Causal models as a scientific framework for next-generation ecosystem and climate-linked stock assessments

Running title: structural causal model for stock assessment

J. Champagnat[1,2*], C.C. Monnahan[3], J.Y. Sullivan[4], James T. Thorson[3], S.K. Shotwell[4], L. Rogers[3], A.E. Punt[1]

* Corresponding author: juliette.champagnat@ifremer.fr

[1] School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA.

[2] MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, Sète, France

[3] Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA 98115, USA

[4] Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Juneau, AK 99801, USA

Abstract

Rapid changes in marine ecosystems highlight the need to account for time-varying productivity in stock assessment models used to support fisheries management. Common approaches incorporate annual variation or regress processes like recruitment, natural mortality, or growth on environmental covariates. While the latter represents a step towards biological realism, it often fails accounting for interactions among covariates and may yield biased inferences when key drivers are correlated or unmeasured. We introduce a novel framework, Structural Causal Enhanced Stock Assessment Modelling (SCEAM), that integrates a Dynamic Structural Equation Model (DSEM) into a state-space stock assessment. SCEAM encompasses and extends the full range of existing time-varying approaches within a single framework, enabling direct comparison among them. We applied SCEAM to walleye pollock in the Gulf of Alaska to improve recruitment forecasting. When we compared three causal models of increasing complexity to recruitment modelled as random deviations around a mean, a first order autoregressive process, or regressed on a single covariate, we found that a causal model with intermediate complexity best balanced fit, parsimony, and predictive skill. This configuration reduced unexplained variance of recruitment by 69% and improved one-year-ahead forecasts. Key predictors included juvenile body condition and juvenile and larval catch rates. Our study represents the first application of a structural causal model embedded within a fisheries population model. SCEAM offers a unified, hypothesis-driven approach to integrating multiple non-independent covariates. We therefore propose that SCEAM can serve as a general scientific and statistical framework for building next-generation ecosystem- and climate-linked fisheries stock assessment models.

Keywords: structural causal models; fisheries stock assessment; recruitment; walleye pollock; dynamic structural equation models

Table of content

1. Introduction

Stock assessments rely on modelling key biological (e.g., recruitment, growth, natural mortality, reproduction) and fishery processes (e.g., selectivity, fishing mortality) that determine population productivity. These processes form the foundation of modern stock assessments, which are traditionally implemented in a single-species framework. Although often treated as time-invariant, both biological and fishery processes frequently vary over time (Vert-pre *et al.*, 2013; Thorson *et al.*, 2015; Szuwalski and Hollowed, 2016). For example, variability in environmental and ecological conditions has been demonstrated to affect survival, growth, and reproductive output in marine populations (Miller and Hyun, 2018; Xu *et al.*, 2018; Feng *et al.*, 2025). Stock assessments can produce biased estimates of stock status, inaccurate projections, and less effective management advice when they ignore time-varying dynamics. Case studies for Pacific cod (*Gadus macrocephalus*) in the Gulf of Alaska (Barbeaux *et al.*, 2020), Atlantic cod (*Gadus morhua*) in the Gulf of Maine (Pershing *et al.*, 2015) and Baltic Sea (Lindegren *et al.*, 2009), and Chilean jack mackerel (*Trachurus murphyi*) in the Southeastern Pacific Ocean (Arcos *et al.*, 2001; Lima *et al.*, 2020) demonstrate the risks of not incorporating environmental and ecological signals into stock assessment and management frameworks. Accounting for time-varying productivity in assessment models will be essential to maintaining reliable stock estimates and precautionary management advice as variability in ecosystem processes intensifies (Szuwalski and Hollowed, 2016).

