North American Journal of Fisheries Management

ISSN: 0275-5947 (Print) 1548-8675 (Online) Journal homepage: http://www.tandfonline.com/loi/ujfm20

Forecasting for Recreational Fisheries Management: What's the Catch?

Nicholas A. Farmer & John T. Froeschke

To cite this article: Nicholas A. Farmer & John T. Froeschke (2015) Forecasting for Recreational Fisheries Management: What's the Catch?, North American Journal of Fisheries Management, 35:4, 720-735, DOI: 10.1080/02755947.2015.1044628

To link to this article: https://doi.org/10.1080/02755947.2015.1044628

North American Journal of Fisheries Management 35:720–735, 2015

ISSN: 0275-5947 print / 1548-8675 online
DOI: 10.1080/02755947.2015.1044628

ARTICLE

Forecasting for Recreational Fisheries Management: What's the Catch?

Nicholas A. Farmer*
National Marine Fisheries Service, Southeast Regional Office, 263 13th Avenue South, St. Petersburg, Florida 33701, USA

John T. Froeschke
Gulf of Mexico Fishery Management Council, 2203 North Lois Avenue, Suite 1100, Tampa, Florida 33607, USA

Abstract
The Magnuson–Stevens Fishery Conservation and Management Reauthorization Act of 2006 required regional fishery management councils to implement annual catch limits (ACLs) for nearly all stocks under U.S. federal management. Since 2011, the number of stocks requiring ACLs (and monitoring) has increased nearly 10-fold, with strict accountability measures requiring either in-season quota closures or shortening of subsequent seasons to avoid ACL overages. Robust forecasts of landings can also provide a projected baseline for evaluation of proposed management alternatives. We compared generalized linear models (GLMs), generalized additive models (GAMs), and seasonal autoregressive integrated moving average (SARIMA) models in terms of fit, accuracy, and ability to forecast landings of four representative fish stocks that support recreational fisheries in the southeastern United States. All models were useful in developing reliable forecasts to inform management. The GAMs provided the best fit to the observed data; however, the modeling approaches of the SARIMA model and GLM provided the best forecasts for most scenarios. The SARIMA model and GLM also provided the best predictions of the seasonal trend in landings, a desirable feature for in-season quota monitoring. The SARIMA model was more sensitive and the GLM was less sensitive to recent trends, providing a useful "bookend" for forecasts. The time span of input data affected forecast accuracy from all model types considered. This study suggests multiple forecasting models should be investigated and performance metrics carefully selected and evaluated, as no single model is likely to perform best for all stocks of interest.

The Magnuson–Stevens Fishery Conservation and Management Reauthorization Act of 2006 requires regional fishery management councils to specify annual catch limits (ACLs) at a level such that overfishing does not occur. Annual catch limits are required for all stocks under U.S. federal management except for stocks with annual life cycles and those managed by international agreements in which the USA participates. This provision was implemented in 2010 or earlier for stocks subject to overfishing and in 2011 for all other federally managed stocks. This requirement resulted in a nearly 10-fold increase in the number of ACLs that must be monitored (from 2012 forward) relative to previous years (NMFS 2014). To address this challenge, methods for forecasting fisheries landings and projecting season lengths to avoid ACL overages are needed. Reliable forecasting methods are needed especially for recreational fisheries in the southeastern region of the United States. In this region, recreational landings comprise the majority of total landings for many species (Coleman et al. 2004); however, most have limited in-season harvest information available (i.e., data available in 2-month "waves" after a 45-d delay for each wave).

*Corresponding author: nick.farmer@noaa.gov
Received December 18, 2014; accepted April 20, 2015

Forecasting fish landings is a critical element in the management of fisheries stocks because it can inform strategy development and policy decisions on timelines necessary for effective management (Stergiou and Christou 1996; Makridakis et al. 2008). Forecasts can be used to apply in-season or postseason accountability measures and also to provide a baseline for forecasting the impacts of proposed management actions. To date, forecasting applications in fishery management applications are limited. Ward et al. (2014) evaluated a suite of models across more than 2,000 vertebrate taxa and provided some general guidance. In the U.S. South Atlantic and Gulf of Mexico, Hanson et al. (2006) evaluated three models used to forecast annual landings of Atlantic Menhaden *Brevoortia tyrannus* and found that multiple regression and artificial neural networks could be used for this long-term commercial fishery. Growth and production of brown shrimp *Penaeus aztecus* are also forecasted in the Gulf of Mexico based on environmental conditions in estuaries (Adamack et al. 2011). To be useful, appropriate methodologies need to be developed and evaluated, weighing the tradeoffs of model complexity, performance, and the ability to inform management (Tsitsika et al. 2007). Approaches to forecasting fish landings are varied but generally fall into four broad categories: (1) using the previous year's landings, (2) population dynamics models, (3) correlation-based regression models, and (4) time series models.

Population dynamics models are advantageous because they attempt to characterize factors affecting abundance, productivity, and growth potential of a stock (Hilborn and Walters 1992; Buckland et al. 2004; Newman et al. 2006). Unfortunately, these models are data intensive and require substantial time, effort, and resources to develop (Ward et al. 2014). Due to these limitations, stock assessment models are only developed every 3–5 years for economically important species in the southeastern United States. For many federally managed species, adequate data are unavailable and resources are insufficient to develop population dynamics assessment models (Carruthers et al. 2014; Berkson and Thorson 2015). Moreover, when forecasting is the primary objective, population dynamics models are not necessarily superior to other less-intensive methods as they require estimates of many parameters and have a tendency to overfit, limiting their forecasting performance (Clark 2004).

Correlation-based regression models (e.g., linear models) have been used successfully to predict menhaden landings in U.S. Atlantic waters and the Gulf of Mexico since at least 1975 (Schaaf et al. 1975) and were used for more than three decades to produce annual forecasts of landings (Hanson et al. 2006). However, landings for many species follow nonlinear trajectories where the response variable may be more appropriately modeled using non-Gaussian error distributions (Ward et al. 2014). Generalized linear models (GLMs) (Nelder and Wedderburn 1972) are extensions of linear models that can accommodate response variables following exponential family distributions (e.g., Poisson, negative binomial) and may be superior to linear models for modeling fish landings data. Generalized additive models (GAMs) (Wood 2006) extend the GLM by allowing nonparametric relationships between the response and explanatory variables (Wood 2003). Rigorous routines for model selection and validation may prevent overfitting that occurs with these models (Zuur et al. 2010). Most correlation-based methods do not account for time explicitly in the model, although some methods may provide this capability (e.g., generalized estimating equations). If covariates are used, a determination of future values of covariates is required to develop a forecast. In some cases, this can be quite realistic (e.g., landings restriction due to closed season); however, in other cases it may be difficult or impossible to predict (e.g., environmental conditions).

Time-series models are conceptually simple and popular tools for forecasting. Seasonal autoregressive integrated moving average (SARIMA) models can be constructed using only the information contained in the series (Dennis et al. 1991; Holmes 2001; Ives et al. 2010) and aim to describe the autocorrelation in these data (Hyndman and Athanasopoulos 2014; Ward et al. 2014). More simply, this can be thought of as a multiple regression model with lagged values as covariates. These models are flexible and assume that future conditions are similar to the past conditions that generated these observed data. The SARIMA models assume that the time series is stationary with stable variance throughout the time period. Unfortunately, these assumptions are frequently violated with fisheries data, although this can often be resolved through differencing and/or transformation, especially to capture seasonal trends (Box et al. 2013).

The purpose of this study was to evaluate a suite of approaches to produce short-term forecasts at 2-month intervals (i.e., waves) necessary to inform fisheries management decisions for U.S. federally managed species in the Gulf of Mexico and Atlantic Ocean. Specifically, we considered approaches that could be fit with minimal data (e.g., landings data) and applied to a range of species with varied life histories and fisheries characteristics. We used four representative fish stocks (or stock complexes) that support recreational fisheries currently managed by the South Atlantic Fishery Management Council (SAFMC) or the Gulf of Mexico Fishery Management Council (GMFMC) and compared the performance of GLMs, GAMs, and SARIMA models in terms of model fit, accuracy, and forecasting ability. The goal of these approaches was to develop reliable methods for predicting the timing of in-season closures to avoid exceeding an ACL and predicting total annual landings in the absence of a quota closure.

METHODS

Recreational Fisheries Catch Data

Recreational landings data were obtained from the NMFS Southeast Fisheries Science Center (SEFSC) ACL data set (accessed May 2013), which provided aggregated landings data from 1986 to 2012 from the Marine Recreational Fisheries Statistics Survey (MRFSS), the Southeast Headboat Survey (HBS), and the Texas Parks and Wildlife Department (TPWD) Creel Survey. Landings data from the various surveys are provided in both numbers and weight (pounds). The ACL data set provides improved quality assurance and quality control on the raw data generated by each of these surveys; for example, the ACL data set implements a hierarchical procedure to backfill missing weight estimates from MRFSS (now MRIP; http://www.st.nmfs.noaa.gov/recreational-fisheries/index). In short, samples are aggregated upward (i.e., wave, mode) to ensure adequate sample size (i.e., $\geq$30).

The MRFSS (http://www.st.nmfs.noaa.gov/recreational-fisheries/index) intercepts collect data on port agent observed landings (A catch) and angler-reported landings (B1 catch) and discards (B2 catch) in numbers by species, 2-month wave (e.g., wave 1 = January–February, … wave 6 = November–December), area fished (inland, state, and federal waters), mode of fishing (charter, private and rental, shore), and state (North Carolina to Louisiana). These dockside intercepts are expanded using effort data collected via telephone surveys (private and rental: random digital dial during each wave; for-hire: weekly 10% random sample). In 2012, MRFSS was nominally replaced by the Marine Recreational Information Program (MRIP). In 2013, the MRFSS survey methodology was modified by MRIP, resulting in some changes that are still being calibrated by SEFSC. Thus, MRIP values from 2013 forward were not considered for this modeling exercise.

