SCRS/2000/053 Col. Vol. Sci. Pap. ICCAT, 53: 164-179. (2001)

FITTING A SURPLUS-PRODUCTION MODEL WITH NUMBERS- VS. WEIGHT-BASED INDICES OF ABUNDANCE TOGETHER WITH REMOVALS DATA IN WEIGHT: AN EVALUATION ON SIMULATED FISHERIES SIMILAR TO BLUE MARLIN IN THE ATLANTIC OCEAN

Michael H. Prager[1] and C. Phillip Goodyear[2]

SUMMARY

Fishery management based on maximum sustainable yield (MSY), e.g., management at ICCAT, usually considers "yield" as biomass rather than numbers of fish, and often uses surplus-production models, either alone or in conjunction with other models, to estimate MSY and stock status. Recent ICCAT assessments of billfishes have used a nonequilibrium, age-aggregated, logistic surplus-production model, fit using an observation-error estimator (as implemented in the ASPIC software). An assumption of such models is that indices of population abundance used in fitting are measured in the same metric, either biomass or numbers, as the catch data. Because of available data, however, ICCAT billfish assessments have used indices of abundance (CPUE) in numbers but removals in biomass. This discrepancy is expected to bias estimates of MSY and related benchmarks.

The impact of the discrepancy on estimates of management benchmarks was evaluated using simulated fishery data. The simulation model for generating data was constructed around the life history characteristics of Atlantic blue marlin (Makaira nigricans)*, and explicitly included sex, size, and age structure on a monthly basis. Growth was sexually dimorphic, with females attaining larger asymptotic mean sizes, and size varied about mean size at age. Simulations and analyses were performed over a range of growth patterns derived from the literature. Annual recruitment was determined from a stochastic Beverton-Holt stock-recruitment function. Natural mortality was age-specific. The logistic surplus-production model was fit to the simulated data sets with ASPIC.*

Estimates of management benchmarks differed only slightly when numbers- rather than biomass-based indices of abundance were used in fitting. Estimates from CPUE in biomass were generally more precise, but not necessarily closer to true values. Estimates of the ratio of the final-year biomass to biomass at MSY tended to be slightly more accurate with numbers-based than with biomass-based abundance indices. This result presumably stems from offsetting biases. Because results varied by the growth pattern, a better understanding of growth in blue marlin (and similar species) would probably allow better approximation of likely biases arising from the use of numbers-based abundance indices.

RÉSUMÉ

La gestion de la pêche fondée sur la production maximale soutenable (PME), à savoir la gestion pratiquée à l'ICCAT, considère en général la "production" comme étant la biomasse, plutôt que le nombre de poissons, et utilise souvent des modèles de production excédentaire, soit seuls, soit avec d'autres modèles, pour estimer la PME et l'état du stock. Les dernières évaluations ICCAT des istiophoridés ont utilisé un modèle logistique de production excédentaire, non-équilibré et agrégé par âge, ajusté au moyen d'un estimateur observation-erreur (tel qu'il est activé dans le logiciel ASPIC). Un postulat de ces modèles est que les indices de l'abondance de la population qui servent pour l'ajustement sont mesurés selon la même jauge, la biomasse ou le nombre de poissons, que les données de capture. L'état de disponibilité des données, toutefois, a fait que les

---

[1] National Marine Fisheries Service, Southeast Fisheries Science Center, 101 Pivers Island Road, Beaufort, NC 28516. Email: Mike.Prager@noaa.gov

[2] Key Biscayne, Florida, USA.

évaluations ICCAT des istiophoridés utilisent des indices d'abondance (CPUE) numériques, mais une ponction exprimée en biomasse. Cette divergence ne peut que biaiser les estimations de la PME et des étalons y relatifs.

L'impact de cette divergence sur les estimations des étalons de gestion a été évalué au moyen de données simulées sur la pêche. Le modèle de simulation qui a fourni ces données était élaboré autour des caractéristiques du cycle vital du makaire bleu (Makaira nigricans), *et comprenait de façon explicite le sexe, la taille et la structure démographique sur une base mensuelle. La croissance montrait un dimorphisme sexuel, les femelles atteignant une taille asymptotique moyenne plus élevée; la taille variait autour de la taille moyenne à l'âge. Les simulations et les analyses ont porté sur tout un éventail de modes de croissances extraits des travaux publiés. Le recrutement annuel a été déterminé au moyen d'une fonction stochastique stock-recrutement de Beverton-Holt. La mortalité naturelle était spécifique de l'âge. Le modèle logistique de production excédentaire a été ajusté par ASPIC aux jeux de données simulées.*

Les estimations des étalons de gestion ne différaient que légèrement lorsque l'on utilisait le nombre de poissons pour l'ajustement, plutôt que les indices d'abondance basés sur la biomasse. Les estimations de la CPUE en termes de biomasse étaient en général plus précis, mais pas forcément plus proches des valeurs réelles. Les estimations du ratio de la biomasse de la dernière année et de la biomasse au niveau de la PME tendaient à être légèrement plus précises avec une base numérique qu'avec les indices d'abondance basés sur la biomasse. Ce résultat découle probablement d'une compensation des biais. Du fait que le mode de croissance rendait les résultats variables, le fait de mieux appréhender la croissance du makaire bleu (et d'espèces similaires) permettrait sans doute de mieux aborder les biais éventuels découlant de l'utilisation d'indices d'abondance basés sur le nombre de poissons.

RESUMEN

La ordenación de pesquerías basada en el Rendimiento Máximo Sostenible (RMS), por ejemplo la ordenación de ICCAT, considera generalmente el "rendimiento"como biomasa en lugar de como número de peces, y a menudo utiliza modelos de producción excedente, en solitario o junto con otros modelos, para estimar el RMS y el estado del stock. Recientes evaluaciones de marlines llevadas a cabo por ICCAT han utilizado modelos de producción excedente de no equilibrio, agregados por edad y logísticos, ajustados utilizando un estimador de observación de error (como el implementado en el programa ASPIC). Un supuesto de tales modelos es que los índices de abundancia de población utilizados en el ajuste se miden de la misma forma, ya sea biomasa o números, que los datos de captura. Sin embargo y debido a los datos disponibles, las evaluaciones de marlines de ICCAT han utilizado índices de abundancia (CPUE) en números, pero supresiones en biomasa. Se espera que esta discrepancia cause sesgos en las estimaciones de RMS y los elementos de comprobación asociados.

El impacto de la discrepancia en las estimaciones de elementos de comprobación de ordenación fue evaluado utilizando datos simulados. El modelo de simulación para generar los datos fue construido en torno a las características del ciclo vital de la aguja azul del Atlántico (Makaira nigricans), *e incluía explícitamente sexo, talla y estructura de edad de forma mensual. El crecimiento era sexualmente dimórfico, con hembras que alcanzan mayores tallas asintóticas medias; y la talla variaba en torno a la talla media por edad. Las simulaciones y los análisis se llevaron a cabo sobre una diversos esquemas de crecimiento derivados de la literatura científica. El reclutamiento anual se determinó a partir de una función estocástica de stock-reclutamiento Beverton-Holt. La mortalidad natural era específica de la edad. El modelo logístico de producción excedente fue ajustado a las series de datos simulados con ASPIC.*

Las estimaciones de los elementos de comprobación de ordenación se diferenciaban sólo ligeramente cuando se utilizaban en el ajuste índices de abundancia basados en números en lugar de basados en biomasa. Las estimaciones a partir de CPUE en biomasa eran por lo general más precisas, pero no necesariamente más cercanas a valores verdaderos. Las estimaciones de la relación de la biomasa del último año con la biomasa en RMS tendían a ser ligeramente más exactas con índices de abundancia basados en números que con índices de abundancia basados *en biomasa. Este resultado se deriva probablemente de sesgos compensatorios. Dado que los resultados variaban según el esquema de crecimiento, es probable que una mejor comprensión del crecimiento de la aguja azul (y especies similares) permita una mejor aproximación de los sesgos probables que surgen de la utilización de índices de abundancia basados en números.*

KEYWORDS

Stock assessment, Simulation, Accuracy, Variance analysis, Fish catch statistics, Tuna fisheries

INTRODUCTION

A common objective of marine fishery management is to achieve stock sizes and fishing mortality rates that produce maximum sustainable yield in biomass (MSY). This objective is shared by ICCAT. Surplus-production models are often used to estimate MSY, and also to estimate stock and fishery status relative to benchmarks related to MSY. In recent ICCAT billfish assessments, such a model has been fit with computer program ASPIC (Prager 1995), which implements an non-equilibrium, continuous-time, observation-error estimator for the logistic (Schaefer 1954, Schaefer 1957, Pella 1967, Schnute 1977, Prager 1994) production model.

