SCRS/2008/094 Collect. Vol. Sci. Pap. ICCAT, 64(2): 491-506 (2009)

EVALUATION OF THE PERFORMANCE AND ROBUSTNESS OF VPA-BASED STOCK ASSESSMENT AND MSY-BASED MANAGEMENT STRATEGY TO PROCESS ERROR: THE ATLANTIC BLUEFIN TUNA CASE STUDY

J.-M. Fromentin[1] and L.T. Kell[2]

SUMMARY

An integrative simulation framework was built to evaluate the consequences of variability attributable to changes in carrying capacity or the stock's migration pattern of Atlantic bluefin tuna on the ICCAT stock assessment and management procedures. We also evaluated the performances of stock assessment methods with respect to their ability to provide good estimates of MSY, F_{MSY} and B_{MSY} and tested the robustness of the current ICCAT management strategy to uncertainty about the true dynamics and historical exploitation levels. The results clearly indicate that the VPA performances were seriously impaired if the long-term variations in catches are due to changes in migration/availability. There is further considerable confounding between the underlying dynamics and increasing effort that makes it difficult to draw any inference about the actual dynamics based on commercial catch data. Reference points based on F (e.g., $F_{0.1}$) were less biased and more precise than those based on yield and/or SSB and were also more robust to uncertainty about the true dynamics than absolute values of F and SSB. However, $F_{0.1}$ cannot indicate past and current levels of exploitation relative to F_{MSY} when there is uncertainty about the dynamics. Therefore, while reference points such as $F_{0.1}$ may be good proxies for F_{MSY}, the MSY concept may be difficult to make operational. A size limit strategy generates greater yields than the more sophisticated $F_{0.1}$, but lower SSB. However, the *performances of both strategies were strongly dependent on the period over which they are implemented relative to the intrinsic cycle of the population. We finally stressed that the performances and robustness of management strategies also depend on concrete objectives, such as fleet composition, gear selectivity and economic constraints.*

RESUME

Un modèle de simulation a été élaboré pour évaluer les conséquences de variations dues à des changements de la capacité de charge ou des routes migratoires du thon rouge atlantique sur les procédures d'évaluation et de gestion de la CICTA. Nous avons également évalué les performances des méthodes d'évaluation pour produire des estimations fiables de PME, F_{PME} et B_{PME} et testé la robustesse de la stratégie de gestion de la CICTA aux incertitudes liées à la dynamique sous-jacente et aux niveaux d'exploitation historiques. Les résultats indiquent clairement que les performances de la VPA sont sérieusement altérées si les variations à long terme des captures résultent de changements migratoires. De plus, les incertitudes sur la dynamique sous-jacente interfèrent avec l'effort lorsque ce dernier est croissant, ce qui rend difficile de déduire la véritable dynamique sur la base des seules données de pêche. Les points de référence basés sur F (par ex, $F_{0,1}$) sont moins biaisés et plus précis que ceux basés sur la biomasse ou les captures et sont, de surcroît, plus robustes à l'incertitude sur la dynamique sous-jacente que les valeurs absolues de F ou SSB. Cependant, $F_{0,1}$ ne peut indiquer correctement les niveaux d'exploitation passés et présents par rapport à l'exploitation à F_{PME} en présence d'incertitude sur la dynamique. Si des points de référence comme $F_{0,1}$ peuvent être de bonnes approximations de F_{PME}, le concept de PME peut donc être difficile à implémenter. Une stratégie de gestion basée sur une taille limite conduit à des captures plus élevées qu'une stratégie sophistiquée basée sur $F_{0,1}$, mais à de plus faibles SSB. Cependant, les performances des deux stratégies dépendent fortement de la période à laquelle elles sont implémentées par rapport au cycle intrinsèque de la population. Nous concluons en rappelant que les

[1] IFREMER, Centre de Recherche Halieutique Méditerranéen et Tropical, avenue Jean Monnet, BP 171, 34203 Sète cedex, France. E-Mail: jean.marc.fromentin@ifremer.fr

[2] CEFAS, Lowestoft Laboratory, Pakefield Road, Lowestoft, Suffolk, NR33 0HT, UK. E-Mail: laurence.kell@cefas.co.uk

performances et la robustesse des stratégies de gestion dépendent aussi des objectifs concrets, comme la composition des flottilles, la sélectivité des engins et les contraintes économiques.

RESUMEN

Se elaboró un modelo de simulación para evaluar las consecuencias de las variaciones debidas a cambios en la capacidad de transporte o de las rutas migratorias del atún rojo atlántico en las evaluaciones de stock y los procedimientos de ordenación de ICCAT. También se ha evaluado el rendimiento de métodos alternativos de evaluación de stock para producir estimaciones fiables de RMS, F_{RMS} y B_{RMS} y se comprobó la robustez de la estrategia de ordenación de ICCAT frente a las incertidumbres vinculadas con la dinámica subyacente y con los niveles históricos de explotación. Los resultados indican claramente que los rendimientos de la VPA se ven fuertemente alterados si las variaciones a largo plazo en las capturas se deben a cambios migratorios o en la disponibilidad. Además, las incertidumbres sobre la dinámica subyacentes interfieren con el esfuerzo cuando este último es creciente, lo que hace que resulte difícil deducir la dinámica real basándose sólo en datos de pesca. Los puntos de referencia basados en F (por ejemplo $F_{0.1}$) presentan menos sesgos y son más precisos que los que se basan en la biomasa o en las capturas y, además, son más robustos frente a la incertidumbre sobre la dinámica subyacente que los valores absolutos de F o SSB. Sin embargo, $F_{0.1}$ no puede indicar correctamente los niveles de explotación pasados y presentes con respecto a la explotación F_{RMS} cuando existe incertidumbre en cuanto a la dinámica. Por tanto, aunque los puntos de referencia como $F_{0.1}$ pueden ser buenas aproximaciones de F_{RMS}, el concepto RMS podría ser difícil de implementar. Una estrategia de ordenación basada en límites de talla se traduce en capturas más elevadas que una estrategia sofisticada basada en $F_{0.1}$, pero también produce una SSB más baja. Sin embargo los rendimientos de las dos estrategias dependen en gran medida del periodo en el que se implementan con respecto al ciclo intrínseco de la población. Se concluye recordando que los rendimientos y la robustez de las estrategias de ordenación dependen también de los objetivos concretos, como composición de las flotas, selectividad de los artes y restricciones económicas.

KEYWORDS

Thunnus thynnus, stock assessment, VPA, biological reference points, MSY, management strategy, simulation model

1. Introduction

A recent study on historical catches from the ancestral Atlantic bluefin tuna trap fisheries (i.e. time-series of 50-400 years) depicted spectacular long-term fluctuations in catches (Ravier and Fromentin, 2001). Further studies postulated that such long-term fluctuations could arise from resonant effects, attributable to a combination of stochastic recruitment and long life (Bjørnstad *et al.*, 2004, Fromentin, 2002) or from modifications in migration patterns due to changes in sea temperature (Ravier and Fromentin, 2004). Whatever are the underlying processes, significant fluctuations in carrying capacity or in migration patterns are likely to affect our perception of stock status (Kell *et al.*, 2000), especially as time-series used in stock assessment represent only about a third of full cycle.

Therefore an integrative simulation framework (Kell *et al.*, 2003, Kell *et al.*, 2007) was built and used to evaluate the consequences of variability attributable to either changes in carrying capacity or the stock's migration pattern in the stock assessment and management procedures of the International Commission for the Conservation of Atlantic Tunas (ICCAT). We further evaluated the performances of stock assessment methods with respect to: (i) their ability to provide estimates of MSY, F_{MSY} and B_{MSY}, and (ii) assessing stock status and exploitation level relative to these MSY targets. We then tested the robustness of the current ICCAT management strategy (i.e. based upon MSY) to uncertainty about the true dynamics and historical exploitation levels, and contrast it with a simpler management strategy, with fewer data and less analytical requirements, based on size limits.

This work has been recently published in two articles (Kell and Fromentin, 2007, Fromentin and Kell, 2007) and the purpose of this manuscript is to highlight the main results that may be of interest for the next BFT stock assessment.

