A model-based approach to estimate latent commercial cobia landings in Atlantic states, with implications for quota monitoring

by

Kyle Dettloff

National Marine Fisheries Service
Southeast Fisheries Science Center

Fisheries Statistics Division
75 Virginia Beach Drive
Miami, FL 33149

May 2019

INTRODUCTION

The Southeast Fisheries Science Center is responsible for providing projections to the Southeast Regional Office (SERO) concerning when commercial Annual Catch Limit (ACL) fisheries in the South Atlantic and Gulf of Mexico are expected to reach their quotas. One of these fisheries is Atlantic Group Cobia, which encompasses state and federal waters from Georgia to New York, and has been managed by NMFS since 2015. Federal dealers are typically required to report purchases electronically, usually resulting in a relatively brief (roughly two week) time-lag between when fish were landed and when data become available to federal analysts. However, some states, such as Virginia, require reporting by fishers instead of dealers. In these cases, data are not available to the SEFSC until landings are incorporated by the Atlantic Coast Cooperative Statistics Program (ACCSP), which usually occurs in May of the following calendar year. Therefore, to provide accurate in-season closure projections, an approach is required to estimate real-time landings by accounting for landings that have occurred but are not yet known to SEFSC. This process can be especially impactful for Atlantic cobia, as the majority of landings occur in Virginia. The method presented here outlines the current process used by SEFSC to estimate these latent landings based only on available electronic dealer reports.

METHODS

To predict current season landings, cumulative daily landings from each of the three most recent prior seasons with complete data (e.g., 2015, 2016, and 2017 to predict 2018) were compiled from ACCSP and matched with total landings that were available in the SEFSC Commercial Landings Monitoring (CLM) database on each day of the season. CLM landings were then blocked by state into those coming from North Carolina, Virginia, and a combined category for other states. Stratification to this level led to the most robust models, as the majority of state under-reported landings come from NC and VA, with rates of under-reporting in other states being more variable from year to year and often negligible. A categorical term representing the day of week was included in models to account for the weekly cyclical nature by which landings are received in CLM.

For each of the prior three seasons, the R *nlme* package (Pinheiro et al. 2018) was used to fit a weighted generalized least-squares (WGLS) model to ACCSP landings, with cumulative CLM landings from each state grouping and day of week included as covariates, as follows:

$$E[ASSCP_i] = \hat{\beta}_1 VA_i + \hat{\beta}_2 {VA_i}^2 + \hat{\beta}_3 NC_i + \hat{\beta}_4 {NC_i}^2 + \hat{\beta}_5 OT_i + \hat{\beta}_6 {OT_i}^2 + \hat{\beta}_j DAY_{ji} \;\; \{\text{eq. } 1\}$$

$$where\ VA = Virginia, NC = North\ Carolina, OT = other\ states, DAY = day\ of\ week\ (j = 1, 2, \ldots, 7)$$

The inclusion of quadratic state terms provided an improved fit over simply including first-order terms, while any higher order terms led to models with less robust predictions.

Weighting was done according to a power variance function structure to account for higher variation in model residuals at higher values of cumulative landings:

$$s^2(\boldsymbol{y}) = |\boldsymbol{y}|^{2t} \;\; \{\text{eq. } 2\}$$

$$where\ \boldsymbol{y} = vector\ of\ current\ season\ ACCSP\ landings$$
$$t = estimated\ variance\ function\ constant$$

To account for autocorrelation in cumulative landings from sequential days, a moving-average correlation structure was assumed. The optimal MA order *q* was selected by minimizing the BIC for the time-series model ARIMA(0, 0, *q*), or equivalently MA(*q*) of the form:

$$y_t = \mu + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q} \;\; \{\text{eq. } 3\}$$

This model was fit to ACCSP landings with all relevant regressors included, using the function 'auto.arima' in the R *forecast* package (Hyndman and Khandakar 2018). As Atlantic cobia landings tend to increase dramatically in early May, the relationship between ACCSP and CLM landings is only important from May through season end, and including earlier dates can lead to less informative models. Therefore, the function 'cpt.mean' in the R package *changepoint* (Killick and Eckley 2016) using the "At Most One Change" (AMOC) method (Hinkley 1970) was used to estimate when this change in landing intensity occurs each year, and only landings from dates following the estimated breakpoint were modeled. Note that autoregressive terms were excluded from the correlation structure in the GLS model as their inclusion led to erroneous predictions.

