Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets

Mark N. Maunder[1,2] and Kevin R. Piner[3]

[1]Inter-American Tropical Tuna Commission, 8901 La Jolla Shores Drive, La Jolla, CA 92037-1508, USA

[2]Center for the Advancement of Population Assessment Methodology, Scripps Institution of Oceanography, La Jolla, USA

[3]NOAA Fisheries, Southwest Fisheries Science Center, 8901 La Jolla Shores Dr, La Jolla, CA 92037-1508, USA

Abstract

Contemporary fisheries stock assessment models often use multiple diverse data sets to extract as much information as possible about all model processes. This has led to the mindset that integrated models can compensate for lack of good data (e.g. surveys and catch-at-age). However, models are, by definition, simplifications of reality, and model misspecification can cause degradation of results when including additional data sets. The process, observation, and sampling components of the model must all be approximately correct to minimize biased results. Unfortunately, even the basic processes that we assume we understand well (e.g. growth and selectivity) are misspecified in most, if not all, stock assessments. These misspecified processes, in combination with composition data, result in biased estimates of absolute abundance and abundance trends, which are often evident as "data conflicts". This is further compounded by over-weighting of composition data in many assessments from misuse of data weighting approaches. The law of conflicting data states that since data is true, conflicting data implies model misspecification, but needs to be interpreted in the context of random sampling error, and down weighting or dropping conflicting data is not necessarily appropriate because it may not resolve the model misspecification. Data sets could be analyzed outside the integrated model and the resulting parameter estimates for population processes and their uncertainty used in the integrated model (e.g. as a prior), but these analyses typically involve more uninformed assumptions, implicit or explicit, that are potentially misspecified leading to biased results. Model misspecification and process variation can be accounted for in the variance parameters of the likelihoods (observation error), but it is unclear when this is appropriate. The appropriate method to deal with data conflicts depends on if it is caused by random sampling error, process variation, observation model misspecification, or system dynamics model misspecification. Diagnostic approaches are urgently needed to test goodness of fit and identify model misspecification. We recommend external estimation of the sampling error variance used in likelihood functions, including process variation in the integrated model, and internal estimation of the process error variance. The

required statistical framework is computationally intensive, but practical approximations are available, computational algorithms are being improved, and computer power is always increasing.

Introduction

The overarching goals of fisheries management have historically been the optimization of yield and sustainability of the stock. This goal led to the concept of maximum sustainable yield (MSY) and its proliferation throughout the fisheries literature (Larkin 1977; Punt and Smith 2001). Contemporary fisheries management objectives are much more complex (Hilborn and Walters 1992), but often the main objectives relate back to optimizing benefits to humanity (e.g. yield or profit) on a sustainable basis while mitigating adverse effects (e.g. bycatch), which parallels MSY to some extent. Attainment of these fisheries goals have been judged via ever more complex quantitative analysis.

Fisheries stock assessments (Hilborn and Walters 1992; Quinn and Deriso 1999, Haddon 2001) have been the gold standard in determining MSY and are conducted for most economically valuable species that have the appropriate data available. With increasing computer power and the popularization of integrated stock assessment modeling (Maunder et al. 2009), the complexity of modern stock assessment modeling is rapidly increasing (e.g. Hampton and Fournier 1998; Methot and Wetzel 2013; Bull et al. 2005; Begley 2004). Considerable research is going towards improving the methods used to generate model predictions (Quinn 2003; Maunder and Piner 2014) as well as the methods used to statistically compare predictions to data (Punt and Hilborn 1997; Maunder 2011; Francis 2013) and estimation algorithms (e.g. Skuag and Fournier 2006; Fournier et al 2012; Kristensen et al. 2014). The complexity of modern stock assessment models has been driven in part by the desire to incorporate a broad range of data types collected with different sampling assumptions (Maunder and Punt 2013; Punt et al. 2013).

A major issue in contemporary stock assessment modelling is the apparent conflicting signals from different data sets given the models structure. Although, conflict can and does occur between all data types, conflict between indices of relative abundance and composition data is particularly prevalent and concerning (Francis 2011; Lee et al. 2014). Conflicts between data, which are often a symptom of model misspecification and evident as model misfit, can affect the estimates of important parameters and derived quantities. The solution to data conflict often is the eliminating of one of the conflicting data sources, or nearly the equivalent, its statistical down-weighting in the model (e.g. Sharma et al. 2014), but this is dealing with the symptoms and not the underlying cause (Wang et al. 2015).

In this paper we argue that the practice of data elimination does not address the more important issue highlighted by internal conflict in the models. As we will discuss in detail, data conflict may be indicative of system dynamics model misspecification. System dynamics control the population dynamics and misspecification of these important model processes will lead to biased estimates of the population dynamics and resulting management information. We highlight how composition data often are the cause of conflict and the different ways both observation and process error, as well as model structure misspecification, can lead to conflict between composition data and indices of abundance. We offer recommendations on building a stock assessment that systematically identifies root cause of data conflicts and how to solve them.

Causes of data conflict

Data conflict occurs when two different data sets, given the model structure, provide information about a model state or process that disagrees. For example, when an index of relative abundance supports a high abundance while catch-at-length data supports a low abundance, the data are in conflict about abundance. At first glance it might be easy to conclude one of the two data sources is not true, however, if the initial examination and development of data series were good, data should generally be considered true. Therefore, conflicts in the data imply that either the model is misspecifed causing the conflict (Table 1) or the precision of the data has been overstated leading to a false impression of data conflict.

Apparent data conflict in modern integrated stock assessment models can occur for three reasons 1) random sampling error, 2) misspecification of the observation model (model processes relating dynamics or states to data), and 3) misspecification of the system dynamics model (the population dynamics model). Random sampling error is variation due to taking a sample rather than a census of the population and can be decreased by increasing the sample size. This differs from error in the observation model for which larger sample size generally will not reduce the error. Data conflicts have to be interpreted in the context of the random sampling error. The differences in information content apparent in the data sets might just be a consequence of one or more of the data sets having a low sample size (and therefore a large confidence interval around the estimate) so that the differences are caused by random sampling error and the data are consistent in that they can come from the same underlying system dynamics. Therefore, it is important to get the assumptions about the random sampling error correct to interpret apparent conflicts in the data. For example, data from an index of relative abundance may be inconsistent with the underlying population dynamics model if the error is assumed to be multiplicative, but consistent with the model if it is assumed to be additive (Figure 1). Of course, large observation errors also mean that the data themselves are not informative about model process.

The observation model describes the relationship between the data and the underlying system dynamics model. An example of an observation model is the proportional relationship, represented by the catchability coefficient, between an index of relative abundance and the absolute abundance represented by the population dynamics model. The observed relative abundance may differ from the true relative abundance due to random sampling error or because the catchability changes (Figure 2). Another example of an observation model misspecification is using the wrong component (e.g. age structure) of the population to represent an index of relative abundance. For example, abundance trends from a relative index that represents spawners may not be the same as that from an index that represents recruits (Figure 3). Misspecification of the observation model is important because misfit to data linked by the observation model process may result in biased estimates of important system dynamic process (Piner et al. 2011).

Systems dynamics model misspecification occurs when the underlying processes governing the population dynamics are not correctly specified. Extending the example of indices of abundance from recruits and spawners to system dynamic errors, Figure 3 shows that if recruitment is assumed to be proportional to spawners (low steepness of the Beverton-Holt stock-recruitment relationship) then the indices are expected to show similar trends, but if recruitment is independent of spawners (high steepness of the Beverton-Holt stock-recruitment relationship) then the index representing recruitment is likely to differ from that representing spawners. Misspecification in the observation or system dynamics models could occur because the wrong value of a fixed parameter is used, the wrong model structure is used, or because the model processes (e.g. as represented by a parameter value) are assumed to be time invariant when they actually change over time. For example, systematic changes in the environment may cause recruitment trends to differ from the abundance trends causing data conflict (Figure 4).

