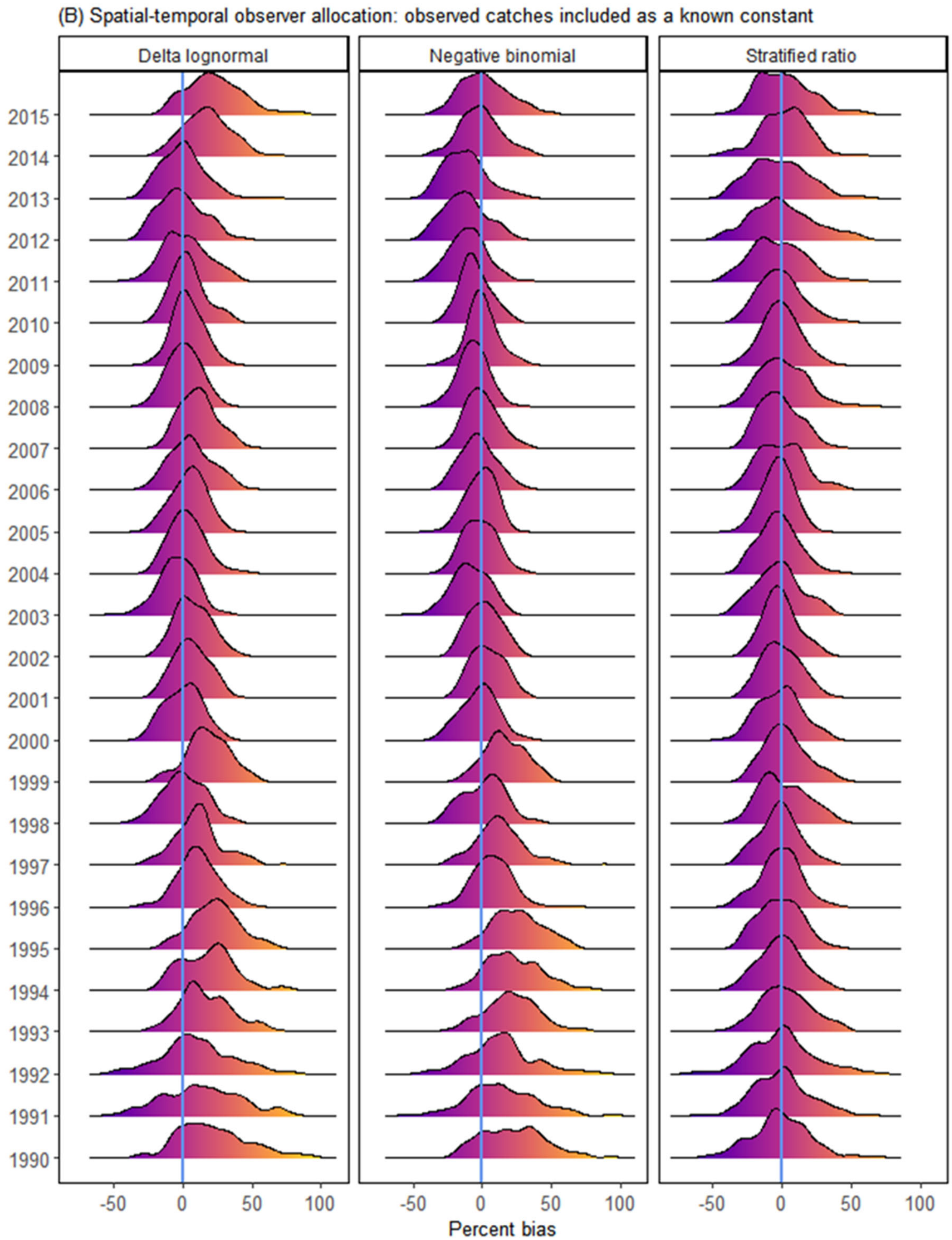


Figure 16. Continued.