Using each model fit, a weighted average of estimated coefficients was calculated based on the inverse of each model's respective variance, calculated as the average of the model residuals adjusted for the number of parameters as follows:

$$\hat{\sigma}^2{}_h = \frac{\sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i{}'\hat{\boldsymbol{\beta}}\right)^2}{n - p} \;\; \{\text{eq. } 4\}$$

$$where\ h = number\ of\ prior\ seasons\ modeled$$

These weighted coefficients were used to generate predictions for the current season based on available CLM landings (provide in table 1):

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{w}} = \frac{\sum_{h=1}^{3} {\hat{\sigma}^2}_h^{\,-1} \widehat{\boldsymbol{\beta}}_h}{\sum_{h=1}^{3} {\hat{\sigma}^2}_h^{\,-1}} \ \{\text{eq. 5}\}$$

The standard error of each predicted value can be calculated using the delta-method (not accounting for the correlation or variance structure of the model):

$$\hat{\sigma}^2{}_i = \boldsymbol{x}_i{}'\boldsymbol{V}\boldsymbol{x}_i \ \{\text{eq. 6}\}$$

These standard errors can be used to estimate (1-α) % prediction intervals for model-predicted values using the following equation, where *t* represents the quantile function of a *t*-distribution:

$$y_i \pm \hat{\sigma}_h * t_{\left(\frac{\alpha}{2},\ n-p\right)} * \sqrt{\frac{\hat{\sigma}^2{}_i}{\hat{\sigma}^2{}_h} + 1} \ \{\text{eq. 7}\}$$

RESULTS AND DISCUSSION

Table 1: Three-year averaged GLS model coefficient estimates and standard errors from models fit to 2015, 2016, and 2017 South Atlantic Cobia Atlantic Group Landings. Inverse of model variances were used to calculate weighted estimates (equation 5).

Coefficient	Estimate	SE	Weighted Estimate	SE_w
VA	4.24	0.312	4.28	0.573
VA^2	$-1.74\text{x}10^{-4}$	$3.1\text{x}10^{-5}$	$-1.58\text{x}10^{-4}$	$6.5\text{x}10^{-5}$
NC	0.837	0.372	1.17	0.902
NC^2	$3.7\text{x}10^{-5}$	$1.8\text{x}10^{-5}$	$1.3\text{x}10^{-5}$	$6.6\text{x}10^{-5}$
OTHER	0.553	0.643	1.18	1.85
$OTHER^2$	$4.88\text{x}10^{-4}$	$1.62\text{x}10^{-4}$	$4.34\text{x}10^{-4}$	$2.05\text{x}10^{-4}$
Mon	107.8	1718.8	-1398.2	4125.7
Tue	82.1	1719.0	-1444.0	4175.5
Wed	114.3	1720.8	-1418.7	4194.5
Thu	107.8	1720.0	-1418.0	4172.1
Fri	87.8	1718.6	-1430.9	4156.8
Sat	156.7	1717.2	-1369.2	4179.4
Sun	196.9	1715.1	-1316.5	4143.6

Figure 1: Comparison of model predicted time-series of landings with realized values. Forecasted values are based off the modeled relationship between prior season (2015) ACCSP landings and CLM landings, which is applied to current season (2016) CLM landings. The dashed blue line represents GLS model predictions for 2016 against actual ACCSP landings in red. Error bar represents a 95% prediction interval for GLS model at season close (note it encompasses the true ACCSP value). Simple linear model predictions are provided for comparison as a dashed pink line. The dashed horizontal line displays the ACL weight, intersected by the model predicted closure date (vertical blue line) and the actual date the quota was reached (vertical red line). Note models are only fit to landings after the empirical breakpoint (early May).