Landings of headboats (i.e., recreational vessels where customers pay "by the head") are calculated using a combination of logbook reports and dockside sampling, and adjustments to landings are made based on underreporting and misreporting determined through dockside validation by port agents. Fishing records from the Southeast Headboat Survey (http://www.sefsc.noaa.gov/laboratories/beaufort/sustainable/headboat/) contain trip-level information on number of anglers, trip duration, date, area fished, landings (number of fish), and releases (number fish) by species.

The TPWD Creel Survey (https://tpwd.texas.gov/fishboat/fish/didyouknow/creel.phtml) generates estimates of landings for private and rental boats and charter vessels fishing off the Texas coast. The TPWD conducts a stratified random angler-intercept survey at specified boat-access sites throughout the year. Landings are reported by TPWD in numbers by "high-use" (May 15–November 20) and "low-use" (November 21–May 14) time periods, area fished (state and federal waters), and mode (charter, private and rental). The high- and low-use landings estimates provided by TPWD are re-estimated by NMFS personnel to correspond to the MRFSS 2-month waves.

Landings time series for three recreationally important stocks and one incidentally caught stock complex with relatively simple management histories were assembled. Landings for Vermilion Snapper *Rhomboplites aurorubens* and Gray Snapper *Lutjanus griseus* managed by the GMFMC as well as Red Porgy *Pagrus pagrus* and the "grunts" complex managed by the SAFMC were computed as the sum of MRFSS, HBS, and TPWD landings by year and wave. The SAFMC grunts complex contains White Grunt *Haemulon plumierii*, Margate *H. album*, Sailor's Choice *H. parra*, and Tomtate *H. aurolineatum*; most of these stocks are incidentally caught on trips targeting other species. During the years considered for this analysis, none of these stocks were subject to quota closures.

Management histories were reconstructed for all four species to account for the timing of federal recreational quota closures and closed seasons. For projection purposes, all recreational landings were assigned to 2-month waves. Model inputs for each species were expressed as landings (in pounds whole weight) per open day (landings were assumed equal for all open federal days within a wave). Expressing landings as a daily rate was important for determining the date a catch limit might be exceeded and also for handling any closures in the management history of the stock. As states adopted compatible seasonal regulations for the species of concern, all landings were assumed to occur within the federal season; thus, the federal open days by wave were used as the divisor for computing wave-specific landings per day. To reduce prediction bias associated with reductions in landings due to fisheries closures in the Gulf of Mexico following the Deepwater Horizon–BP oil spill in April 2010, values for April–December 2010 in the Gulf of Mexico were recomputed as the average of 2009 and 2011 values for the same time period. No adjustments were made for more spatiotemporally discrete events such as hurricanes and red tides.

Modeling Approach

Time series of recreational harvest for each species were fitted using GLMs (Hardin and Hilbe 2007), GAMs (Wood 2006), and SARIMA models (Box et al. 2013). Projected landings per day by wave were projected in weight (pounds) instead of numbers because the ACLs for these stocks are specified in pounds.

GLM.—Long-term and seasonal trends in the landings-per-day time series were captured using a GLM, fit with Proc GENMOD in SAS version 9.2 software (2000; SAS Institute, Cary, North Carolina). Mean landings per day (lb) were dependent upon a linear predictor of year and a quadratic predictor of wave, which were linked via a log-link function with a negative binomial response error distribution (Nelder and Wedderburn 1972). Residual diagnostics and Akaike's information criterion (AIC; Akaike 1974) values were used to select the final model configurations.

GAM.—Generalized additive models were also fit to each time series. Mean landings per day (lb) were predicted using a cubic-spline smoother (s) for the main effects (year) and a tensor product spline (te) (De Boor et al. 1978) for the interaction term (wave, year) (Gasper et al. 2013). The GAMs were fit using the mgcv library (Wood 2006) in R version 3.0.2 (R Development Core Team 2013). Backward selection was used to determine whether predictors or interactive effects could be removed without compromising model performance. Akaike's information criterion and a log-likelihood ratio test were used to determine whether more complex models were warranted (Froeschke et al. 2012).

SARIMA model.—Time series exhibiting a long-term trend and a seasonal trend may be well-suited to a SARIMA model (Box et al. 2013). In a SARIMA$(p,d,q)\times(P,D,Q)$ model, the autoregressive component (p) represents the lingering effects of previous observations, the integrated component (d) represents temporal trends, and the moving average component (q) represents lingering effects of previous random shocks (or error). The SARIMA models were implemented using Proc ARIMA in SAS version 9.2 (SAS Institute). All possible combinations of single-difference SARIMA models for landings per day by wave were considered (Table A.1 in the Appendix). A single-difference SARIMA model only considers a maximum of one differencing term in the annual and one differencing term in the seasonal component. All SARIMA models were fit using conditional least squares. Stationarity tests were used to guide differencing selection. Final SARIMA model selection was guided by the examination of autocorrelations, inverse autocorrelations, partial autocorrelations, cross-correlations, residual diagnostics, and AIC.

Model Evaluation and Performance

Time series of three different lengths (i.e., 1999–2011, 2004–2011, and 2007–2011) were compared in terms of model fit and forecasting performance. Exploring time series of varying lengths is important as stocks vary in the period for which reliable landings data exist, and this approach permits a mechanism to examine trade-offs with model complexity across time series of different lengths that are not confounded by individual species effects. Although data were available prior to 1999, preliminary projections suggested model performance was occasionally improved by truncating the time series but not by extending it to before 1999. To evaluate forecast utility, we evaluated the proportion of variation explained by the covariates (R^2), and the mean error (i.e., observed versus fitted values) for the final year of data. For Atlantic stocks, we also removed the terminal year from the time series (i.e., "drop-one" approach), refit the model to 2004–2010 data, and predicted landings for 2011 to provide a more robust evaluation of forecast performance. This was accomplished by using the fitted model to forecast beyond the data that were used to build the model and more closely simulate how these models would be used in practice by resource managers. The deviance between the forecast and the actual landings in the final year provided an additional estimate of accuracy. This drop-one approach was only applied to Atlantic stocks due to the confounding effect of having up to 36.6% of the Gulf of Mexico EEZ closed to fishing in 2010 because of the Deepwater Horizon–BP oil spill. Finally, a variation on the drop-one approach was applied to all four stocks by plotting cumulative landings time series to evaluate model fits from 1999–2011, 2004–2011, and 2007–2011 data relative to observed values in 2011 and model forecasts relative to observed values in 2012. A simple approach of using the previous year's landings as a forecast was also explored for all scenarios. As SARIMA uses a Gaussian error structure and permits negative forecast values, all SARIMA-based predictions of negative landings within a wave were converted to zeroes for these comparisons.

RESULTS

Most stocks exhibited long-term trends as well as seasonal periodicity in landings. Landings were typically lowest during winter (i.e., waves 1 and 6) and peaked during summer (i.e., waves 3 and 4). Model statistics are provided in Table 1. For the longer time series (i.e., 1999–2011 and 2004–2011), a SARIMA$(0,1,1)\times(0,1,1)$ structure fit the data best of the different SARIMA models considered, meaning the data were differenced at the previous time step and the seasonal time step, and a moving average term was used on both to fit the data. In the shortest time series evaluated (i.e., 2007–2011), a SARIMA$(1,1,0)\times(0,1,1)$ structure fit the data best of the different SARIMA models considered, indicating an autoregressive term did better at capturing the trend with a limited set of data than a moving average.

Gulf of Mexico

From 1999 to 2011, Vermilion Snapper landings peaked during the summer each year and total annual landings increased during the time series (Figures 1, 2). All modeling approaches captured this pattern after appropriate model fitting and selection routines. For Vermilion Snapper, R^2 increased with shorter time series for all models and the GAM model provided the best fit to the observed data (Table 1). Examining the mean error during the final year of data indicated the SARIMA model fits were much closer to the observed values, and the lowest mean error in the final year was provided by the shortest time series (2007–2011; SARIMA model: 162.66 lb/d; GAM: 735.01 lb/d; GLM: 607.86 lb/d). This time series was nonstationary and had increasing rates of landings toward the end of the period. Of the models considered, the SARIMA model most closely captured this pattern in the observed data and only underestimated landings by 5% for the 2007–2011 input time series. The other modeling approaches resulted in much higher underestimation of total landings (19–53%), and the GLM showed the greatest fluctuation in accuracy depending upon input time series.

TABLE 1. Goodness of fit (R^2), mean error in terminal year across waves (ME; lb/d), total percent error in final projected cumulative landings (TE), and mean error in projected year across waves (Drop1) for different stocks, modeling approaches, and time series. Drop1 denotes forecasts where the terminal year of data are removed, the model refit, and the terminal year fit is compared with the observed values; PrevYr = previous year, NA = not applicable.

