# Title:

Unlocking the triad of age, year, and cohort effects in stock assessment: a proof-of-concept study

# Abstract:

Many demographic processes vary by age and over time, and accounting for this variation within fisheries management remains a key challenge for many contemporary stock assessments. Although there is evidence for time, age, and cohort specific effects on various components within stock assessment (e.g., selectivity, growth), methods are lacking to simultaneously estimate autocorrelation over time, among ages, and by cohort while also quantifying residual variation. Drawing from previous research on separable cohort models, we reintroduce the idea of “triple-separability”, which simultaneously estimates autocorrelation for time, age, and cohort effects, and reduces to two-dimensional autocorrelation when one of the processes is fixed at zero. Utilizing eastern Bering Sea walleye pollock (*Gadus chalcogrammus*)as a case-study, we illustrate differences in predicted weight-at-age values from models with and without a triple-separable assumption and show that traditional model selection tools can be used to identify the relative evidence and magnitude of age, time, and cohort effects. We recommend that the method be integrated as a routine option within next-generation stock assessments, and note that it generalizes widely-used options in existing assessment frameworks (i.e., WHAM and SAM).

# Introduction:

Fisheries often experience dynamic shifts in demographic (e.g., growth, natural mortality, movement) and removal processes (e.g., fleet structure, selectivity), which can be facilitated by environmental forcing and management regulations (Eigaard *et al.*, 2014; King *et al.*, 2015). Although it is well recognized that these processes are correlated and vary across, space, time, ages, and cohorts (Sampson and Scott, 2012; Taylor and Methot, 2013; Thorson and Minte-Vera, 2016), these processes are commonly represented as invariant within stock assessments. We first briefly discuss how these processes can arise within fisheries, and then provide examples for the consequences of assuming invariant age, time, and cohort-dependent effects. Correlated age processes can arise when individuals of the same age-classes experience habitat conditions that result in consistent ontogenetic variation in demographic processes. Likewise, variable annual processes can arise when individuals in a given year experience similar environmental conditions (e.g., marine heatwaves; Barbeaux *et al.*, 2020), or when the fishery experiences changes in fishing practices (Martell and Stewart, 2014). Finally, cohort-specific effects can result from large recruitment events that alter targeting strategies from fishers and/or density dependence in growth (Rose *et al.*, 2001; Goethel *et al.*, 2022). However, as noted previously, these processes are often assumed to be invariant within stock assessments, despite evidence suggesting otherwise, and failing to appropriately account for these processes may result in biased estimates of model results. For example, Taylor and Methot (2013) found that disregarding individuals exhibiting differential growth trends resulted in biased estimates of spawning biomass and depletion. Similarly, failing to account for time-varying selectivity, and age-and-time varying natural mortality has been illustrated to result in biased estimates of quantities [reference points, spawning biomass, recruitment](https://www.zotero.org/google-docs/?jlEIGP) (Deroba and Schueller, 2013; Martell and Stewart, 2014). Consequently, it is necessary for stock assessment models to be parameterized by realistic and sensible model structures that adequately represent the underlying stock and fishery dynamics.

Parameterizing model structures that are both rationally and biologically motivated is often a key challenge within stock assessment modeling. However, recent advances in computational tools have aided in such model parameterizations (Kristensen *et al.*, 2016), allowing for the efficient computation of latent variables, which are well-suited to represent correlations in biological and fishery processes (Nielsen and Berg, 2014; Stock and Miller, 2021). However, methods are currently lacking that allow for the simultaneous estimation of correlated processes among ages, years, and cohorts. Thus, drawing from research on separable cohort models, we reintroduce the idea of “triple-separability”, which was initially applied to estimate age, time, and cohort-dependent processes in selectivity, to account for the increased vulnerability of more abundant cohorts (*sensu* Vasilyev, 2000). In the present study, we extend this work by estimating correlations among ages, years, and cohorts for weight-at-age represented as a Gaussian Markov Random Field (GMRF) using eastern Bering Sea walleye pollock (*Gadus chalcogrammus*) as a case-study. Our method described here is widely can be extended to represent other biological and fishery parameters that are commonly estimated within integrated stock assessments (i.e., selectivity, natural mortality).