2. The simulation model

2.1 Testing the robustness of VPA stock assessment method to process errors

The simulation framework (Kell *et al.*, 2003, 2007) was used to build: (i) an operating model (OM), that represented alternative plausible hypotheses about stock and fishery dynamics, allowing integration of a higher level of complexity and knowledge than used within stock assessment models; (ii) an observation model that describes how pseudo-data are sampled from the operating model; (iii) an assessment procedure to derivate estimates of stock status from the pseudo-data (**Figure 1**). Within this framework, we considered three different types of variable. In the operating model, the variables are controlled, i.e. fixed by the operator, so decisions have to be taken on biological parameters. In the observation model, the variables are observed, i.e. deduced from the sampling model, whereas in the assessment procedure, the variables are estimated through the stock assessment model.

The objective of the present study is to use a simulation framework to evaluate whether stock management and assessment procedures based on Virtual Population Analysis (VPA) can capture the dynamics of stocks where ecological processes may generate long-term fluctuations in catches and in particular, how they affect our perception of the stock and ability to manage it. To do so, an operating model was constructed on the basis of the classical age-structured equation, assuming an annual spawning (1 cohort per year), a fecundity being proportional to the weight, 50% maturity at age 4, full maturity at age 5+, a lifespan of 20 years, the natural mortality vector and the von-Bertalanffy equation used in the ICCAT working group. The population dynamics of the OM was further based on a Beverton and Holt SSB/recruitment relationship (Beverton and Holt, 1957) for which steepness was set within a range of values that make biological sense (i.e. 0.75 and 0.9). The expected virgin biomass was arbitrarily set at 10^6 t because the study was intended to provide strategic rather than tactical advice.

Operating models

The purpose of the generic operating model is not to reproduce the entire complexity of bluefin tuna biology and ecology, which is not the objective of this work and would be highly subjective because our knowledge of bluefin tuna ecology relies mainly on a variety of unproven hypotheses, but to define plausible hypothesis about population dynamics, then to implement the processes of interest, i.e. changes in carrying capacity and migration pattern.

The Operating model for Carrying Capacity (**H_K**) assumes that the long-term fluctuations observed in trap catches reflect changes in the carrying capacity or virgin biomass (Fromentin, 2002), which is assumed to vary with respect to time, as:

$$\lambda_t = \varpi \sin(\theta + \delta t)$$

where □, □ and δ were chosen to reflect the observed long-term trend in trap catches (**Figure 1**).

The Operating model for migration (H_M) assumes that long-term variations in trap catches may result from changes in migratory patterns that would affect the proportion of mature bluefin tuna entering the Mediterranean Sea each year to reproduce (Ravier and Fromentin, 2004). Knowing that most fishing effort is in the Mediterranean, such changes induce concomitant variations in the fraction of the population available to fishing gears. Therefore, such an hypothesis may be modeled as:

$$N_{a+1,t+1} = A_t \times N_{a,t} \times e^{(-M_t - F_{a,t})} + (1 - A_t) \times N_{a,t} \times e^{(-M_t)},$$

where A_t is the proportion of the population available to fishing ($1 - A_t$) can thus be equated to the unexploited or cryptic biomass). It was assumed that A_t was a sinusoidal function of time:

$$A_t = \varpi \sin(\theta + \delta t)$$

where □, □ and δ were again chosen to reflect the long-term trend in trap catches (**Figure 1**).

As the dominant signal, i.e. low frequencies, was the process of interest, the smoothed time-series of trap catches (**Figure 1**) was used to estimate the time-varying population parameters. In both models, the yield was predicted using an age-structured forward projection, with a selectivity pattern similar to the present exploitation (Anon., 2003). The sum of squares of the difference in the observed trend was scaled so that the mean yield and its amplitude was a given fraction of MSY, and predicted yields were then minimized.

Experimental treatments

As the actual population is unknown, decisions have also to be taken with respect to (i) the amplitude of the trend, (ii) the steepness of the stock/recruitment relationship, (iii) the historical fishing mortality (relative to MSY), and (iv) the starting point (its value and its phase relative to the trend).

The amplitude was fixed so that it was similar under both hypotheses (H_K and H_M) and in agreement with the variability observed in catches over the 1910–1930 (Ravier and Fromentin, 2002) and 1950–1990 periods (i.e. a rough CV of ~40%).

Four scenarios were chosen corresponding to historical *F*s: (1) constant effort corresponding to a fishing mortality equal to 50% of F_{MSY}, (2) equal to $F_{MSY,}$ (3) equal to 150% of F_{MSY}, and (4) a linear increase in effort corresponding to an increase in fishing mortality from 50% of F_{MSY} to 250% of F_{MSY} over a full population cycle, i.e. 110 years. In addition, yield in year 1 was at four different phases of the cycle, i.e. at (1) the maximum, (2) the middle of the decreasing yield, (3) the minimum, or (4) the middle of the increasing yield.

Monte Carlo simulations, where process and observation errors are modeled as random variables, were also considered as part of the experimental design.

2.2 Testing the performances of two different Management Strategies

The Management Procedure (MP) is the specific combination of: (i) the sampling regime, (ii) the stock assessment method, (iii) the biological reference points, and (iv) the management strategies. Here the MP is based on the ICCAT management regime applied to Atlantic bluefin tuna (Anon., 2003).

Sampling regime

The sampling regime corresponds to the collection of commercial catch data and the derivation of catch numbers-at-age and catch per unit effort (CPUE). These data were generated by the Observation Error Model in which growth, maturity and natural mortality-at-age were sampled without error from the OM. Modeled values were the same as used in the 2002 stock assessment and did not vary between years. However, catch-at-age was sampled with random error (from a multinomial distribution) based on the study of Arrizabalaga *et al.* (2005) which used Monte Carlo simulation of monthly catch-at-size data of some fleets to estimate measurement errors in the whole catch-at-age. These data could then be used to estimate the correlations between ages and the mean-variance relationship for each age to derive the covariance matrix for sampled catch-at-age. However, we fixed the CV for all ages at 20% to avoid high variances at some ages (mostly caused by a lack of monitoring).

Assessment method and biological reference points

The stock assessment model used is ADAPT-VPA as used to perform bluefin tuna stock assessments by ICCAT (Porch, 1997). ADAPT-VPA uses virtual population analysis (VPA) to recreate historical numbers and fishing mortality at age from the total catch-at-age data, conditional upon numbers (or fishing mortality) at age of the oldest age in each cohort, where the latter is estimated using CPUE from the fishery. It is assumed that catch and natural mortality are known without error, that there is no immigration or emigration, and that the stock is homogeneous.

Biological reference points (BRP) chosen for management were all proxies for F_{MSY}, i.e. $F_{0.1}$, F_{max}, $F_{30\%SPR}$ and $F_{40\%SPR,}$ or the corresponding values of MSY and B_{MSY} (calculated from the yield- and spawner-per-recruit curves times the mean recruitment). $F_{30\%SPR}$ and $F_{40\%SPR}$ are the fishing mortalities that correspond to values of spawner-per-recruit that are 30% and 40% respectively of the virgin biomass (i.e. biomass at zero fishing mortality).

Management strategies

Two contrasting management regimes were considered:

(i) a harvest control rule (HCR) based on $F_{0.1}$ (mimicking the ICCAT harvest control rule);
(ii) an alternative simple regime based on a change in selection pattern of immature fish (i.e. younger than 5 years), to allow 75% of fish to spawn at least once.

For each experimental treatment, the management strategies were run for 15 years into the future (i.e. years 111 to 125) and population parameters and biological reference points were re-estimated using ADAPT-VPA (i.e. there were two stock assessments, one taking place in year 110 and one in year 125). The period 15 years was chosen because it corresponds to the generation time of Atlantic bluefin tuna. ICCAT has previously expressed concern about the quality of BFT catch and effort data (e.g. Anon., 2005). Various possible causes for misreporting of catches (including non-reporting by members and non-member countries) have been postulated. One of the main reasons for misreporting appears to be related to the implementation of quotas for East Atlantic and Mediterranean bluefin tuna in 1996 and 1998. It was subsequently believed that this resulted in over-reporting prior to the period 1996–1998, and underreporting since. Although little quantitative information is available to characterize misreporting precisely, ICCAT proposed an alternative catch scenario, based on 15% over-reporting for the period 1993–1997 and 15% under-reporting from 1998 onwards, in order to conduct sensitivity trials during the last stock assessment (Anon., 2003). We therefore added an Implementation Error Model to reflect the fact that current harvest control rules may be poorly endorsed, using the same scenario as that applied in the 2002 stock assessment.

For each treatment, 1000 Monte Carlo simulations were run, with random variables as stated above. Stochastic runs including recruitment and observation error were performed for 1000 simulations.