System and observation processes that affect composition data

Maunder and Piner (2014) use the estimation of absolute abundance from an index of relative abundance and catch or from composition data to illustrate how stock assessments rely on knowledge of growth, recruitment, natural mortality, selectivity, and sampling processes. But then go on to explain how these are poorly known for many, if not most, species. In the following paragraphs we focus on model processes that affect the observed size/age composition observed. Although indices of abundance (relative or absolute) are affected by model process, data conflicts are often traced to issues of misfit to some form of composition data (Francis 2011; Lee et al. 2014). In addition to providing information on recruitment and selectivity, composition data in integrated models provide information on abundance trends and absolute abundance (Maunder and Piner 2014). However, the information on abundance from composition data often is too informative because of the complex system and observations process related to that data results in misfit that greatly affects parameter estimates (Francis 2011; Lee et al. 2014).

Because unmodeled temporal variation in biological (a system model process) and fishery processes (either a system model process, an observation model process, or both) can lead to misfit and its resulting misinformation on abundance, it is important to understand why composition data may change over time to identify which process errors should be modeled. Spatial distribution and recruitment are likely to show considerable temporal variability, but other processes such as selectivity, fecundity, weight-at-length, growth (in length), survival, and maturity may vary as well. We start with the biological processes influencing population age/size structure and then discuss the fisheries processes that influence the age/size structure observed.

Recruitment is perhaps the most important biological process because it is one of the most variable. The strength of a cohort is usually determined at a very young age with inter-annual variability often on a scale of orders of magnitude. The relative strength of a cohort can persist for several years and may be observed in multiple years of composition data; therefore relative recruitment strength should be estimated reasonably well in integrated assessment models. Age composition data with low aging error is much more informative than length composition data about recruitment because cohort specific modes in length composition data are often obscured because lengths from multiple cohorts overlap, particularly at older ages. In some applications, strong cohorts observed in age data or modes in length composition data are not consistent from one year to the next due to other processes as described below. Reliable estimates of relative recruitment strength may also be dependent upon correct specification of the spawner-recruitment process, as well as other model processes relating to the composition data.

Natural mortality affects the population composition by altering the descending slope of the numbers at age curve. The higher the natural mortality rate the steeper the descending slope. The majority of stock assessments assume natural mortality to be constant over age and time despite growing evidence this is unlikely to be true (Vetter 1988). Few data sources provide direct information on natural mortality making it difficult to estimate (Lee et al 2011). Mispecification of the magnitude of natural mortality or how it varies with age is likely to bias the predicted composition. Temporal variation in natural mortality will likely influence the underlying population composition resulting in misfit if not included in the model. This effect may be observed when initially strong cohorts (as seen at young ages) are not strong at latter ages due to changes in natural mortality. In general, survival of younger fish is more likely to be impacted by changes in the environment or predation which are the major sources of variation in natural mortality. When fish are captured at a relatively large size, these impacts will occur at an age before the fish are vulnerable to the fishery and are incorporated into the definition of recruitment to the fishery, which may minimizes it impact on the model. However, for small-sized species such as sardine and anchovies, variability in natural mortality of adults may impact the underlying population composition and thus the catch composition impacting the model fits. In addition to life history, the selectivity of the gear (e.g. scientific surveys catching small fish) will influence the impact of variation in natural mortality on model fit.

Variation in both the rate of growth and the average maximum size is common (Thorson and Minte-Vera in press), affecting the size structure of the population. Most assessment models that fit to size composition data assume that length-at-age is normally distributed and does not change over time (Francis in press). Thus, any changes in growth often lead to a misfit to the composition data unless compensated for by erroneously estimating other population processes. These changes in growth may be density dependent, and when coupled with size selective gear, bias early estimates of cohort strength. It is generally thought that temporal variation in length-at-age is most problematic for young fish, but the differences may persist to adulthood. Weight-at-age is generally more variable than length at age (Thorson and Minte-Vera in press) and is particularly important when fitting to weight composition data. It is worth noting that growth is important beyond the issue of data conflict and fitting to composition data. Catch is often recorded in biomass, but numbers underlie the stock assessment model (Francis in press). Therefore, specification of the appropriate growth is required to remove the correct number fish from the population. Age-length data provide direct information on growth so it is easier to estimate if these data are available and in some cases using empirical weight at age may be appropriate (Taylor et al. submitted; Kuriyama et al. submitted). Maunder et al. (2015) recommend that when there is an adequate amount of growth information available, estimation of annual variation in the asymptotic length should be the default assumption using the growth increment approach (see Methot and Wetzel 2013). However, appropriate approaches are needed to ensure fish do not decrease in size as they grow over time. It might be reasonable to fix the level of the temporal variation at a value based on meta-analysis of data-rich stock assessments. Sex-specific growth should be considered because it may be of magnitude to substantially impact the assessment model.

Of all the fisheries processes which influence the age or size of fish we measure, selectivity of the gear is perhaps the most crucial as it is the model process directly linking estimated dynamics to composition data. Selectivity as represented in stock assessment models includes both contact selectivity (e. g. the probability that a fish goes through a gill net without being caught) and population selectivity (also known and availability; e.g. spatial distribution) (Sampson 2014). In general, the contact selectivity of the gear will not change over time unless the characteristics of the gear changes (e. g. mesh size or fishing depth) or if the gear performs differently in different environments. Availability is influenced by the spatial distribution of the fishing fleet and the stock (both horizontally and vertically). Research has shown that even if the contact selectivity is asymptotic, dome shape selectivity can result if there is limited movement between sub populations and/or differences in exploitation rates (Sampson and Scott

2011) and Butterworth et al. (2014) found dome shape selectivity to be common in application. Sampson and Scott's (2011) results are based on population selectivity where all fisheries are combined (e.g. as used in a VPA). Waterhouse et al. (2014) obtained the same results for fisheries defined by area (e.g. as used in integrated models). Beyond the shape of the selectivity process, selectivity is often assumed to be time-invariant. However, the same spatial differences in exploitation may result in selection patterns with temporal variation (Sampson 2014; Waterhouse et al. 2014). Cohort targeting also implies temporal changes in selection (Stewart and Martell 2014). Whatever the cause, assuming too inflexible a selectivity pattern will result in misfit to the composition data (Piner et al. 2011) and ultimately data conflict (Lee et al. 2014; Ichinokawa et al. 2014; Wang et al. 2014). Maunder et al. (2014) suggest that it is prudent to model dome-shaped and time-varying selectivity for all fisheries using nonparametric methods (e.g. Thorson and Taylor 2014; Nielsen and Berg 2014), particularly if a survey with constant asymptotic selectivity is available. But they also caution that it is not clear whether selectivity can adequately account for spatial structure in the population age composition (e.g. Hurtado-Ferro et al. 2014).

Beyond the shape or temporal variability of gear selectivity, it may be important to know if selectivity is length-based, age-based or potentially both. Temporal variation in growth rates can cause temporal variation in selectivity if selectivity is length-based, but modelled as age-based (Stewart and Martell 2014; Crone and Valero 2014). Most age-based stock assessment models assume that variation of length-at-age is normally distributed and length-based removals do not change the population distributions. However, very selective fisheries with high fishing mortality (e.g. some crab and lobster fisheries) can distort the length-at-age distribution (Punt et al. 2013; Punt et al. in press) causing model misspecification and data conflict. Variation in growth can interact with the sampling processes for age composition data when the selectivity is length based (Francis in press). For example, smaller fish of a given age might escape a trawl or gillnet by passing through the mesh. Therefore, a cohort that has slower growth might have lower selectivity (or availability) at a given age changing the sampled age composition. Alternatively, if selectivity/availability is age based as might be expected with ontogenetic shifts in spatial distribution, the implied size based selectivity may change with changes in growth influencing the length composition (Francis in press).