# Methods:

We first describe the construction of a computationally efficient matrix that accommodates partial correlations along ages, years, and cohorts arising from a GMRF process. We then illustrate our method by conducting a 2x2x2 factorial experiment where we implement the GMRF process onto a weight-at-age matrix for walleye pollock, and use traditional model selection tools to assess support for incorporating partial correlations along ages, years, and/or cohort axes.

*Specification of GMRF process*

We seek a computationally efficient way to construct the inverse-covariance (i.e., precision) matrix that results in a two-dimensional process with indexed by age and year , where has partial correlations along a year, age, and cohort axis. The inverse-covariance would then be used to specify a hyperdistribution for the process:

where is a vector that follows a multivariate normal distribution with a mean vector of 0s, and is the inverse of the precision matrix (i.e., covariance). Specifying the covariance matrix as the inverse of the precision matrix when evaluating the multivariate normal probability density function allows for sparsity, which facilitates efficient computation to perform Laplace approximation for models with a large number of random effects (Kristensen *et al.*, 2016).

To do so, we take inspiration from Simultaneous Autoregressive (SAR) processes in spatial statistics (Ver Hoef *et al.*, 2018a, 2018b). Specifically, we construct a square matrix that represents the partial effect of on preceding ages and/or years (Fig. 1). The square matrix is analogous to the spatial weights matrix described in the SAR literature, and represents a “neighborhood” structure. Note that ages and/or years do not depend on themselves and thus, matrix has zeros on the diagonals. Furthermore, matrix **B** does not need to be symmetrical as it does not directly appear in the inverse of the covariance matrix (see Eq. 3 below). In a simplified example of matrix **Y** indexed with ages and (e.g., where the second element of ***y*** corresponds to age 2 and year 1), we then construct matrix **B** as:

where is autocorrelation among ages in a given year, is autocorrelation among years for a given age, and is autocorrelation along a cohort. We then construct the precision matrix as:

where **I**is an identity matrix, and is a positive diagonal matrix that determines the variance of the process. To demonstrate the process specified from Eq. 2 and 3, we construct the covariance and visualize random multivariate normal draws using different specified values for partial correlations to provide intuition on scenarios with strong age, year, or cohort effects (Fig. 2). We note two alternative ways to specify , where both involve specifying that is a diagonal matrix. We call these the “conditional variance” and “marginal variance” forms:

1. Conditional variance form: In the following, we specify that , where would be an estimated parameter representing the variance for conditional upon previous ages and years. This construction then results in a heteroskedastic (and potentially nonstationary) process, i.e., where varies among ages and years, but has the benefit that there are no restrictions on partial correlations .
2. Marginal variance form: We could instead calculate values for such that , where would be an estimated parameter representing the marginal variance for , which is stationary (i.e., the same value for all ages and years). This then implies bounds on , and requires some extra code to implement (see Appendix A for details).

*Factorial Experiment*

To demonstrate the implementation of the GMRF, we conducted a full-factorial experiment (2x2x2) applied to weights across ages and years for walleye pollock. The full-factorial experiment estimated all possible combinations of partial correlations along the age, year, and cohort axes. Thus, the null model in this experiment did not estimate any parameters to represent partial correlations. We first estimated a time-invariant mean weight-at-age ( for each age-class:

where is asymptotic maximum length-at-age, which was fixed at 1 given confounding with . represents the theoretical mean length-at-birth, and *k* is the Brody growth coefficient. Equation 5 describes an isometric relationship between length and weight, where represents the average condition factor. Unobserved values of weight-at-age ( were then treated as latent random variables, and were assumed to arise from a multivariate GMRF process:

where is a vector of mean weight-at-age estimated from Equation 5, and is the inverse of the precision matrix describing correlations along ages, years, and cohorts, which was specified according to Equation 3. Predicted values of weight-at-age () were assumed to be normally distributed:

where are the estimated unobserved weights for age *a* and time *t*, and is the standard deviation from the observed weight-at-age matrix. Finally, we compared the results of the 8 different model variants, using standard convergence diagnostics (i.e., invertible Hessian matrix, gradient components), estimated parameter values, and Akaike Information Criterion (AIC) values, to assess the relative evidence and magnitude of age, year, and cohort effects.