Stock	Model	R^2 (1992–2011)	R^2 (1999–2011)	R^2 (2004–2011)	R^2 (2007–2011)	ME (1992–2011)	ME (1999–2011)	ME (2004–2011)	ME (2007–2011)	% TE (1992–2011)	% TE (1999–2011)	% TE (2004–2011)	% TE (2007–2011)	ME_Drop1 (2004–2010)	% $TE_{fit2010}$_Drop1 (2004–2010)	% $TE_{predict\,2011}$_Drop1 (2004–2010)
Gulf of	SARIMA	0.66	0.69	0.73	0.86	895	518	513	163	29	17	17	−5	NA	NA	NA
Mexico	GAM	0.70	0.79	0.81	0.91	1,140	1,031	995	735	−36	−33	−32	−23	NA	NA	NA
Vermilion	GLM	0.75	0.75	0.75	0.91	1,660	1,067	1070	608	−53	−34	−34	−19	NA	NA	NA
Snapper	PrevYr	NA	NA	NA	NA	1,097	1,097	1,097	1,097	−34	−34	−34	−34	NA	NA	NA
Gulf of	SARIMA	0.64	0.60	0.66	0.85	1,002	840	1,307	59	−25	−21	−33	2	NA	NA	NA
Mexico	GAM	0.70	0.79	0.81	0.91	480	591	2,475	650	18	22	93	24	NA	NA	NA
Gray	GLM	0.50	0.52	0.63	0.65	2,034	1,892	900	463	76	71	34	17	NA	NA	NA
Snapper	PrevYr	NA	NA	NA	NA	385	385	385	385	16	16	16	16	NA	NA	NA
Atlantic grunts	SARIMA	0.51	0.42	0.37	0.31	212	51	368	169	−25	−6	−42	−19	735	7	−67
complex	GAM	0.70	0.79	0.81	0.91	68	80	4	84	7	9	−1	−10	218	17	−26
	GLM	0.45	0.45	0.50	0.64	290	332	119	66	33	38	13	−8	213	52	24
	PrevYr	NA	NA	NA	NA	104	104	104	104	−12	−12	−12	−12	104	86	−12
Atlantic Red	SARIMA	0.40	0.61	0.66	0.65	36	17	11	58	−12	−5	−1	−25	6	6	4
Porgy	GAM	0.72	0.70	0.84	0.85	1	43	43	30	0	−22	−22	−15	129	−20	−66
	GLM	0.63	0.60	0.66	0.85	5	114	20	52	2	58	10	−27	67	51	34
	PrevYr	NA	NA	NA	NA	18	18	18	18	−9	−9	−9	−9	18	46	−9

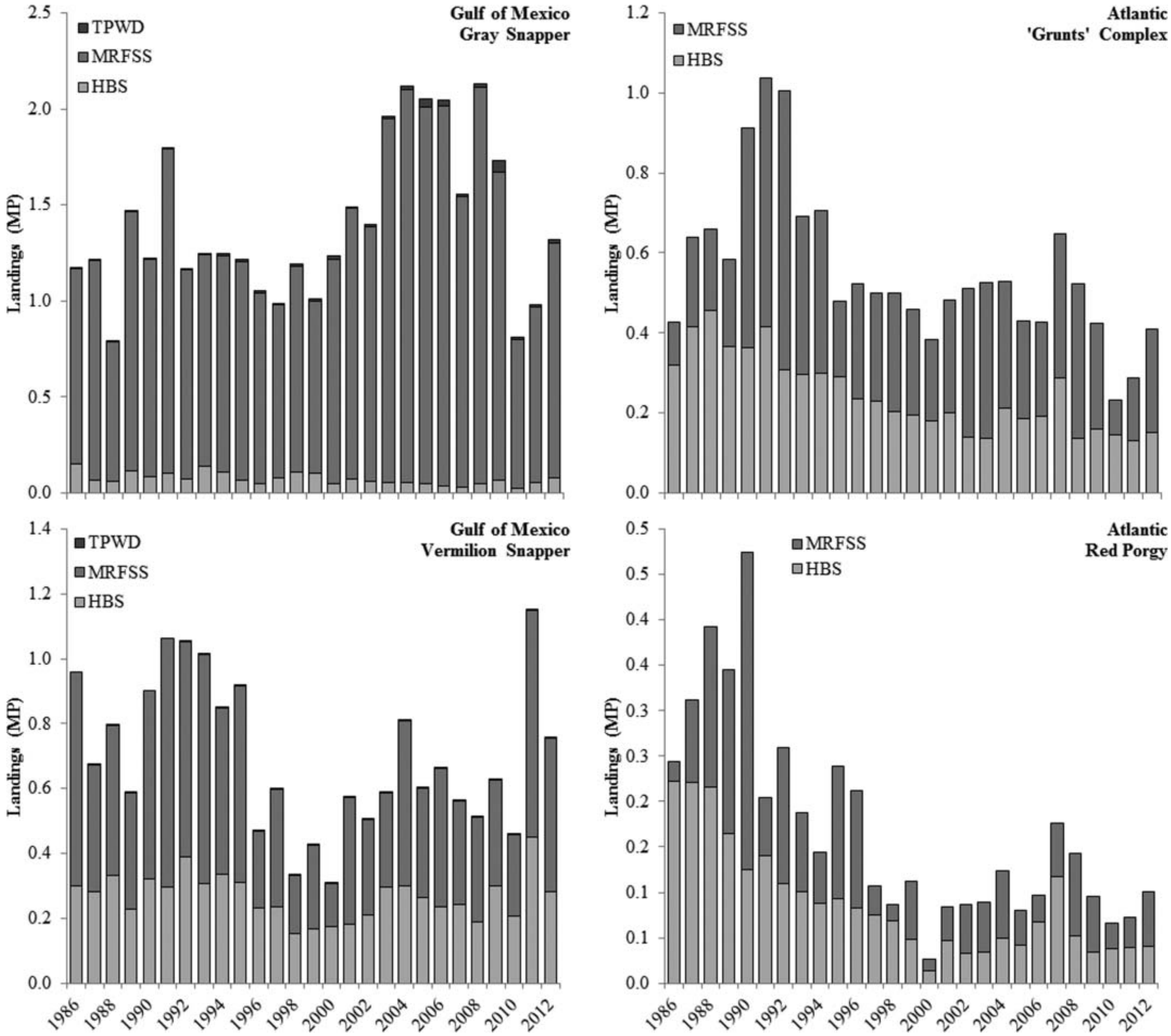

FIGURE 1. Time series of recreational landings data from 1986 to 2012, in millions of pounds (MP) whole weight, for Gulf of Mexico Gray Snapper and Vermilion Snapper, and Atlantic grunts complex and Red Porgy, by data source (Texas Parks and Wildlife Department Creel Survey [TPWD], Marine Recreational Fisheries Statistics Survey [MRFSS], and Southeast Headboat Survey [HBS]).

From 1999 to 2011, Gray Snapper landings peaked during the summer each year and total annual landings increased and then decreased during the time series (Figures 1, 3). As with Vermilion Snapper, all three models captured this pattern and R^2 increased with shorter time series for all three model approaches; the GAM was the best fit to the observed data (Table 1). In terms of explained variance, the GLM and SARIMA model were comparable although the SARIMA model had lower mean error in the final year of the times series. The lowest mean error in the final year was provided by the shortest time series (2007–2011; SARIMA model: 59.31 lb/d; GAM: 650.30 lb/d; GLM: 463.01 lb/d). The time series was nonstationary; however, the greatest annual landings of Gray Snapper occurred in the middle third of the time series. Fits of regression models to the final year in the time series were highly dependent on the input time series selected, and in most cases were outperformed by the previous year's landings. In the shortest time series considered (i.e., 2007–2011), the SARIMA model provided the best fit, overestimating cumulative landings by only 2%. The SARIMA model produced negative landings predictions for some waves (Figure 3).

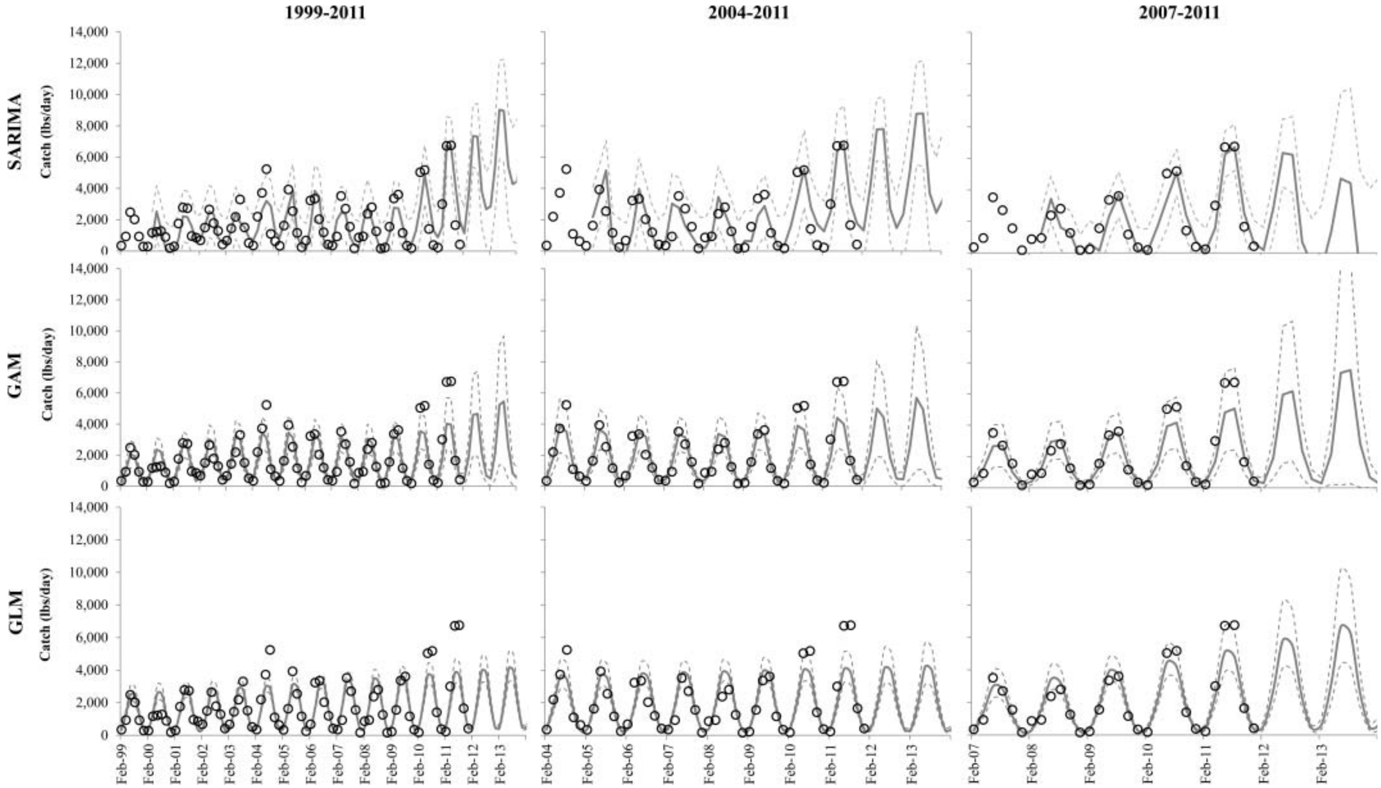

FIGURE 2. Three statistical models (solid gray line) and their 95% confidence limits (dashed gray line) were fit to landings data (lb/d) of Gulf of Mexico Vermilion Snapper from 1999 to 2011 (open circles) to evaluate model fits across model types (GLM, GAM, and SARIMA model) and times series (1999–2011, (2004–2100, and 2007–2011).