The theory of production modeling allows expressing stock size and yield in terms of either biomass or numbers; however, it assumes that a consistent metric is used. In modeling fish populations, the choice has usually been to use biomass, and hence to estimate MSY as a measure of biomass (more correctly, biomass per unit time). The most recent ICCAT stock assessments of blue and white marlins violated this assumption in using catches in weight, but an abundance index based on catch per unit effort in numbers (ICCAT 1998).When modeling age-structured populations, this issue is of theoretical importance because MSY is a function of yield per recruit (YPR) as well as changes in recruitment per spawner as the population size changes due to fishing (Goodyear 1996). An important way that fishing changes YPR is by reducing survival to older (and larger) age classes and thus to reduce the mean weight of an individual at the time it is caught. An abundance index based solely on numbers of individuals cannot capture this effect, and it consequently understates changes in population biomass from increased fishing. It seems unlikely, then, that an abundance index based on numbers per unit effort should be a good surrogate for the biomass-based population abundance index assumed when fitting a production model in biomass metric. However, the nature of the error introduced into the past marlin stock assessments because of this substitution has not been studied, and the problem is complex enough so that it is not easily soluble through theoretical considerations.

Given the nature of data available on historical catches of billfishes in the Atlantic, it seems likely that representative indices of abundance in biomass will remain unavailable for most, if not all, major fisheries. Because age-structure data are not available, the surplus-production model is the most viable candidate for stock assessments of these stocks, and numbers-based abundance indices will continue to be required as surrogates for biomass-based indices. Reliability of such assessment results will be uncertain unless the effect of this substitution can be adequately characterized. In this study, we examine this issue by comparing the accuracy and precision of estimates of management benchmarks resulting from fits of the surplus production model to simulated data for both types of indices. We also address a second point: the potential bias and variance effects introduced into estimates by using a penalty term in estimation of a stock-production model. Such terms can be used to avoid solutions in which the ratio of the first year’s biomass to the carrying capacity parameter K is estimated as greater than unity.

METHODS

The general approach of this study was to simulate catch and abundance data with a population model structured about biological characteristics of Atlantic blue marlin. The surplus-production model was then fit to the simulated data with ASPIC, and the resulting estimates of management benchmarks were compared to known true values from the simulations. The age-structured population model employed in this analysis is described in Goodyear (1989); its application in the context of MSY estimation is described in (Goodyear 1996). The simulation model is implemented monthly with separate sexes and includes consideration of the variability of size at age. The number of explicit ages considered was 100, to minimize any possible effects of artificially truncating age structure. Unless otherwise stated here, we adopted the conventions employed in Goodyear and Prager (2000).

Biological Characteristics

Growth – It is clear from the varied estimates of growth rates in the literature that mean sizes at age and asymptotic maximum size (L_∞) of blue marlin are poorly known. Goodyear and Prager (2000) adopted separate models for male and female size at age based upon the available data, but did not attempt an evaluation of uncertainty in growth because it was not an important issue in the main subject of their study. However, because different growth rates will affect proportional change in mean weights between different mortality schedules, growth is a major concern for the present analysis. Several authors have fit von Bertalanffy growth equations to male and female blue marlin (Wilson 1984, Cyr 1987, Boggs 1989, Wilson *et al* 1991). In each case, males were smaller at age than females. In the above studies, the estimated von Bertalanffy growth coefficient K tended to be larger for females than for males, but estimated values of K for the two sexes overlapped among studies. For the purposes of the present investigation, we established five growth treatments (labeled K1–K5), with different values of K for each. Those values of K were derived from probability points on the mean of log-transformed estimates of K, pooled separately for males and females, from several sources (Wilson 1984, Cyr 1987, Boggs 1989, Wilson *et al* 1991). The values selected for K were: K1=0.0985 (p=0.05); K2=0.1426 (p=0.25); K3=0.1845 (p=0.50); K4=0.2385 (p=0.75); and K5=0.3453 (p=0.95). Asymptotic maximum sizes (L_∞) were 207.9 cm LJFL for males and 282.6 cm LJFL for females (as in Goodyear and Prager 2000); von Bertalanffy parameter t_0 was set to zero. Resulting growth curves by sex are shown in Figures 1–5.

Other major aspects of growth included in the simulations include the assumption that within-year growth after age 0 did not vary seasonally; the coefficient of variation of size at age was assumed to be 0.12 for both sexes. Length–weight relations were taken from Prager *et al*. (1995).

Fecundity – The limited data relating to reproduction in blue marlin in the Atlantic are discussed in Goodyear and Prager (2000). Data available in the literature are too sparse to characterize the relation between blue marlin length and fecundity, so we adopted the relation used by Goodyear and Prager (2000), in which fecundity is a constant multiple of body weight for blue marlin above 45 kg. It is assumed that the abundance of males is not limiting.

Natural Mortality **–** Natural mortality M is difficult to estimate, and especially to separate from total mortality Z. Goodyear and Prager (2000) discuss the available data relating to natural mortality in blue marlin and conclude that it is not well characterized. Here, we chose the value of M=0.10/yr for older fish, a value selected partly because of the relationship between natural mortality and the mean weight of individuals in the stock. Compared to high natural mortality, low natural mortality favors larger numbers of age classes in the unfished population, and therefore tends to favor heavier unfished mean weights. As a consequence, fishing mortality tends to have a greater proportional effect on mean weight at lower values of M. Because the current study was designed to explore relative effects of abundance indices based on weight versus numbers, we chose a value of M lower than the median explored in the analyses of Goodyear and Prager (2000) to accentuate any difference found. In accordance with the belief that mortality declines with size, we assumed that M declines from an initial value of 0.5/yr for the youngest recruits to 0.10/yr for the oldest fish in the population (Figure 6).

Recruitment **–** Annual recruitment in the simulated populations was determined from population fecundity at the beginning of each year, using a Beverton-Holt stock-recruitment function. The stock-recruitment relation controls mean recruitment for any given adult stock size. Because absolute recruitment levels are not a major concern in this analysis, mean recruitment at MSY was arbitrarily set to 1,000,000 fish at 3 months of age.

The stock-recruitment curve also controls the population response to fishing mortality. In the context of the current analysis, the main effect of the stock-recruitment curve is to increase recruitment per spawner as fishing mortality increases from zero. For any particular fishing mortality rate, the equilibrium mean weight of a fish in an age-structured population is a joint function of growth and mortality, but is not influenced by recruitment. Hence, at least under equilibrium conditions, the stock-recruitment curve determines only average recruitment and does not affect the relationship between CPUE in numbers and CPUE in weight. Consequently, we limit the current analysis to a single Beverton-Holt stock-recruitment curve with an unfished slope at the origin of 10.

Recruitment also exhibits stochastic annual variability due to environmental factors not related to the adult stock size. We acknowledged this characteristic by adding a random component to mimic the egg-recruitment survival variability seen in nature. This was accomplished by multiplying the predicted (mean) recruitment from the stock-recruitment relationship by $exp(R*\mathrm{CV} - 0.5*\mathrm{CV}^{2})$, where R is a random normal deviate with mean of zero and a variance of 1.0; and CV is the coefficient of variation of the log of the random multiplier, assumed to be 0.15 for this analysis. An example of the magnitude of this variation taken from one of the simulations for growth treatment K1 is illustrated in Figure 7. Separate analyses were performed with and without this source of recruitment variability.

Simulations

Initial conditions **–** The initial year of each simulation was designated 1955, and began with the population at its unfished stable age distribution, determined by M and the stock–recruitment curve. The population was then simulated through 1998, for a total of 44 years.

Fishing mortality **–** Total annual fishing mortality rates (F) for each growth treatment were adjusted so that the trend in relative biomass approximated the trend estimated in the last ICCAT assessment of blue marlin (ICCAT 1998). Fishing selectivities were held constant across all simulations. Fish first recruited to the fishery at age 1, and all ages older than age 0 were equally susceptible to fishing. Total annual fishing mortality was attributed to a single fishery; simulated catch data were saved for analysis with ASPIC.