3. Results

3.1 Robustness of VPA stock assessment method to process errors

Operating models with constant effort

To illustrate the contrast in dynamics under the two hypotheses for a full cycle of 110 years, the results from a single scenario, with constant effort (equal to F_{MSY}) and a steepness of 0.75 are presented (**Figure 2**). Interpretation of the results under H_K (where virgin biomass is the fluctuating control variable) are relatively straightforward (**Figure 2a, c, e**). As expected, CPUE, yield and recruitment (R) vary in synchrony with virgin biomass (note that the catch is confounded with CPUE because effort is constant). An increase/decline in R induces a subsequent increase/decline in SSB and yield with a lag of about 15 years (**Figure 2c, e**). Therefore, for a given SSB there are two values of recruitment (and *vice versa*). The fact that the yield and CPUE (or catch) are slightly delayed simply reflects changes in the composition of the catches caused by variations in R, because older fish weigh more (**Figure 2c**). Under H_M when availability is the fluctuating control variable (**Figure 2b**), the results are more complex. Although effort is constant, F fluctuates owing to the varying proportion of the population available to fishing: it increases as the proportion of the population available to fishing increases (**Figure 2f**). Variations in fishing mortality and availability are, therefore, synchronous (**Figure 2b, f**). Also in contrast to the situation with H_K, synchrony is seen between SSB and R (**Figure 2f**). Furthermore, the phases of SSB and R are now in opposition to the control variable (and F). SSB and R are high when the availability to fishing is low, and *vice versa*. This results from the stock not being in equilibrium and, while fishing mortality is a function of current availability, year-class strength is a function of historic availability. Consequently, there are two dissimilar and contrasting values of yield and SSB for a given F. Variations in CPUE (or catch) and yield result from both the fraction being harvested (i.e. availability) and the population size (or SSB), so CPUE and yields peak after the maximum SSB, but before F does. The yield is, however, more dependent on SSB than on CPUE (or catch), which generates a slight delay between both variables (**Figure 2d**).

Operating Models with increasing effort

The response of the two Operating Models to a linear increase in effort (leading to an increase in F from 50% of F_{MSY} to 250% of F_{MSY}) are explored in **Figure 3** for a steepness of 0.75 and two contrasting starting points, i.e. initially close to a maximum (continuous line) and a minimum (broken line) in carrying capacity or availability. Increasing effort alters the long-term fluctuations in carrying capacity or availability that are observed under constant effort (especially in the second part of the cycle which is generally masked; **Figure 3a, b, c, d**). Differences are still seen between H_K and H_M, but also within each hypothesis, depending on the starting conditions. Under H_K, there is a lag of about 15 years between yield and SSB when effort increases (whereas these are in phase under constant effort), which is due to year-class effects (**Figure 3a**). Differences in starting conditions further generate distinct yield patterns under H_K (**Figure 3c**). In one case, yields are relatively stable through time, whereas in the second case they clearly display a dome-shaped curve (with more variability). Under H_M, differences in starting conditions are more acute and induce strong differences in the range of F. When the simulation starts (and ends) at low availability, fishing mortality is about F_{MSY} at the end of the series, whereas it exceeds 250% of F_{MSY} when starting at high availability (**Figure 3f**). Again this is because availability and SSB are in opposite phase. As F is proportional to availability (see **Figure 2**), F remains relatively low for initial and final low availability, because high values of effort are partially compensated for by low values of availability. For initial high availability, greatest effort coincides with greatest availability and generates very high F (**Figure 3b, f**). This makes the final decreasing slope in yield more abrupt (**Figure 3d**). If final SSBs are different (75% and 40% of B_{MSY} when starting at low and high availability, respectively), there are even stronger differences in the terminal trends in SSB, slightly positive in the former case, strongly negative in the latter (**Figure 3b**). Finally, there is considerable confounding between the dynamics and effort.

Consequences for stock assessment

To evaluate the consequences of these two Operating Models on our perception of stock status, catch-at-age and CPUE were sampled from the Operating Model (without observation error) for the most recent 30 years. We then used the same ADAPT-VPA used by the ICCAT working group (Porch, 1997) to reconstruct the historical stock parameters of interest (i.e., SSB and F). The performance of VPA to both Operating Models is firstly evaluated for the scenarios with a constant effort (equivalent to F_{MSY}), a steepness of 0.75 and for four different starting points. In the case of H_K, VPA is always able to reconstruct the time-series of SSB and fishing mortality accurately (both with respect to trends and absolute values, **Figure 4a, c**). However, the reconstruction of the historical time-series under H_M is more problematic (**Figure 4b, d**), in particular because SSB, yield (and CPUE) and fishing mortality are not synchronous within the OM (see above). Consequently, there is in most cases a mismatch between the time-series of auxiliary information (CPUE), catch, SSB and fishing mortality over the stock assessment period, which induces strong biases in both trend and absolute values (that can even reach up to 500%, **Figure 4d**). The trends in CPUE can, fortuitously, match those of SSB and F over a short period (such as the 30 years of the stock assessment period). In those cases (1 in 4 of the scenarios), the estimates are more satisfactory (**Figure 4b, d**). The other scenarios (with different fishing mortality patterns and higher steepness) are in agreement with the above findings.

Let's consider the 64 scenarios (2 hypotheses, 2steepness, 4 historical *Fs*, 4 starting points) to evaluate the main sources of the bias (**Figure 5**). To do so, we computed the ratio of the VPA to the OM quantities (SSB and F) of all scenarios (a value close to 1 indicating no bias) through box-plots. The results by hypothesis are clearly dissimilar (**Figure 5a**), being unbiased (i.e. equal to 1) for all the cases (no variability) under H_K, but strongly biased under H_M, with high variability between cases (> 300% in some cases for SSB). The results by steepness, constant historical *Fs* relative to F_{MSY} and effort (constant *versus* increasing) did not display any special patterns (**Figure 5b, c, d**). The results are generally unbiased, except for some cases when F=50% or 150% of F_{MSY}. Differences in steepness, historical *Fs* and effort (i.e., constant or increasing) do not affect the performances of the VPA. The boxplots split by phases (or starting points, **Figure 5e**) exhibit some strong differences, since the estimates of the VPA (especially those of SSB) are more biased for some phases ($\pi/2$ and $3\pi/2$) than others. Note that for all the scenarios, the estimates of SSB are more biased than those of F.

3.2 Performances of different Biological Reference Points (BRPs)

The performance of the MP in providing proxies of MSY, F_{MSY} and B_{MSY} is evaluated in **Figure 6**. For each hypothesis and each phase, a range of proxies for F_{MSY} were calculated (i.e., $F_{0.1}$, F_{max}, $F_{SPR30\%}$, $F_{SPR40\%}$). Here again, ratios close to 1 indicate good performance. In contrast to Figure 5, the differences between hypothesis (H_M and H_K) are not so critical, especially for BRP based on F (**Figure 6a, 6b**). In other words, the performances of the BRPs appear to be more robust to uncertainty in the dynamics than are the stock assessment estimates (**Figures 4 and 5**). This is because BRP are based on equilibrium calculations and selectivity patterns, which are biased by other factors than the underlying dynamics. The biggest difference between scenarios is for initial conditions (i.e. phases between exploitation and natural long-term cycle), i.e., is availability or carrying capacity currently increasing/decreasing or at a top/bottom. However, the variability remains much higher under H_M than under H_K, especially for yield- and SSB-based BRP (**Figure 6d, 6f**). In general, $F_{0.1}$ (and secondarily $F_{40\%SPR}$) gives better and more consistent results than the other BRPs ($F_{0.1}$ values are indeed always around 1 and display little variance among scenarios; **Figure 6**). $F_{0.1}$ therefore appeared to be the best proxy for F_{MSY}, so subsequent analyses are only presented for *F*.

Figure 7 presents the $F_{0.1}$-based reference points from the MP (i.e., $B_{F0.1}$, $F_{0.1}$, $Y_{F0.1}$) divided by the corresponding MSY-based quantity from the OM (F_{MSY}), i.e. $F_{0.1}$ relative to F_{MSY}. The corresponding proxies for B_{MSY} and MSY are derived from yield- and spawner-per-recruit assuming that recruitment was equal to the mean of the last five years. The performances of $F_{0.1}$-based reference points in providing a good proxy for F_{MSY} is always better under H_K than under H_M, especially for increasing *F* scenarios (**Figure 7e, 7f**). For some phases, biomass- and yield-based reference points ($B_{F0.1}$ and $Y_{F0.1}$) are strongly biased (up to 200%), and can further display large variability (especially for increasing *F* under H_K, **Figure 7f**). In contrast, $F_{0.1}$ reference points are more consistent among starting points and hypotheses. They are always precise (i.e., they display little variability) and are only slightly biased (but consistently underestimated). Misreporting has little effect on the performances of all $F_{0.1}$-based reference points.