Atlantic Ocean

Red Porgy annual landings were relatively stable from 1999 to 2012 with the exception of a trough in 2000 and a peak in 2007 (Figure 1). Red Porgy displayed a distinct seasonal pattern with landings rates peaking during summer each year (Figure 4), and this pattern was captured by all models. Model fits improved with shorter time series and the GAM provided the best fit to the observed data (Table 1). Examining the mean error during the final year of data indicated the SARIMA model fits were much closer to the observed values. The lowest mean error in the final year was provided by the middle time series (2004–2011; SARIMA model: 10.51 lb/d; GAM: 43.33 lb/d; GLM: 20.41 lb/d). Of the models considered, the SARIMA model most closely captured the interannual pattern in the observed data, and model fits to the 2004–2011 time series underestimated 2011 cumulative landings by only 1%.

From 1999 to 2011, landings of the grunt species complex peaked during summer each year and total annual landings were relatively stable over the study period (Figures 1, 5). Similar to the other species examined, the explained variance in annual landings of grunts for all models increased with shorter time series and the GAM provided the best fit to the observed data (Table 1). In terms of explained variance, the GLM and SARIMA model were comparable, although the SARIMA model was more accurate than GLM when the fitted and observed values were compared in the final year of the times series. The lowest mean error in the final year was provided by the GAM in the middle time series (2004–2011; SARIMA model: 368.10 lb/d; GAM: 4.38 lb/d; GLM: 119.34 lb/d). There was a spike HBS and MRFSS landings in wave 3 in 2007 that was not captured by any models.

Forecast and Summary

The trend for the grunts complex was dynamic (see Figure 5). The drop-one-scenario model fits to the final year were excellent for the SARIMA model (only 7% error), but predictions from the SARIMA model were poor (a 67% underestimate). Both the SARIMA model and the GAM overweighted the long-term decline in landings (Figure 6). For the Atlantic grunts complex the most accurate prediction was provided by the previous year's landings. For Red Porgy, the SARIMA model provided the best model fit to the final year of data (a 6% overestimate) and the best forecast accuracy (a 4% overestimate). Landings levels in 2011 for both stocks were within the long-term range of previous landings levels.

Examination of model fits to cumulative observed landings for 2011 indicated that in 9 of 12 scenarios, model fits from

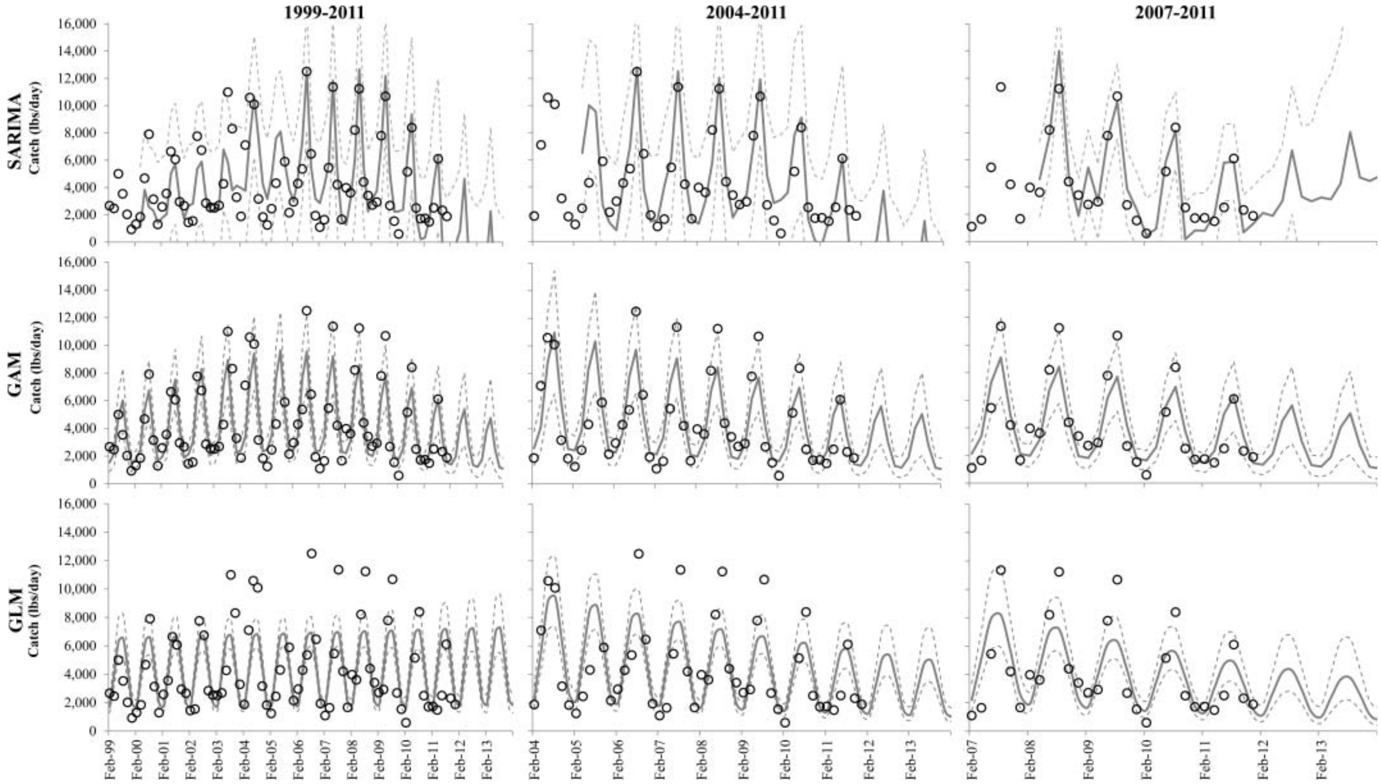

FIGURE 3. Three statistical models (solid gray line) and their 95% confidence limits (dashed gray line) were fit to landings data (lb/d) of Gulf of Mexico Gray Snapper from 1999 to 2011 (open circles) to evaluate model fits across model types (GLM, GAM, and SARIMA model) and times series (1999–2011, (2004–2100, and 2007–2011).

the SARIMA model were superior to the GLM and GAM (closer to observed values; Figure 7). The SARIMA model CIs were much larger than those for fitted GLM or GAM models (Figure 7). For SARIMA models, the CI contained the observed values in all 12 scenarios examined, whereas the CIs estimated using GLM and GAM did not always contain the observed values. A comparison of the percent deviation from the observed cumulative landings trend by wave, across stocks and time series, indicated that GLM provided the best overall model fits (0.2% ± 36.2% error; mean ± SD), followed by SARIMA model (5.7% ± 76.7%). The GAM and the previous year's landings provided similar overall predictive error (13.9% ± 31.1% and 13.2% ± 43.3%, respectively). One undesirable feature of the SARIMA model is that declining trends in landings during a given wave may be forecast as negative landings, as observed with Gulf Gray Snapper (Figure 3). In this study, negative forecasts were replaced with zeroes; however, it may be preferable to substitute the most recent year's landings for that wave to avoid underestimating harvest. This approach reduced mean error from SARIMA model predictions by wave across stocks and time series by nearly one-half (from 5.7% to 2.9%).

Total landings in 2012 for Gulf Vermilion Snapper were 23% lower than 2011 landings, whereas total 2012 landings for the other stocks evaluated were 35–42% higher than 2011 values. Examination of model forecasts to cumulative observed landings for 2012 indicated that in 5 of 12 scenarios, mean forecast values of the SARIMA model were closest to observed values predictions (Figure 8). In five of the remaining seven scenarios, the GLM provided the best predictions (Figure 8). For Gulf Gray Snapper, the SARIMA model provided the best prediction using the 2007–2011 time series (8% error). For Gulf Vermilion Snapper, the GLM provided the best prediction using the 2004–2011 time series (−1% error). For the Atlantic grunts complex, the best predictions were obtained from SARIMA model and GLM using the 1999–2011 time series (−4% and +4% error, respectively, in the cumulative landings prediction). For Red Porgy, the best prediction was from the GLM using the 1999–2011 time series (−16% error). The CIs for SARIMA models were much larger than the CIs for fitted GLMs or GAMs, indicating greater uncertainty (Figure 8). The SARIMA model tended to be more responsive to short-term trends in landings that deviated from the long-term average trend. For SARIMA models, the CIs contained the observed values in all 12 scenarios examined, whereas the CIs estimated using GLM and GAM did not always contain the observed values. Overall, SARIMA model fits to seasonal