Methods have been developed to account for time-variation in population processes and typically fall into two categories. The first, termed "empirical" by Punt et al. (2014), allows parameters to vary stochastically over time using structures such as independent and identically distributed (iid) or autoregressive (AR) processes (first two columns in Fig. 1). This approach captures variability from unobserved sources without explicitly modelling them and has been implemented in several modern stock assessment models (Nielsen and Berg, 2014; Cadigan, 2016; Stock and Miller, 2021). Despite being very flexible, this approach requires caveats regarding which parameters are considered time-varying (Punt, 2023). The second approach, called "mechanistic" by Punt et al. (2014), explicitly links population processes to environmental covariates (regression, Fig. 1) which we refer to as a regression paradigm. While being attractive in principle and having more biological realism than the empirical approach, the mechanistic approach has been less widely used (Punt *et al.*, 2014; Karp and Vieser, 2024). It also has several potential challenges: (i) dealing with multiple correlated covariates, (ii) outside sample extrapolations (Kell *et al.*, 2005; Punt *et al.*, 2014, p 20) because estimated relationships are based on historical observations, and (iii) relationships breaking over time (Myers, 1998; McClatchie *et al.*, 2010). Furthermore, regression analyses are not sufficient to establish mechanistic links because they ignore relationships among potential covariates and the population process that are known to lead to biased estimates of mechanistic relationships, particularly with missing confounding variables (Grace and Irvine, 2020; Byrnes and Dee, 2025).

Recently, causal analysis has been advocated as a more reliable scientific framework for determining mechanistic links as well as overcoming some of these covariate dilemmas (Grace and Irvine, 2020; Arif and MacNeil, 2023; Siegel and Dee, 2025). The goal of causal inference is to quantify the effect of a variable, called the exposure, on another variable called the outcome. The structural causal modelling framework (SCM, Pearl, 2009) allows the use of observational data and derives causal effects based on a set of assumptions about the data-generating process (e.g., the exposure affects the outcome and not the other way around, Arif and MacNeil, 2023). SCM uses causal diagrams to represent how a system works; that is, all the hypotheses about the relationships in a study system, including all known confounding variables (Grace and Irvine, 2020). This requires expert scientific knowledge regarding the studied system before statistical modelling (McElreath, 2018; Grace and Irvine, 2020; Hernán and Robins, 2023). We believe that SCMs present desirable features for linking covariates to stock assessment: (i) it allows a broader and more rigorous use of multiple non-independent environmental variables sampled through data collection processes than traditional regression and correlative approaches; (ii) it actively integrates the process knowledge generated by ecological research; (iii) it transparently represents hypotheses through causal diagrams and provides a comprehensive communication tool towards fisheries stakeholders.

While causal questions are common in ecology, causal inference approaches are not yet widely adopted, although recent studies and reviews are encouraging its (Laubach *et al.*, 2021; Thorson *et al.*, 2021; Arif and MacNeil, 2023; Siegel and Dee, 2025). Several applications in the marine sciences have linked species abundance to environmental and/or anthropogenic drivers using Structural Equation Modelling (SEM; Kuczynski *et al.*, 2017; Boyce *et al.*, 2021). SEM is a part of the SCM framework and can be defined as a multivariate method testing direct and indirect effects on pre-assumed causal relationships (Lefcheck, 2016). Another approach used to explore causal relationships between time series is Empirical Dynamic Modelling (EDM, Sugihara *et al.*, 2012) which allows for testing non-linear relationships. This method and its expansion (Convergence Cross Mapping, Clark *et al.*, 2015) have also been used to understand causes of species abundance fluctuations, and to project it in the future (Deyle *et al.*, 2013; Ye *et al.*, 2015). Recent applications demonstrated EDM abilities to compute reference points (Giron-Nava *et al.*, 2021; Tsai *et al.*, 2024), but to our knowledge, SEM, EDM, and CCM have not been embedded in an age-structured population model used as the basis for stock assessment.

State-space models combine observed variables with unobserved latent states modelled as random variables with a specified probability distribution and for which process errors can be estimated, and they provide a natural way to represent natural resource dynamics. They are being increasingly used in fisheries science, notably for stock assessment (e.g., Cadigan, 2016; Aeberhard *et al.*, 2018; Cadigan *et al.*, 2024). Time series of stock abundance, its age/size structure, fishing mortality rates, and other processes can be stochastic latent states in a state-space stock assessment model (SSAM), informed by indirect observations such as commercial catches and survey indices, which are subject to observation errors. A common method to estimate SSAM parameters is maximum

marginal likelihood estimation, which requires high dimensional integrals to be evaluated. Recent advances in algorithms and software (e.g., TMB, Kristensen *et al.*, 2016) allow efficient fitting of state-space models. Because of their flexibility SSAMs have modelled population time-varying processes using either empirical or mechanistic frameworks (e.g., WHAM, SAM). We suggest instead that a causal framework should be used to model time-varying processes, which would allow the inclusion of multiple non-independent covariates in a rigorous and transparent way. In this framework, a time-varying process from a stock assessment would be modelled as an outcome of exposure variables linked through a causal diagram (far right column in Fig. 1). Dynamic Structural Equation Models (DSEM, Thorson *et al.*, 2024) were recently developed to unify causal inference and time series forecasting in a computationally efficient framework that can also impute missing values for covariates. DSEM is compatible with most SSAMs because it presents a state-space structure and has versions available in TMB and RTMB. Moreover, it provides computationally efficient options for short-term forecasts and end-of-century projections, both of which are useful for different purposes in stock assessments. DSEM is therefore a convenient statistical framework to incorporate causal diagrams into SSAMs.

