

Proposal for the methodological review of sdmTMB for index standardization

Kelli F. Johnson, Northwest Fisheries Science Center
Eric J. Ward, Northwest Fisheries Science Center

Background and purpose

Raw data on catch and effort of marine species can be misleading because changes in catches over time often represent an amalgamation of many factors beyond just changes in abundance. Standardization of these data in an attempt to remove the impact of auxiliary factors is standard procedure and is often done using generalized linear mixed effects models (GLMMs) with catch rates as the response variable. Most recently, standardization of catch rates presented to the PFMC has been accomplished using the vector autoregressive spatiotemporal (VAST) framework that can account for correlations in space as well as time (Thorson 2019). VAST fits GLMMs using Template Model Builder (Kristensen et al. 2016) and Integrated Nested Laplace Approximation software (Rue, Martino, and Chopin 2009). Since the approval for the use of VAST by the SSC, the VAST framework has grown in both its use in various regions throughout the US and globally and its range of available output. Unfortunately, the increasing complexity of the framework has led to a steep learning curve regarding its use. This learning curve has been slightly mitigated through the use of VASTWestCoast, a wrapper written around VAST to specify default settings specific to the US west coast, but constant effort is needed to maintain VASTWestCoast given the constantly-changing source code used to implement VAST.

Concurrent to the development of VAST, scientists at the Northwest Fisheries Science Center, the Alaska Fisheries Science Center, and the Pacific Biological Station (Fisheries and Oceans Canada) have developed, tested, published, and implemented for management a similar framework called sdmTMB that also uses GLMMs but has the sole purpose of standardizing catch-rate data (Anderson, Keppel, and Edwards 2019). Both sdmTMB and VAST are coded as C++ template files in Template Model Builder, which allows for fast estimation of complex models, such as GLMMs, and maximizes the Laplace approximation of the marginal likelihood. Estimation for both sdmTMB and VAST are performed using `stats::nlminb()` and have the following shared default settings: 10000 maximum iterations and 10000 maximum objective function/gradient evaluations. By default, VAST starts each model from 2 random starting values, while sdmTMB uses 1 (nlminb loops). Similarly, by default VAST uses 1 Newton optimization step with `stats::optimHess()` after estimation, while sdmTMB defaults to 0. Many one-off comparisons between these two frameworks have already been completed and provide support suggesting their similarity, i.e., estimated indices are identical when fit to data for some species or simulated data.

We believe that the use of sdmTMB for index standardization will lead to three major benefits

1. increased speed of fitting models due to a less-steep learning curve and a faster estimation of parameters,
2. increased collaboration with Canadian scientists who use sdmTMB to standardize Canadian data on stocks that migrate between the US-Canada border, and
3. increased robustness of parameter estimates because of the ease of turning parameters on and off in sdmTMB which decreases the likelihood that models are overparameterized when data are sparse, particularly for non-target species.

Methods

Paired fitting of sdmTMB and VAST to both simulated data and data sampled by the West Coast Groundfish Bottom Trawl Survey of species that are well sampled and species that are less well sampled will be used to highlight the similarity of estimates between the two frameworks when the frameworks are parameterized identically. Additionally, the utility of sdmTMB will be highlighted for when data are sparse and the parameterization must be simplified, e.g., removing spatial effects and focusing on temporal effects. This was done in 2021 for vermilion during the data-preparation stage but not presented to the STAR panel.

In addition to summaries of the comparison exercises, detailed explanations of both frameworks and all relevant references will be provided prior to the review. Because both tools rely on the same underlying model and conduct estimation using the same framework, the details will focus on the small differences between them.

References

- Anderson, Sean C., A. E. Keppel, and A. M. Edwards. 2019. “A Reproducible Data Synopsis for over 100 Species of British Columbia Groundfish.” DFO Can. Sci. Advis. Sec. Res. Doc. 2019/041. https://www.dfo-mpo.gc.ca/csas-sccs/Publications/ResDocs-DocRech/2019/2019_041-eng.html.
- Kristensen, Kasper, Anders Nielsen, Casper W. Berg, Hans Skaug, and Bradley M. Bell. 2016. “TMB: Automatic Differentiation and Laplace Approximation.” *Journal of Statistical Software* 70 (1): 1–21. <https://doi.org/10.18637/jss.v070.i05>.
- Rue, Håvard, Sara Martino, and Nicolas Chopin. 2009. “Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71 (2): 319–92. <https://doi.org/https://doi.org/10.1111/j.1467-9868.2008.00700.x>.
- Thorson, James. 2019. “Guidance for Decisions Using the Vector Autoregressive Spatio-Temporal (VAST) Package in Stock, Ecosystem, Habitat and Climate Assessments.” *Fisheries Research* 210 (February): 143–61. <https://doi.org/10.1016/j.fishres.2018.10.013>.