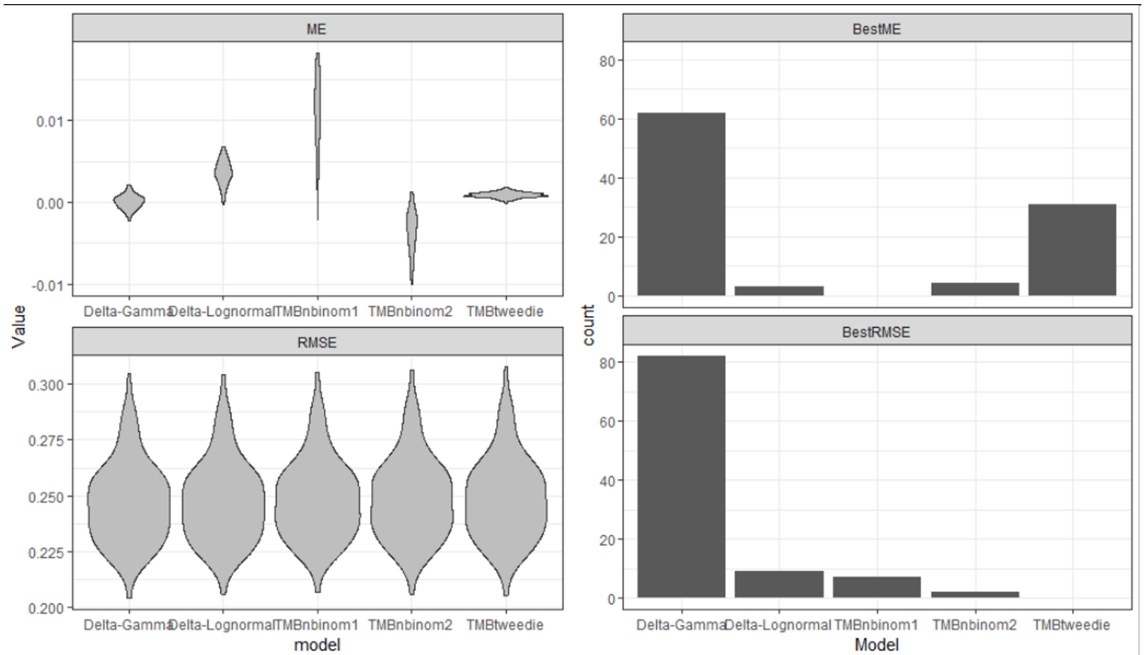


Figure 17. Violin plots of cross-validation mean error (ME) and root mean squared error (RMSE) across 100 draws of the realistic spatio-temporal coverage scenario for each observation error model (left), along with the number of times each model type was chosen as the best by ME or RMSE across the draws.