Abundance indices – Two simulated abundance indices were constructed for each simulated data set, one in units of numbers of individuals and the second in units of biomass. The first was derived by multiplying an arbitrary catchability coefficient (q=10^{-6}) by the time-averaged total number of individuals alive at age 1 and older each year. The second index was derived by multiplying q by the time-averaged average total biomass of ages 1 and older. (Time averaging was within each year.) Resulting values were saved for use in fitting the production model. As an example, a plot of these two indices for a simulation with growth treatment K1 is presented in Figure 8. These indices are referred to as indices of NPUE (numbers per unit effort) and WPUE (weight per unit effort), respectively. When abundance without regard to metric is considered, we use the terminology CPUE.

Transformation of Simulated Data for Model Fitting

Measurement Error –Simulated measurement error was added to catch and CPUE data at the time input files for production modeling were created. Each input file contained catch in weight and either NPUE or WPUE for each year. Measurement error was normally distributed (catch) or lognormally distributed (CPUE). Four treatments of error were added, corresponding to CV’s of 0.01, 0.05, 0.125, and 0.25, but in all cases random errors were constrained to within ±2? . In generating each file, the same level of error was applied to catch and CPUE.

Penalty Term – In fitting a logistic production model to the more flexible realized dynamics of an age-structured simulated population, one must consider that biomass at MSY of a logistic population is always half the maximum population size (B_{MSY}=K/2). This constraint is not shared by either real populations or the age-structured model used here to generate data. This discrepancy could potentially influence the fitting process, because the simulations began with a nearly unfished stock, and the beginning biomass (B_1) was always larger than, and in many cases much larger than, twice B_{MSY}. Such a condition would not be expected in a stock adhering to the dynamics of the logistic production model.

When a data series that begins with strongly declining pattern of abundance is fit with a logistic stock-production model, the estimate of first-year biomass B_1 tends to be in the region $B_1 > K$; a biomass greater than K results in negative surplus production, and for that reason it often fits the observed stock decline well. However, such solutions have been questioned as being inconsistent with production-model theory or as simply unrealistic for a particular stock. It is not known, moreover, whether, such solutions provide unbiased estimates of MSY and other benchmarks. Several investigators have devised modified fitting procedures for production models to avoid or discourage solutions in which $B_1 > K$. Punt (1990) described the "Butterworth–Andrew" estimator that includes the constraint B_1=K, and he compared it to unconstrained estimation for simulated populations similar to Cape hake off southern Africa. Punt concluded that, for the stock he studied, the constrained model provided lower estimated expected discrepancy (similar to the PE measure used here) and that estimation of B_1 tended to provide overestimates of fishing effort rate corresponding to the Total Allowable Catch (TAC). It is not clear how applicable Punt's (1990) results would be to stocks like the billfishes, which have quite different life histories from Cape hake. Prager (1994) described a fitting approach in which B_1 is not fixed, but instead a penalty term is added to the objective function in fitting to discourage solutions in which $B_1 > K$. That approach is implemented in the ASPIC software (Prager 1995), as is Punt's (1990) approach.

A secondary objective of the present study was to examine effects of using the Prager (1994, 1995) penalty term when fitting a logistic surplus production model of blue marlin. Therefore, we fit each simulated data set with and without the penalty term invoked, to examine its effect on accuracy and precision of estimates of the three management benchmarks.

Alternative Trajectory - Because results might be sensitive to the particular population trajectory used, and because we mainly used an approximation of the single trajectory previously estimated for blue marlin, we were concerned that the generality of our results might be severely limited. We tested this possibility by simulating an alternative trajectory in which the population recovered from an initial overfished state.

Summary of Production-Model Fits – The five growth treatments and choice of either stochastic or deterministic first-year survival result in 10 unique combinations. For each combination, 100 separate simulations were performed for the main trajectory, resulting in 1,000 population simulations. For each simulation, four levels of measurement error and two types of CPUE (NPUE and WPUE) were considered, giving 8,000 different simulated data sets available for fitting. As each was fit both with and without the penalty term, 16,000 model fits were made for the main trajectory. For the alternative trajectory, we only generated simulated data sets with stochastic recruitment, and thus an additional 8,000 fits were made, giving a total of 24,000 production-model fits.

Management Benchmarks Evaluated – To evaluate effects of the above factors on estimation quality, we compared production-model estimates of three management benchmarks to their true values. The three benchmarks were maximum sustainable yield (MSY), final-year fishing mortality F relative to F at MSY ($F./F_{MSY}$), and final-year biomass (B) relative to B at MSY ($B./B_{MSY}$). Estimation quality was represented by percent error (PE) in each estimate, where PE=100*(estimate - true) / true, PE of zero thus indicating an unbiased estimate. Results were analyzed graphically through box-and-whisker diagrams (Tukey 1977) using S-Plus software (Mathsoft 1999).

RESULTS

Properties of Simulations

In each simulated data set, underlying population values of the three management benchmarks depended on the growth treatment. For the main trajectory and growth treatments K1 through K5, true values of MSY were approximately 3,893, 6,629, 8,765, 11,450, and 15,860 t/year, respectively. Corresponding true values of final-year relative biomass $B./B_{MSY}$ in simulations with deterministic recruitment were 0.81, 1.30, 0.68, 0.43, and 0.25, respectively. Values of this benchmark for simulations with stochastic recruitment varied, but were distributed tightly about those values (Figure 8.5). Corresponding true values of final-year relative fishing mortality rate $F./F_{MSY}$ were 0.75, 0.53, 0.85, 1.87, and 2.44.

For the alternative, true values of MSY were the same as for the main trajectory. Corresponding mean true values of $B./B_{MSY}$ ranged from 1.37 to 1.85, while corresponding true values of $F./F_{MSY}$ ranged from 0.25 to 0.57. The ranges in the above stock-status measures (from the two trajectories) represent a wide variety of final-year stock conditions, in which in the final year some stocks were overexploited and some were not; also, some were being fished at an unsustainable rate, and some were not.

Estimates of Benchmarks—General (Main Trajectory)

Measurement error did not have a meaningful effect on the median bias in estimates of any of the management benchmarks for any growth treatment; thus results from all levels of measurement error were pooled within each growth treatment for further analysis. The same graphical format is used to display distribution of PE in estimates of each benchmark (Figures 9-20); that format is explained in detail here. A group of four figures is used to display error in estimates of a single benchmark (e.g., Figures 9-12 for MSY): within the group of four figures, simulations with deterministic recruitment are placed above and simulations with stochastic recruitment are placed below. Likewise, simulations fit without the penalty term are placed on the left, and those with the penalty term, on the right.

Each figure within the group of four, then, corresponds to a single benchmark and to a single combination of either a penalty term or not, and either stochastic or deterministic recruitment. Each such figure is divided into two, with the left panel giving results of fitting to NPUE, and the right panel, of fitting to WPUE. Results for the growth treatments K1 through K5 are arranged vertically, with one box-and-whisker unit for each combination of growth treatment and CPUE type. Within each box-and-whisker unit, the filled circle represents the median PE from 192 estimates of that benchmark. The rectangle of each box-whisker unit encloses the interquartile range, the outer whiskers denote the range of most other observations and the open circles are “outside values” by the criteria of Tukey (1977, p. 44). To increase legibility, we have not always displayed the entire range of PE in estimates; in some cases, a few far-outside values have been repositioned near the smallest or largest value that can be displayed on the graph.

Estimates of MSY (Main Trajectory)

For the main trajectory, estimates of MSY based on NPUE were more variable than estimates based on WPUE, but those differences were not large (Figures 9-12). Adding stochasticity to recruitment increased variability in the estimates, but it did not change the median bias appreciably (Figures 9 and 10; 11 and 12). Most estimates of MSY from either index were positively biased: median PE ranged from about +10% to more than +50%, depending upon growth treatment. There is no clear pattern in magnitude of relative bias (as reflected in median PE) with respect to the use of NPUE versus WPUE; in some cases, NPUE gave less biased results, while in others, WPUE gave less biased results. Bias in MSY increased with increasing growth treatment level, which means the bias increased with increasing values of the von Bertalanffy growth coefficient K.

