Diagnostic Template

**Name**: Leave-One-Out analysis, also known as jack-knife analysis

**Goal**: A leave-one-out analysis, or jack-knife analysis, is used to determine if any single data source (or data point) is having undue influence on the model estimates and causing tension with other data in terms of estimating parameters. Note: could also use leave-m data out analyses or the delete-m jackknife

**Description:**

* Leave-One-Out analyses, where individual data sets are removed one at a time and the model is refit to the remaining data, can evaluate the stability of the assessment, i.e., whether the addition or the change of a data source leads to different model estimates.
* Leave-One-Out analysis can also be conducted to identify data points with high leverage and to evaluate the predictive capability of the assessment model (Brooks and Deroba 2015).
* Leave-one-out analyses of indices of abundance are often conducted to determine if any single index is driving model estimates of derived quantities (e.g., SSB, recruitment). Each index of abundance is removed one at a time, and results can identify indices that may be giving conflicting abundance trend signals compared to the remaining indices. Additionally, groups of indices may be removed simultaneously (e.g., all recreational CPUE indices, all commercial CPUE indices, all fishery-dependent CPUE indices, and all fishery-independent surveys). A full leave-one-out analysis would include removing all information associated with a survey (index of abundance, including associated age or size composition data) one survey at a time and refitting the model.

**How to:** Leave-One-Out analysis can be thought of as sensitivities and requires the following steps: (i) run the integrated assessment model, (ii) remove one data source at a time and refit (i); and (iii) repeat (ii) for each data source. Within Stock Synthesis, this analysis can be conducted by multiplying the likelihoods for a given fleet times zero such that they do not inform parameter estimates. To do this, the “lambdas” for a given fleet-data combination located at the bottom of the control file should be zero, where by default they are one (Methot et al. 2020). For surplus production models that are fit to more than one abundance index, individual indices are removed one at a time. It may also be possible to conduct a ‘leave-one-in’ analysis, fitting the model to only one index at a time.

Trends in year-specific estimates of abundance, biomass, recruitment, and mortality can then be compared to determine if any one data source is greatly impacting the model. The effects of removing composition data on growth can also be examined. The expected outcome is that no one data source will be driving the assessment results, and that the results will generally be in agreement when each data source is removed. If removing a dataset leads to dramatically different results, it suggests that the dataset should be reexamined to determine if the sampling procedures are consistent and appropriate (e.g., an index may only be sampling a sub-unit of the stock and resulting abundance signals may only reflect a local sub-population and not the trend in the entire stock).

**Regional Practices and Examples:**

Leave-One-Out analyses are handled differently among regions.

In the southeast, leave-one-out analyses (referred to as jack-knife analyses in the region) are standard sensitivity runs conducted for Gulf of Mexico reef fish assessments. and often focus on removing one index of abundance at a time, or a group of indices. Full leave-one-out analyses are usually not conducted because the other data sets are considered fundamentally necessary to stabilize the assessment. Generally, a leave-one-out analysis is not used as a pass/fail criterion for a base assessment model, and no model adjustments are made based on the results. It is primarily used to provide managers with an understanding of how sensitive the model outcomes are to the indices of abundance that are included in the model, given the high volume of indices incorporated into southeast assessments (range: 4 [Gulf yellowedge grouper, Gulf tilefish] to 18 [Gulf red snapper]). For example, the removal of the video index for the Gulf vermilion snapper assessment had a noticeable impact on both SSB and recruitment in the last few years of the assessment (Figure 1). While this was discussed in detail during the assessment process and at the Gulf SSC, ultimately the base SEDAR67 model (including all indices of abundance) was used to set catch advice.

In the northeast, leave-one-out analyses are conducted in a similar manner as the southeast and are generally limited to considering the impact of removing abundance indices from the assessment. While not mandated in the Terms of Reference, leave-one-out analyses are often completed in assessments that incorporate several indices of abundance in order to evaluate the sensitivity of model estimates to the included indices. However, leave-one-out analyses have not been used to inform dataset weightings or which indices are included in the assessment model.

In the northwest and southwest, leave-one-out analyses vary among assessments where some assessments will emphasize investigating the removal of a single data type (e.g., the index) and more data-rich assessments will perform a full leave-one-out analysis. Few assessments to date have explored the removal of individual data points within a data set. When used for management, surplus production models often perform both leave-one-out and leave-one-in analyses to better understand conflicts in trend information (Fig. 2, Dick et al. 2014). Highly migratory species assessments use jackknifing in the development of abundance indices but not in the formal assessment. Coastal pelagic species do not include jackknifing nor leave-one-out analyses.

