Incorporating various Gulf of Mexico Integrated Ecosystem Assessment products into the Stock Synthesis Integrated Assessment Model framework

Michael J. Schirripa, Richard D. Methot, et al.

SEDAR33-DW10

6 May 2013

Please cite as:

Schirripa, M.J., R.D. Methot, et al. 2013. Incorporating various Gulf of Mexico Integrated Ecosystem Assessment products into the Stock Synthesis Integrated Assessment Model framework. SEDAR33-DW10. SEDAR, North Charleston, SC. 17 pp.

Incorporating various Gulf of Mexico Integrated Ecosystem Assessment products Into the Stock Synthesis Integrated Assessment Model framework

Michael J. Schirripa, Richard Methot,

&

SEDAR 33 IEA Working Group

Abstract

This paper is part of a set that represent the first attempt of the Gulf of Mexico (GOM) Integrated Ecosystem Assessment (IEA) to work with its state and academic partners to integrate ecosystem level information into a single species stock assessment. The specific purpose of this work is demonstrate how various ecosystem indicators and environmental covariates produces via the GOM IEA process can be brought into the single species assessment modeling process and to provide guidance as how to most correctly incorporate them into the single species assessment models, namely Stock Synthesis (SS). Several methods are outlined for deviations in both recruitment deviations as well as natural mortality. The important assumptions regarding the Ecopath-Ecosim model used to generate estimates of historic natural mortality are discussed.

Introduction

The Gulf of Mexico Integrated Ecosystem Assessment (GOM IEA) Program is producing several ecosystem indicators and products for use in SEDAR 33. One of the more quantitative ways these products can be used is to directly account for some of the variation in the data presented to the single species assessment model, error that would otherwise be viewed by the model as random. By explicitly accounting for ecosystem processes in the assessment model we give the model the ability to account for variations in key life history events other than changes solely in spawning stock biomass or fishing mortality.

It has long been accepted that environmental factors play a major role in accounting for year-to-year variability in exploited fish stocks (Hjort, 1914; Cushing, 1982). These environmental factors can include changes in air and ocean temperatures (Chavez et al., 2003), upwelling (Ware and McFarlane, 1995), or timing of the spring transition (Logerwell et al., 2003). One of the main ways that the environment influences fish population dynamics is by modulating annual recruitment, usually in the form of young of the year survival. Explicitly including environmental variables underlying this modulation into stock assessments can help determine whether changes in recruitment are a result of changes internal (i.e. parental stock size or spawning–stock biomass) or external (i.e. changes in recruit survivorship) to the population structure. At least three situations can benefit from the inclusion of environmental data: (i) environmental variability causes a large deviations in recruitment, but conventional fishery and survey data are not adequate to capture this variability clearly, so including environmental data helps the model estimate the correct time-series of recruitment; (ii) fish recruit to the fishery at a young age, but there are no surveys of young fish to estimate the recent levels of recruitment to improve forecasts; (iii) there is a long-term signal in the environment that affects recruitment, but this trend is confounded with a one-way decrement in the spawning biomass. In this regard, including environmental data can decrease the variance in parameter estimation and help to determine the true stock–recruitment relationship and subsequent management benchmarks.

One of the most elusive yet most important parameters of any age-structured stock assessment is the rate of annual rate of natural mortality (*M*). *M* is average rate at which a fish would die in the absence of fishing. However, because this rate is nearly impossible to measure in the field, it is most often assumed to be constant over time, and its level is usually based on life history correlates or the oldest fish of that species ever observed. While it is recognized that even though a constant rate is assumed, there is still annual variation in this rate that is usually too difficult to quantify. In this situation the assumption of a constant rate may not be misplaced. However, if in reality there is a trend in *M*, or episodic events that can cause spikes in this rate that are not properly accounted for in the stock assessment model of that species, this could have significant implications regarding the estimated status of the stock relative to management benchmarks.