For this trajectory, estimates of MSY made with the penalty term tend to be less variable, no matter which type of index was used (compare Figures 11 and 12 with Figures 9 and 10). However, the reduction in variance was much more pronounced when WPUE was used. Use of the penalty term had little effect on median bias when NPUE was used, use of the penalty term markedly increased bias when WPUE was used. This pattern was consistent for simulations with either deterministic or stochastic recruitment. Median magnitude of bias was generally less than 50% in analyses without the penalty term, but for growth model K5, median bias approached 70% to 80% when the penalty term was used in conjunction with a WPUE index of abundance.

Estimates of $B./B_{MSY}$ (Main Trajectory)

Median estimates of final-year relative biomass $B./B_{MSY}$ were centered within about 20% of the true value and tended to exhibit negative bias (Figures 13-16). As with estimates of MSY, estimates of $B./B_{MSY}$ were more variable when NPUE was used to index abundance than when WPUE was used. The bias does not appear consistently larger in either WPUE-based estimates or those based on NPUE. Adding stochasticity to recruitment increased dispersion in the estimates, but no other effect was apparent. The greater variability in estimation using NPUE suggests a greater potential for difference between true and estimated values when NPUE is employed in any particular analysis.

Use of the penalty term decreased variability in estimates of $B./B_{MSY}$ from both indices (compare Figures 13 and 15; Figures 14 and 16). For estimates from the NPUE index there was little, if any, corresponding increase in bias; however, for estimates from the WPUE index, bias increased noticeably, if not greatly.

Estimates of $F./F_{MSY}$ (Main Trajectory)

Median estimates of final-year relative fishing mortality rate $F./F_{MSY}$ tended to be centered within about 20% of the true values (Figures 17–20). Estimation variability appeared to decrease with increasing growth rate (K1 ? ?K5). Because estimates were particularly variable with growth treatment K1 and stochastic recruitment, the decrease of variability with increasing growth rate appears stronger with stochastic recruitment. Also, there is tendency towards increasing negative bias with increasing growth rate (Figures 17-20). As with the results for $B./B_{MSY}$, NPUE seems to slightly outperform WPUE in terms of bias but at the cost of greater variance, but the differences are small.

Use of the penalty term in fitting reduced variability in estimates of $F./F_{MSY}$ slightly and increased bias slightly, but it had less effect on estimates of $F./F_{MSY}$ than it did on estimates of the other benchmarks. In all cases, average effect of the penalty term was small.

Alternative Trajectory

Estimation results with and without the penalty term were almost identical; this occurred because the penalty term only has an effect when the model estimates $B_1 > K$, but in the alternative trajectory, the starting biomass was low. Results are presented only for estimation without the penalty term.

Estimates of MSY (Figure 21) based on NPUE were slightly negatively biased for the slower-growth treatments (K1 and K2), had little bias for the central growth treatment K3, and were slightly positively biased for the faster-growth treatments (K4 and K5). Compared to results from the main trajectory (Figure 10), these results exhibited less median bias and somewhat less variability. Estimates of MSY based on WPUE exhibited consistent positive median bias (Figure 21) and about the same level of variability as those based on NPUE. This is in slight contrast to results from the main trajectory; there, MSY estimates from WPUE were slightly less variable than those from NPUE and of about the same median bias (Figure 10). However, difference between trajectories were not dramatic.

Estimates of $B./B_{MSY}$ based on NPUE and WPUE (Figure 22) were less variable than those from the main trajectory (Figure 13). In both cases, biases were small, and the estimates based on WPUE exhibited slightly more negative median bias than those based on NPUE (Figures 13, 22).

Estimates of $F./F_{MSY}$ based on NPUE (Figure 23) were more apparently variable than those based on WPUE. Bias differed between indices, but neither was consistently better than the other (Figure 23). Variability of estimation (Figure 23) appeared larger than from the main trajectory (Figure 17); however, the alternative trajectory had on average much lower underlying ("true") values of this benchmark, so that errors appear larger when expressed as PE. Patterns of median bias were slightly different from those obtained on the main trajectory, but bias was not markedly different in scale.

DISCUSSION

Our results indicate few systematic problems arising from the use of a numbers-based abundance index (NPUE) as a surrogate for the biomass-based abundance (WPUE) index when fitting a surplus-production model to removals in weight. We noticed a possible slight improvement in median relative bias when using NPUE to estimate stock status, as embodied in benchmarks $F./F_{MSY}$ and $B./B_{MSY}$. However, precision of estimates was often reduced slightly, in particular when estimating $B./B_{MSY}$. Such decreased precision would result in increased probability of a meaningful error in estimation from any particular set of fishery data.

Use of the penalty term in fitting increased precision in estimates of all three management benchmarks. However, when using WPUE, the added precision came at the cost of a slight increase in median bias in estimates of $B./B_{MSY}$, an increase not evident when using NPUE. In consequence, our results suggest that the penalty term may provide more reliable results for surplus production model fits that employ number-based abundance indices. These results confirm those of Punt (1990).

The success of estimation using mixed metrics is surprising and raises two immediate questions: (1) Why should using mixed metrics result in successful production-model estimation? (2) Why should using mixed units in some cases be superior to using consistent units? At present, answering these questions requires speculation. The answer to (1) may well be that NPUE and WPUE are often highly correlated (e.g., Figure 8), so that the two types of index provide similar information. In particular, the change in time slope of CPUE from negative to positive in Figure 8 occurs at about the same time for both indices. Answering question (2) is more difficult. We suspect that our results are due to interaction of offsetting biases. Model misspecification is present when one uses a surplus-production model on a complex and highly age-structured population, and model misspecification is a commonly studied cause of bias in parameter estimates (Kotz *et al.* 1982, p. 230). Use of the incorrect metric in the CPUE index would constitute a second specification error, and would also be expected to bias estimation. We conjecture that for our simulated populations, and perhaps for real populations similar to them, these two biases combine so as to reduce overall bias.

In conclusion, it appears that use of NPUE in place of WPUE is an acceptable method for populations similar to blue marlin in the Atlantic. This conclusion was not found to depend on any of the following factors examined: specific level of observation error in the data, presence of moderate stochasticity in recruitment, choice of growth pattern, or general trend of population time trajectory. In addition, we believe it should be independent of the stock–recruitment relationship and of the natural mortality rate. When using WPUE, it appears that use of the penalty term (described above) in estimation can provide a favorable mean–variance tradeoff for stocks similar to those examined here. Use of the penalty term does not seem beneficial overall when using NPUE for fitting.

ACKNOWLEDGEMENTS

We thank the Billfish Foundation for supporting C.P.G.'s contribution to this work and the Southeast Fisheries Science Center of the National Marine Fisheries Service, NOAA, for supporting M.H.P.'s contribution.

REFERENCES

BOGGS, C.H. 1989. Vital rate statistics for billfish stock assessment. Pages 225-233 i*n* R. H. Stroud (ed) Planning the future of billfishes: research and management in the 90's and beyond. Marine Recreational Fisheries 13, Part 1 National Coalition For Marine Conservation, Savannah Georgia, USA. CYR, E. C. 1987. Age growth and reproduction of blue marlin, Makaira nigricans, from South Carolina billfish tournament collections. MS Thesis University of South Carolina, USA.41 pp.

GOODYEAR, C. P. 1989. LSIM - a length-based fish population simulation model. NOAA Tech. Memo. NMFS-SEFC-219, iii plus 55 pages.

GOODYEAR, C. P. 1996. Variability of fishing mortality by age: Consequences for MSY. N. Am. J. Fish. Manage.16:8-13.

GOODYEAR, C.P. and M. H. Prager. 2000 (this volume). Fitting Surplus-Production Models with Missing Catch Data using ASPIC: evaluation with simulated Data on Atlantic blue marlin. ICCAT Col. Vol. Sci. Pap., 53, document SCRS/00/52.

ICCAT. 1998. Report of the third ICCAT billfish workshop. ICCAT Col. Vol. Sci. Pap., 47: 1-128.

KOTZ, S., N. L. Johnson, and C. B. Read (eds.). Encyclopedia of statistical sciences, Volume 1. John Wiley & Sons, New York, USA. 480 p.

MATHSOFT. 1999. S-Plus 2000 programmer's guide. Data Analysis Products Division, Mathsoft, Inc., Seattle, Washington, USA. 868 p.