Beyond gear selection, the variability and the absolute magnitude of fishing mortality can impact the underlying stock composition. Low to moderate levels of fishing mortality should only alter the underlying population composition slowly over time and with relatively constant selection those changes will have a similar impact on similar aged fish. In contrast, cohort targeting might have a larger impact reducing the abundance of the targeted cohort relative to the other cohorts. High fishing mortality that changes markedly from year to year, particularly in short-lived species or fisheries that target a few cohorts/ages, may alter the population composition. However, changes in fishing mortality that impact the underlying population structure is likely to have minor or long term trends, and be a much smaller impact than selectivity impacts on the sampling of the population's composition. Catch data is included in stock assessment models so the impact of fishing mortality on composition data through modification of the population composition should be adequately estimated in the model as long as the absolute abundance is well estimated.

Sampling error also contributes to the variability in observed composition data. Sampling error arises because not all fish caught are measured or aged. The proportion of fish aged or measured varies substantially among stocks are can vary over time. When the sample size is low, a different (hypothetical) sample can produce a different catch composition. Even if the sample size is large enough that the random error should be small, correlation in the fish measured (e g. fish caught in the same school are similar lengths) may cause high errors (pseudo-replication) and more temporal variation in the composition data than expected (Francis 2011). Therefore, the actual sample size should not be used as the effective sample size for most stocks. The calculation of the error (e. g. the effective sample size) should take the sampling design and data correlation (Francis 2011; Francis 2014; Francis et al. in press) into consideration. For example, a two stage bootstrap that resamples the set and the fish in the set might be appropriate to estimate sampling error.

Diagnosing data conflict

It could be argued that all model diagnostics are dealing with symptoms associated with lack of fit and resulting data conflict. Diagnostics that identify misfit to data likelihood components or which data components are in conflict can be used as a starting place to identify what is the potential problem. However, as we explained above, in integrated models the root cause of the data conflict may not be the model process directly linking the data to the dynamics because all data components are linked via the underlying population dynamics.

Residual analysis is perhaps the most basic method of assessing goodness of fit. Overly large or temporal trends in residuals are indicative of lack of fit and model misspecification. The distribution of residuals (e.g. the distributional form and the variance) should be consistent with the sampling error (likelihood) assumptions. The distribution of expected likelihood values can be derived using artificial data based on the model assumptions and each data component compared to that obtained from the observed data. Any model misspecification including fixed parameters, model structure, and assumed sampling distributions (likelihood functions) can result in the likelihood values from the observed data falling outside the distribution from the simulated data. Initial application of this method for stock assessment models has been problematic, but a similar approach has recently been proposed for wildlife applications (Besbeas and Morgan 2014).

Retrospective analysis is another method of detecting bias and model misspecification (Hurtado-Ferro et al. 2014). By itself it does not identify the causes of the retrospective pattern, although assessment authors often alter data or model process until the retrospective pattern disappears. Alternative diagnostics looking at model self-consistency have also been developed that diagnose misspecification and potentially the data component that is most affected (Piner et al. 2011).

Likelihood profiling of individual data components across a parameter (e.g. average recruitment, which scales recruitment) provides one of the best diagnostic to evaluate the influence of data associated with model structure on estimated dynamics (Maunder 1998; Maunder and Starr 2001; Francis 2011; Lee et al. 2014; Ichinokawaa et al. 2014). For example, it has been shown using the likelihood profiling diagnostic that there was no data conflict about virgin recruitment (the absolute biomass scaling parameter) when the model is correctly specified, but un-modeled temporal variation in selectivity or a mispecificed selectivity curve can cause data conflict (Wang et al. 2014). However, despite identifying which data components are in conflict, in many cases it may not be possible to identify the true cause of model conflict. More recently a production model diagnostic (Maunder and Piner 2014) has been proposed to evaluate data conflict. This diagnostic evaluates data conflict in information about both absolute abundance and abundance trends.

Dealing with data conflict

There are various methods that have been or could be used to deal with data conflicts. The most common method is to eliminate conflicting data sets or perhaps a more complex "down weighting" of one or more of the conflicting data sets. Eliminating data is simply eliminating its contribution to the total model fit used to estimate model parameters. Dropping conflicting data may result in the use of alternative models (e.g. different parameter values) conditioned on the different competing data. The results of these competing models may then be combined in some form of model averaging to arrive at a single management result or a distribution of management results (e.g. Schnute and Hilborn 1993).

In contrast to the elimination of data, down weighting is achieved by changing the variance parameter (e.g. increasing the standard deviation in a normal distribution based likelihood or decreasing the sample size in a multinomial based likelihood). Other methods have been used such as multiplying the log-likelihood for a particular data set by a weighting factor called lambda, but we recommend against using these methods because they are ad hoc statistical methods. Down weighting may be applied by increasing the observation error for that data component until its sampling error is consistent with the fit to the data (i.e. there is apparent conflict, but it is within the range of the assumed random sampling error). The "weighting" can be done automatically by estimating the variance parameter (McAllister and Ianelli 1997; Deriso et al. 2007; Maunder 2011) and the result ("down weighting" or "up weighting") will be dependent on how consistent the data set is with the system model, itself, and the other data sets. However, the correlation among data should be taken into consideration, particularly for composition data to ensure the correct weighting is applied (Francis 2011; Francis 2014; Francis et al. in press). The extensive use of the McAllister and Ianelli (1997) approach, which does not account for the correlation, may have resulted in many stock assessments being biased.

Neither elimination nor down-weighting of data should be considered successful resolution of the of the data conflict because they may not solve the underlying problem, which is analogous to treating the symptom but not the disease (Wang et al. 2015). In down weighting data sets, the additional of observation error does not change the dynamics of the system and therefore it does not address the possible problem of model misspecification (observation or system dynamics). Eliminating data is problematic for the same reasons. Severe underweighting of data may introduce the additional problem that parameters associated with that data set (e.g. selectivity for composition data) when estimated may be distorted so that the model better fits to other less directly related data sets.

In contrast to elimination or down weighting of data, process error can be modelled explicitly rather than accounting for misspecification in the variance parameters of the likelihood functions. Typically, random process error that is assumed to come from a common distribution is modelled. Systematic process error is generally dealt with using model structure sensitivity analyses. The methodologies used to deal with random process error come under a variety of names, but most commonly referred to as random effects or state-space models. The most commonly modelled process is recruitment (Maunder and Deriso 2003), but other processes such as fishing mortality (Nielsen and Burg 2014) or survival (Millar and Meyer 2000) have been modelled. Integration across the process error to create a true (marginal) likelihood is the appropriate approach for statistical inference, but it is computationally intensive and approximations using penalized likelihood are typically used (Maunder and Deriso 2003). AD Model Builder (ADMB), the most commonly used programming language to develop complex stock assessment models, has the facility to integrate across process error (Skaug and Fournier 2006; Fournier et al. 2012) and has been used in some applications (e.g. Nielsen and Burg 2014). Template Model Builder (TMB, Kristensen et al. 2014), which is based on ADMB, but with efficient space matrix capabilities, automatic likelihood separability determination, in addition to Laplace approximation is also appropriate for these models. Commonly used general stock assessment programs (e.g. MULTIFAN-CL (Fournier et al 1998), Stock Synthesis (Methot and Wetzel 2013)) do not have the capability to integrate across the process error, however practical approximations are available (Thompson and Lauth 2012; Thorson et al. 2015). The penalized likelihood approach (Fournier and Archibald 1984; Maunder and Watters 2003) with the distributional assumption standard deviation used in the penalty fixed at a reasonable value may be adequate for some applications.

The difficulty in dealing with data conflict arises because the actual misspecified process is often unknown or the process (including variation) is not estimable. More work is needed to know if modeling a substitute but incorrect model process is better than eliminating or downweighting data or accounting for the process error in the estimation of observation error. As an example, is accounting for time varying natural mortality or growth via the fishery selectivity process a reasonable modeling approach. Almost certainly these compromises exist, as we know very little about variation in natural mortality compared to selectivity. However, we do know that misfit due to misspecification of one process can be accounted for by changes in another model process (Piner et al. 2011), but it does not necessarily improve important model results. It seems equally unlikely that accounting for misspecification with observation error will improve model results.