METHODS

All methods described here will pertain to the Stock Synthesis (SS) Integrated Assessment Modeling framework (Methot and Wetzel, 2013). An important reference document to accompany this document is the SS User's Manual, which can be found on the NOAA Fisheries Toolbox webpage (http://nft.nefsc.noaa.gov/downloads/SSv3.24f-Documentation.zip). The first analysis that should be considered before working with the environmental data is to determine the behavior of the targeted deviations (either recruitment of natural mortality for this paper) without the influence of the environmental data in the model. The patterns that emerge will then be the result of all other observational data in the model (CPUE, length/age compositions, etc), which can be highly informative.

When using environmental covariate to modulate recruitment deviations it is important to demonstrate that these deviations are significantly related to a chosen environmental index. An important initial exploratory analysis will extract the deviations the model and use them as the dependent variable in a statistical analysis conducted outside the assessment model. Then there are two methods for using the environmental index in the SS model. It needs to be kept in mind however that when using the "model method" (described below) SS assumes a linear relation between the deviations and the environmental covariate. However, when using the "data method" (described below) there is increased flexibility in this assumption.

The other approach that can be used to determine if the deviation are significantly related to an environmental covariate is a more classical hypothesis testing approach. Without the environmental data in the model, the null hypothesis is that the deviations and the environmental data should have a slope equal to zero. If, on the other hand, a significant relation does exists then the slope parameter estimated within the SS model should be significantly different than zero as well as reduce the negative log-likehood by at least two or more units.

Testing the significance of an environmental covariate with natural mortality is not as straight forward as there is no separate slope parameter estimated within the SS model. In this case, one is generally left with estimating the natural mortality deviations in the absence of the environmental data and then applying a statistical test (i.e. linear regression) externally to test for a relation. Because both the independent and dependent variables have observation error associated with them, a function regression may be the best model choice.

Recruitment

The standard approach to include environmental data into population models has been through the addition of a parameter to the standard stock–recruitment function, which allows recruit survival to deviate annually from the mean levels predicted by the simpler function (Hilborn andWalters, 1992, p. 285). Because this approach modifies the structure of the existing stock–recruitment model by adding a parameter, this method is referred to here as the model method. The shortcoming of the model method is that it relies on the unrealistic assumption that the environmental index is measured without error.

Furthermore, because this method accounts for the portion of the overall recruitment standard deviation caused by the environment through the additional parameter in the modified stock–recruitment function, the environmentally caused deviation no longer contributes to the overall recruitment standard deviation, usually an assumed (i.e. not estimated) value in the stock assessment model. This necessitates that the assumed overall recruitment standard deviation value be reduced to reflect variation attributable only to forces other than the environment. However, because the level of the overall recruitment standard deviation scales the log-bias adjustment (so that the expected arithmetic mean recruitment is equal to the mean from the stock–recruitment function), this reduced value causes an incorrect log-bias adjustment for the estimates of both the annual as well as the virgin recruitment values.

The Model Method. The model method directly adjusts the level of recruitment expected from the stock–recruitment function as

$$\hat{R}_t = f\,(\mathrm{SSB}_t) \times \exp(\beta E_t),$$

where β is the (slope) parameter relating the environmental time-series (E_t) to the recruitment deviation. For years where recruitment residuals were estimated, the level of total recruitment was given by

$$R_t = \hat{R}_t \exp(-0.5\sigma_R^2) \exp(\tilde{R}_t),$$

where σ_R is the standard deviation for recruitment in log space, and ~R_t is the lognormal recruit deviation in year *y*. Note that because a portion of the deviation in recruitment is being accounted for by an environmental covariate, when using this method the overall sigma-r term in the model should be reduced to reflect this accounting.

The Data Method. Another approach that seeks to overcome the shortcoming of the model method is to use the environmental time-series in the same manner that an age-0 survey is used. With this approach, the environmental data are considered an index of recruitment variability and, as such, are used to tune the time-series of annual recruitment deviations from the fitted stock–recruitment curve (Brandon et al., 2007). Because the environmental time series is fitted as part of the stock-assessment-model objective function, and hence contributes to the total maximum likelihood component, this approach is referred to as the data method. Similar to a survey, the data method allows the environmental data to have annual observation error associated with it and, unlike with the model method, missing years are treated only as missing years of data. The environmental effect is assumed to occur after any density-dependence on recruitment has taken effect.