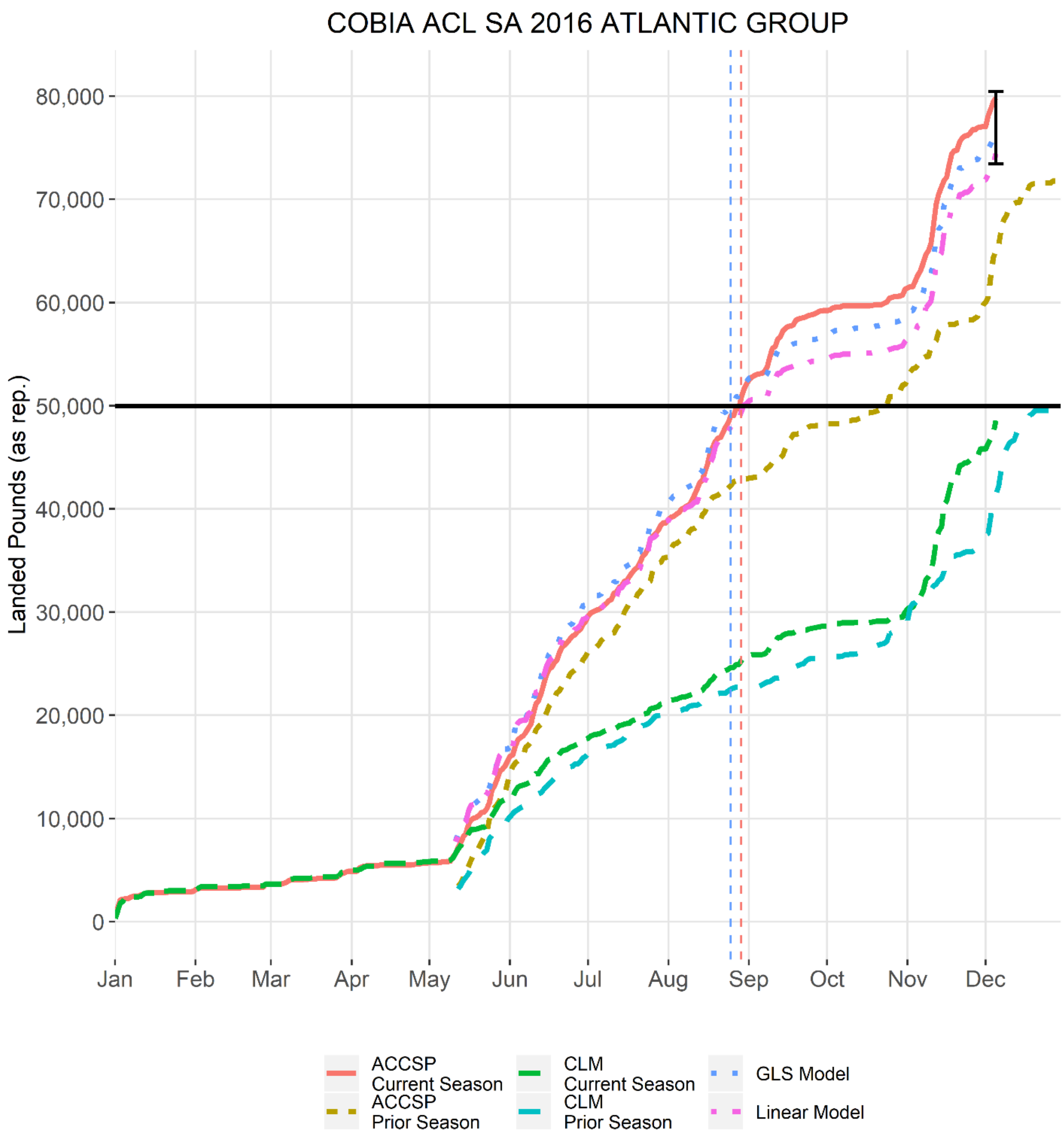

Table 2: Comparison of in-season CLM landings with final ACCSP landings and GLS model predicted landings. The Atlantic Group Cobia fishery was established in 2015, so model predictions are only available beginning in 2016. *In 2017, a closure date was recommended using a simple linear model to account for missing state landings. No correction was used in 2015 or 2016.