We finally evaluated the performance of relative indicators defined by the ratio of *F*, SSB, and Yield to their corresponding $F_{0.1}$-based reference points (again by dividing these ratios to corresponding ratios from the OM) e.g., $\frac{\bar{F}:F_{0.1}}{F:F_{MSY}}$ ($\bar{F}$ being the mean value of *F* of the MP over the past five years). In contrast to previous results, *F*-based quantities exhibit the greatest biases and, most often, the largest variations (**Figure 8a to 8f**). Both hypothesis and starting point are important in determining the accuracy of these estimates, which are again more biased and much less precise under H_M than under H_K, because VPA performs poorly under H_M (see above). Here again, misreporting has little effect on the performance of *F*-based quantities, but increasing historical *Fs* have. In summary, the *F*-based quantities (i.e., *F* relative to $F_{0.1}$) lead to unreliable estimates (especially under H_M), and are more biased and more variable than $F_{0.1}$-based reference points.

3.3 Performances of different management strategies

The evaluation of reference points is best performed as part of a Management Procedure that includes the harvest control rule and stock assessment method. Therefore, the performances of an $F_{0.1}$-based Management Procedure was then evaluated and compared to an alternative where selection pattern (rather than *F*) was the management variable. The $F_{0.1}$ MP is an attempt to implement ICCAT management objectives in a harvest control rule intended to achieve MSY. In contrast, the alternative MP solely relies on size limit regulation and does not modify fishing effort (the size limit set at age-at-maturity with a 25% of tolerance). Performances of both MP were based upon summary statistics from the OM after 15 years of implementation (one generation time) and were evaluated by comparison with the *status quo* (no regulation). The results are depicted by phases when considering constant historical *Fs* equivalent to 150%F_{MSY}, which is a scenario closer to the current fishing pressure than the others. As MSY depends on the selectivity pattern of the fleets (Powers, 2005), yield is expressed relatively to maximum possible yield and SSB relatively to virgin biomass.

Relative yields under both management strategies vary between 20% and 60% of the maximum yield (**Figure 9a, 9b**). Differences appear to be mostly due to management strategy and starting point, and less to the underlying hypothesis (H_M or H_K). Although higher yields are seen under H_K, the range of yields remains similar under both hypotheses. Regarding the MP, expected yields are always highest under the size limit strategy (up to 60% of maximum yield), second highest under *status quo* and lowest under the $F_{0.1}$ strategy (where they do not exceed 40% of the maximum yield (**Figure 9a, 9b**). However, the performance of a given management strategy also depends on when it is implemented relative to the intrinsic cycle of the population. The expected SSBs are in general similar to or slightly higher under the $F_{0.1}$ than under the size limit strategy (**Figure 9c, 9d**). Depending on starting points, SSBs are at 40% (about B_{MSY}) or 20% (about $B_{75\%R}$) of virgin biomass. Under the size limit strategy, SSB is always greater than the *status quo* and, in all the cases but one, SSB remains greater than 20%Virgin (a potential recovery level). The *status quo* generally leads to the lowest SSBs which are sometimes under the 20%Virgin limit. When considering increasing historical *Fs*, results display comparable patterns under H_K, but not for under H_M, for which both yields and SSB are clearly lower.

4. Conclusion

This simulation study has displayed several results of interest for BFT stock assessment:

- The underlying mechanism causing the long-term fluctuations in catches is of primary importance, as H_K and H_M lead to contrasting results, especially regarding F (constant in one case and fluctuating in the other) and regarding the differences of phase between SSB, R, yield and F.
- There is considerable confounding between the underlying dynamics and increasing effort that makes it difficult to draw any inference about the actual dynamics based on commercial catch data.
- The phases between exploitation and natural cycle induce contrasting outputs, especially for H_M, as the terminal F can vary by 250% for the same effort.
- The VPA is generally able to reconstruct accurately the historical stock parameters under H_k, but not under H_M.
- The performances of VPA further depend on the starting point (i.e. when the exploitation starts along the natural long-term cycle).
- The performances of the biological reference points appear to be more robust to uncertainty in the dynamics than the stock assessment estimates.
- $F_{0.1}$ is the best proxy for F_{MSY} and performs better under H_K than under H_M.
- $F_{0.1}$ appears to be more accurate (i.e., less biased and more precise) than reference points based on biomass or yield ($B_{F0.1}$ and $Y_{F0.1}$).
- However, the ability of F-based quantities to predict exploitation level relative to F_{MSY} was poor.
- A size limit strategy generates greater yields than the more sophisticated $F_{0.1}$, but lower SSBs in some cases. More important than the underlying process (H_M or H_K) was the current phase of the cycle which always strongly affected the performances of both management strategies.

Interestingly, reference points based on F (e.g., $F_{0.1}$) were less biased and more precise than those based yield and/or SSB and were also more robust to uncertainty about the true dynamics than absolute values of F and *SSB*. However, $F_{0.1}$ cannot indicate past and current levels of exploitation relative to F_{MSY} when there is uncertainty about the dynamics. Therefore, while reference points such as $F_{0.1}$ may be good proxies for F_{MSY}, the MSY concept may be difficult to make operational when trends in yield can occur, either through variations in carrying capacity (H_K), migration (H_M), or effort.

ICCAT has stressed that there are potential biases in VPA estimates, due to strong large in the catch-at-age data and CPUE indices (Anon., 2003, 2007). Our study further indicates that the VPA performances could be seriously impaired if the long-term variations in catches are due to changes in migration/availability. These results highlight the risks of basing scientific advice of the East Atlantic and Mediterranean BFT solely on VPA estimates, as this was done until 2002. There is a need therefore for alternative indicators, as was done in 2006 (Anon., 2007), however how to use such indicators as part of a management procedure needs to be further evaluated.

An alternative to an F_{MSY} strategy could be simply setting a size limit (with or without constraints on effort) or a time/area closure. This would require less knowledge about stock dynamics, and might provide an alternative that is more robust to uncertainty about biological processes. We did not investigate the potential of a time/area closure, but the performances of a size limit regulation were rather similar to those of a F_{MSY} strategy (although leading to lower SSB in some cases). However, the performances of both strategies were strongly dependent on the period over which they are implemented relative to the intrinsic cycle of the population.

We finally stressed that the performances and robustness of management strategies also depend on concrete objectives, such as fleet composition, gear selectivity and economic constraints. Indeed, a size limit strategy (or a time/area closure one) will reduce effort or yield for certain fleets more than others, while an $F_{0.1}$ strategy mostly implies an equal cut in effort by all fleets.

References

Anon., 2003. Report of the 2002 Atlantic Bluefin Tuna Stock Assessment Session. Collect. Vol. Sci. Pap. ICCAT, 55(3): 710-937.

Anon., 2005. Report of the 2004 data exploratory meeting for the East Atlantic and Mediterranean bluefin tuna. Collect. Vol. Sci. Pap. ICCAT, 58(2): 662-699.

Anon., 2007. Report of the 2006 Atlantic Bluefin Tuna Stock Assessment Session. Collect. Vol. Sci. Pap. ICCAT, 60(3), 652-880.

Arrizabalaga, H., 2005. The effect of eastern Atlantic and Mediterranean bluefin tuna sampling error on the catch at age. Collect. Vol. Sci. Pap. ICCAT*,* 58(2): 791-797.

Bjørnstad, O., Nisbet, R.M., Fromentin, J.-M., 2004. Trends and cohort resonant effects in age-structured populations. *Journal of Animal Ecology* 73**,** 1157-1167.

Fromentin, J.-M., 2002. Can stochastic variations in recruitment induce long-term fluctuations in the carrying capacity? Collect. Vol. Sci. Pap. ICCAT, 54(4): 985-991.

Fromentin, J.M., Kell, L.T., 2007. Consequences of variations in carrying capacity or migration for the perception of Atlantic bluefin tuna population dynamics. *Canadian Journal of Fisheries and Aquatic Science* 67**,** 627-836.