The data method treats the environmental data as if it were a survey of annual recruitment deviations. This approach is similar to using the environmental index as if it were a survey of age-0 recruitment abundance. By focusing on the fit to the deviations, it removes the effect of spawning biomass on recruitment. In this method, the likelihood of the deviations are expressed as

$$\text{Likelihood} = 0.5 \sum_t \left(\frac{\ln(E_t) - \ln(\hat{D}_t)}{\sigma_t} \right)^2,$$

where E_t is the environmental observation at time *t*, D_t the deviation from the fitted stock–recruitment function at time *t*, and σ_t the standard deviation of the observation error of the environmental time-series.

Stock Synthesis Coding Methodology

The section below should be considered as a supplement to the SS Users Manual, not a standalone explanation. It is written to include some content that may not be specifically pointed out in the manual. As such, it should be considered along with the manual and not as a replacement for it. For further considerations on the two methods, see Schirripa et al. 2009.

Model Method. When using the model method recruit deviations are modeled as a strict function of an environmental input (i.e. no error). The recruitment deviation data will go into the SS data file in the environmental data section. Before entering the environmental data, it should first be transformed into z-scores (z-score = (observation - mean of all observations)/SD all observations). This will ensure the values are centered on zero.

Next, an environmental link parameter needs to be turned on in the Spawner-Recruitment section of the control file. This parameter creates a multiplicative adjustment to the target parameter, so $P_y = P * \exp(env\_link * env\_data_y)$. The environmental link function should be the same number as the environmental time series the user wishes to associate with the deviations. Make sure that the link parameter is associated with recruitment deviations rather than spawning stock biomass (also an option). Because no true observational data is added to the model, but one more parameter is added, the model method lends itself well to hypothesis testing. That is, if the environmental observations and the addition of the one parameter do in fact account for more of the unexplained variation, this can be quantified via the change negative log likelihood of the overall model fit. When using this method, the higher the value of the estimated environmental link parameter, the more sensitive the recruitment deviations are to the environmental data.

Data Method. When using the data method recruit deviations are modeled the same way that any other CPUE time series, only instead of the CPUE time series indexing biomass, the environmental data indexes recruit deviations. In this way, the environmental data is input into the data file as a survey with a year and standard deviation association. It is not necessary to associate and estimate a catchability coefficient (q) for the environmental "survey" unless a non-linear relation between the recruitment deviations and the environmental data is desired. It is, however, necessary to specify that the environmental data is indexing recruit deviation by setting the size selectivity for this survey equal to one of the "special selectivity" patterns (Table 1).

Natural Mortality

The Ecopath-with-Ecosim (EwE) is one method in which to generate estimates of natural mortality. Several considerations are necessary in order to ensure that the estimates of *M* from EwE are properly aligned in time with those of SS. Time varying estimates of instantaneous, per year, *M* are an emergent property of the EwE that can be regulated by two processes: (1) non-predation ('other') mortality (M0), a catch-all rate including all mortality not elsewhere included, and is internally computed from:

$$MO_i = P_i * \frac{(1 - EE_i)}{B_i}$$

Where *P* is the total production rate of group *i*, *EE* is the 'ecotrophic efficiency', and *B* represents the biomass; and (2) predation mortality (*M2*, including any cannibalism), which serves to link the predators and prey as:

$$M2_i = \sum_{j=1}^{n} \frac{Q_j * DC_{ji}}{B_i}$$

where the summation is over all *n* predator groups *j* feeding on group *i*, Q_j is the total consumption rate for group *j*, and DC_{ji} is the fraction of predator *j*'s diet contributed by prey *i*. Q_j is calculated as the product of B_j, the biomass of group *j* and $(Q/B)_j$, the consumption/biomass ratio for group *j*.

Predator-Prey Definitions. The first consideration we address is the predator-prey matrix specified for the target species. It is important to determine if the EwE model specifies whether each stanza of the species has predators or not. If there are specified predators, then top down events and process will affect annual variations in *M*; if not, we would expect bottom-up process only to drive annual variations. It is important to ensure that assumptions in the ecosystem model match those known to occur in reality and of the SS model. If no predators are present on the target species we should not expect large changes in natural mortality.