**PIFSC:** PIFSC does not typically use jackknife-type analyses for characterizing uncertainty in stock assessments. This is primarily due to the fact that there are few relative abundance indices available for most stock assessments. However, many PIFSC assessments include sensitivity analyses where one or more abundance index or size composition likelihood components are excluded from the model fitting process. This sensitivity information is provided to characterize the effects of removing one or more data sources on model diagnostics and results.

At PIFSC, we are currently investigating the use of k-fold cross validation to estimate relative model weights for ensemble modeling, where we note that the k-fold cross validation procedure is similar to the delete-m jackknife procedure. Here the idea is to quantify the individual model fits to the hold-out data subset using a specific distribution that can be described by a log-likelihood function. The sum of the log-likelihood fits over the k-fold data subsets can provide a measure of predictive accuracy for each model in the ensemble and these measures can be used to compute model weights (Hauenstein et al. 2018) that are similar to the AIC weights for multimodel inference described in Burnham and Anderson (2002). This approach is very general and can be applied with any model fitting algorithm (e.g., maximum likelihood, Bayes, random effects, machine learning, etc…) provided there is an appropriate log-likelihood for the predicted data.

**Examples:**

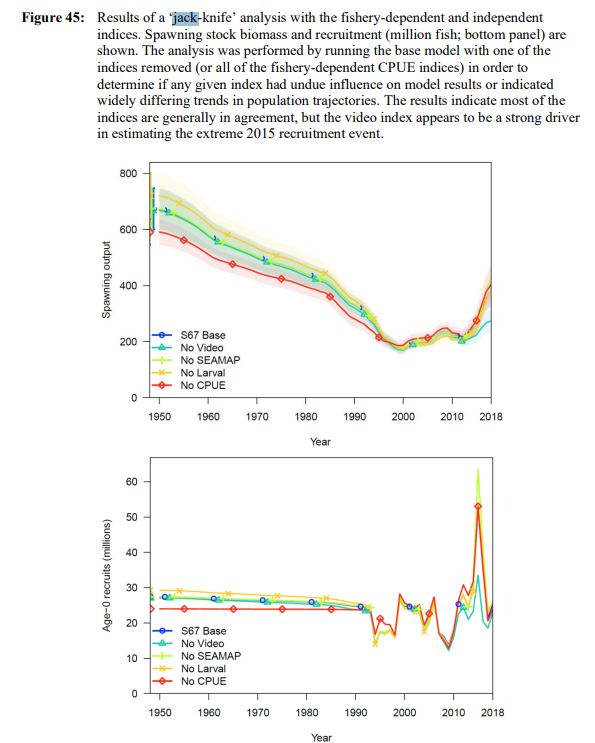
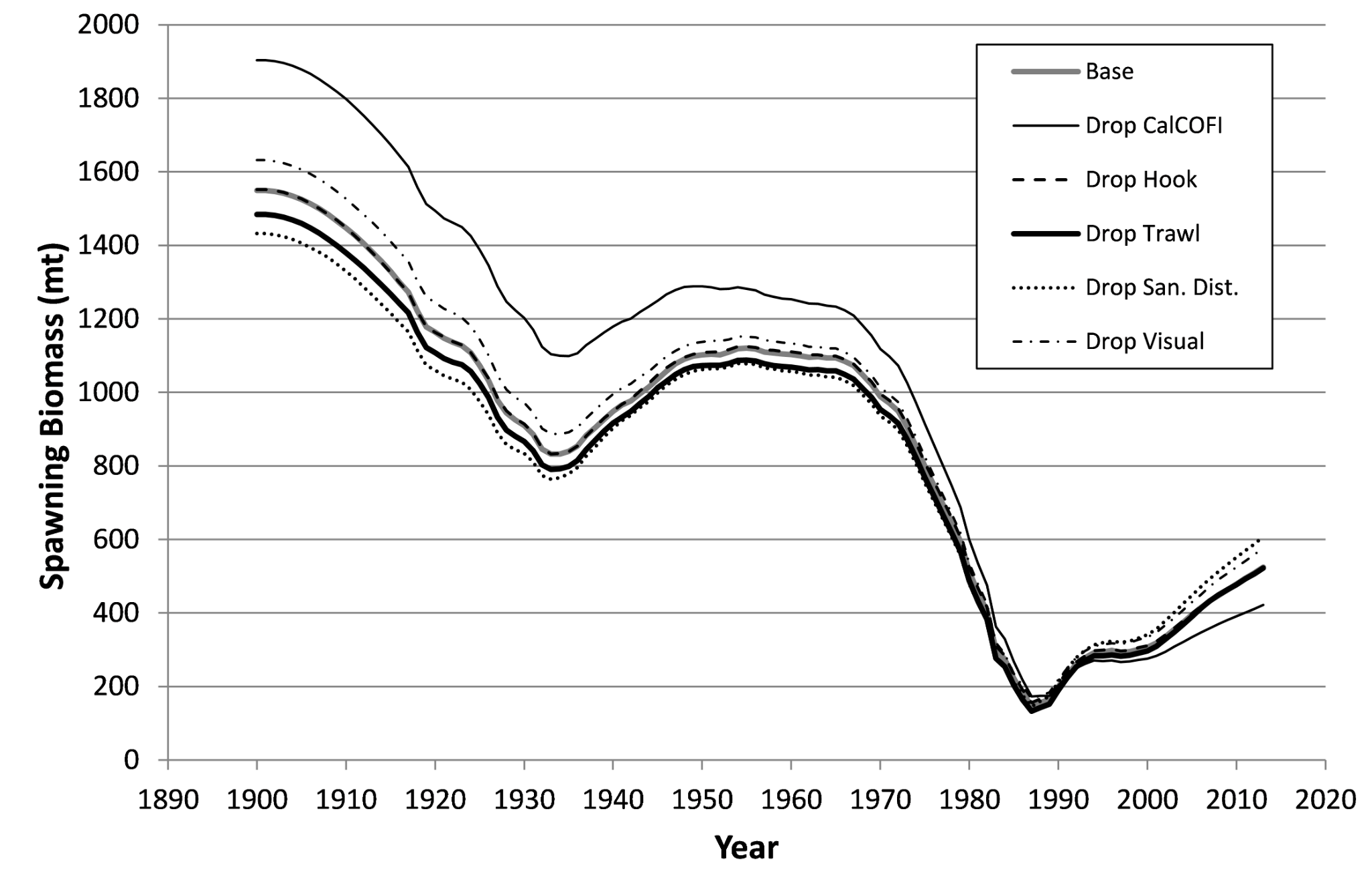