Kell, L.T., Die, D., Restrepo, V., Fromentin, J.-M., Ortiz de Zarate, V., Pallares, P., 2003. An evaluation of management strategies for Atlantic tuna stocks. *Scientia Marina* 67, 353-370.

Kell, L.T., Fromentin, J.-M., Gauthiez, F., Restrepo, V., 2000. A simulation framework to evaluate management strategies for Atlantic tunas: a preliminary example based on East Atlantic bluefin tuna. Collect. Vol. Sci. Pap. ICCAT, 51(6): 2095-2116.

Kell, L.T., Fromentin, J.M., 2007. Evaluation of the robustness of MSY-based management strategies to variations in carrying capacity or migration pattern of Atlantic bluefin tuna *Canadian Journal of Fisheries and Aquatic Science* 67, 837-847.

Kell, L.T., Mosqueira, I., Grosjean, P.,Fromentin, J-M., Garcia, D., Hillary, R., Jardim, E., Mardle, S., Pastoors, M.A., Poos, J.J., Scott, F. and Scott, R.D., 2007. FLR: an open-source framework for the evaluation and development of management strategies. *ICES Journal of Marine Science* 64: 640-646

Porch, C.E., 1997. A user's manual for VPA-2BOX Version 2.0.

Powers, J.E., 2005. Maximum Sustainable Yield and Bycatch Minimization ''to the Extent Practicable''. *North American Journal of Fisheries Management* 25, 785-790.

Ravier, C., Fromentin, J.-M., 2001. Long-term fluctuations in the Eastern Atlantic and Mediterranean bluefin tuna population. *ICES Journal of Marine Science* 58, 1299-1317.

Ravier, C., Fromentin, J.-M., 2002. Eastern Atlantic bluefin tuna: what we learnt from historical time-series of trap catches. Collect. Vol. Sci. Pap. ICCAT, 54(2): 507-516.

Ravier, C., Fromentin, J.-M., 2004. Are the long-term fluctuations in Atlantic bluefin tuna (*Thunnus thynnus*) population related to environmental changes? *Fisheries Oceanography* 13, 145-160.

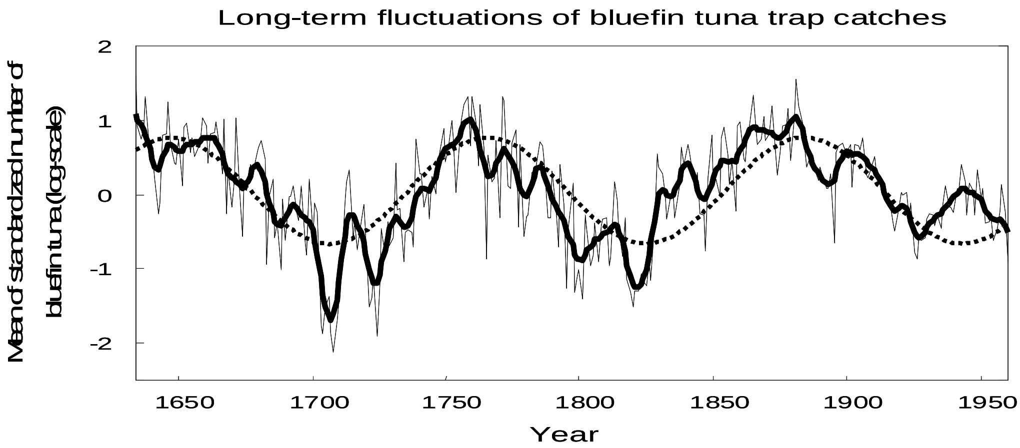

Figure 1. Historical catch time-series from Atlantic and Mediterranean trap fisheries (thin line) with the smoothing time series (bold line, after Ravier and Fromentin, 2001) and fitted sign wave (dotted line).

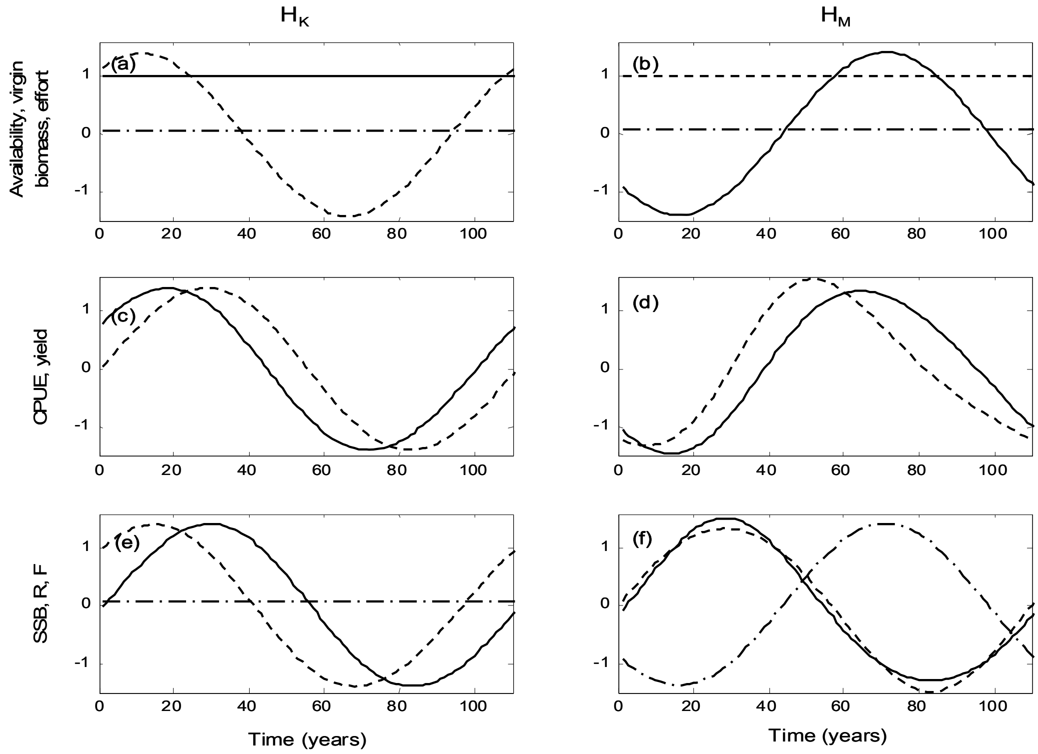

Figure 2. Outputs from the Operating Model (OM) for a full cycle (110 years) with constant fishing mortality (F) equivalent to F at maximum sustainable yield (F_{MSY}) and a steepness of 0.75. (a) Standardized time-series of availability (thick line), virgin biomass (broken line) and effort (dotted line) for the carrying capacity hypothesis (H_K); (b) same as (a) for the migration hypothesis (H_M); (c) standardized time-series of capture per unit effort (CPUE, thick line) and yield (broken line) for H_K; (d) same as (c) for H_M; (e) standardized time-series of spawning stock biomass (thick line), recruitment (broken line), F (dotted line) for H_K; (f) same as (e) for H_M.