# Results

All model variants assessed in the present study appeared to have positive definite Hessian matrices, with maximum gradient components that were < 0.001. Furthermore, uncertainty estimates of parameters appeared fairly reasonable. In addition, the model that estimated partial correlations along the age, year, and cohort axis (model y\_a\_c) had the lowest AIC score (-836.50). Conversely, there appeared to be substantially less statistical support for the null model (highest AIC score; -431.64), which did not estimate any partial correlations (Fig. S1A).

# Discussion:

* Discussion on: 1) general results, and 2) how Wald tests and AIC can be used for model selection
* Discuss about the benefits of implementing triple-separability, and potential consequences of failing to account for such processes
  + Describe to the audience that this can be practical within contemporary stock assessments and can be implemented in growth, natural mortality, selectivity processes, survival, maturity (e.g., tell folks what the main use of this new method is).
* Potential limitations and caveats of this method?
* Link to current literature (e.g., state-space models, GMRFs, 2DAR1 processes that are currently being implemented)
* Finally, discuss future research (e.g., simulation testing, real-world examples, and implement on other components of stock assessment)

# References

Barbeaux, S. J., Holsman, K., and Zador, S. 2020. Marine Heatwave Stress Test of Ecosystem-Based Fisheries Management in the Gulf of Alaska Pacific Cod Fishery. Frontiers in Marine Science, 7: 703.

Deroba, J. J., and Schueller, A. M. 2013. Performance of stock assessments with misspecified age- and time-varying natural mortality. Fisheries Research, 146: 27–40.

Eigaard, O. R., Marchal, P., Gislason, H., and Rijnsdorp, A. D. 2014. Technological Development and Fisheries Management. Reviews in Fisheries Science & Aquaculture, 22: 156–174.

Goethel, D. R., Rodgveller, C. J., Echave, K. B., Shotwell, S. K., Siwicke, K. A., Malecha, P. W., Cheng, M., *et al.* 2022. Assessment of the Sablefish Stock in Alaska: 182.

King, J. R., McFarlane, G. A., and Punt, A. E. 2015. Shifts in fisheries management: adapting to regime shifts. Philosophical Transactions of the Royal Society B: Biological Sciences, 370: 20130277.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. 2016. TMB: Automatic Differentiation and Laplace Approximation. Journal of Statistical Software, 70. http://arxiv.org/abs/1509.00660 (Accessed 8 January 2023).

Martell, S., and Stewart, I. 2014. Towards defining good practices for modeling time-varying selectivity. Fisheries Research, 158: 84–95.

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

Rose, K. A., Cowan, J. H., Winemiller, K. O., Myers, R. A., and Hilborn, R. 2001. Compensatory density dependence in fish populations: importance, controversy, understanding and prognosis: Compensation in fish populations. Fish and Fisheries, 2: 293–327.

Sampson, D. B., and Scott, R. D. 2012. An exploration of the shapes and stability of population-selection curves: Shapes and stability of population-selection curves. Fish and Fisheries, 13: 89–104.

Stock, B. C., and Miller, T. J. 2021. The Woods Hole Assessment Model (WHAM): A general state-space assessment framework that incorporates time- and age-varying processes via random effects and links to environmental covariates. Fisheries Research, 240: 105967.

Taylor, I. G., and Methot, R. D. 2013. Hiding or dead? A computationally efficient model of selective fisheries mortality. Fisheries Research, 142: 75–85.

Thorson, J. T., and Minte-Vera, C. V. 2016. Relative magnitude of cohort, age, and year effects on size at age of exploited marine fishes. Fisheries Research, 180: 45–53.