Compensatory Natural Mortality. A second consideration is the EwE representation of compensatory natural mortality. Compensatory juvenile mortality is represented through changes in Z for juvenile stanzas associated with changes in foraging time and predator abundances. Compensatory changes in natural mortality rate (M) can be simulated by combining two effects: non-zero feeding time adjustment, and either high EE from Ecopath or high proportion of 'other' mortality being sensitive to changes in predator feeding time. With these settings, especially when vulnerabilities of prey to a group are low, decreases in biomass lead to reduced feeding time, which leads to proportional reduction in natural mortality rate (Christensen and Walters, 2004).

Reproduction. In EwE fecundity is assumed proportional to body weight above a weight at maturity, and size-numbers-dependent monthly egg production is used to predict changes in recruitment rates of age 0 fish. Compensatory juvenile mortality is represented through changes in Z for juvenile stanzas associated with changes in foraging time and predator abundances, as in split-group calculations (Christensen and Walters, 2004).

Top-Down versus Bottom-Up Control. Natural mortality can come in the form of either top down, bottom up, or both combined. Top-down (consumer) regulatory processes are said to occur when, for example, predators keep prey populations at levels below the population size that would be observed in the absence of predators. If on the other hand, factors such as food and/or habitat availability are the main drivers explaining population fluctuations, a population is said to be regulated by bottom-up (resource) regulatory processes. This distinction is important when attempting to understand the results.

In the original Ecosim formulations (Walters et al., 1997, 2000) the consumption rate for a given predator feeding on a prey was thus predicted from the effective search rate for predator–prey specific interactions, base vulnerabilities expressing the rate with which prey move between being vulnerable and not vulnerable, prey biomass, predator abundance (numbers for split pool groups as discussed later, and biomasses for other groups). The model as implemented implies that 'top-down versus bottom-up' control is in fact a continuum, where low v's implies bottom-up and high v's top-down control. The input vulnerability rates (υij) in EwE are scaled to range from 0 to 1, with 0.3 serving as default for mixed control, and 0 implying bottom-up, 1 implying top-down control. The actual vulnerabilities (vij) used in the computations are rescaled from the entered υij's as: vij = exp[2 × (exp(υij) − 1)] (Christensen and Walters, 2004).

Stock Synthesis Coding Methodology

Model Methods. Allowing a parameter, such as natural mortality, in SS to deviate freely on an annual basis is quite straight forward. Any parameter with a "long parameter line" (see SS Manual) can be allowed to deviate annually. Driving these deviations will be the sum total of all observational data within the model. To allow a parameter to freely deviate on an annual basis one first defines the type of deviation that is desired in element number nine in that parameter line ("use dev"). There are three numerical options (1) multiplicative, (2) additive, and (3) random walk additive. To allow for a fully freed-up estimate of an annual deviation one ensures that element number 8 in the parameter line is set to "0" (indicating NOT to associate it with any environmental data input). The next and final step is define the start (element #10) and end (element #11) years the deviations will be estimated. Results of the fitted deviations can be viewed in the Report.sso file.

Figure 1 uses data from the previous gag assessment to demonstrate how annual deviations in natural mortality would manifest if allowed to deviate freely and in the absence of any environmental covariates. Even without including environmental data, signs of increased natural mortality can be seen in adult gag in 2005, which coincides with a strong red tide even that year.

The second method of allowing a parameter to deviate is by the blocking of years. With this option the user can define years in which deviations are expected to be similar such that only one deviation need be estimated for the entire block. Blocks can be 1 year to nyears-1 in size. When using blocks, as with any time varying option, the deviation can be estimated via multiplicative, additive, or simple replacement.

The third option to allow a parameter to deviate is to use a smoothed function. SS has the option of two smoothed functions. The first is a random walk function in which Nyears-1 parameters are estimated and a standard deviation is defined. The defined standard deviation dictates how large of a deviation the “walk” is allowed to stray from the base parameter value. The second smooth function is a logistic function in which three additional parameters are estimated (the width, the inflection, and offset). To use the smoothed function one must define the environmental data time series identification number to be fit using element 8 in the parameter line.