PELLA, J. J. 1967. A study of methods to estimate the Schaefer model parameters with special reference to the yellowfin tuna fishery in the eastern tropical Pacific ocean. Ph.D. Dissertation, University of Washington, Seattle, USA. 156 p.

PRAGER, M. H. 1994. A suite of extensions to a non-equilibrium surplus production model. Fish. Bull. 92:374–389.

PRAGER, M. H. 1995. User's manual for ASPIC: a stock-production model incorporating covariates, program version 3.6x. U. S. NMFS, Southeast Fisheries Science Center, Miami Lab., Document MIA-92/93-55, 4th edition, Miami, Florida, USA.

PRAGER, M. H., C. P. Goodyear, and G. P. Scott 1996. Application of a surplus production model to a swordfish-like simulated stock with time-changing gear selectivity. Trans. Am. Fish. Soc.125:729-740.

PRINCE, E. D., D. W. Lee., J. R. Zweifel, and E. B. Brothers. 1991. Estimating age and growth of young Atlantic blue marlin Makaira nigricans from otolith microstructure. Fish. Bull. 89(3):441-459.

PUNT, A. E. 1990. Is B_1=K an appropriate assumption when applying an observation error production-model estimator to catch–effort data? S. Afr. J. Mar. Sci. 9:249–259.

SCHAEFER, M. B. 1954. Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. Bull. I-ATTC 1(2):27–56.

SCHAEFER, M. B. 1957. A study of the dynamics of the fishery for yellowfin tuna in the eastern tropical Pacific Ocean. Bull. I-ATTC 2: 247–268.

SCHNUTE, J. 1977. Improved estimates from the Schaefer production model: theoretical considerations. J. Fish. Res. Board Can. 34: 583–603.

TUKEY, J. W. 1977. Exploratory data analysis. Addison-Wesley Publishing Company, Reading, Massachusetts, USA. 688 p.

WILSON, C. A. 1984. Age and growth aspects of the life history of billfishes. Ph.D. Dissertation. University of South Carolina, USA..180 pages.

WILSON, C. A., J. M. Dean, E. D. Prince, and D. W. Lee. 1991. An examination of sexual dimorphism in Atlantic and Pacific blue marlin using body weight, sagittae weight, and age estimates. J. Exper. Mar. Biol. Ecol..151:209-225.

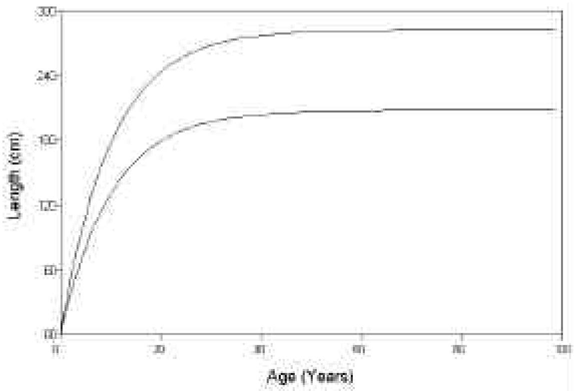

Figure 1. Male (lower curve) and female (upper curve size at age for growth treatment K1.

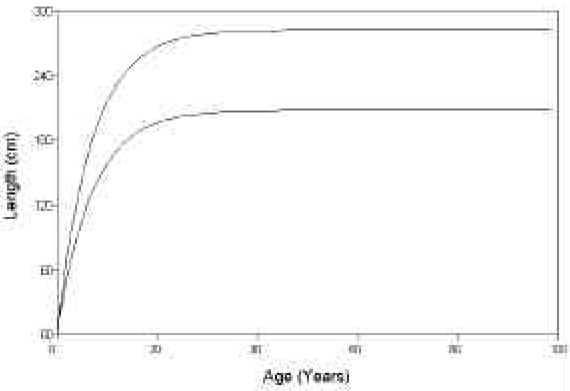

Figure 2. Male (lower curve) and female (upper curve size at age for growth treatment K2.

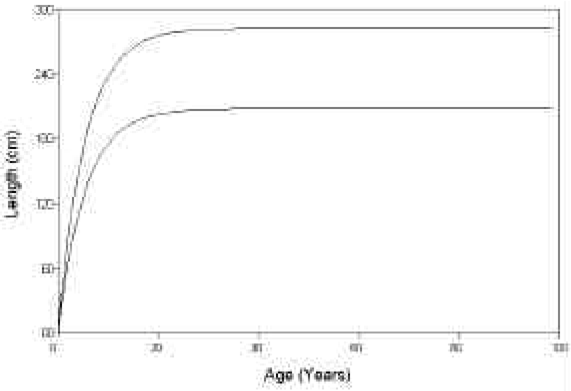

Figure 3. Male (lower curve) and female (upper curve size at age for growth treatment K3.

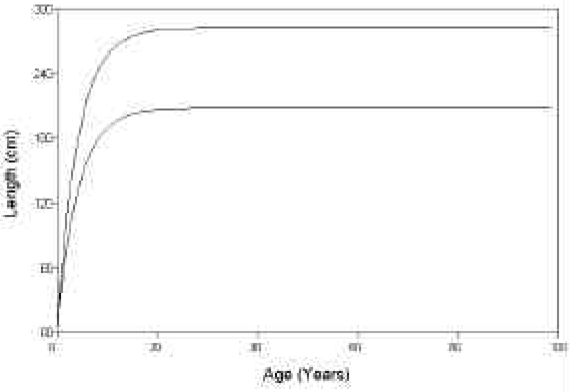

Figure 4. Male (lower curve) and female (upper curve size at age for growth treatment K4.

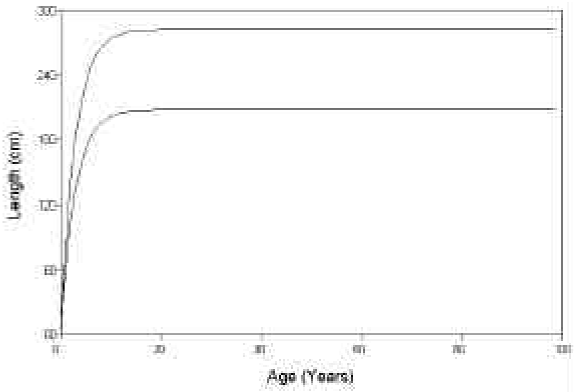

Figure 5. Male (lower curve) and female (upper curve size at age for growth treatment K5.

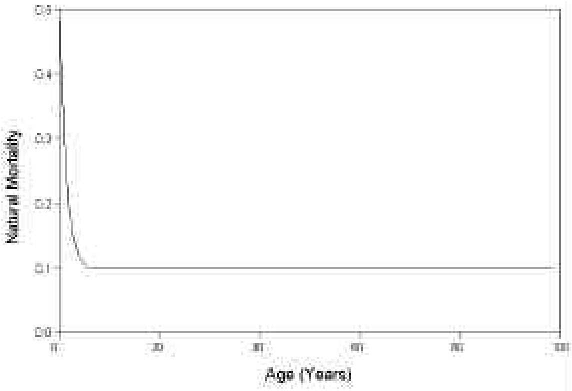

Figure 6. Natural mortality rates at age assumed for these analyses.

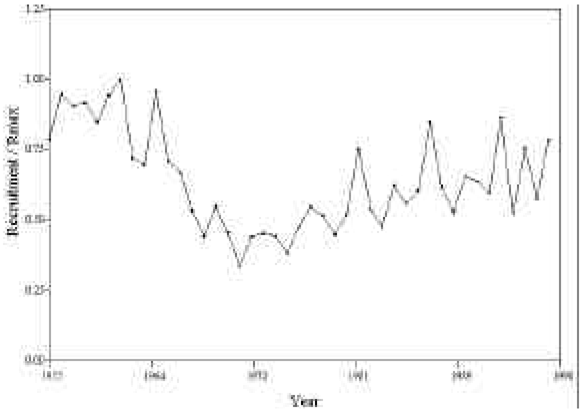

Figure 7. Time series of recruitment for one sample of simulations for growth treatment K1.