Vasilyev, D. A. 2000. TRIPLE-SEPARABLE VPA (TSVP.4) OR A STONE TO BRIDGE THE GAP BETWEEN SEPARABLE COHORT MODELS AND NONSEPARABLE ONES. ICES Conference and Meeting.

Ver Hoef, J. M., Peterson, E. E., Hooten, M. B., Hanks, E. M., and Fortin, M.-J. 2018a. Spatial autoregressive models for statistical inference from ecological data. Ecological Monographs, 88: 36–59.

Ver Hoef, J. M., Hanks, E. M., and Hooten, M. B. 2018b. On the relationship between conditional (CAR) and simultaneous (SAR) autoregressive models. Spatial Statistics, 25: 68–85.

# Figures

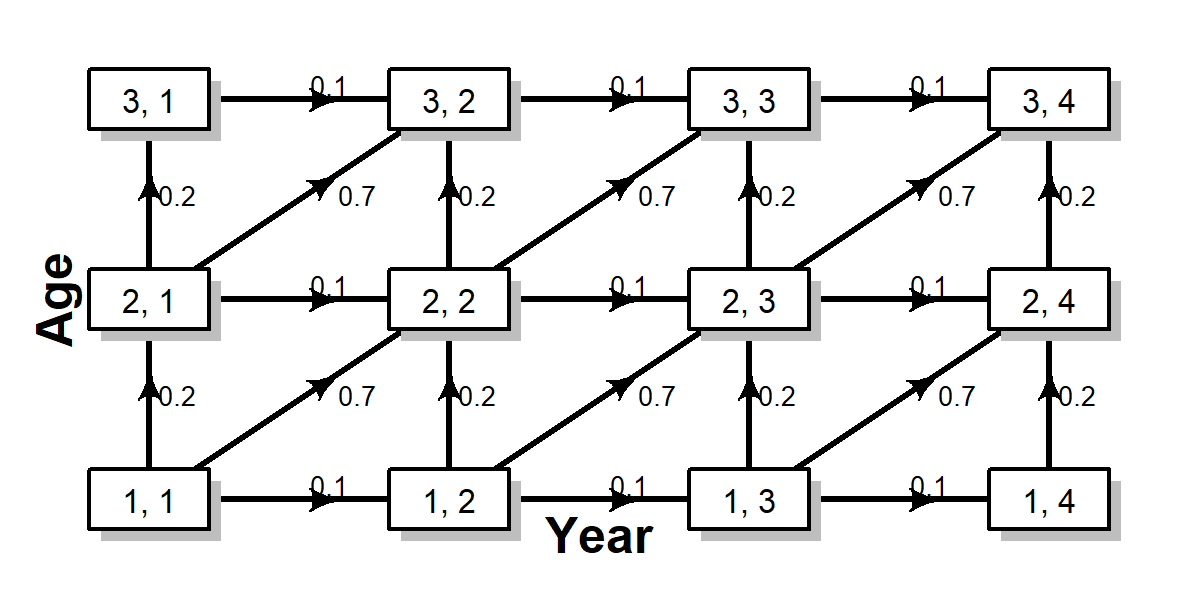


Figure 1: Diagram illustrating the formulation of matrix . Here, partial effects of age (y-axis), year (x-axis), and cohort (ascending diagonal) are indexed by ages and years, and are represented by a weak partial effect of year , intermediate partial effect of age , and a strong partial effect of cohort

Chart, treemap chart

Description automatically generated

Figure. 2 – Examples of random values of process arising from a multivariate normal GMRF process resulting from strong year (left panel), age (middle panel), and cohort (right panel) effects.



Figure 3. Estimated weight-at-age (kg) for the 2x2x2 factorial experiment across model variants (panels). The letters y, a, and c denote whether partial correlations were estimated for years, ages, and cohorts (i.e., model a\_c estimated partial correlations for ages and cohorts). Panel A is a heatmap where the shading reflects the anomaly of estimated values relative to the mean. Panel B shows the trends in weight-at-age over time, where the numbers denote respective ages. For visualization purposes, only a subset of ages are depicted.