Recommendations

Recommendations for dealing with data conflict can be divided into those aimed at avoiding data conflict (or facilitate its interpretation) and those to diagnose and fix data conflict (Table 2 and Figure 5). Some recommendations can be made for modelling in general and some specific to stock assessment modelling.

We recommend that, whenever possible, the variance parameter for random sampling (observation) error should be estimated from the empirical data outside the stock assessment model. For example, by bootstrapping the sampling design of the composition data or estimating the variability of CPUE observations around a smoothed fit to a CPUE series (Clark and Hare 2006; Francis 2011; Lee et al. 2014). Estimating the variance parameter inside the model will confound observation error, process error, and model misspecification and make it difficult to identify model misspecification. The residuals of the fit to the data should be evaluated to determine if they match the assumptions of the random sampling distribution that is used to construct the likelihood function (e.g. does the variance of the residuals equal the sampling variation). Specific attention should be paid to systematic patterns in the residuals that may indicate model misspecification. If the apparent data conflict is simply due to random sampling variation and the likelihood function is correctly specified then the data should be retained in the model and not down weighted. Robust likelihood functions should be considered to avoid the influence of outliers, particularly for composition data (Fournier et al. 1998; Chen et al. 2003).

Conflict due to misspecification of the observation model when identified is best addressed by adding more appropriate model structure. If this is not feasible, the use of additional observation error to account for the process error not modeled can be used. Although this down weighting of the data does not use the information to the fullest, it does allow subjective interpretation of model fit to the data. As a last resort, the removal of this data eliminates the possible bias caused by the data while losing all of its potential information. Since the system model is not misspecified, this does not result in bias. However, it is difficult to know whether it is the observation model or the system model that is misspecified. The decision to drop the data is essentially a bias variance tradeoff. Special consideration should be given to situations where the data is the only source of direct information about parameters estimated in the model. Because removal of the data (or highly down weighting them) may cause the model to estimate these parameters in an unrealistic way to better fit other data. For example, in a situation where a composition data set influences estimates of absolute abundance due to selectivity misspecification, this is the only data set with direct information about selectivity of that fishery. Removing the composition data may cause the model to estimate the selectivity of this fishery to better fit other data sets unrelated to the fishery selectivity. In this case, it may be better to fix the selectivity at some reasonable value if the composition data is removed from the model, because biased selectivity may influence management advice for that fishery.

The most serious cause of data conflict and perhaps the most difficult to diagnose is mispecification of the system dynamics model. This mispecification does not only degrade the fit to the data but influences the whole dynamics of the system. Therefore, removal of the conflicting data set does not necessarily fix the problem, although it may improve diagnostics to the point that it appears it does. Similarly, down weighting the conflicting data explicitly, or by internal estimation of the variance parameter, or model averaging of multiple models created by removing different data sets, may not be appropriate. It is common to apply sensitivity analysis to evaluate the influence of different assumptions, especially of the system dynamics model. The sensitivity analysis can include changes to fixed parameter values or alternative model structure to evaluate their influence on management quantities. However, it can also be used to determine if the model fit, as measured by the likelihood, improves in general or for a specific data component.

It is also worth discussing the definition of “true” data as might be considered in the law of conflicting data. At one level it might be any data that is not fabricated. For example, catch data may not be considered true data in some cases due to substantial missreporting or underreporting. However, it is possible that some model could be developed to represent the level of misreporting. At another level it could be similar to Francis’ (2011) definition of representative data. For example, a survey may not be representative because it only partially covers the stock spatial distribution. This later definition highlights a confusion among sampling error, observation error, and process error. This is why we use sampling error to represent the uncertainty due to the fact that a census is not taken and the estimates are based on a subsample of the population and a different random subsample will produce a different estimate. The fact that environmental variation might cause a different proportion of the stock to be in the survey error each year could be attributed to observation error or process error depending on your point of view. This is often considered part of the variation in survey catchability. For example, Francis et al. (2003) estimated annual survey catchability variability to have a CV of about 0.2 and termed this process error, but suggested adding this to the CV in the likelihood, which was based on the sampling error. This is really process error in the observation model. For clarity, we recommend using the terms sampling error, observation model, and system model, and when discussing process error it should be stated if it is process error in the observation model or the system model. A better defining of errors also helps understand how the errors should be accounted for. For example, environmental influences on the proportion of the population within the survey area are likely to be correlated so modelling a random walk in the survey catchability (observation model process error) is more appropriate than simply inflating the CV of the likelihood function.

Stock assessment modelling requires more information than available for most applications. Our lack of understanding of complexity of the actual population inevitably leads to model simplification, misspecification, and resulting internal model conflict. Following the above advice can lead to the construction of an internally consistent model that adequately represents the available data and minimizes data conflict. However, because diagnosing which of many confounded model processes lead to the initial model conflict is difficult; several equally plausible models with the same data and diagnostics can likely be produced. It is these alternative, but equally well fitting, models that can form the basis of providing between-model variability to describe assessment uncertainty and used in model averaging. In routine applications, assessment scientists should strive to derive alternative, but similarly plausible models. This differs from the suggestion of Francis (2012) who suggests that conflicting data should be dropped from the analysis.

Ideally models would incorporate the correct model processes to provide reasonable representations of all data sources and eliminating misspecification. If elimination or down weighting the data is to continue, more research should be done to justify when these approaches are appropriate to deal with data conflict due to model misspecification. This includes the estimation of observation error in the assessment model (e.g. for composition data using the Francis (2011) method) because including the poorly modeled process error in with the observation model error is essentially the same as down-weighting that data component. Because isolating the model misspecification causing data conflict may be difficult, more research should focus on the use of the “wrong” model process to eliminate data conflict. However, isolating the actual cause of internal model conflict and including appropriate model structure should be the goal.

Acknowledgements

This paper benefitted from extensive discussions with Chris Francis while he was a visiting scientist at the Center for the Advancement of Population Assessment Methodology (CAPAM). This research addresses the data weighting component of the CAPAM Stock Assessment Good Practices Program. The views and opinions in this paper should not be construed as reflecting those of the author’s institutions.

References

Begley J, HowellD(2004)An overview ofGadget, theGlobally applicableArea-Disaggregated General Ecosystem Toolbox ICES CM 2004/FF:13.

Besbeas, P. and Morgan, B.J.T. 2014. Goodness-of-fit of integrated population models using calibrated simulation. Methods in Ecology and Evolution 5: 1373–1382.

Bull B, Francis RICC, Dunn A, McKenzie A, Gilbert DJ, Smith MH (2005) CASAL (C++ algorithmic stock assessment laboratory): CASAL user manual v2.07-2005/08/21. NIWA Technical Report 127.

Butterworth, D. S., Rademeyer, R. A., Brandão, A., Geromont, H. F., Johnston, S. J. 2014. Does selectivity matter? A fisheries management perspective. Fisheries Research, 158: 194-204.

Carvalho, F., Ahrens, R., Murie, D., Ponciano, J.M., Aires-da-Silva, A., Maunder, M.N., and Hazin, F. 2014. Incorporating specific change points in catchability in fisheries stock assessment models: An alternative approach applied to the blue shark (Prionace glauca) stock in the south Atlantic Ocean. Fisheries Research 154: 135-146.

Chen, Y., Jiao, Y., and Chen, L. (2003). Developing robust frequentist and Bayesian fish stock assessment methods. Fish and Fisheries, 4(2), 105-120.

Clark, W.G., and Hare, S.R. 2006. Assessment and management of Pacific halibut: data, methods, and policy. Scientific Report 83. International Pacific Halibut Commission, Seattle, Wash.

Crone, P. R., Valero, J. L. 2014. Evaluation of length- vs. age- composition data and associated selectivity assumptions used in stock assessments based on robustness of derived management quantities. Fisheries Research, 158: 165-171.

Deriso, R. B., Maunder, M. N., and Skalski, J. R. 2007. Variance estimation in integrated assessment models and its importance for hypothesis testing. Canadian Journal of Fisheries and Aquatic Sciences, 64: 187–197.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., and Sibert, J. (2012) AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software, 27: 233-249.

Francis, R.I.C.C., 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. 68, 124–1138.