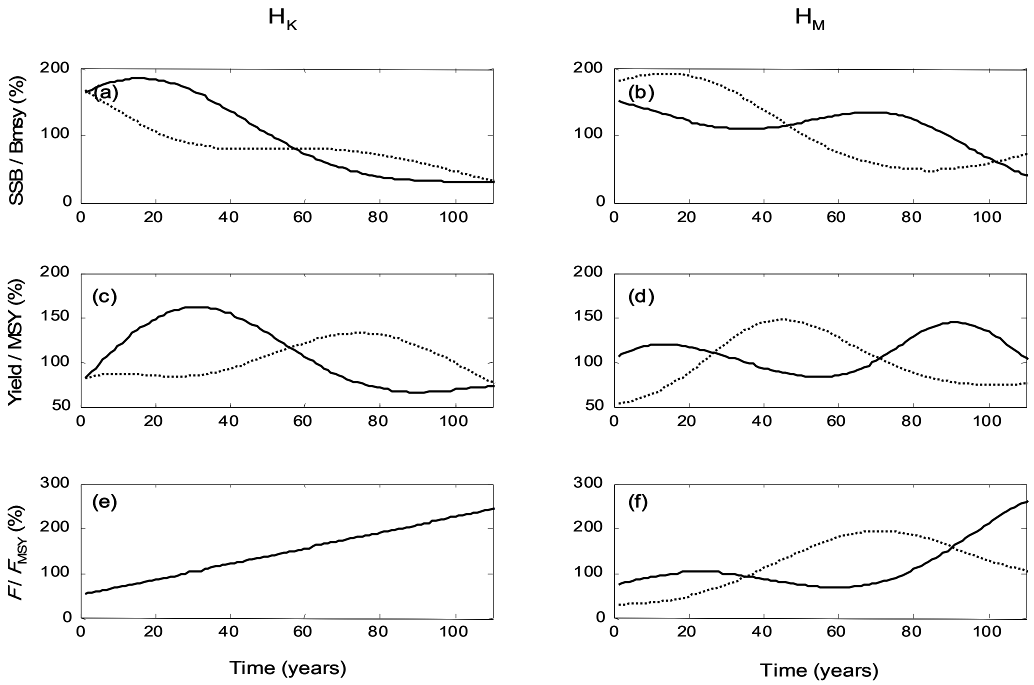

Figure 3. Outputs from the Operating Model (OM) for a full cycle (110 years) with steepness of 0.75, increasing fishing mortality (F, from 50% F at maximum sustainable yield, F_{MSY}, to 250%F_{MSY}) and two different starting points, i.e. starting at high (thick line) or low (broken line) carrying capacity or availability. (a) Time-series of spawning stock biomass (SSB) relative to spawning stock biomass at MSY (B_{MSY}) under carrying capacity hypothesis (H_K); (b) SSB under the migration hypothesis (H_M); (c) time-series of yield (relative to MSY) under H_K; (d) yield under H_M; (e) time-series of F (relative to F_{MSY}) under H_K; (f) F under H_M.

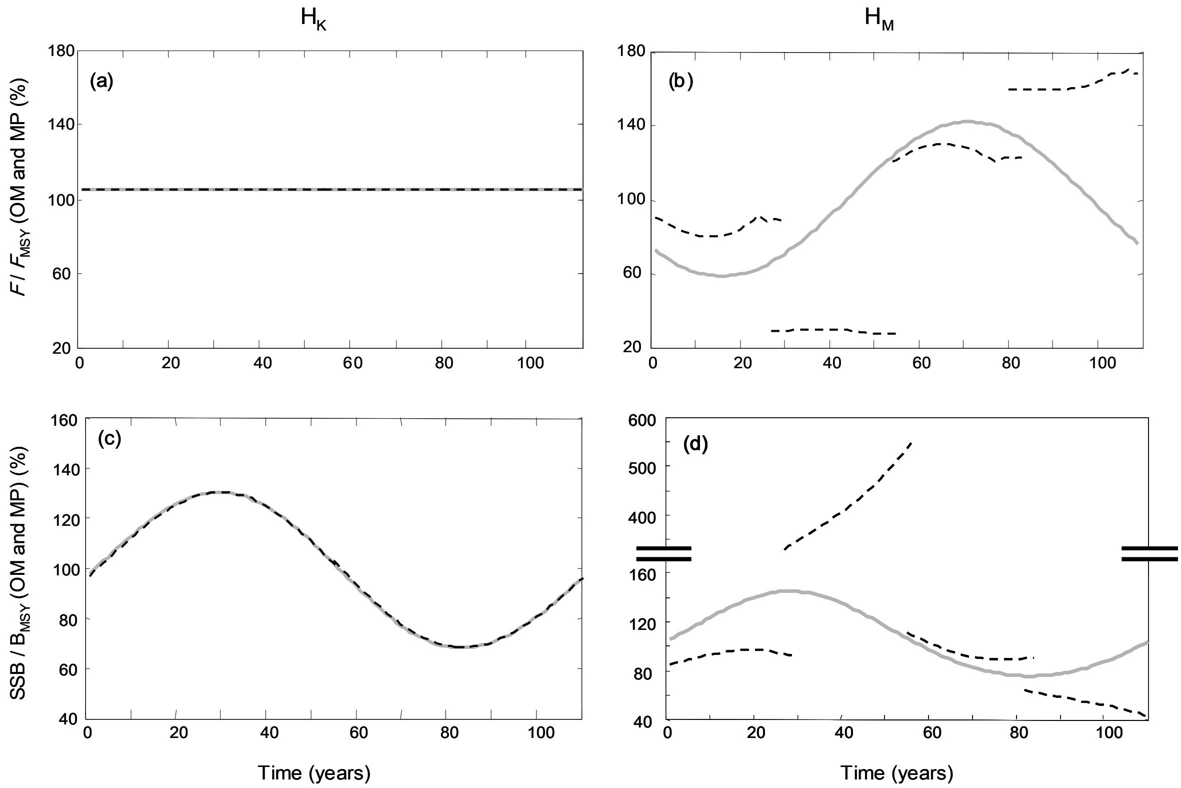

Figure 4. Fishing mortality (F) and spawning stock biomass (SSB) from the Operating Model (thick grey line) and estimated by the virtual population analysis (VPA, dotted line) for both carrying capacity hypothesis ((a) and (c)) and migration hypothesis ((b) and (d)), considering four different starting points (scenarios with constant historical F = F at maximum sustainable yield, F_{MSY}, and steepness = 0.75).

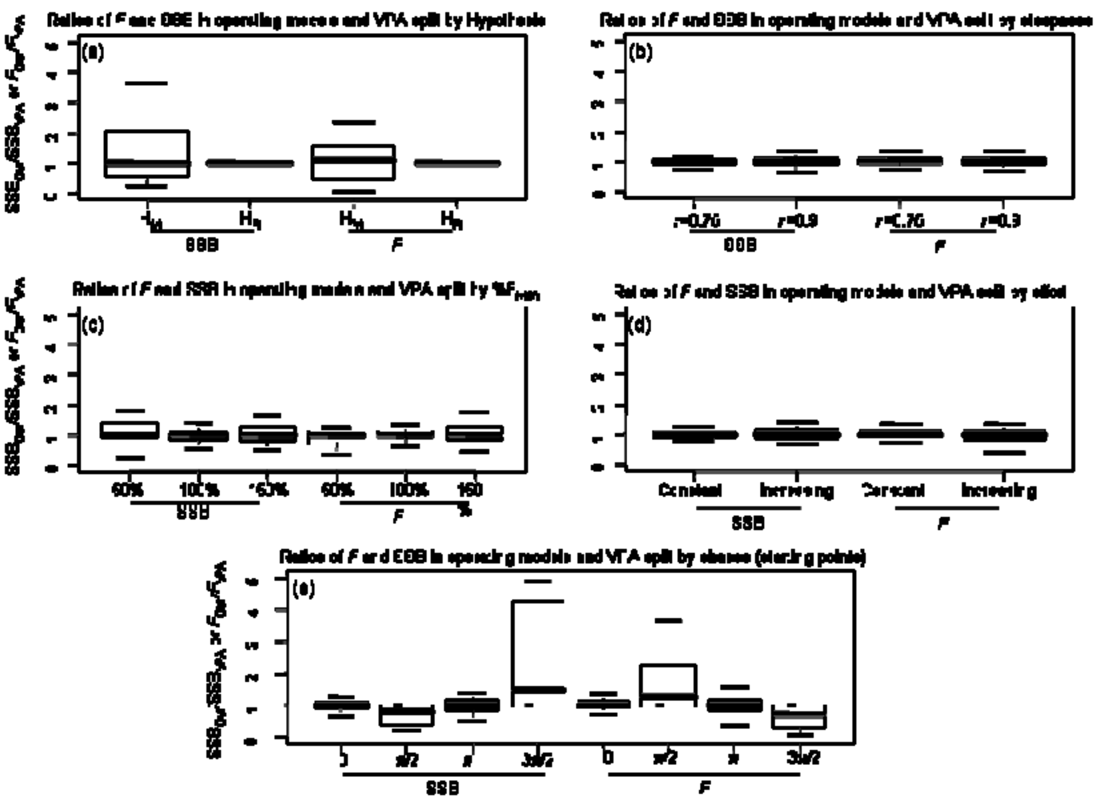

Figure 5. Boxplots of the ratios of spawning stock biomass (SSB) and fishing mortality (F) in Operating Models (OM) and virtual population analysis (VPA) for all the scenarios (ratios close to 1 indicate no bias or a good performance of the VPA). (a) scenarios split by hypothesis (carrying capacity, H_K, versus migration hypothesis, H_M); (b) split by steepness (0.75 and 0.9); (c) split by historical fishing mortality intensity (i.e., 50% of F at maximum sustainable yield (F_{MSY}), 100%F_{MSY}, 150%F_{MSY}); (d) split by constant versus increasing effort and (e) split by starting points (i.e. phases between exploitation and natural long-term cycle).