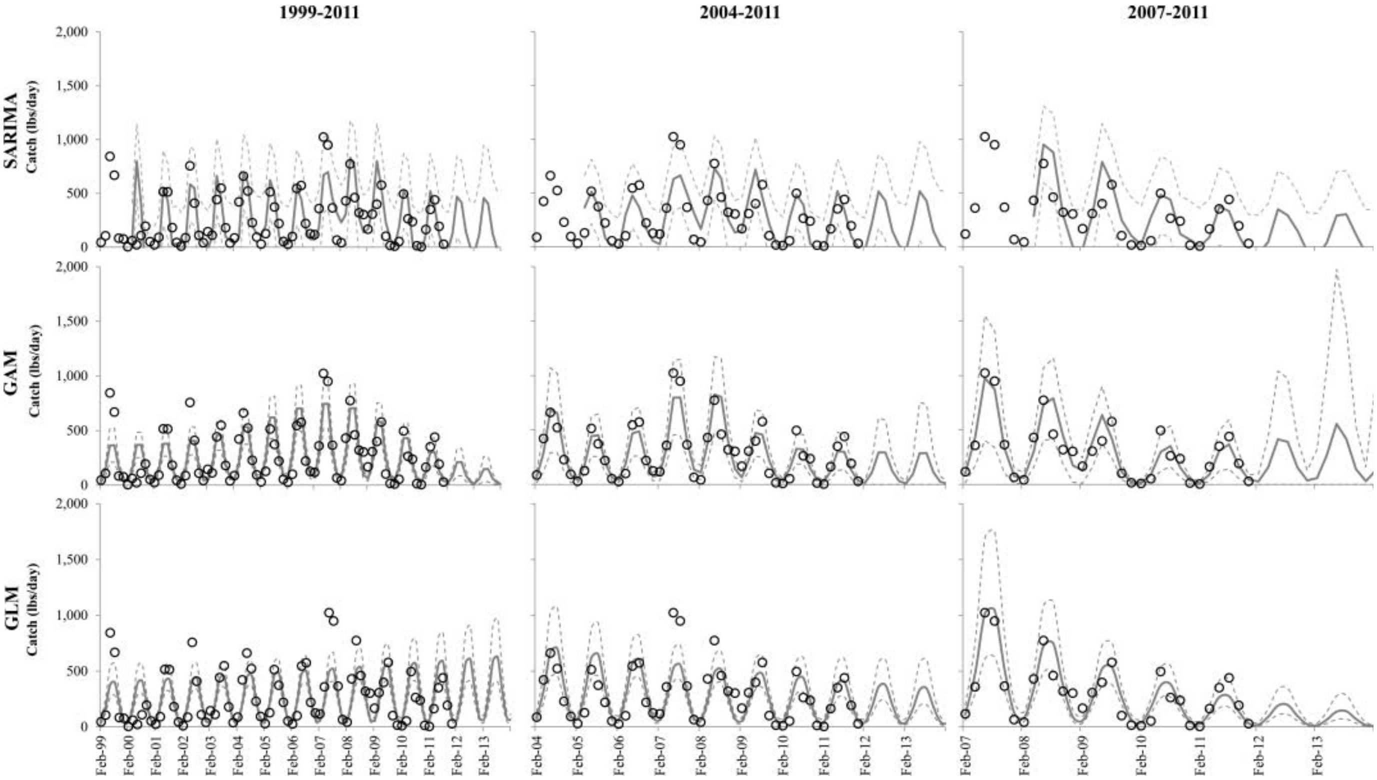

FIGURE 4. Three statistical models (solid gray line) and their 95% confidence limits (dashed gray line) were fit to landings data (lb/d) of Atlantic Red Porgy from 1999 to 2011 (open circles) to evaluate model fits across model types (GLM, GAM, and SARIMA model) and times series (1999–2011, (2004–2100, and 2007–2011).

patterns were less biased, but all model fits became more similar as the length of the input time series was reduced. In the 12 scenarios explored, at least one regression-based approach provided a superior prediction relative to using the previous year's landings. The graphical representation of the track of predicted landings relative to observed landings in Figure 8 can be used to evaluate the in-season monitoring performance of the various models. When the predicted landings track is above the observed landings for a given model, the model would predict an earlier quota closure date than necessary, resulting in an ACL underage that would reduce near-term economic benefits to the fishery. When the predicted landings track is below the observed landings for a given model, the model would predict a later quota closure date than necessary, resulting in an ACL overage that would trigger postseason AMs. There was substantial variability within and across models with regard to the cumulative landings track relative to model predictions.

A graphical comparison of the model-fitting and forecasting performance of GLM, GAM, and SARIMA model across the four stocks illustrates the tradeoffs in terms of model fit, explained variance, and forecasting performance (Figure 9). In terms of fitting the model to the observed data, the flexibility of the GAM provided superior fits for each stock relative to the SARIMA model and GLM. However, in terms of predictive performance, as indicated by fits to the terminal year of the time series and accuracy of drop-one scenario forecasts, the SARIMA model and GLM were generally superior to the GAM.

DISCUSSION

Federal requirements implemented in the amended Magnuson–Stevens Act (U.S. Congress 2006) require specification (and monitoring) of ACLs for most federally managed stocks. Resources are insufficient to develop population-dynamics-based landings projection models for most managed stocks (Hilborn and Walters 1992; Hanson et al. 2006). Thus, other methods must be identified to predict landings rates to ensure landings remain within prescribed ACLs (Carruthers et al. 2014). Given the large number of stocks that must be monitored (Berkson and Thorson 2015), routines must be robust to widely varying temporal patterns that characterize recreational landings patterns for most species (Ward et al. 2014). Similar to previous efforts with Atlantic Menhaden and Gulf Menhaden *B. patronus*, this study suggested statistical forecasting could be a viable approach to predicting landings (Hanson et al. 2006; Ives et al. 2010). A major goal of recreational fisheries management is to prevent catch limit overages. This can be accomplished by in-season closures or postseason

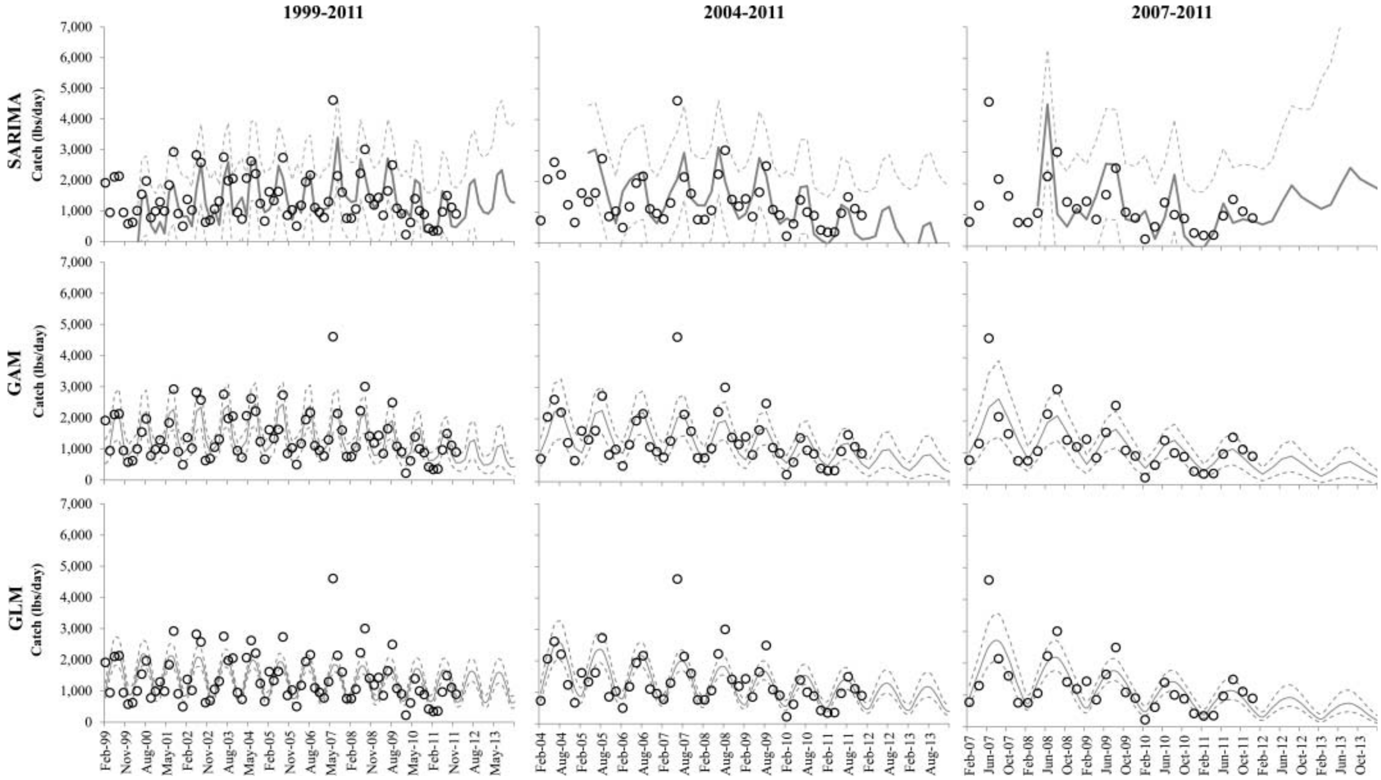

FIGURE 5. Three statistical models (solid gray line) and their 95% confidence limits (dashed gray line) were fit to landings data (lb/d) of the Atlantic grunts complex from 1999 to 2011 (open circles) to evaluate model fits across model types (GLM, GAM, and SARIMA model) and times series (1999–2011, 2004–2100, and 2007–2011).

adjustments to the regulations or season length in the following year. Accurate forecasts of recreational landings are critical to the application of both of these accountability measures.

Our study suggests semiautomated model-fitting and selection routines for the SARIMA model or GLM be used to develop short-term (i.e., 1 year) forecasts to inform management decisions; however, the quality and time span of input data can affect the accuracy of model forecasts. Longer time series tended to include up and down fluctuations in landings, whereas cutting the regression input time series omitted these fluctuations. By fitting to a shorter time series, the short-term trend tended to be better captured at the expense of long-term fluctuations in landings. No single model or time series performed best across all stocks of interest; thus, performance metrics need to be carefully selected and evaluated across multiple models. Our projections implicitly integrated the highly correlated terms of landings and effort by expressing landings rates as landings per open day. Changes in management regulations, environmental conditions, or economic conditions that might lead to changes in landings per unit effort would lead to increased uncertainty in forecasts; if these changes have occurred or are anticipated, they can be incorporated as covariates in the models or the data inputs can be parsed by time period or by region to better represent anticipated future conditions.