Year	Quota Weight	CLM Total Landings	Closure Date	ACCSP Total Landings	% ACL	Actual Date Quota Reached	Model Predicted Closure Date	Model % ACL	Model RMSE
2015	60,000	50,054	December 31st	73,326	122%	November 30th	--	--	--
2016	50,000	48,722	December 6th	80,269	161%	August 29th	August 25th	97%	2.0%
2017*	50,000	28,407	September 5th	58,392	117%	August 21st	August 5th	90%	7.1%
2018	50,000	25,376	September 5th	46,901	94%	--	September 3rd	94%	7.9%

Observe the significantly earlier recommended closure dates when accounting for state under-reporting vs. managing strictly on in-season landings available to CLM. For both 2015 and 2016, the GLS model predicted closure dates much closer to when the fishery actually reached the quota, even when modeled on a single year level. It should be noted that the estimated closure dates provided by these models assume that all fishing stops on the predicted date. Therefore, if anything, predicted closure dates should be viewed as liberal since fishing can still occur in state waters following a federal closure.

While the implementation of a simple linear model in 2017 resulted in a 60% reduction in overfishing compared to the prior season (117% vs. 161% of the ACL), the GLS model would have predicted an earlier (and more accurate) recommended date. It is expected that the three-year weighted average model introduced in 2018 will result in even more robust predictions in future years.

REFERENCES

Hinkley, D.V. 1970. Inference about the change-point in a sequence of random variables, Biometrika 57(1): 1-17.

Hyndman R., G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O'Hara-Wild, F. Petropoulos, S. Razbash, E. Wang and F. Yasmeen. 2018. forecast: Forecasting functions for time series and linear models. R package version 8.4, http://pkg.robjhyndman.com/forecast.

Killick R., K. Haynes and I.A. Eckley. 2016. changepoint: An R package for changepoint analysis. R package version 2.2.2, https://CRAN.R-project.org/package=changepoint.

Pinheiro J., D. Bates, S. DebRoy, D. Sarkar and R Core Team. 2018. nlme: Linear and nonlinear mixed effects models. R package version 3.1-137, https://CRAN.R-project.org/package=nlme.

APPENDIX 1: Model Output

SA Cobia Atlantic Group: Generalized Least-Squares Models

	ACCSP LANDINGS		
	2016	2017	2018
VA	4.876*** (0.234)	3.975*** (0.188)	3.854*** (0.447)
VA^2	-2.01e-4*** (2.69e-5)	-1.05e-4*** (1.53e-5)	-2.16e-4*** (4.44e-5)
NC	1.288*** (0.053)	1.619*** (0.107)	-0.398 (0.632)
NC^2	-6.22e-6*** (9.93e-7)	-1.25e-5*** (2.16e-6)	1.31e-4*** (3.03e-5)
OTHER	2.162** (0.877)	1.545*** (0.341)	-2.046*** (0.594)
$OTHER^2$	2.64e-4 (2.52e-4)	4.59e-4*** (7.10e-5)	7.41e-4*** (9.81e-5)
Mon	-1,061*** (397)	-3,943*** (608)	5,327* (2,887)
Tue	-1,132*** (394)	-4,004*** (609)	5,382* (2,887)
Wed	-1,105*** (394)	-3,991*** (607)	5,439* (2,891)
Thu	-1,121*** (394)	-3,969*** (610)	5,412* (2,889)
Fri	-1,112*** (395)	-3,984*** (610)	5,359* (2,887)
Sat	-1,033*** (396)	-3,944*** (611)	5,448* (2,884)
Sun	-992** (397)	-3,865*** (611)	5,448* (2,880)
Observations	233	211	122
Log Likelihood	-1,542.4	-1,431.7	-821.8
Akaike Inf. Crit.	3,124.9	2,901.5	1,681.7
Bayesian Inf. Crit.	3,192.7	2,964.0	1,732.8

Note: Standard errors in parenthesis *p<0.1; **p<0.05; ***p<0.01

Model output formatted by stargazer:

Hlavac, M. 2018. stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.1. https://CRAN.R-project.org/package=stargazer.