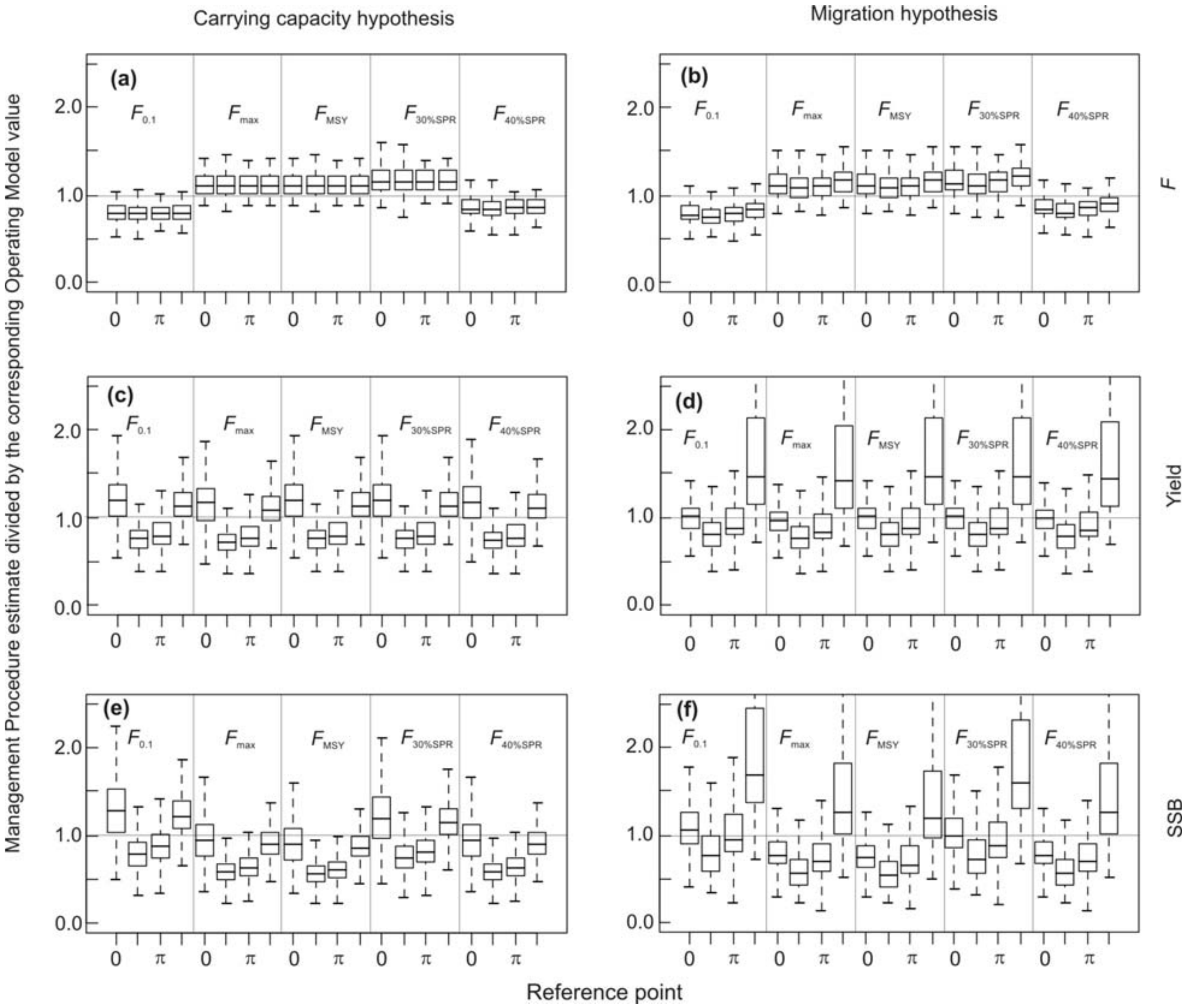

Figure 6. A comparison of Maximum sustainable yield (MSY) based biological reference points (BRP) estimated by the Management Procedure and divided by the corresponding true (i.e., OM) values of F_{MSY}, MSY and B_{MSY}. (a) BRP based on F under the carrying capacity hypothesis (H_K); (b) BRP based on F under the migration hypothesis (H_M); (c) BRP based on yield under H_K; (d) BRP based on yield under H_M; (e) BRP based on spawning stock biomass (SSB) under H_K; (f) BRP based on SSB under H_M.

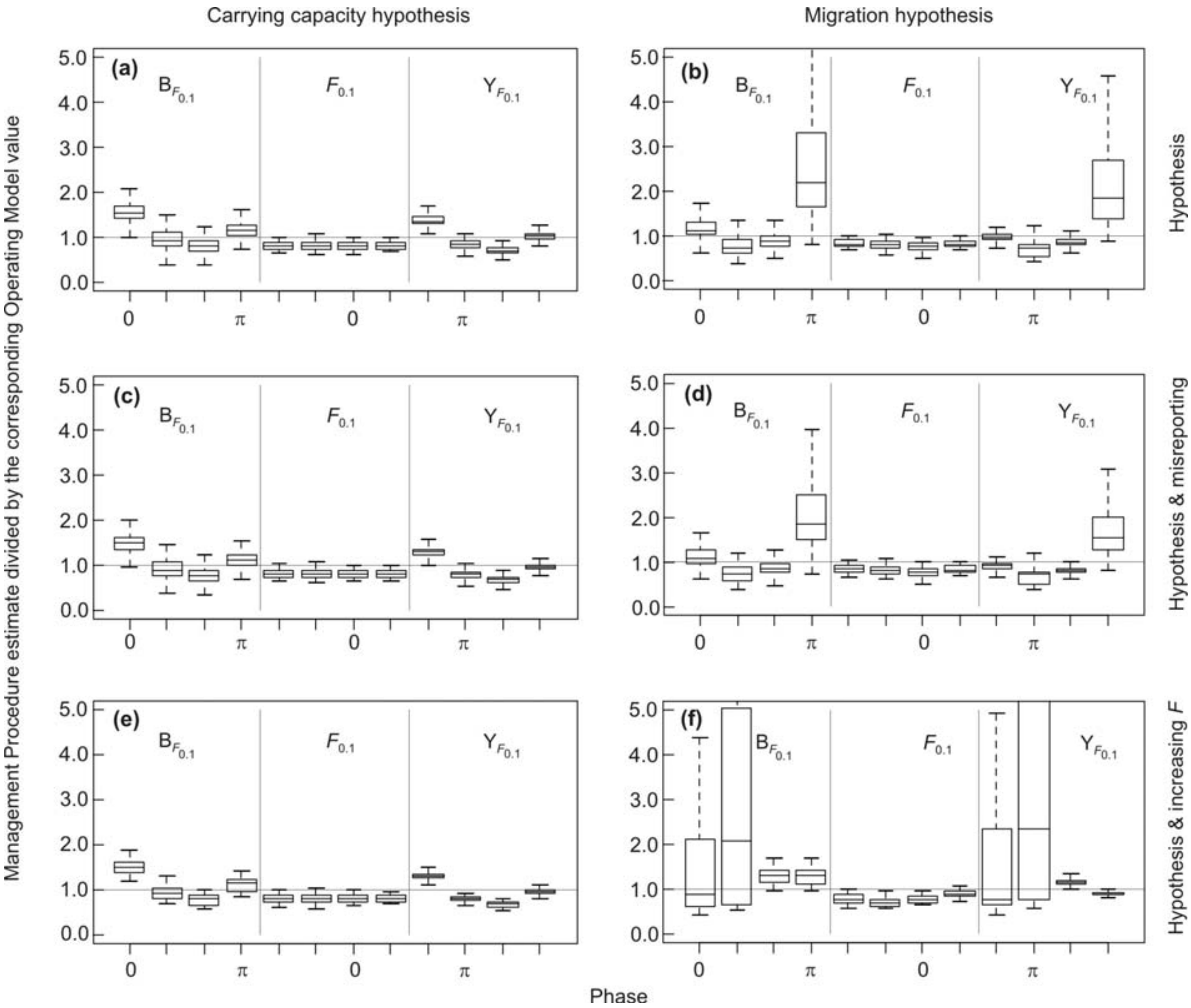

Figure 7. A comparison of $F_{0.1}$-based reference points (estimated by the Management Procedure) divided by the corresponding true (i.e., OM) F_{MSY}. (a) all scenarios with constant historical Fs under the carrying capacity hypothesis (H_K); (b) all scenarios with constant historical Fs under the migration hypothesis (H_M); (c) same as (a) with misreporting; (d) same as (b) with misreporting; (e) all scenarios with increasing historical Fs under the H_K; (f) all scenarios with increasing historical Fs under the H_M.