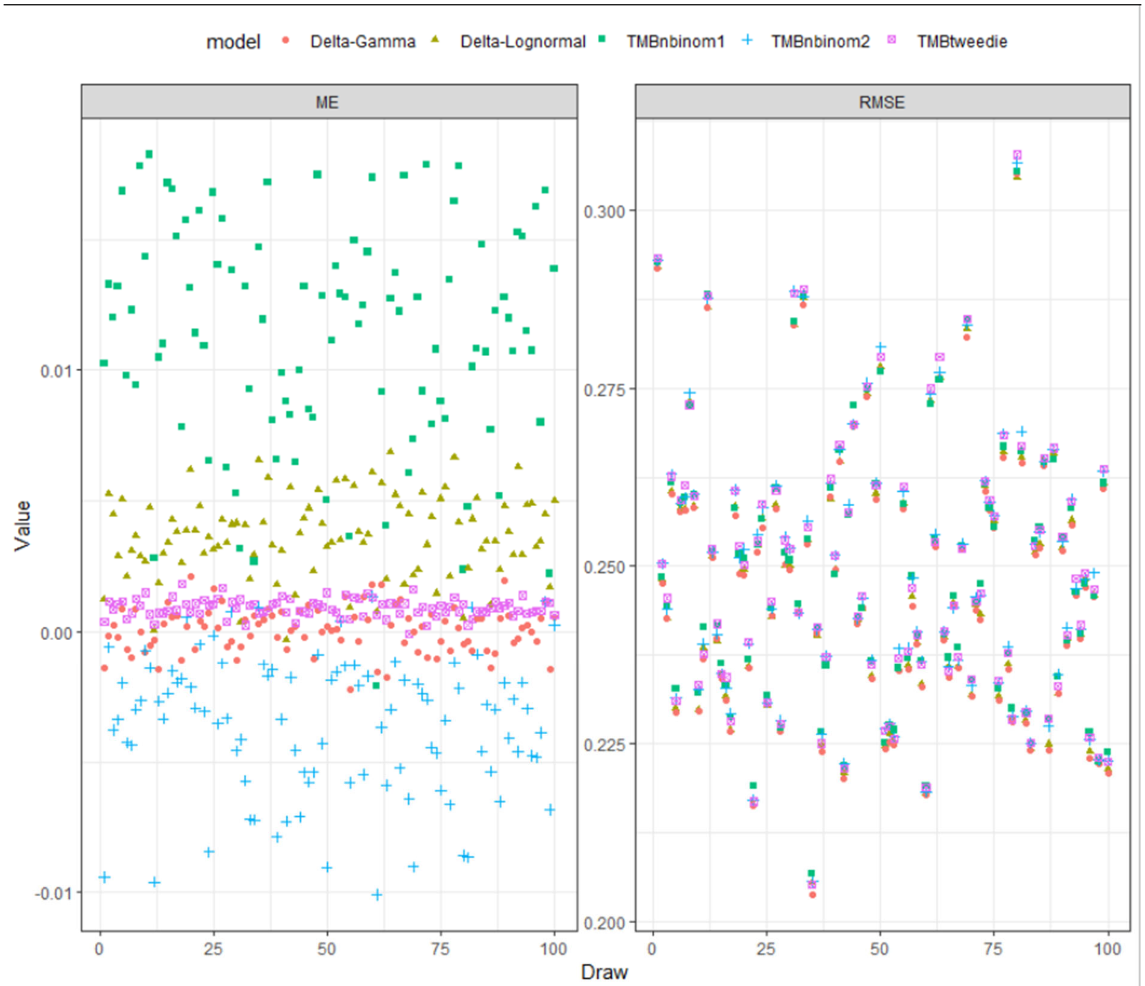


Figure 18. ME and RMSE values by model and draw.

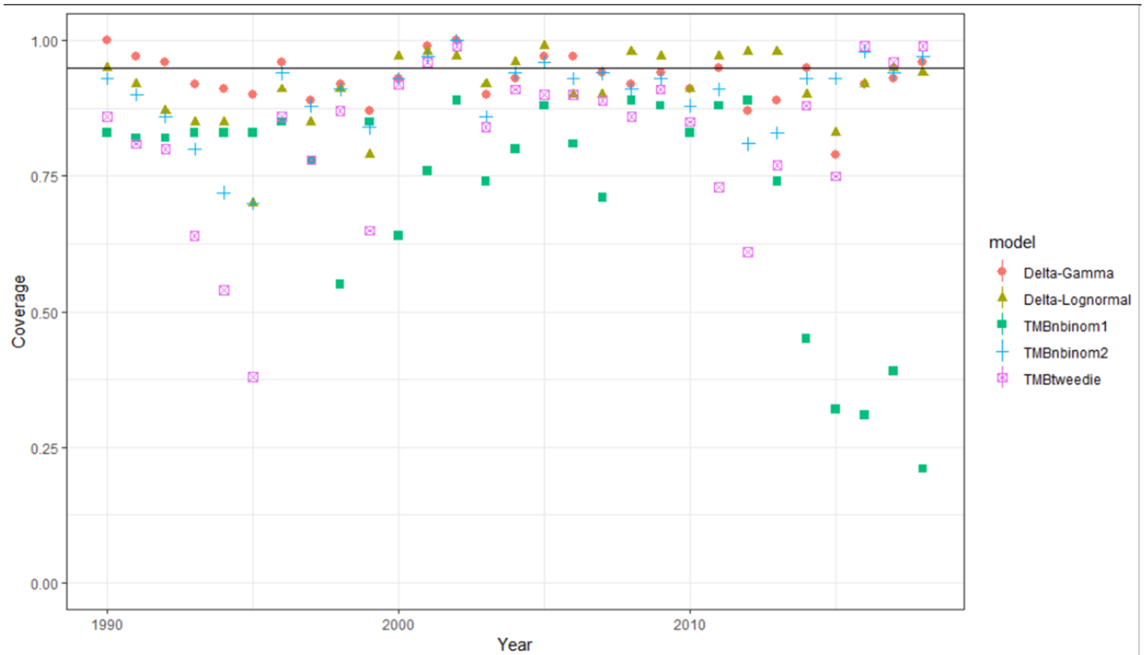


Figure 19. Fraction of 100 draws in which the true total annual bycatch was included in the 95% confidence interval for the realistic spatiotemporal coverage scenario. The horizontal line indicates the correct value of 95%.