This study is the first demonstration (to our knowledge) of combining population models and SCM to address the needs of fisheries management in a changing environment. Specifically, we show how coupling DSEM and a SSAM (i) generalizes the regression paradigm while addressing several of its important limitations, and (ii) provides accurate forecasts of population. Our Structural Causal Enhanced Assessment Model (SCEAM) facilitates testing against status quo methods (empirical – iid, AR1, and regression, Fig. 1) to facilitate comparison of models of increasing complexity. As a case study, we apply SCEAM to estimate and forecast recruitment of the walleye pollock (*Gadus chalcogrammus*) stock in the Gulf of Alaska.

2. Materials and methods

SCEAM consists of combining two models: DSEM and a population dynamics model. We first describe DSEM and its functionality and explain the conceptual steps for coupling it to a population model. Second, we introduce the GOA pollock case study with an overview of the stock assessment and the environmental and ecosystem variables used. Then, we describe the process of building causal diagrams and how they were implemented within the stock assessment to build SCEAM. Finally, we detail how we evaluated model performance against traditional approaches and a simple simulation test to explore the statistical properties of SCEAM.

2.1 DSEM overview

DSEM estimates simultaneous and lagged effects in multivariate time series analysis (Thorson *et al.*, 2024). This is defined through a generalized linear mixed model (GLMM) for a $T \times J$ matrix Y, where y_{tj} is the measurement in time t in {1,...,T} for variable j in {1,...,J}. This measurement

matrix can include missing values, and DSEM estimates a $T \times J$ matrix of latent states X for all modelled times and variables. The observation module for y_{tj} allows several data distributions (normal, Poisson, gamma - called families hereafter) to be used (top insert Fig. 1).

For each latent state (also called random effect), the user can specify the time series structure (autoregressive process) and the causal relationship (simultaneous or lagged) with other variables. This defines K relationships between variables encoded as coefficients in the path matrix P_k of dimension $J \times J$. The lag information about each relationship is stored in a $T \times T$ matrix L_k. These two matrices are used to build a joint path matrix P_{joint} of dimension $TJ \times TJ$ representing relationships between all variables at all time-steps (Eq. 1):

$$P_{joint} = \sum_{k=1}^{K}(L_k \otimes P_k), \qquad \text{Eq. 1}$$

where $\otimes$ is the Kronecker product.

Latent states are then estimated as a Gaussian Markov random field (GMRF, Rue and Held, 2005, Eq. 2):

$$vec(X) \sim GMRF(0, Q_{joint}), \qquad \text{Eq. 2}$$

where $Q_{joint} = \left(I - P_{joint}^T\right)V^{-1}(I - P_{joint})$ is a sparse $TJ \times TJ$ precision matrix constructed based on P_{join} , V is the sparse $TJ \times TJ$ matrix of exogenous variation for each variable, and the sparsity of Q_{joint} is limited by $P_{joint}^T P_{joint}$ and V.

Inference is done by maximizing the marginal log-likelihood approximated by integrating the random effects (X) out using the Laplace approximation (Skaug and Fournier, 2006), and asymptotic uncertainty is calculated using the generalized delta method as implemented in TMB (Kristensen *et al.*, 2016). Estimated parameters (fixed effects) include time series process error variances (σ_j^2) and correlations (ρ_j) in V, effect sizes of the causal relationships between two variables at lag (β) in P_{joint}, and, optionally, observation errors depending upon the specified family for the GLMM. The matrix of latent states X can naturally be projected into the future because of the lags, autocorrelation, and causal relationships X, either by augmenting X with additional rows for future years (for short-term forecasts) and/or sampling future years from the conditional-GMRF distribution (for end-of-century projections).