Francis, R.I.C.C. (2014). Replacing the multinomial in stock assessment models: A first step. Fisheries Research, 151, 70-84.

Francis, R.I.C.C. (in press). Growth in age-structured stock assessment models. Fisheries Research. http://dx.doi.org/10.1016/j.fishres.2015.02.018

Francis, R.I.C.C., Hurst, R.J., and Renwick, J.A. 2003. Quantifying annual variation in catchability for commercial and research fishing. Fish Bull. 101: 293–304.

Francis, C., A. Aires-da-Silva, M. N. Maunder, K. M. Schaefer, D. W. Fuller. (In Press). Estimating fish growth for stock assessments using both age–length and tagging-increment data.

Goethel, D.R., Quinn II, T.J., Cadrin, S.X., 2011. Incorporating spatial structure in stock assessment: movement modeling in marine fish population dynamics. Rev. Fish. Sci. 19, 119–136.

Goodyear, P.C. 1996. Variability of fishing mortality by age: consequences for maximum sustainable yield. North American Journal of Fisheries Management 16: 8-13.

Haddon M (2001) Modelling and quantitative methods in fisheries. Chapman & Hall/CRC Press, Boca Raton

Hampton, J., Fournier, D.A., 2001. A spatially disaggregated, length-based, age structured population model of yellowfin tuna (Thunnus albacares) in the western and central Pacific Ocean. Mar. Freshwat. Res. 52, 937–963.

Hilborn, R., and Walters, C. J. 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty. Chapman and Hall, New York.

Hurtado-Ferro, F., Punt, A. E., Hill, K. T. 2014. Use of multiple selectivity patterns as a proxy for spatial structure. Fisheries Research, 158: 102-115.

Hurtado-Ferro, F., Szuwalski, C. S., Valero, J. L., Anderson, S. C., Cunningham, C. J., Johnson, K. F., Licandeo, R., McGilliard, C.R., Monnahan, C.C., Muradian, M.L., Ono, K., Vert-Pre, K.A., Whitten, A.R., and Punt, A. E. (2014). Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES Journal of Marine Science 72: 99-110.

Ichinokawaa, M., Okamura, H., Takeuchi, Y. 2014. Data conflict caused by model mis-specification of selectivity in an integrated stock assessment model and its potential effects on stock status estimation. Fisheries Research, 158: 147-157.

Kristensen, K., Thygesen, U.H., Andersen, K.H. & Beyer, J.E. (2014). Estimating spatio-temporal dynamics of size-structured populations. Canadian Journal Fisheries and Aquatic Sciences 71, 326–336. doi: 10.1139/cjfas-2013-0151

Kuriyama, P. Hurtado-Ferro, F., Ono, K., Hicks, A.C., Taylor, I.G., Licandeo, R.R., Johnson, K.F., Anderson, S.C., Monnahan, C.C., Rudd, M.B., Stawitz, C.C., Valero, J.L. (submitted) An empirical weight-at-age approach reduces estimation bias compared to modeling parametric growth in integrated, statistical stock assessment models when growth is time varying. Fisheries Research.

Larkin PA (1977) An epitaph for the concept of maximum sustainable yield. Transactions of the American Fisheries Society 106: 1–11.

Lee, H-H, Maunder, M.N., Piner, K.R., and Methot, R.D. (2011) Estimating natural mortality within a fisheries stock assessment model: an evaluation using simulation analysis based on twelve stock assessments. Fisheries Research, 109: 89–94.

Lee, H-H., Maunder, M.N., Piner, K.R., and Methot, R.D. (2012) Can steepness of the stock-recruitment relationship be estimated in fishery stock assessment models? Fisheries Research 125-126: 254-261.

Lee, H. H., Piner, K. R., Methot, R. D., Maunder, M. N. 2014. Use of likelihood profiling over a global scaling parameter to structure the population dynamics model: An example using blue marlin in the Pacific Ocean. Fisheries Research, 158: 138-146.

Maunder, M.N. (2001) Integrated Tagging and Catch-at-Age ANalysis (ITCAAN). In Spatial Processes and Management of Fish Populations, edited by G.H. Kruse,N. Bez, A. Booth, M.W. Dorn, S. Hills, R.N.

Lipcius, D. Pelletier, C. Roy, S.J. Smith, and D. Witherell, Alaska Sea Grant College Program Report No. AK-SG-01-02, University of Alaska Fairbanks, pp. 123-146.

Maunder, M.N. (2002). The relationship between fishing methods, fisheries management and the estimation of MSY. Fish and Fisheries, 3: 251-260.

Maunder, M.N. (2003) Is it time to discard the Schaefer model from the stock assessment scientist's toolbox? Fisheries Research, 61: 145-149.

Maunder, M.N. (2011) Review and evaluation of likelihood functions for composition data in stock-assessment models: Estimating the effective sample size. Fisheries Research, 109: 311–319

Maunder, M.N. and Deriso, R.B. (2003) Estimation of recruitment in catch-at-age models. Can. J. Fish. Aquat. Sci. 60: 1204-1216.

Maunder, M.N. and Piner, K.R. (2014) Contemporary fisheries stock assessment: many issues still remain. ICES Journal of marine Science doi: 10.1093/icesjms/fsu015

Maunder, M.N. and Punt A.E. (2013) A review of integrated analysis in fisheries stock assessment. Fisheries Research 142: 61– 74.

Maunder, M.N. and Watters, G.M. (2003). A general framework for integrating environmental time series into stock assessment models: model description, simulation testing, and example. Fishery Bulletin, 101: 89-99.

Maunder M.N., Schnute, J.T., and Ianelli, J. (2009) Computers in Fisheries Population Dynamics. In Megrey, B.A. and Moksness, E. (eds.). Computers in Fisheries Research. Springer, pp: 337-372.

Maunder, M. N., Crone, P. R., Valero, J. L., Semmens, B. X. 2014. Selectivity: Theory, estimation, and application in fishery stock assessment models. Fisheries Research, 158: 1-4.

Maunder, M.N., Crone, P.R., Valero, J.L., and Semmens, B. X. 2015. Growth: theory, estimation, and application in fishery stock assessment models. CAPAM Workshop Series Report 2: 55 pp.

McAllister, M.K., Ianelli, J.N., 1997. Bayesian stock assessment using catch-age data and the sampling/importance resampling algorithm. Can. J. Fish. Aquat. Sci. 54, 284–300.

Methot, R. D., and Wetzel, C. 2013. Stock Synthesis: a biological and statistical framework for fish stock assessment 557 and fishery management. Fisheries Research, 142: 86–99.

Millar, R. B., and Meyer, R. (2000). Bayesian state-space modeling of age-structured data: fitting a model is just the beginning. Canadian Journal of Fisheries and Aquatic Sciences, 57(1), 43-50.

Nielsen, A., Berg. C. W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fisheries Research, 158: 96-101.

Piner, K.R., H.H Lee,.M. N. Maunder, and R. D. Methot. (2011). A simulation-based method to determine model misspecificaton: Examples using natural mortality and population dynamics models. *Mar. Coast. Fish*.3:336-343.

Punt, A. E., and Hilborn, R. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. Reviews in Fish Biology and Fisheries, 7: 35–63.

Punt AE and Smith ADM (2001) The gospel of maximum sustainable yield in fisheries management: birth, crucifixion and reincarnation. In: Reynolds JD, Mace GM, Redford KR, and Robinson JR (eds.) Conservation of Exploited Species, pp. 41–66. Cambridge: Cambridge University Press.

Punt, A. E., Huang, T-C., and Maunder, M. N. 2013. Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusc species. ICES Journal of Marine Science, 70: 16–33.

Punt, A. E., M. Haddon, R. McGarvey. (In Press). Estimating growth within size-structured fishery stock assessments: What is the state of the art and what does the future look like?. Fisheries Research.

Quinn, T. J., II. 2003. Ruminations on the development and future of population dynamics models in fisheries. Natural Resource Modeling, 16: 341–392.