Modified Data Method. It is not possible at present in SS to vary natural mortality using data in the same manner as is done with recruitment. However, a modified data method can be designed which can achieve similar results. The first step in designing the modified data method is to create a set of one-year blocks that encompass the time period of the environmental data (n_blocks = n_years) and to use this block design for the natural mortality parameter. Then, each year’s “new” natural mortality value in the custom set up is given a Bayesian prior and standard deviation equal to the environmental data value and standard deviation. In this way, the new annual value of natural mortality will be indexed by the environmental data, as in the data method, but not as strictly dictated by the data, as in the model method. This method is superior to the using a smoothed function if natural mortality is believed to be driven by episodic events, such as red tides or the inadvertent release of contaminants, rather than by more auto-correlated events, such as low frequency ecosystem regime shifts.

DISCUSSION

Perhaps one of the most controversial aspects of including environmental data into a single species model is the notion that inclusion of the data will result in a better model fit. This is not necessarily the case. Great consideration must be given to the structure of the model and the auxiliary data and potentially competing observation data before such a determination can be made. For instance, by definition when using the data method, the additional data can only result in a higher likelihood (unless the data is fit mathematically perfect). The consequences of the environmental process should already be imbedded in the rest of the observational data; so generally speaking, indexing these processes should not necessarily improve the model fit. This is especially true if one regresses the free deviations against an environmental covariate, and then inserts this relationship back into the model. The inclusion of environmental data is often used to verify the signal that already exists in the data. This is especially true with parameters that are routinely allowed to be estimated, such as recruitment deviations. In the case of recruitment deviations the key improvement will come by allowing the model another way to explain variation in recruitment other than changes in spawning stock biomass. In the case of natural mortality, the model is given a means to account for observed changes in stock size that are not necessarily accompanied by higher landings during the same time period. changes However, in the case of natural mortality, a parameter that is routinely held constant, we may well expect to see an improvement in fit if that parameter is allowed to be freely estimated year to year. This might even be expected whether or not an environmental driver is included or not because one is adding parameters to the model rather than observation data.

There are ways though that the inclusion of environmental data can improve the model in general. For instance, given a satisfactory fit to the environmental driver, one might expect to see smaller standard deviations on some of the estimated parameters and/or derived quantities, resulting in higher precision about those estimates. But perhaps the most useful manner that environmental data can improve the assessment is to provide observational data outside the bounds of fishery dependent data. For instance, if the environmental driver in question has been recorded during a time period before reliable length compositional data was collected, recruitment deviations may well be able to be estimated much further back in time. Conversely, if the environmental driver data can be collected in real time, then estimates of recruitment deviations in the current assessment year, or years before the recruits are seen in the fishery data, maybe very useful in forecasts of the future condition of that stock. Environmental regime shifts are another critical way that environmental covariates can be of great help. If a dramatic drop in recruitment values are observed and the model is given no other means to account for this drop, then the model is left to attempt to reduce its estimates of spawning stock biomass to account for the change, which may well not be the case.

REFERENCES

Chrsitensen, V. and C.J. Walters. 2004. Ecopath with Ecosim: methods, capabilities and limitations. Ecological Modeling, 172: 109-139.

Chavez, F. P., Ryan, J., Lluch-Cota, S. E., and Niquen, C. M. 2003. From anchovies to sardines and back: multidecadal changes in the Pacific Ocean. Science, 299: 217–221.

Cushing, D. 1982. Climate and Fisheries. Academic Press, London. 373 pp.

Hjort, J. 1914. Fluctuations in the great fisheries of northern Europe viewed in the light of biological research. Rapports et Proce`s-Verbaux des Re´unions du Conseil Permanent International pour l'Exploration de la Mer, 20: 1–228.

Logerwell, E. A., Mantua, N., Lawson, P. W., Francis, R. C., and Agostini, V. N. 2003. Tracking environmental processes in the coastal zone for understanding and predicting Oregon coho (Oncorhynchus kisutch) marine survival. Fisheries Oceanography, 12: 554–568.