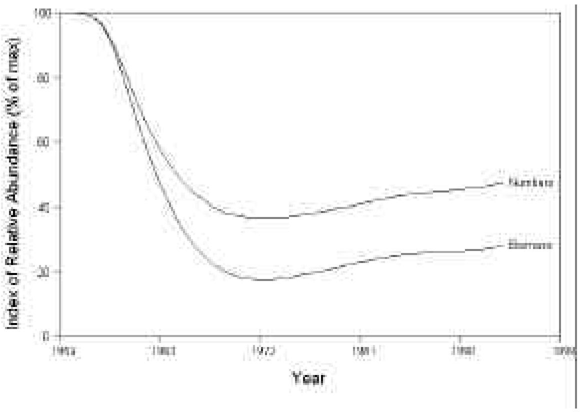

Figure 8. Comparison of relative abundance indices based on numbers vs. on biomass for the simulation depicted in Figure 7.

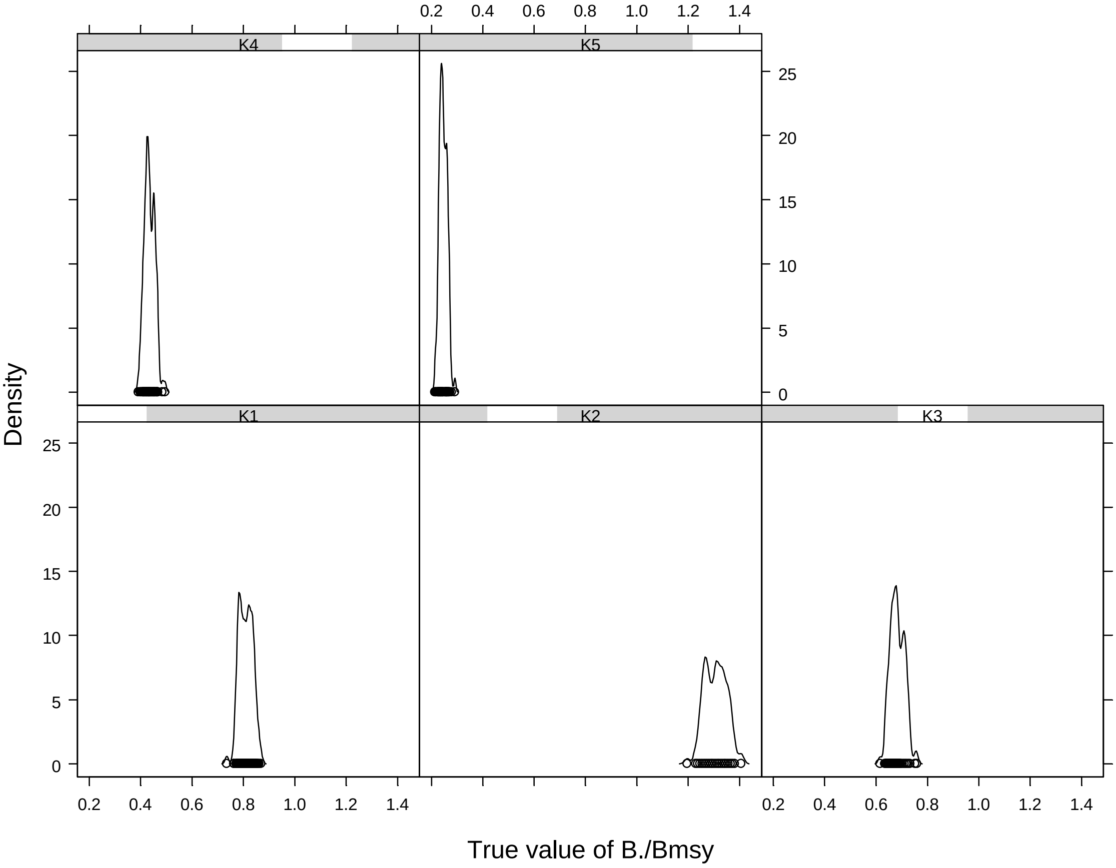

Figure 8.5. Distribution of relative final-year biomass (B./B_{MSY}) realized in simulated populations. A separate distribution is shown for each growth treatment, K1–K5.

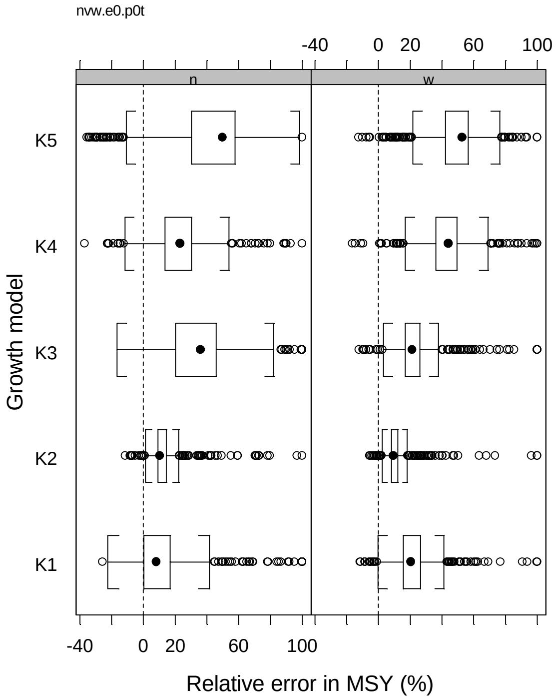

Figure 9. Distribution of percent error in MSY estimates for trials with deterministic recruitment and no penalty term.

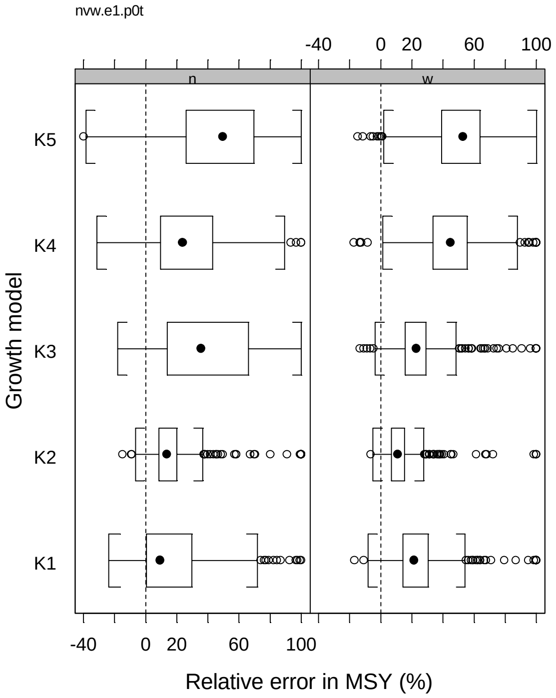

Figure 10. Distribution of percent error in MSY estimates for trials with stochastic recruitment and no penalty term.

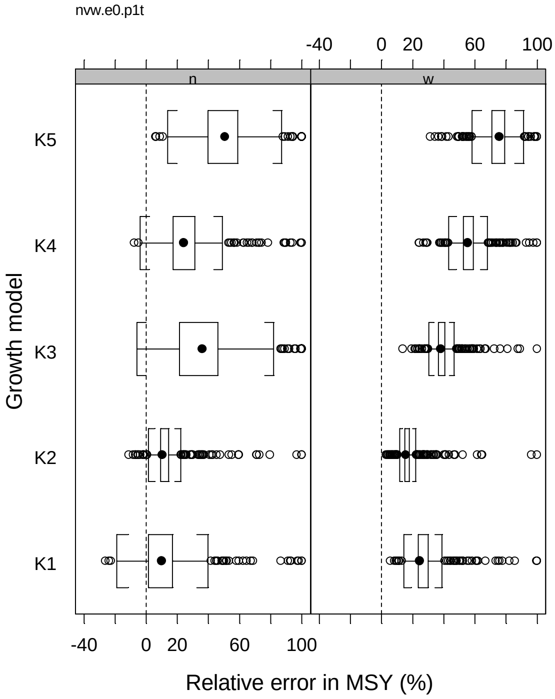

Figure 11. Distribution of percent error in MSY estimates for trials with deterministic recruitment and with the penalty term.