In general, SARIMA models performed well across a range of time series and would serve as an appropriate starting point for forecasting landings. The SARIMA model mean forecasts were generally unbiased in fits to observed data although confidence limits were consistently greater than those produced from GLMs or GAMs. The SARIMA models can accommodate but do not require additional covariates for either model building or forecast, a distinct advantage over the GLM and GAM. For in-season quota monitoring, the manager's goal is to close the fishery before the landings exceed the quota, but without forgoing harvest up to the quota. Thus, the predicted trajectory of cumulative landings is more important than the final projected total. Comparisons of the SARIMA model, GAM, and GLM forecasts fit to the 2011 cumulative observed landings time series indicated the SARIMA model approach best fit the cumulative landings time series for most scenarios. However, for some stocks, GLM performed better than the SARIMA model and was less sensitive than the SARIMA model or GAM to recent trends, providing a useful "bookend" for forecasts.

The SARIMA model forecasts should be treated with skepticism when they generate negative landings values, as they are likely overfitting a recent trend. Negative forecast values from any landings forecast model should minimally be replaced with zero values, as negative landings are not

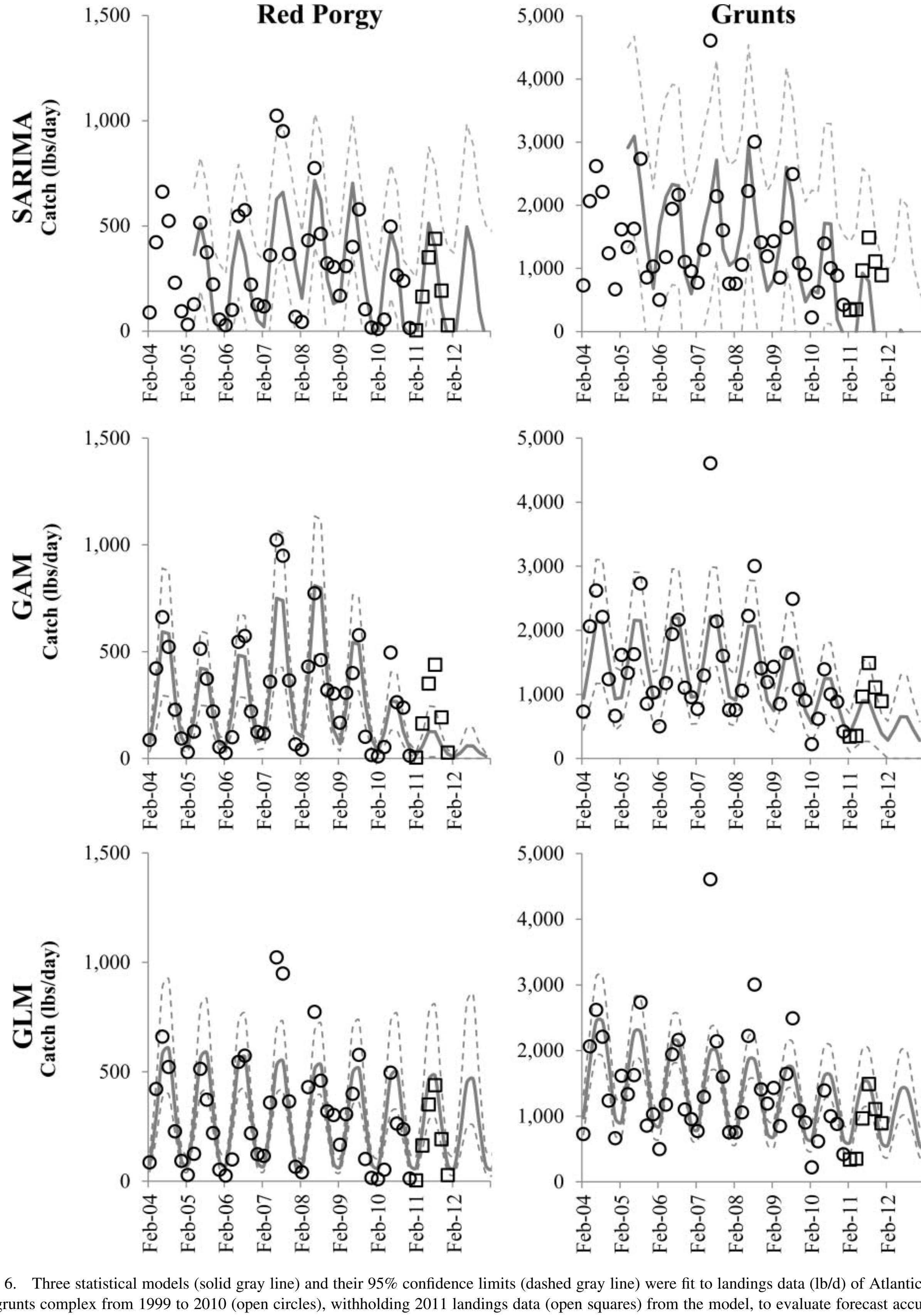

FIGURE 6. Three statistical models (solid gray line) and their 95% confidence limits (dashed gray line) were fit to landings data (lb/d) of Atlantic Red Porgy and the grunts complex from 1999 to 2010 (open circles), withholding 2011 landings data (open squares) from the model, to evaluate forecast accuracy across model types (GLM, GAM, and SARIMA model) and times series (1999–2011, 2004–2100, and 2007–2011).

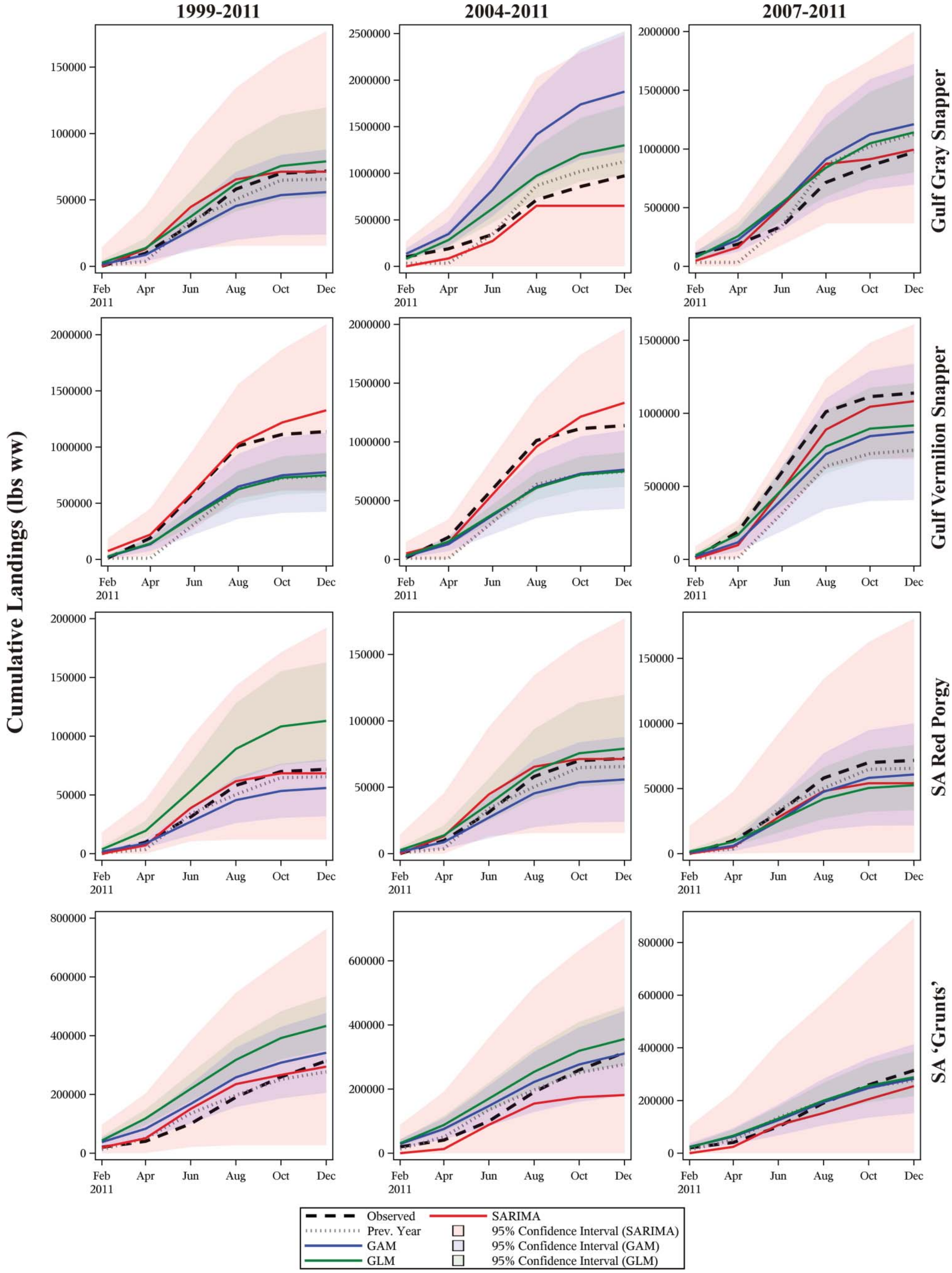

FIGURE 7. Cumulative landings plots showing SARIMA model (red), GAM (blue), and GLM (green) model fits and 95% confidence limits (shaded areas) relative to observed cumulative landings (lb wet weight) for 2011 based on 1999–2011, 2004–2011, and 2007–2011 time series data for Atlantic Red Porgy, Atlantic grunts complex, Gulf of Mexico Gray Snapper, and Gulf of Mexico Vermilion Snapper.