Quinn, T. J., II, and Deriso, R. B. 1999. Quantitative Fish Dynamics. Oxford University Press, New York, NY.

Sampson, D. B. 2014. Fishery selection and its relevance to stock assessment and fishery management. Fisheries Research, 158: 5-14.

Schnute, J.T., Hilborn, R., 1993. Analysis of contradictory data sources in fisheries stock assessment. Can. J. Fish. Aquat. Sci. 50, 1916–1923.

Sharma, R., Langley, A., Herrera, M., Geehan, J., Hyun, S-Y. 2014. Investigating the influence of length–frequency data on the stock assessment of Indian Ocean bigeye tuna. Fisheries Research, 158: 50-62.

Stewart, I. J., Martell, S. J. D. 2014. A historical review of selectivity approaches and retrospective patterns in the Pacific halibut stock assessment. Fisheries Research, 158: 40-49.

Taylor, I.G., Stewart, I.J., Hicks, A.C., Hamel, O.S. (Submitted) Drowning in data: empirical vs. parametric growth in an integrated stock assessment model. Fisheries Research.

Thompson, G. G., and Lauth, R. R. 2012. Assessment of the Pacific cod stock in the Eastern Bering Sea and Aleutian Islands Area. In: Plan Team for Groundfish Fisheries of the Bering Sea/Aleutian Islands (Compiler), Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands regions. North Pacific Fishery Management Council, Anchorage, AK, pp. 245–544.

Thorson, J. T., C. V. Minte-Vera. (In Press). Relative magnitude of cohort, age, and year effects on size at age of exploited marine fishes. Fisheries Research.

Thorson, J. T., Taylor, I. G. 2014. A comparison of parametric, semi-parametric, and non-parametric approaches to selectivity in age-structured assessment models. Fisheries Research, 158: 74-83.

Thorson, J. T., Hicks, A. C., and Methot, R. D. (2015). Random effect estimation of time-varying factors in Stock Synthesis. ICES Journal of Marine Science, 72(1), 178-185.

H. Skaug and D. Fournier, Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models, Comput. Stat. Data Anal. 56 (2006), pp. 699–709.

Wang, S. P., Maunder, M. N., Piner, K. R., Aires-da-Silva, A. Lee, H. H. 2014a. Evaluation of virgin recruitment profiling as a diagnostic for selectivity curve structure in integrated stock assessment models. Fisheries Research, 158: 158-164.

Wang, S. P., Maunder, M. N., Aires-da-Silva, A. 2014b. Selectivity's distortion of the production function and its influence on management advice from surplus production models. Fisheries Research, 158: 181-193.

Wang, S.P, Maunder, M.N., Nishida, T., Chen, Y.R . 2015. Influence of model misspecification, temporal changes, and data weighting in stock assessment models: Application to swordfish (Xiphias gladius) in the Indian Ocean. Fisheries Research 166: 119–128.

Waterhouse, L., Sampson, D. B., Maunder, M. Semmens, B. X. 2014. Using areas-as-fleets selectivity to model spatial fishing: Asymptotic curves are unlikely under equilibrium conditions. Fisheries Research, 158: 15-25.

Table 1. The law of conflicting data.

The law of conflicting data

Axiom

Data is true

Implication

Conflicting data implies model misspecification

Caveat

Data conflict needs to be interpreted in the context of random sampling error

Significance

Down weighting or dropping conflicting data is not necessarily appropriate because it may not resolve the model misspecification

Table 2. Modeling recommendations to minimize, identify, and correct data conflicts

Preventing data conflict

General

Estimate random sampling variance outside the stock assessment model

Use likelihood functions based on the sampling design to represent the fit to the data

Account for correlations in the data

Stock assessment

Model annual recruitment variation with a distributional penalty. Estimate the variance of the penalty if possible.

Divide data into fleets so that selectivity is relatively constant over time within a fleet

Model fishery selectivity using a flexible and time varying method

Ensure that surveys/fisheries used for developing indices of relative abundance have time invariant selectivity

Diagnose and fix data conflict

General

If the standard deviation of the residuals is not the same as the assumed standard deviation of the sampling distribution (take data correlations into consideration), then the sampling distribution assumptions are wrong, the model is misspecified, or there is unmodeled process variation.

If there is a trend in residuals then the model is misspecified or there is unmolded process variation

Apply likelihood component profiles on main model parameters to identify data conflict

Conduct model structure and parameter sensitivity analysis to identify model misspecification

Consider adding process error to model parameters

Stock assessment

Conduct parameter uncertainty on steepness of the stock-recruitment curve, natural mortality, growth rate, asymptotic length to identify model misspecification

Apply likelihood component profiles on virgin (or average) recruitment

Consider adding process error to growth if there is age-length data and to natural mortality

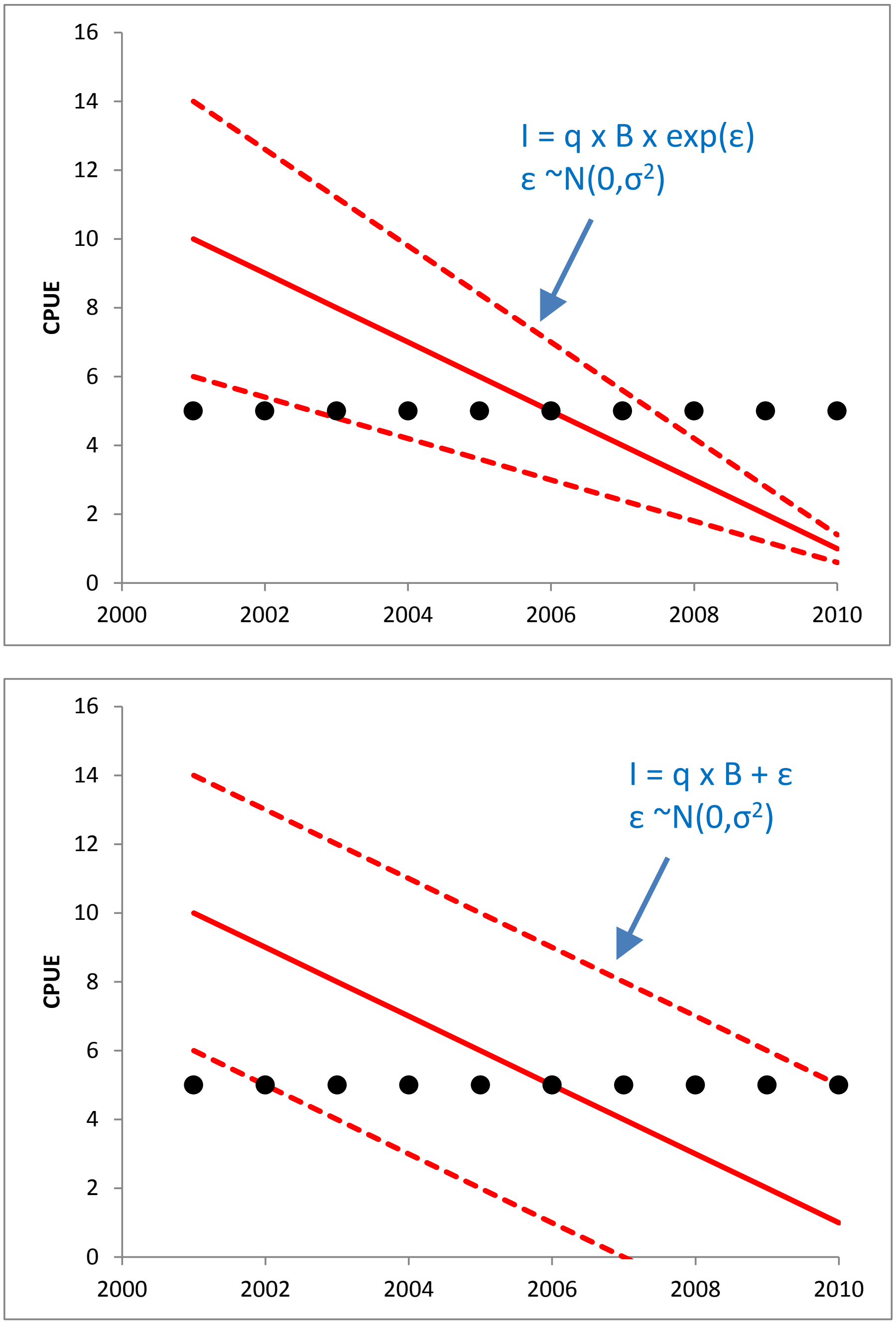

Figure 1. Illustration of how apparent data conflict in catch-per-unit-of-effort (CPUE), a relative index of abundance, (upper panel) could simply be a consequence of large random sampling error (lower panel) and that correct specification of sampling variation is necessary to interpret data conflict. Points = data, solid line = true relative abundance, dashed line = 95 percentiles of the assumed sampling distribution.