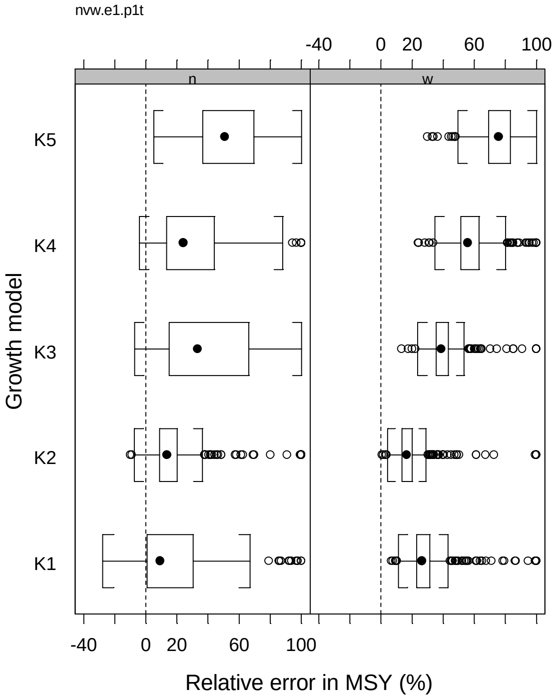

Figure 12. Distribution of percent error in MSY estimates for trials with stochastic recruitment and with the penalty term.

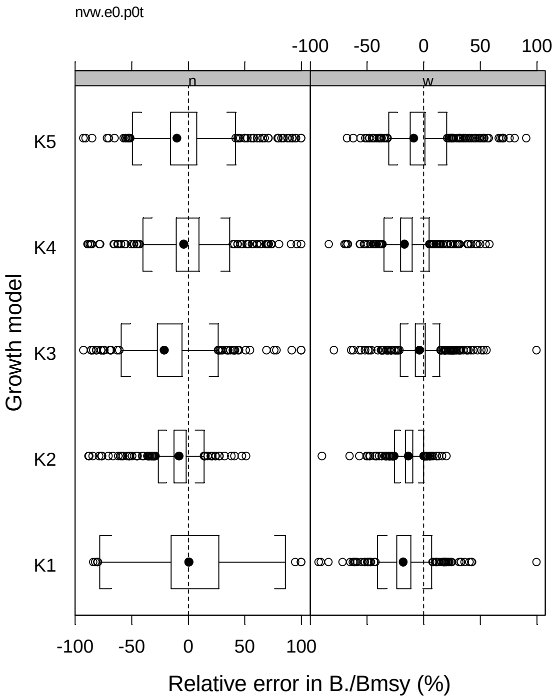

Figure 13. Distribution of percent error in B/B$_{MSY}$ estimates for trials with deterministic recruitment and no penalty term for B$_1$>K.

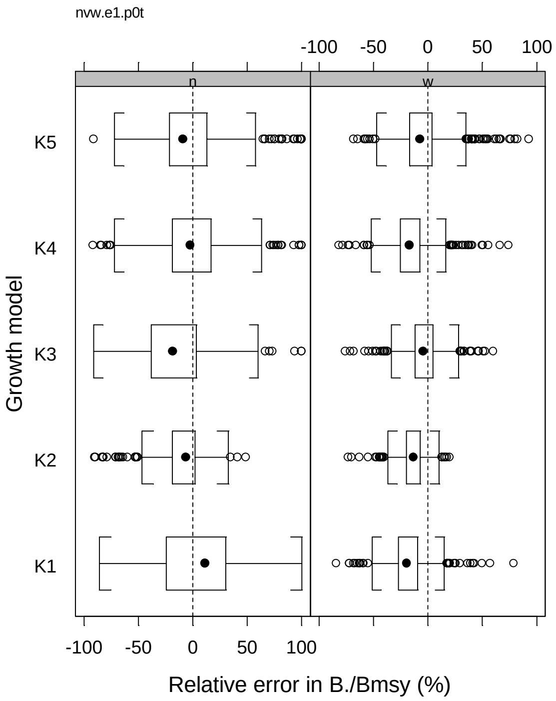

Figure 14. Distribution of percent error in B/B$_{MSY}$ estimates for trials with stochastic recruitment and no penalty term

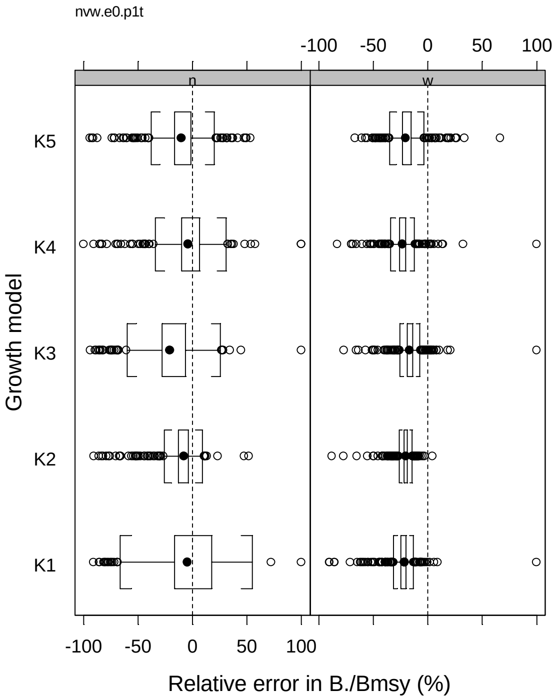

Figure 15. Distribution of percent error in B/B$_{MSY}$ estimates for trials with deterministic recruitment and with the penalty term.

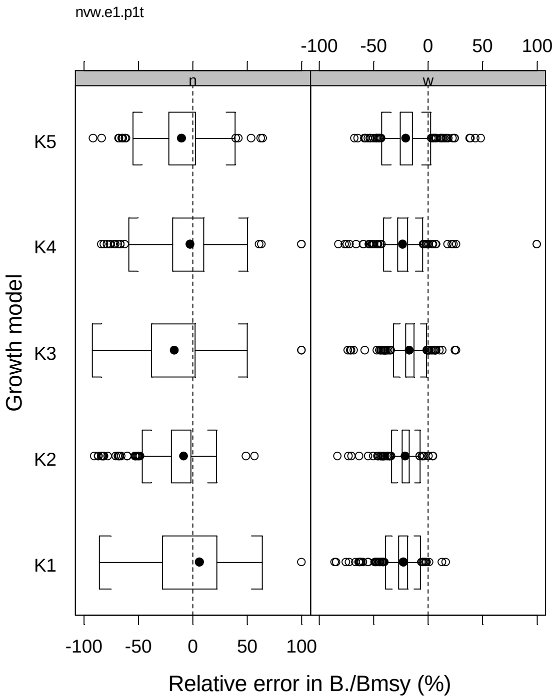

Figure 16. Distribution of percent error in B/B$_{MSY}$ estimates for trials with stochastic recruitment and with the penalty term.

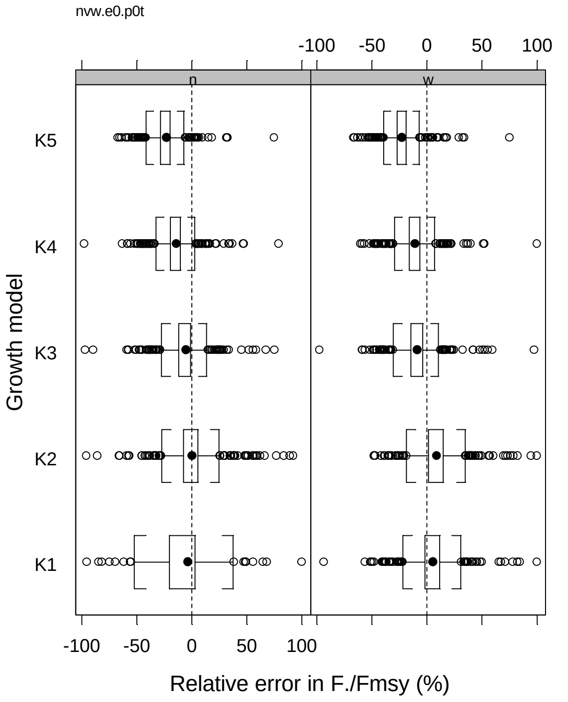

Figure 17. Distribution of percent error in F/F$_{MSY}$ estimates for trials with deterministic recruitment and no penalty term.