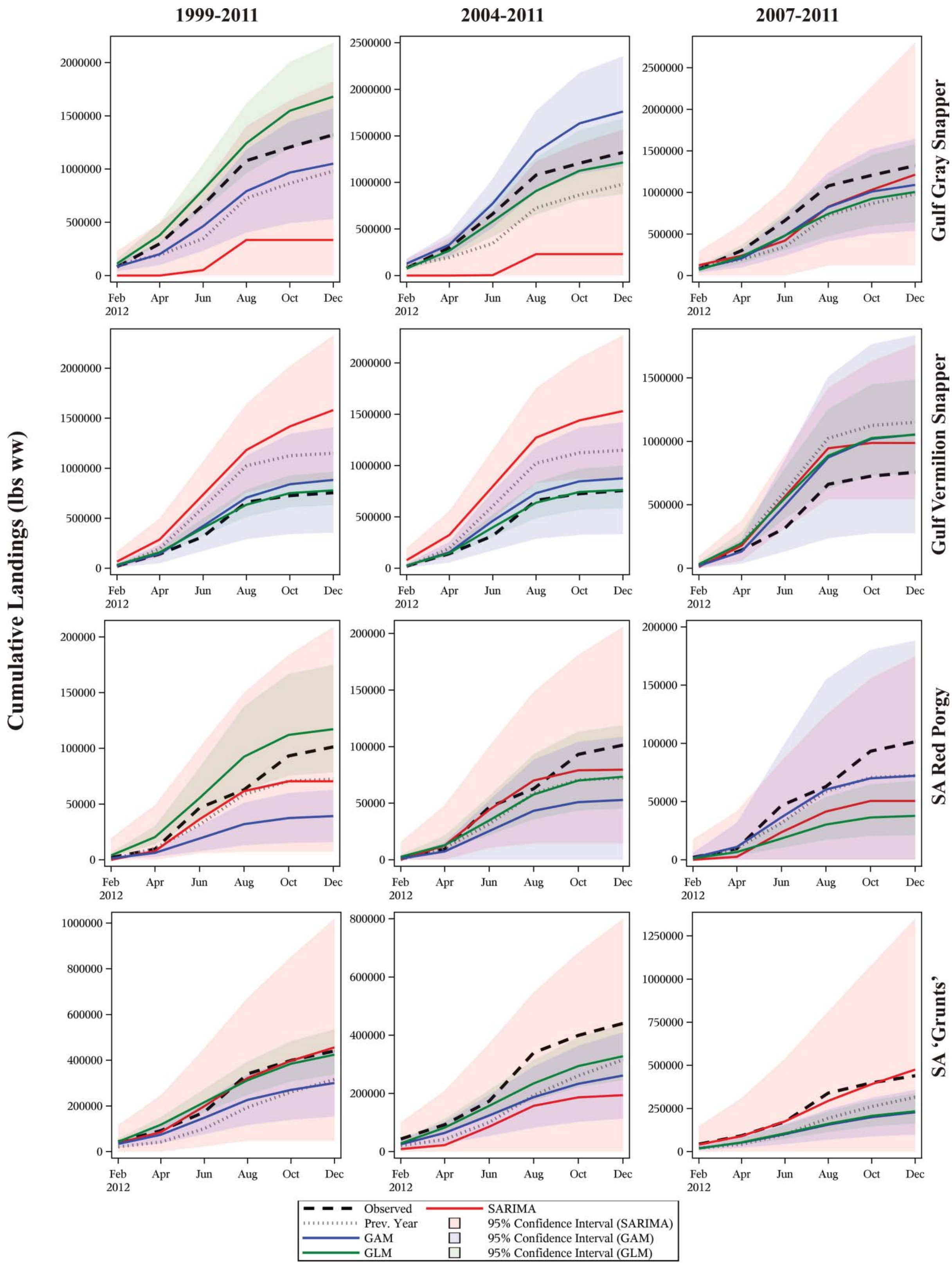

FIGURE 8. Cumulative landings plots showing relative model performance between SARIMA model (red), GAM (blue), and GLM (green) forecasts with 95% confidence limits (shaded areas) relative to observed cumulative landings (lb wet weight) for 2012, based on model fits to 1999–2011, 2004–2011, and 2007–2011 time series data for Atlantic Red Porgy, Atlantic grunts complex, Gulf of Mexico Gray Snapper, and Gulf of Mexico Vermilion Snapper.

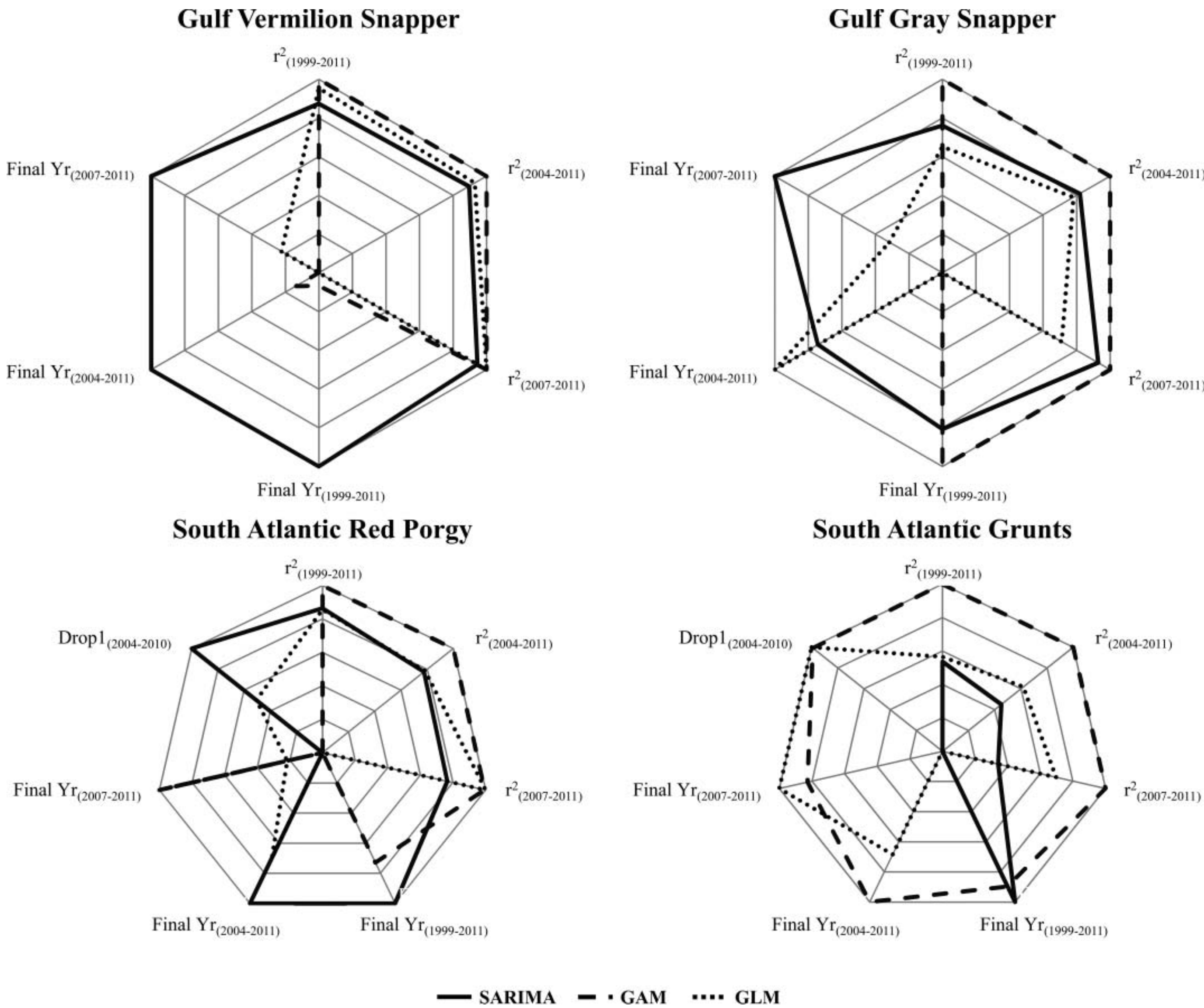

FIGURE 9. Radar plots showing relative model performance between the SARIMA model (solid line), GAM (dashed line), and GLM (dotted line) forecast models with regards to model fitting (R^2) to different time series lengths, mean error in model in the final year for model fits, and mean accuracy of model forecasts under "drop-one" fit scenarios for four recreationally exploited stocks.

possible. In this study, substitution of landings values for the most recent year of fishing improved forecast accuracy over replacement with zero values in most cases. For model projections to 2012, the SARIMA model forecast negative landings rates in 2012 for Atlantic Red Porgy in wave 1 using all three time series, and in wave 6 using the 1999–2011 and 2007–2011 time series. Replacing these forecasts with the previous year's landings resulted in minor improvements in cumulative total forecast accuracy (projected cumulative landings relative to observed cumulative landings) compared with replacement with zeroes (1999–2011: +6%, 2004–2011: +1%, 2007–2011: +4% more accurate). Replacement of the wave 6 landings for Gulf Vermilion Snapper in the 2007–2011 forecast reduced forecast accuracy by 11% compared with substituting zero values. The SARIMA model forecast negative landings rates in 2012 for Gulf Gray Snapper in waves 1, 2, 5, and 6 using the 1999–2011 and 2004–2011 time series. Replacing these forecasts with the previous year's landings resulted in major improvements in cumulative total forecast accuracy compared with replacement with zeroes (1999–2011: +26%, 2004–2011: +42% more accurate). In summary, post hoc replacement of negative SARIMA model values with landings from the most recent year of fishing is recommended.

A strength of GAMs is the ability to fit noisy nonlinear data; however, this flexibility can also permit overfitting of the model to these data if careful model selection and validation routines are not employed (Wood 2006). The GAMs provided the best fit to the observed data in nearly all cases owing to the additional flexibility of this model to accommodate noisy data. However, their tendency to overfit, despite model selection and validation, resulted in reduced forecasting performance in comparison with SARIMA models and GLMs. While overfitting can be addressed in GAMs (Zuur et al. 2010) by controlling the "wiggliness" of the smoothing function, this can be quite arbitrary with small data sets. Alternatively, cross validation could be used, though the appropriateness of this approach in the present study is doubtful given the size of the input data sets.