Methot Jr., R.D. and Wetzel, C. 2013. Stock Synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fisheries Research 142: 86-99.

Schirripa, Michael J., C. Phillip Goodyear, and Richard M. Methot. "Testing different methods of incorporating climate data into the assessment of US West Coast sablefish." ICES Journal of Marine Science: Journal du Conseil 66.7 (2009): 1605-1613.

Ware, D. M., and McFarlane, G. A. 1995. Climate-induced changes in Pacific hake (*Merluccius productus*) abundance and pelagic com- munity interactions in the Vancouver Island upwelling system. Canadian Special Publication of Fisheries and Aquatic Sciences, 121: 509–521.

Table 1. Selectivity options for recruitment deviations when using the data method (from *SS Users Manual*).

SPECIAL SELECTIVITY		
Pattern	N Parameters	Description
30	0	Sets expected survey abundance equal to spawning biomass (population fecundity)
31	0	Sets expected survey abundance equal to exp(recruitment deviation). This is useful if environmental data is used as an index of recruitment variability.
32	0	Sets expected survey abundance equal to exp(recruitment deviation) * SpawnBiomass. So this is recruitment without density-dependence (for pre-recruit survey) because most ecological logic places the density-dependent stage during the juvenile period following the larval stage that is most sensitive to environmental perturbation.
33	0	Sets expected survey abundance equal to age 0 recruitment.
Do not input any size/age comp for surveys using pattern 30-33. The “catchability” coefficient for these selectivity patterns 30-33 have all the general properties of the catchability coefficient for real surveys, e.g. they can be time-varying, use power relationship, etc.		

Table 2. Long parameter line options for natural mortality estiamtes (from *SS Users Manual*).

ENV	A positive value, *g*, causes SS to set the annual working value of this parameter equal to a multiplicative function of Environmental Variable *g*: $parm'(y) = parm * exp(link * env(y,g))$ A negative value, *g*, causes SS to set the annual working value of this parameter equal to a additive function of Environmental Variable *g*: $parm'(y) = parm + link * env(y,-g)$ Where, *link* is the environmental link parameter, parm is the base parameter being adjusted, parm' is the value after adjustment, and *env(y,-g) is* the value of the environmental input g in year y. SS counts the number of parameters that invoke use of an Environmental Variable. After SS finishes reading the section's parameter lines, it then creates/reads additional short parameter line(s) to set up the link parameters. If custom=0, then one short parameter line is used to define the min, max, init, etc, for each of the link parameters. If custom=1, then a separate line is read for each.
USE_Dev	A value of 1 invokes multiplicative: $parm'(y) = parm * exp(dev(y))$ A value of 2 invokes additive: $parm'(y) = parm + dev(y)$ A value of 3 invokes additive random walk: $parm'(y) = parm'(y-1) + dev(y)$ The vector of devs is simply a vector of offsets, there is no inherent sum to zero constraint. However, the fact that they are each penalized by the DEV std.dev. below will tend to make them sum towards 0.0.
DEV min yr	Beginning year for the dev vector for this parameter
DEV max yr	Ending year for the dev vector for this parameter
DEV std.dev.	Standard deviation for elements in the dev vector for this parameter
USE-BLOCK	Block: A positive value identifies which block pattern will be used for time changes to a parameter value. Block patterns are simply numbered sequentially as they are defined near the top of the control file, so the index here must be correct for the order in which they are defined. More than one parameter can use the same block definition. The order of generated block parameters is by the order of the parameters that call for creation of the block parameters, then by the order of the blocks within that pattern. Trend: A negative value for the Use_Block input causes SS to create a parameter time trend instead of blocks. This time trend requires 3 parameters (instead of the normal one parameter per block). The base parameter is the value for the adjusted parameter in year = start year. For subsequent years, the three parameters define a normal distribution of change over time: P1: parameter value for year = end year. Either as logistic offset from base P (if Use_Block=-1), or as direct usage (if Use_Block=-2) P2: inflection year; if HI value for the base parameter is >1.1, then use as year, else use as fraction of range styr - endyr

Table 2 (cont.)