In addition to partial effects (β), the "total effect" of each variable on every other variable accounts for the impact of direct and indirect links and can be computed by summing the product of partial effect sizes along each causal path. Because of the lags, this computation can easily get complex to do "by hand" and we instead obtain total effect by extracting them from the total effect matrix of dimension $TJ \times TJ$ defined as: $(I - P_{joint})^{-1}$.

Finally, DSEM introduced an arrow-lag notation enabling easy specification of a wide range of statistical models including structural equation models (SEMs), dynamic factor analysis (DFA, Zuur *et al.*, 2003), autoregressive integrated moving average (ARIMA e.g., random walk, AR1, Box *et al.*, 2016), and vector autoregression (Thorson *et al.*, 2024). Because this flexible interface allows us to write out a variety of statistical configurations of time-variation, we adopt it as our interface for SCEAM and use it to explore and compare a range of models within the same framework as detailed below.

2.2 Conceptual steps to model coupling

Coupling DSEM to a population model requires the latter to have a state-space structure, which has become more common in the last years (Schaub *et al.*, 2024). For compatibility purposes, the population model should also use TMB (either directly or through RTMB, Kristensen, 2024). We list below the general steps necessary to combine DSEM and a population dynamics model. We use the causal diagram represented on the far-right column in Fig. 1 to illustrate these steps. Further details regarding code to write can be found in Table S1.1.

1. *Define random effects:* Add the vector P containing variables from the population model as additional column to the matrix Y (the DSEM time series data input), where P contains NA values representing missing values (that will later be estimated).
2. *Specify a structural causal model*: Specify the time series structure and causal relationship between time series A, B, C, D and P through the 'arrow interface' (third row in Fig. 1), and the observation error type (family) for each time series.
3. *Combine models:* Add all likelihood components to form a unique joint likelihood. As vector P has been moved to a column of Y, it will be estimated through the DSEM likelihood statement (Eq. 2) so its original likelihood component from the population model can be deleted. Adapt the population model code to connect P (now estimated through DSEM as part of X) with the rest of the population dynamics, in the hindcast and forecast period.
4. *Information workflow:* P is informed by time series of population observations as well as DSEM covariates A, B, C and D. Both data types can have missing values, which results in periods where P estimation is mostly informed by a certain data type. In the most recent years, or during the forecast period, no population observations are available, but P will be informed by previous observations of A, B, C and D because of lags and autoregressive processes. Notably, D in Fig. 1 has a forecast in the future (e.g., from an oceanographic model) which can inform P forecasts.

2.3 Case study

Walleye pollock recruitment in the Gulf of Alaska (*Gadus chalcogrammus*, hereafter "GOA pollock") is used as a case study to illustrate the benefits of coupling DSEM to a population

dynamics model. This stock has high recruitment variability, making it a good candidate for aiming to explain this variation through causal relationships with ecosystem variables (Monnahan *et al.*, 2023). More generally, forecasting recruitment is known to be an important, but complex and difficult task and there is a long history of developing methods to do so (Haltuch *et al.*, 2019; Ward *et al.*, 2024).

2.3.1 Assessment model overview

The stock is assessed using an age-structured population dynamics model covering the period 1970 to 2023. The modelled population includes individuals from age 1 to age 10+ (i.e., a plus group, or all individuals of age 10 and older). Recruitment in year y (billions of age-1 pollock on January 1st, $N_{1,y}$) is modelled as an independent random effect around the mean (Eq. 3):

$$N_{1,y} = e^{ln(\overline{R})+\varepsilon_y} \qquad \text{where } \varepsilon_y \sim N(0, \sigma_R), \qquad \text{Eq. 3}$$

with $ln(\overline{R})$ being the log of mean recruitment of the time series and σ_R the standard deviation of the recruitment process which for an iid model represents total variation in recruitment (i.e., there is no explained variation).

Standard formulations are used to represent the population dynamics, mortality and fishery catch (Fournier and Archibald, 1982; Hilborn and Walters, 1992; Quinn and Deriso, 1999). The model is fitted to time series of catch biomass, survey biomass indices and age- and length-compositions from the fishery and survey (Fig. 2). Several small adjustments were made from Monnahan et al. (2023) to ease the implementation (Supplementary Information S1.1). Model parameters are estimated by maximizing the marginal log-likelihood using the Laplace approximation implementation from TMB (Kristensen *et al.*, 2016).