Figure 1. Results of a leave-one-out analysis with the fishery-dependent and independent indices for Gulf vermilion snapper. Spawning stock biomass and recruitment (million fish; bottom panel) are shown. The analysis was performed by running the base model with one of the indices removed (or all of the fishery-dependent CPUE indices) in order to determine if any given index had undue influence on model results or indicated widely differing trends in population trajectories. The results indicate most of the indices are generally in agreement, but the video index appears to be a strong driver in estimating the extreme 2015 recruitment event. Source: SEDAR 2020.



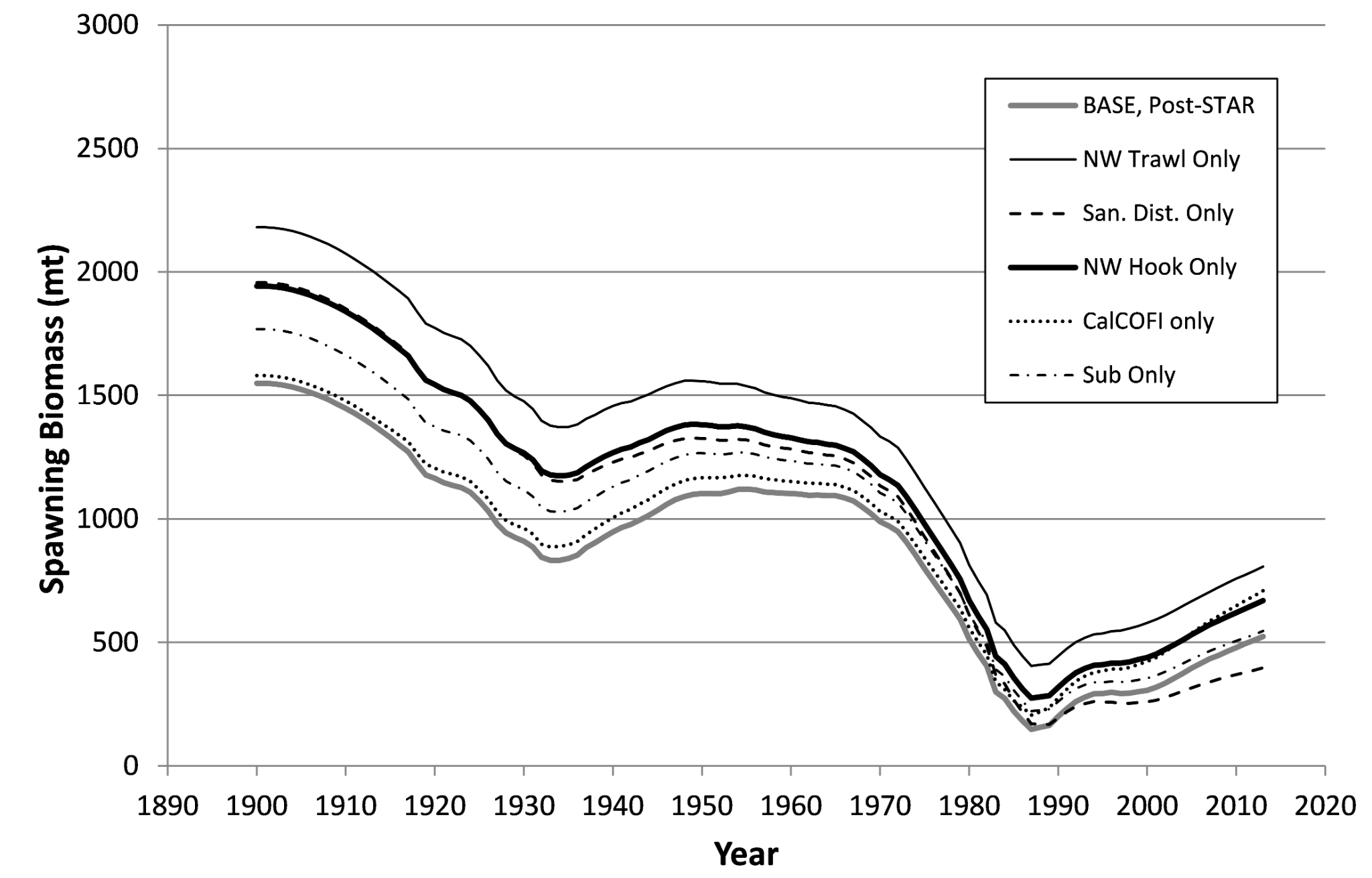


Figure 2. Median spawning biomass for cowcod surplus production model fit by excluding individual indices (‘leave-one-out’, upper panel), or including individual indices (‘leave-one-in’, lower panel). Source: Dick and MacCall, 2014.

**Recommendations:** 

**Key Literature:**

Brooks, E.N. and Deroba, J.J., 2015. When “data” are not data: the pitfalls of post hoc analyses that use stock assessment model output. Canadian Journal of Fisheries and Aquatic Sciences, 72(4): 634-641.

Burnham, K.P. and Anderson, D.R. 2002. Model selection and multimodel inference: A practical information-theoretic approach, 2nd ed. New York, NY: Springer

Dick, E.J. and A. MacCall. 2014. Status and productivity of cowcod, *Sebastes levis*, in the Southern California Bight, 2013. Pacific Fishery Management Council, Portland, OR. Available from <http://www.pcouncil.org/groundfish/stock-assessments/>

Hauenstein, S., Wood, S. N., and Dormann, C. F. 2018. Computing AIC for black-box models using generalized degrees of freedom: A comparison with cross-validation. Communications in Statistics-Simulation and Computation, 47(5): 1382-1396.

Methot Jr, R.D., Wetzel, C.R., Taylor, I.G., and Doering, K. 2020. Stock Synthesis User Manual Version 3.30.16. NOAA Fisheries, Seattle WA. 225 p.

SEDAR (Southeast Data Assessment and Review). 2020. SEDAR67 Gulf of Mexico Vermilion Snapper Stock Assessment Report. 199 p. Available online at: <http://sedarweb.org/sedar-67>