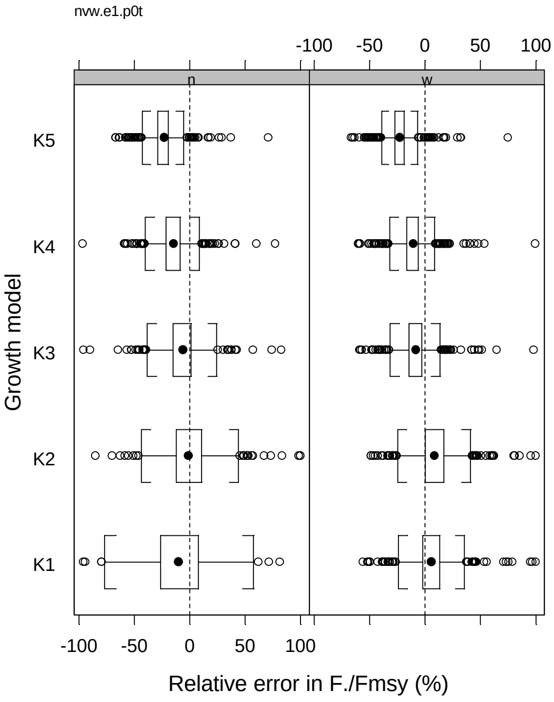

Figure 18. Distribution of percent error in F./F$_{MSY}$ estimates for trials with stochastic recruitment and no penalty term for B_1>K for each growth treatment (K1 through K5)

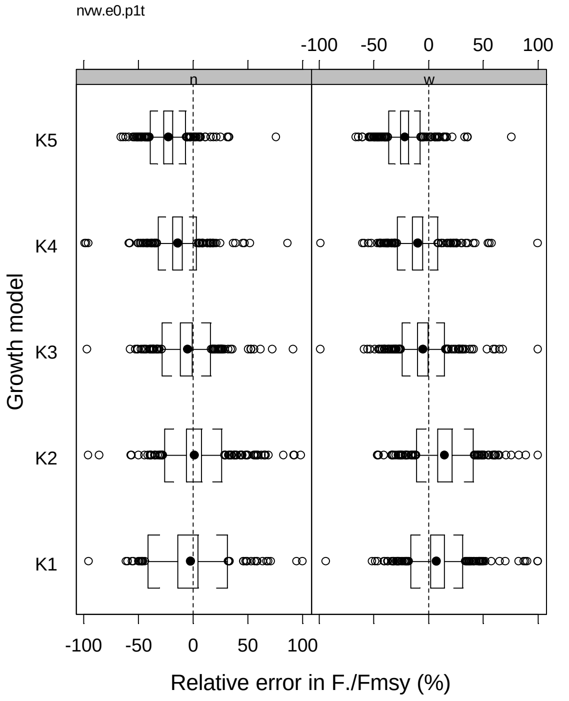

Figure 19. Distribution of percent error in F/F$_{MSY}$ estimates for trials with deterministic recruitment and with the penalty term.

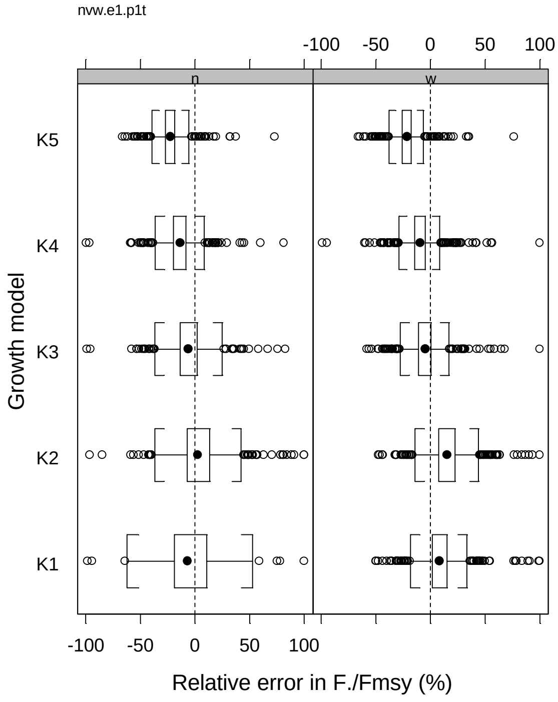

Figure 20. Distribution of percent error in F/F$_{MSY}$ estimates for trials with stochastic recruitment and with the penalty term.

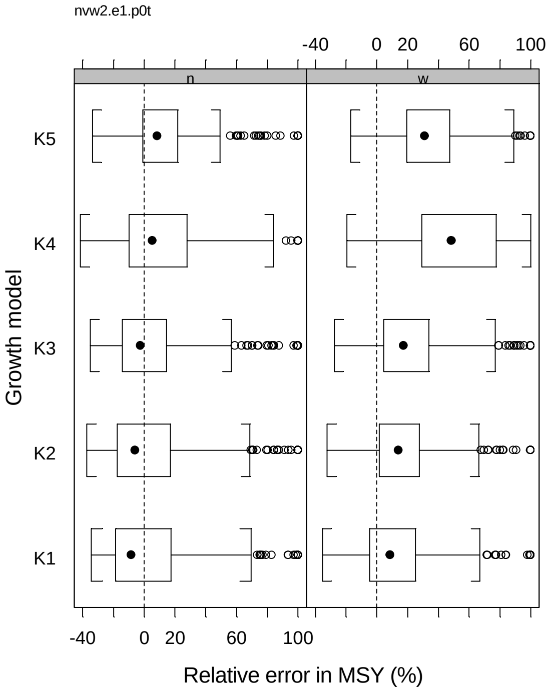

Figure 21. Distribution of percent error in estimates of MSY from trials with stochastic recruitment and alternative trajectory.

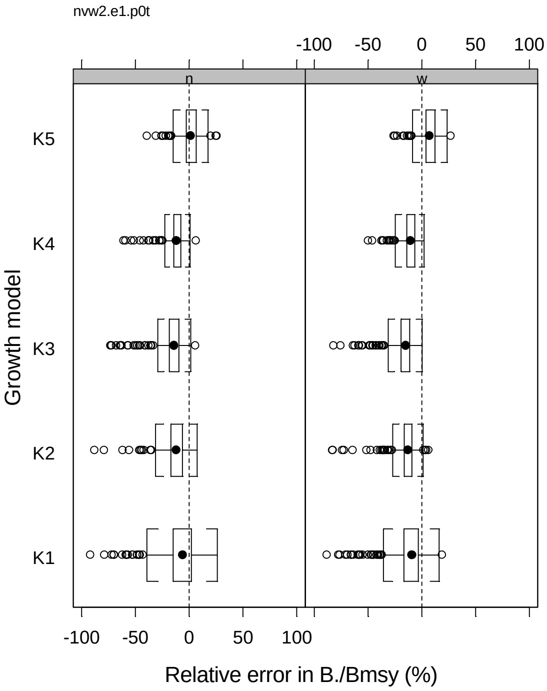

Figure 22. Distribution of percent error in estimates of final-year relative biomass from trials with stochastic recruitment and alternative trajectory.

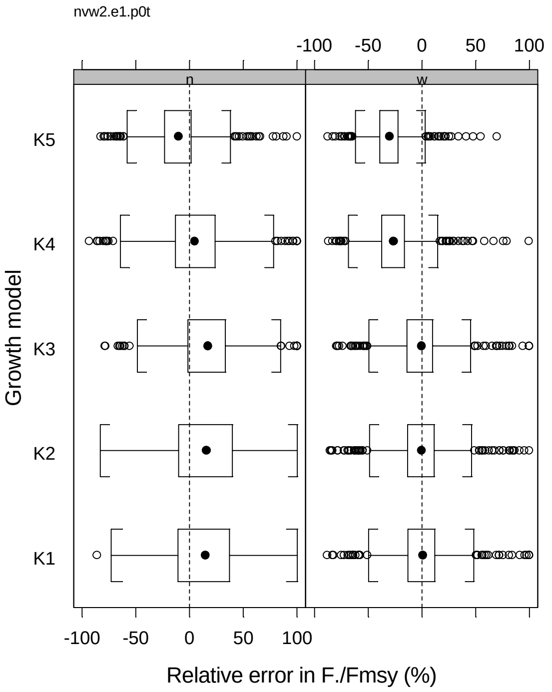

Figure 23. Distribution of percent error in estimates of final-year relative fishing mortality rate from trials with stochastic recruitment and alternative trajectory.