2.3.2 Environmental indicators of stock productivity

The development of "Ecosystem and Socioeconomic Profiles" (ESPs) aims to enable the integration of a broad range of factors within the U.S. stock assessment processes to facilitate Ecosystem-Based Fisheries Management (EBFM) (Shotwell *et al.*, 2023b). ESP is based on data from several national initiatives, scientific literature, ecosystem surveys, process studies, and laboratory analyses to generate a set of standardized indicators that capture potential drivers of the dynamics of a given stock (Shotwell *et al.*, 2023b). The ESP consists of a suite of indicators relevant to stock processes and dynamics and is also a repository for process knowledge regarding recruitment and other important processes that may be important in a stock assessment context. A 'full' ESP report was first conducted for GOA pollock in 2019 (Shotwell *et al.*, 2019), and since then, yearly 'report cards' are produced that contain updated indicator statuses and trends (https://akesp.psmfc.org/). This process knowledge and indicator suite (fully presented in Shotwell *et al.* (2023a), restricted to relevant ones for the study, see Table 1) allowed us to identify a general

conceptual model of how ecosystem processes impact recruitment (Fig. 3a). Note that this diagram is a conceptual representation of these processes among several other possible representations.

2.3.3. Causal diagrams design

Generating causal diagrams for a study system requires gathering knowledge and hypotheses (Grace and Irvine, 2020; Arif and MacNeil, 2023). We started this process relying on the initial conceptual model (Fig. 3a) representing interactions between pollock recruitment and ecosystem processes, and the available data to inform these processes (e.g., ESP time series). Because previous uses of DSEM relied on a limited number of variables (eight as a maximum, Thorson *et al.*, 2024), we reduced the number of variables used in the conceptual model (Fig. 3a) by working with experts to identify the ESP time series with well-established mechanisms linked to recruitment (Table 1). Several iterations were required to identify candidate causal diagrams which performed well, a point we return to in the discussion. We developed three candidate diagrams called 'simple', 'moderate', and 'complex' causal diagrams (Fig. 3b) with the number of causal variables ranging from five to 10. Two differences between panel 3a and 3b are worth highlighting: (i) in the conceptual model (3a) icons represent processes evaluated for the GOA pollock stock as part of the initial ESP evaluation process which can be informed by multiples data sources, whereas in the alternative simplified models (3b) the icons represent the paired down indicator suite described in Table 1 that is highly relevant to recruitment based on expert knowledge; (ii) in the conceptual model (3a) the arrows are colored and represent hypotheses regarding the sign of the relationship, while in the alternative simplified models, no hypotheses regarding the sign is made, only the direction and lags are used as input within the modelling framework, which is why we used black arrows.

2.3.4 Model implementation and alternative configurations

Coupling the pollock stock assessment with the DSEM framework, resulting in SCEAM, was achieved as outlined in Section 2.2. Note that the age composition data were initially weighed using Francis methods (Francis, 2011) before coupling, but not in subsequent fits. All scripts and code used for the analysis are available online at https://github.com/jchampag/GOApollock/tree/dsem.

We used SCEAM to test seven different configurations of time-variation or environmental information to explain recruitment (Fig. 3b). The simplest two involved modelling recruitment as an iid and an AR1 process. Another configuration used linear regression to estimate a slope for a covariate hypothesized to affect recruitment. For this, several physical and biological indicator time series from the ESP dataset were tested as a covariate of recruitment with a lag of one year and that leading to the lowest AIC value was chosen, as is commonly done in the regression paradigm. The fourth configuration illustrates the use of a DFA-like modelling approach, where multiple environmental time series information is reduced to fewer latent states directly affecting

recruitment (Ward *et al.*, 2024), which is possible via the DSEM interface. The DFA structure used the same variables as the Moderate causal diagram and pooled them in two states causally linked to recruitment with a lag of 1 or 2 years. The remaining three configurations relied on hypothesized causal diagrams to inform recruitment, with different levels of complexity (Simple, Moderate, Complex) as developed in section 2.3.3.

All configurations do not link all ESP time series to recruitment; however, all ESP time series (Fig. 2) were kept in all models. We will refer to time series as either "active" when they are linked to recruitment or "background" when the time series is being modelled but is not linked to recruitment. We retained background time series in the model so that all models are fit to the same set of data and AIC can be used to compare among models. However, these background time series do not affect any other portion of the model. Therefore, they could be dropped from the model without affecting estimates for other model components.