As with any model, the reliability of our forecasts was dependent upon both the accuracy and the consistency of the historical data. Recreational fisheries data in the southeastern United States is based upon surveys (i.e., SE Headboat Survey, MRFSS, and the TPWD Creel Survey). Each of these surveys contains uncertainty, and spikes in landings estimates may occur when high landings rates from a limited subsample are expanded. Survey data based upon dockside intercepts extrapolated to a fishing population comprising millions of people is subject to variability, which may reflect sampling issues rather than actual landings trends. Changes in survey methodologies or management regulations may reduce the predictive utility of historical data. Future forecasting modeling should attempt to incorporate uncertainty in wave-specific recreational landings estimates to avoid model overweighting of outliers that may be an artifact of survey design. Additionally, the utility of all of the methods explored in this study is contingent upon the ability of historical trends to represent future landings. Angler behavior is notoriously difficult to predict (Johnston et al. 2010; Hunt et al. 2011), and changes in management regulations (i.e., closed seasons, bag limits, size limits) within or following the historical time series make forecasting future recreational landings even more challenging. Future forecasting modeling could explore the use of management regulation time series as covariates, and also evaluate the utility of economic predictors of recreational fishing effort such as per-capita U.S. Gross Domestic Product or mean fuel prices. Finally, changes in stock size due to rebuilding may also pose a problem, as increasing landings rates may result in higher-than-expected landings. When a stock assessment is available, catchability may be combined with historical and projected abundance at age to produce a time series of exploitable abundance. Exploitable abundance may be a useful predictive covariate for landings forecasting models (N. A. Farmer, unpublished data).

CONCLUSIONS

Recreational landings comprise a substantial proportion of the total landings for many fish species in the southeastern United States, and this pattern is becoming more common worldwide (Coleman et al. 2004; Cooke and Cowx 2004). Coupled with more stringent fishery regulations, the need to predict recreational fish landings will only increase. Although GAM's flexibility consistently provided the best fits to the input data, the SARIMA model most often provided the best fit to the final year in the time series, the most reliable forecast, and the best track to the in-season cumulative landings curve. Given that management agency resources are currently inadequate to develop stock assessments for all managed species (Martell and Froese 2013), developing suites of semiautomated approaches to understanding historical landings and future patterns is essential. Our analysis suggests a simple regression-based modeling approach that avoids overfitting the input data can provide useful forecasts of the seasonal dynamics and magnitude of future recreational landings.

ACKNOWLEDGMENTS

The authors thank A. J. Strelcheck and M. F. Larkin for facilitating the development of the various regression methods explored in this manuscript. The authors also thank A. J. Strelcheck and M. Kilgour for their review of this manuscript.

REFERENCES

Adamack, A. T., C. A. Stow, D. M. Mason, L. P. Rozas, and T. J. Minello. 2011. Predicting the effects of freshwater diversions on juvenile brown shrimp growth and production: a Bayesian-based approach. Marine Ecology Progress Series 444:155–173.

Akaike, H. 1974. A new look at the statistical model identification. IEEE (Institute of Electrical and Electronics Engineers) Transactions on Automatic Control 19:716–723.

Berkson, J., and J. T. Thorson. 2015. The determination of data-poor catch limits in the United States: is there a better way? ICES Journal of Marine Science 72:237–242.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 2013. Time series analysis: forecasting and control. Wiley, New York.

Buckland, S., K. Newman, L. Thomas, and N. Koesters. 2004. State-space models for the dynamics of wild animal populations. Ecological Modelling 171:157–175.

Carruthers, T. R., A. E. Punt, C. J. Walters, A. MacCall, M. K. McAllister, E. J. Dick, and J. Cope. 2014. Evaluating methods for setting catch limits in data-limited fisheries. Fisheries Research 153:48–68.

Clark, J. S. 2004. Why environmental scientists are becoming Bayesians. Ecology Letters 8:2–14.

Coleman, F. C., W. F. Figueira, J. S. Ueland, and L. B. Crowder. 2004. The impact of United States recreational fisheries on marine fish populations. Science 305:1958–1960.

Cooke, S. J., and I. G. Cowx. 2004. The role of recreational fishing in global fish crises. BioScience 54:857–859.

De Boor, C. C. 1978. A practical guide to splines. Springer-Verlag New York.

Dennis, B., P. L. Munholland, and J. M. Scott. 1991. Estimation of growth and extinction parameters for endangered species. Ecological Monographs 61:115–143.

Froeschke, J. T., B. F. Froeschke, and C. M. Stinson. 2012. Long-term trends of Bull Shark (*Carcharhinus leucas*) in estuarine waters of Texas, USA. Canadian Journal of Fisheries and Aquatic Sciences 70:13–21.

Gasper, J. R., G. H. Kruse, and R. Hilborn. 2013. Modeling of the spatial distribution of Pacific Spiny Dogfish (*Squalus suckleyi*) in the Gulf of Alaska using generalized additive and generalized linear models. Canadian Journal of Fisheries and Aquatic Sciences 70:1372–1385.

Hanson, P. J., D. S. Vaughan, and S. Narayan. 2006. Forecasting annual harvests of Atlantic and Gulf menhaden. North American Journal of Fisheries Management 26:753–764.

Hardin, J. J. W., and J. Hilbe. 2007. Generalized linear models and extensions. Stata Press, College Station, Texas.

Hilborn, R., and C. J. Walters. 1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Reviews in Fish Biology and Fisheries 2:177–178.

Holmes, E. E. 2001. Estimating risks in declining populations with poor data. Proceedings of the National Academy of Sciences of the USA 98:5072–5077.

Hunt, L. M., R. Arlinghaus, N. Lester, and R. Kushneriuk. 2011. The effects of regional angling effort, angler behavior, and harvesting efficiency on landscape patterns of overfishing. Ecological Applications 21:2555–2575.

Hyndman, R. J., and G. Athanasopoulos. 2014. Forecasting: principles and practice. OTexts [only textbook]. Available: https://www.otexts.org/fpp. (May 2015).

Ives, A. R., K. C. Abbott, and N. L. Ziebarth. 2010. Analysis of ecological time series with arma (*p, q*) models. Ecology 91:858–871.

Johnston, F. D., R. Arlinghaus, and U. Dieckmann. 2010. Diversity and complexity of angler behaviour drive socially optimal input and output regulations in a bioeconomic recreational-fisheries model. Canadian Journal of Fisheries and Aquatic Sciences 67:1507–1531.

Magnuson–Stevens Fishery Conservation and Management Reauthorization Act of 2006. Public Law 2006(109-479), 109th Congress (12 January 2007). [slip law].

Makridakis, S., S. C. Wheelwright, and R. J. Hyndman. 2008. Forecasting methods and applications. Wiley, New York.

Martell, S., and R. Froese. 2013. A simple method for estimating MSY from catch and resilience. Fish and Fisheries 14:504–514.

Nelder, J. A., and R. W. Wedderburn. 1972. Generalized linear models. Journal of the Royal Statistical Society Series A 135:370–384.

Newman, K., S. Buckland, S. Lindley, L. Thomas, and C. Fernández. 2006. Hidden process models for animal population dynamics. Ecological Applications 16:74–86.

NMFS (National Marine Fisheries Service). 2014. Annual catch limits (ACL) monitoring. NMFS, Southeast Regional Office, St. Petersburg, Florida.

R Development Core Team. 2013. R: a language and environment for statistical computing. R Foundatin for Statistical Computing, Vienna. Available: http://www.R-project.org. (June 2015).

Schaaf, W. E., J. E. Sykes, and R. B. Chapoton. 1975. Forecasts of Atlantic and Gulf menhaden catches based on the historical relation of catch and fishing effort. U.S. National Marine Fisheries Service Marine Fisheries Review 37(10):5–9.

Stergiou, K., and E. Christou. 1996. Modelling and forecasting annual fisheries catches: comparison of regression, univariate and multivariate time series methods. Fisheries Research 25:105–138.

Tsitsika, E. V., C. D. Maravelias, and J. Haralabous. 2007. Modeling and forecasting pelagic fish production using univariate and multivariate ARIMA models. Fisheries Science 73:979–988.

Ward, E. J., E. E. Holmes, J. T. Thorson, and B. Collen. 2014. Complexity is costly: a meta-analysis of parametric and non-parametric methods for short-term population forecasting. Oikos 123:652–661.

Wood, S. 2006. Generalized additive models: an introduction with R. CRC Press, Boca Raton, Florida.

Wood, S. N. 2003. Thin plate regression splines. Journal of the Royal Statistical Society Series B 65:95–114.

Zuur, A. F., E. N. Ieno, and C. S. Elphick. 2010. A protocol for data exploration to avoid common statistical problems. Methods in Ecology and Evolution 1:3–14.

Appendix: Combinations of Single-Difference SARIMA Models

TABLE A.1. Seasonal (s) autoregressive integrated moving average (SARIMA) $(p,d,q)\times(P,D,Q)s$ model combinations evaluated, where the autoregressive component (p) represents the lingering effects of previous observations, the integrated component (d) represents temporal trends, the moving average component (q) represents lingering effects of previous random shocks (or error), and s denotes the seasonal time step. As recreational landings are primarily collected in 2-month waves, s was set to 6. A "1" denotes an active component in the model.

ARIMA$(p,d,q) \times (P,D,Q)s$ model
ARIMA(0,1,1) × (0,1,1)s
ARIMA(1,0,0) × (0,1,1)s
ARIMA(0,0,1) × (0,1,1)s
ARIMA(0,1,1) × (1,1,0)s
ARIMA(1,0,0) × (1,1,0)s
ARIMA(0,0,1) × (1,1,0)s
ARIMA(1,1,0) × (0,1,1)s
ARIMA(1,1,0) × (1,1,0)s