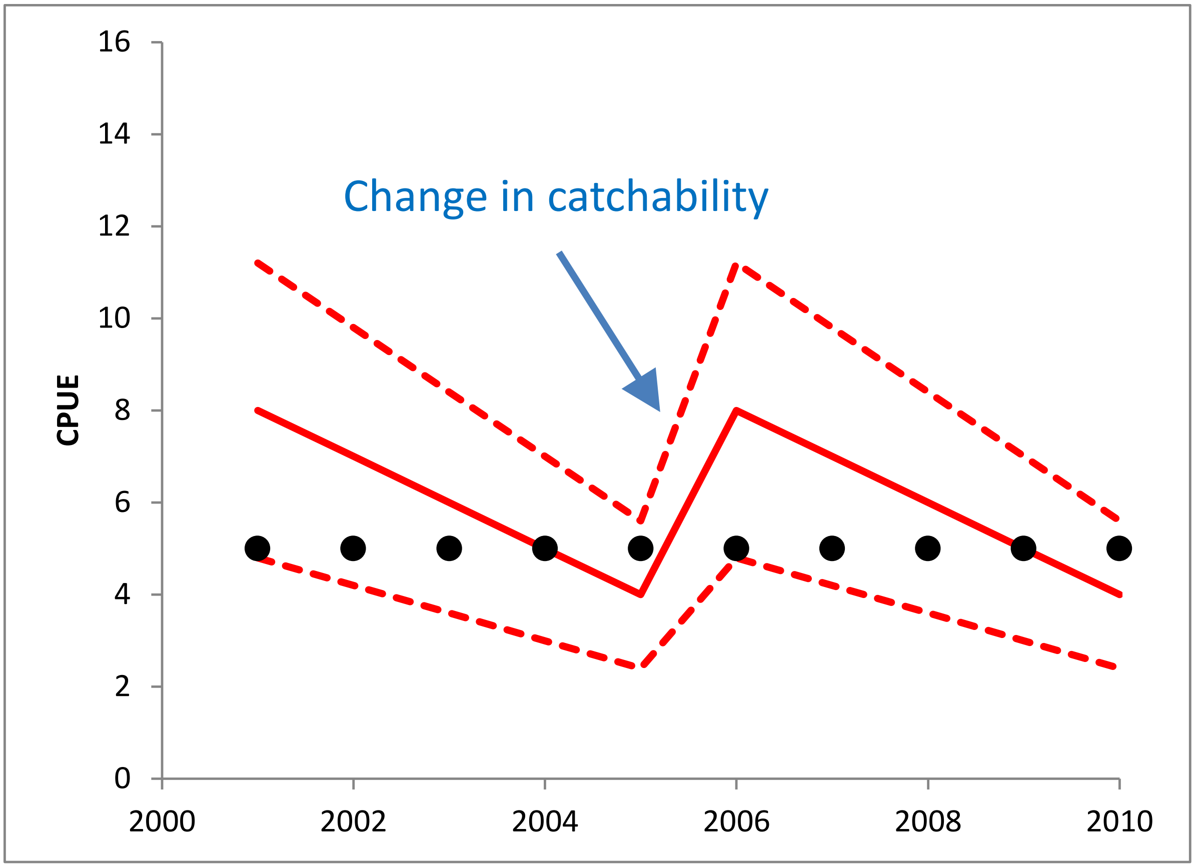

Figure 2. Illustration of how data conflict in catch-per-unit-of-effort (CPUE), a relative index of abundance, could be the consequence of misspecification of the observation model. In this case catchability of the index of relative abundance changed in 2006. Points = data, solid line = underlying CPUE, dashed line = 95 percentiles of the assumed sampling distribution.

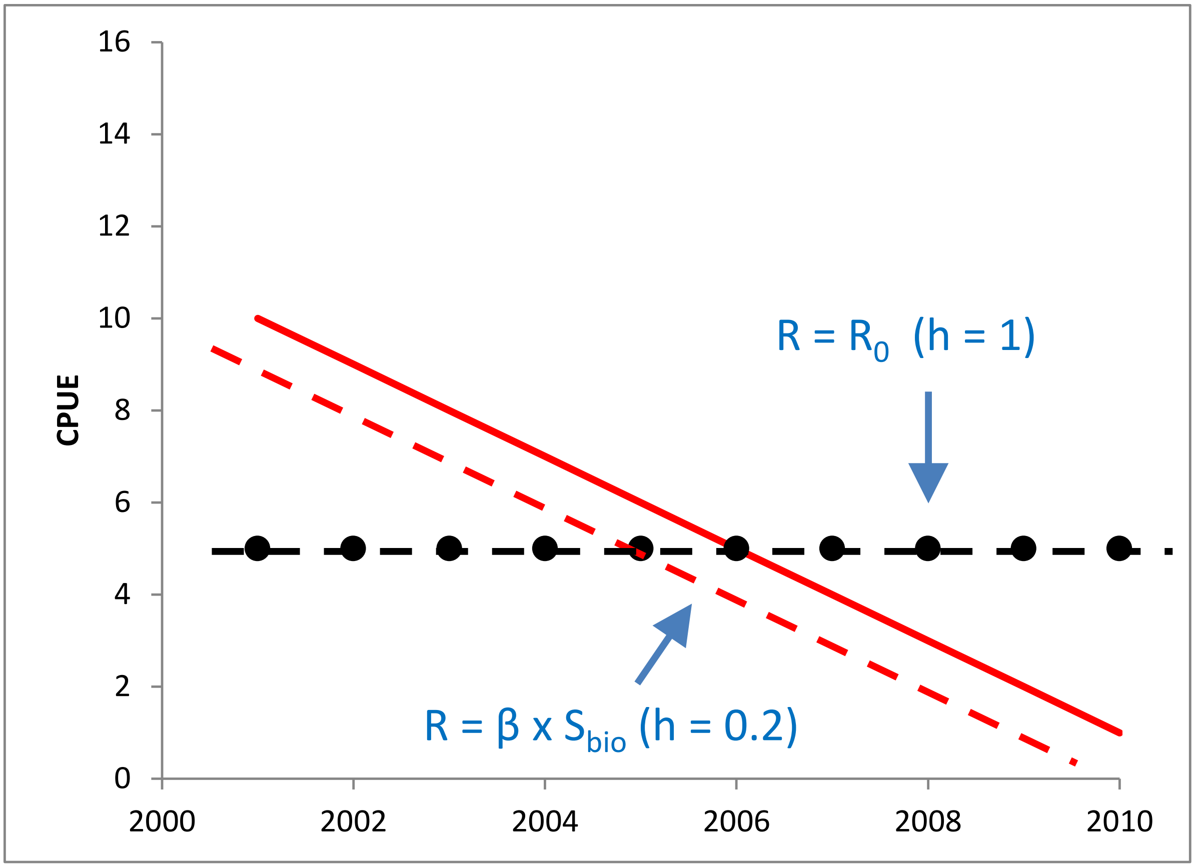

Figure 3. Illustration of how data conflict n catch-per-unit-of-effort (CPUE), a relative index of abundance, could be a consequence of the system dynamics model misspecification. In this example two indices of relative abundance represent different components of the population, spawners and recruits. The two time series conflict, but if the recruitment is assumed to be independent of spawners (high steepness of the stock-recruitment relationship) then the data may not be in conflict with the model. However, if the recruitment is assumed to be proportional to spawners (low steepness) then the data conflicts with the model. Points = data from an index of recruitment, solid line = relative spawner abundance (S_{bio}), dashed grey line = relative recruitment (R) if recruitment is proportional to spawner abundance, dashed black line = relative recruitment if recruitment is constant and independent of spawner abundance, h = steepness of the stock-recruitment relationship.