ESP time series were all modelled as AR1 processes and fitted with normal families with a fixed standard deviation of 0.1 which corresponds to a CV of 10% for standardized variables; explorations of larger CVs indicated the results were insensitive to this choice. To allow comparison among the DFA and other configurations, some changes were made to the DFA configuration, which separate it from a traditional DFA: (i) variables are modelled as AR1 and not iid, (ii) the variance of variables is fixed close to 0 and not at 0, (iii) observation error of variables was chosen to be independent and identical (fixed at 0.1 as stated above).

2.4. Model performance and evaluation

All model configurations converged (small gradient and invertible Hessian) in less than two minutes. We used three metrics to evaluate the performance of each model configuration:

1. *Parsimony:* While making different assumptions about how environmental and biological variables affect recruitment, all configurations are fitted to the same data (Fig. 2) so that the marginal AIC can be compared. Marginal AIC is a measure of expected predictive performance of new data;
2. *Exogenous variance of recruitment*: we compared the unexplained recruitment variance $\sigma^2_{R,unexplained}$ value (contributing to V from Eq. 2), which is a partition of the total recruitment variance $\sigma^2_{R,total}$, not explained by other variables in the model. Explicitly: $\sigma^2_{R,total} = \sigma^2_{R,explained} + \sigma^2_{R,unexplained}$;
3. *Leave-future-out (LFO) cross-validation*: the predictive forecast performance of each model configuration was assessed by conducting a retrospective experiment where we excluded all data after year T and then forecasted recruitment for year T+1. We fitted 10 "retrospective peels" where the last year of data T={2022, 2021,…, 2013} and compared the recruitment forecast with the estimate arising when fitted to all data (T=2023) for the same model configuration (see Fig. S2.3 for an example). The overall predictive skill is computed as the Root Mean Standard Error (RMSE, equation in Supplementary

Information S.2.3). We compared RMSE when peeling only 5 years of data, or 10. LFO cross-validation is a measure of expected predictive performance for a decadal projection, a projection that is critically needed for fisheries management (Tolimieri and Haltuch, 2023).

These three performance metrics were used to select a 'best' configuration for which additional results will be shown.

2.5 Simulation experiment

For the 'best' configuration based on parsimony, residual variance, and cross-validation, a self-test simulation experiment was conducted to check for parameter unbiasedness. The overall idea of self-test simulation is to generate new data based on a fitted model, fit these new data with the same model configuration, and to check relative error between true and new estimated parameters values. Given that SCEAM couples two models, the procedure is more complicated and follows multiple steps: (i) generate new latent states (ESP and recruitment time series) using DSEM function dsem::simulate() to simulate 'true' time series for recruitment deviations and environmental and ecological variables given estimated linkages and variance parameters; (ii) simulate new assessment data conditioned on those new states; (iii) refit simulated data with its matching model; and (iv) calculate relative errors in DSEM parameters and quantities of interest such as total catch, recruitment, and stock spawning biomass. Relative error is calculated as $\frac{\widehat{\theta}_i}{\theta_i} - 1$, where θ_i is the true value for simulated dataset i and $\widehat{\theta}_i$ is the value estimated from fitting the model to simulated data. Given the sparsity (i.e., missing data) of the ESP time series (Fig. 2), we performed two simulation experiments, one where simulated ESP time series are as sparse as real one and a second called 'idealistic' without missing data.

3. Results

3.1 Model performance

The Moderate causal diagram configuration had the lowest marginal AIC by far (Fig. 3b and Table 2). It was better than the two other causal diagrams with simpler ($\Delta AIC = 12$) and more complex structures ($\Delta AIC = 13$), as well as regression ($\Delta AIC = 22$), iid ($\Delta AIC = 35$), AR1 ($\Delta AIC = 31$) and DFA ($\Delta AIC = 58$). It also had the lowest value for recruitment standard deviation ($\sigma_{R,unexplained}$, Table 2). This corresponds to a 69% reduction in unexplained recruitment variation between this configuration and the iid one, for example. Forecasting skill is assessed based on RMSE, indicating how well a configuration predicts one-year-ahead recruitment, with the lower the RMSE, the more accurate the prediction. Moderate and Complex causal diagrams show the best predictive skill (Table 2), with a slightly better value for the Moderate causal diagram when averaged over 10 peels.