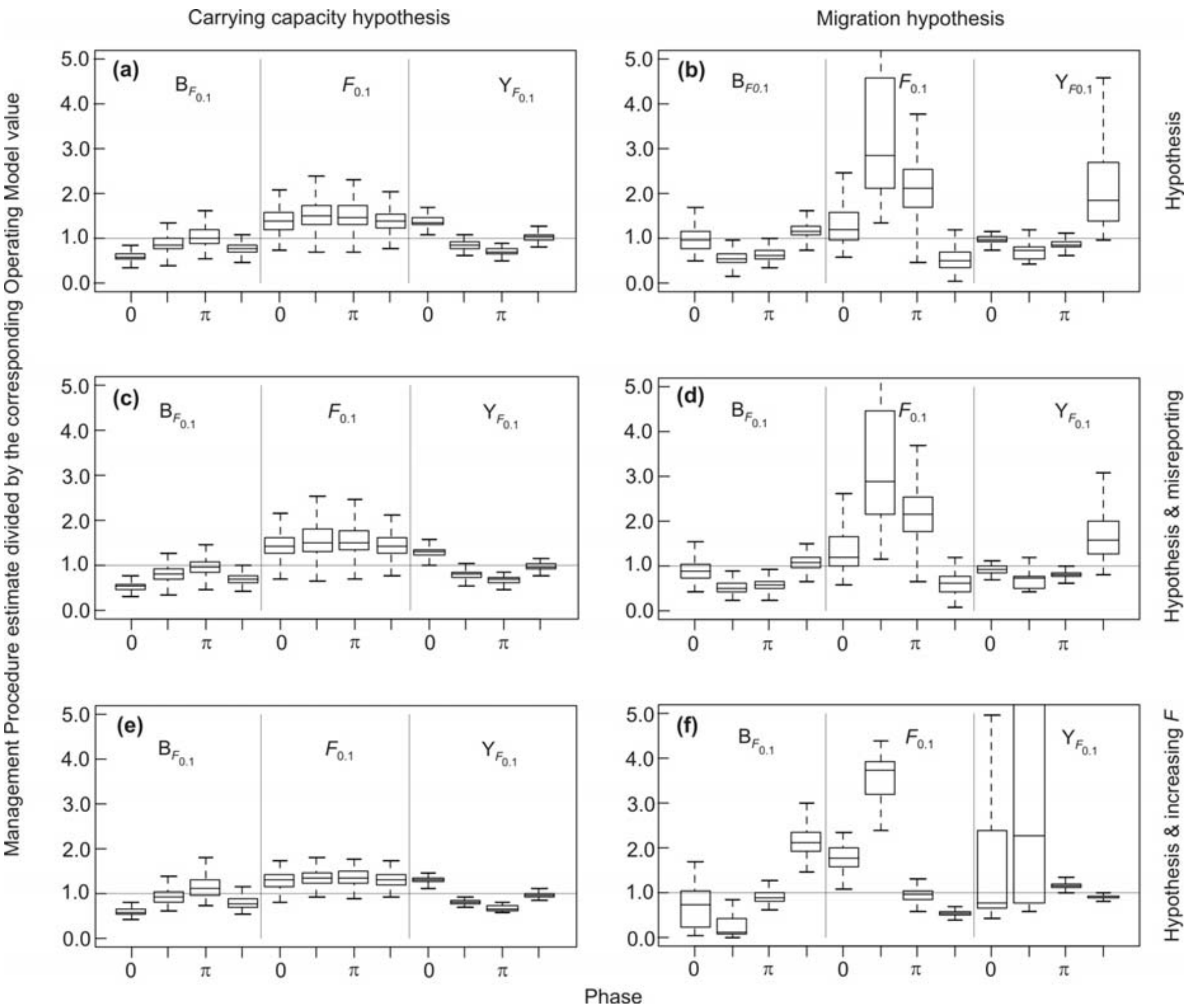

Figure 8. A comparison of the ratios of F, SSB, and Yield to their corresponding $F_{0.1}$-based reference points divided by the corresponding ratios from the Operating Model) e.g., $\frac{\overline{F}:F_{0.1}}{F:F_{MSY}}$ ($\overline{F}$ being the mean value of F of the MP over the past five years). (a) all scenarios with constant historical Fs under the carrying capacity hypothesis (H_K); (b) all scenarios with constant historical Fs under the migration hypothesis (H_M); (c) same as (a) with misreporting; (d) same as (b) with misreporting; (e) all scenarios with increasing historical Fs under the H_K; (f) all scenarios with increasing historical Fs under the H_M.

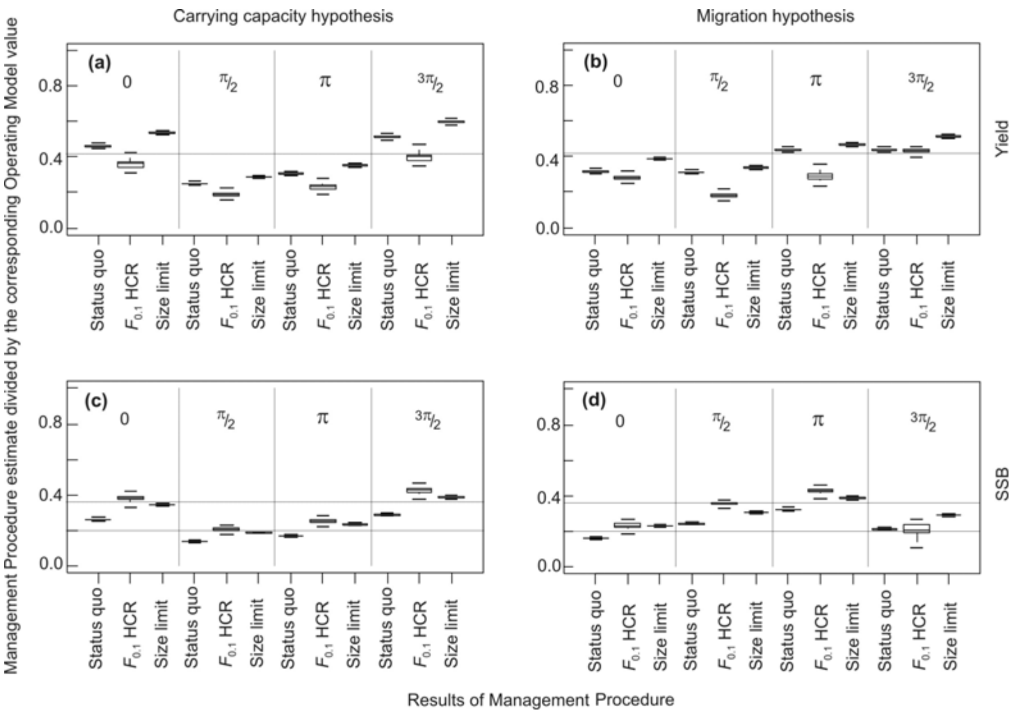

Figure 9. Comparison of a Management Procedure (MP) based on a harvest control rule based on $F_{0.1}$ ($F_{0.1}$ HCR) with a MP based on size limit regulation (size limit), when considering: constant historical Fs equivalent to 150%F_{MSY}. Status Quo assumes no regulation. Yield or spawning stock biomass (SSB) of each MP is given relative to the maximum possible yield and to virgin biomass, respectively. The maximum yield is found by harvesting all fish when production attributable to growth becomes less than that lost to natural mortality; for Atlantic bluefin this is at age 13 (under H_M, MSY is 17 500, but the maximum yield is 42 500). (a) Comparison of MPs in terms of yield under the carrying capacity hypothesis (H_K); (b) same as (a) under the migration hypothesis (H_M); (c) comparison of MP in terms of SSB under the carrying capacity hypothesis (H_K); (d) same as (c) under the migration hypothesis (H_M).