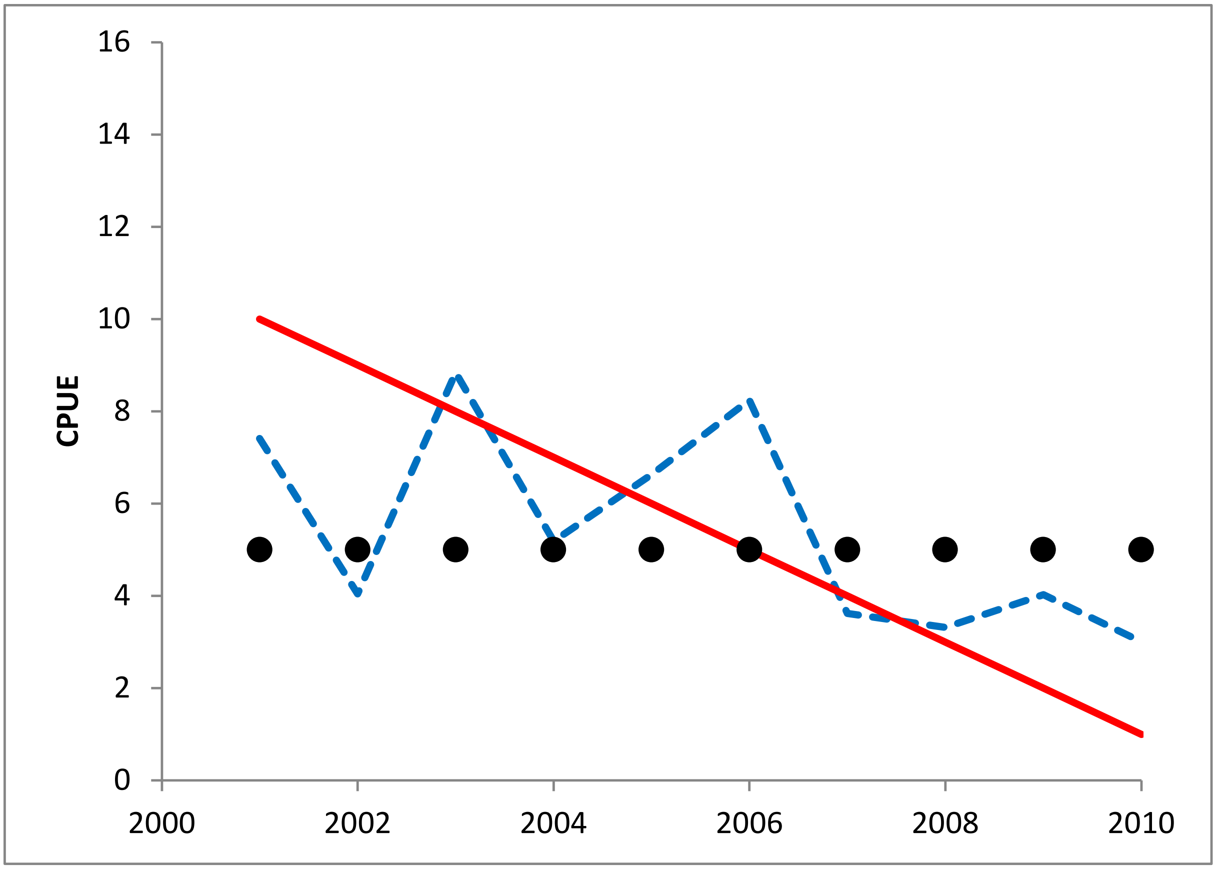

Figure 4. Illustration of how data conflict could be a consequence of the system dynamics model misspecification due to systematic temporal variation in a model process. In this example the two indices of relative abundance represent different components of the population, spawners and recruits, where the underlying relationship is that recruitment is proportional to spawner abundance. The two time series conflict, but if the recruitment is assumed to be dependent on a systematic change in environmenral conditions then the data is not in conflict with the model. Points = observed relative recruitment, solid line = relative spawner abundance, dashed line = relative recruitment based on a systematic change in environmental conditions that influence recruitment

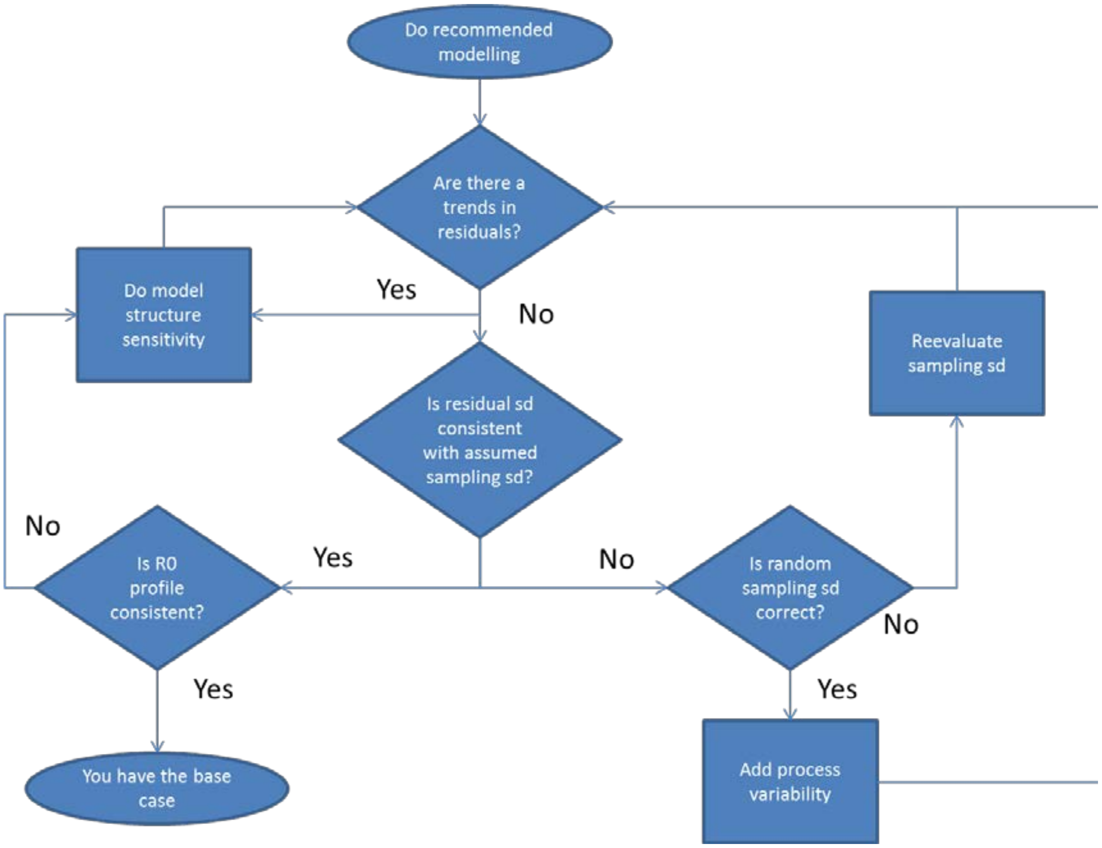
Do recommended modelling
Are there a trends in residuals?
Yes
No
Do model structure sensitivity
Reevaluate sampling sd
Is residual sd consistent with assumed sampling sd?
No
Is R0 profile consistent?
Yes
No
Is random sampling sd correct?
No
Yes
Yes
You have the base case
Add process variability

Figure 5. Flow chart depiction of a simplification of the recommended modelling approach