BLOCK-TYPE	This selects the way in which the block parameter creates an offset from the base parameter. 0 means that parm' = baseparm * exp(blockparm) 1 means that parm' = baseparm + blockparm 2 means that parm' = blockparm 3 means that parm' = is additive offset from previous block. Note that blocks must be contiguous to use this option properly.

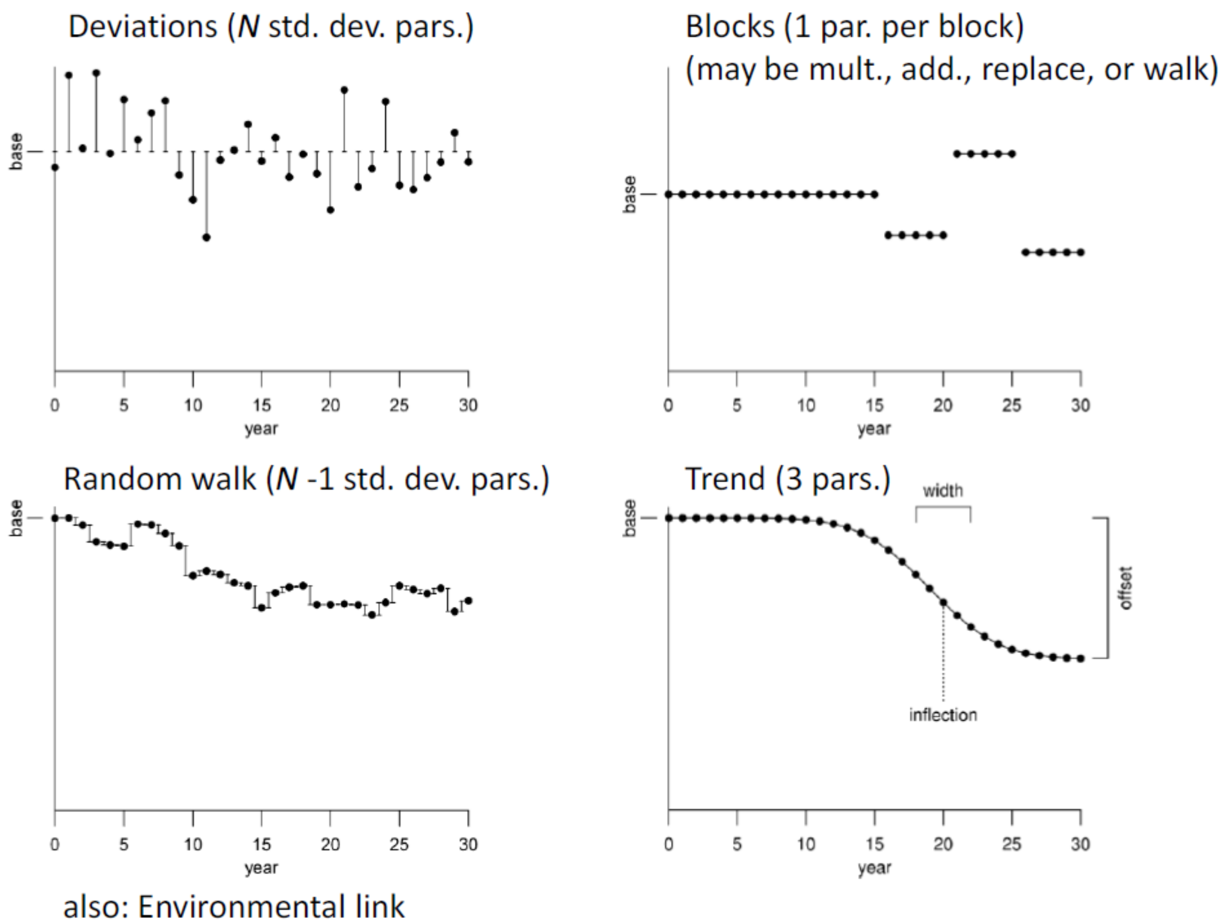
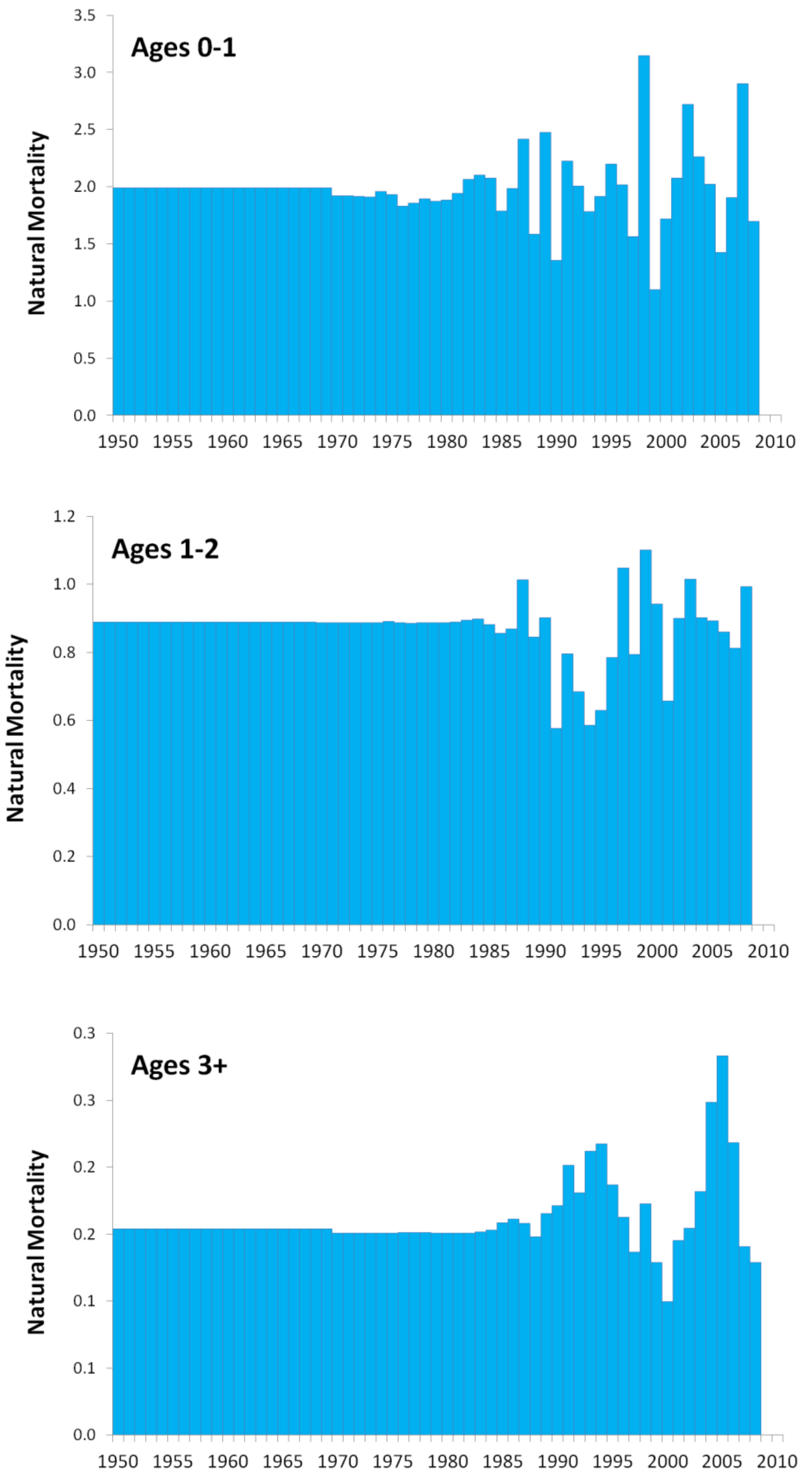


Figure 1. *Options for time-varying parameters in Stock Synthesis for recruitment deviations (upper left) and natural mortality (upper right and lower left and lower right)*



22

Figure 2. *Estimates of natural mortality (M) for gag for three life stanzas in the absence of any environmental data when M is allowed to deviate freely on an annual basis, 1970-2008.*