3.2 Preferred configuration estimates and validation

The three performance indicators agreed that the Moderate causal diagram is our best configuration, and additional results will be shown for this configuration only. The signs of the estimated causal link effect sizes are consistent across the three alternative causal diagrams, but their values and *p*-value (from a two-sided Wald test) are not identical (Fig. 3b). This is also the case for estimated stock assessment parameters that are also close but not identical for the Moderate configuration and the iid one (Table S2.1).

We compare the estimates of recruitment deviations and ESP time series for the Moderate diagram configuration with the iid one to highlight how the imputed data are affected by the causal relationships (Fig. 4). Only the eight "active" ESP time series used in the Moderate configuration are shown (out of the 10 total series). The difference in estimation mainly occurs when there is missing data: the iid configuration fills these gaps with a mean value (and an interval that covers most historical estimates), whereas the Moderate configuration predicts a value aligning with causally connected time series and ultimately recruitment (the prediction of which is also driven by the age composition data). Recruitment deviations are similar for both configurations except for the most recent years (2016 and 2022-2023) and projection period where the estimates from the Moderate configuration are slower to return to mean because of the influence of the environmental and ecosystem data included in the model.

We computed the total effect of each variable on recruitment for the Moderate configuration, which allows quantification of how much a change in a variable would impact recruitment with all else held constant (Fig. 5). Because time series have an AR1 structure, a causal effect of a certain lag also has an impact at other lags, leading to total effects at lags larger than those hypothesized Fig. 3b (we only show lags {0,1,2} but they exist through $T-1$, with an exponential decay in the absolute effect with longer lags). The most influential time series are offshore YOY condition, with negative effects for both lags 1 and 2 and offshore YOY, and larvae with positive effects, particularly for lag 1. Other variables such as Adult Condition, Spring SST, Euphausiids and Wind, achieve a non-zero total effect on recruitment but with a wide uncertainty around this total effect, which prevents making strong inference regarding their effects.

The self-simulation experiment was conducted with 600 repetitions. Among these, 10% did not converge (maximal gradient >0.1) and were excluded from analysis. The computation of the relative errors shows no bias in the estimation of parameters and quantities of interest for management (e.g., catch, recruitment and SSB, Fig S2.1). Another self-simulation experiment based on 'idealistic data' (i.e., simulated variable time series with no NAs) had a 100% convergence rate and also showed no-bias in the estimates (Fig S2.2).

4. Discussion

This study presents the first fisheries stock assessment model to incorporate a structural causal model component. We hypothesized that this causal component could help to explain variation in population productivity and improve forecasting it in the future. DSEM's flexible interface allows us to write out a variety of statistical configurations to model time-variation in a population process and to compare them within the same modelling framework. We applied this approach to the GOA pollock stock, using it to explain recruitment variation using standard statistical approaches such as iid, AR1, linear regression with environmental covariate, DFA and three causal diagrams of varying complexity (Fig. 3b). Model configurations where recruitment was informed by a causal diagram performed better than other traditional approaches in terms of AIC, unexplained recruitment variance reduction, and leave-future-out cross-validation (Table 2). The Structural Causal Enhanced Assessment Model (SCEAM) framework offers a novel way to incorporate multiple correlated covariates and scientific expertise directly into stock assessments. By formalizing ecological hypotheses and embedding them in a flexible statistical framework, SCEAM provides a pathway for addressing key challenges in managing fish stocks under changing environmental conditions. We thus propose SCEAM as a general scientific and statistical framework for building next-generation ecosystem and climate-linked fisheries stock assessment models and progressing toward EBFM.

4.1 Technical challenges of SCEAM

Despite its promise, SCEAM introduces several new challenges in addition to the usual stock assessment caveats (Maunder and Piner, 2015). First, we used the DSEM framework, which currently does not support estimation of non-linear or threshold effects (Thorson *et al.*, 2024). Such effects are common in nature (Samhouri *et al.*, 2017), including pollock, which exhibit non-linear temperature-dependent survival and growth (Laurel *et al.*, 2016, 2018). Although extensions to DSEM to accommodate these effects are possible, they are currently incompatible with missing data and causal loops (Thorson *et al.*, 2024). Another important limitation is related to the assumption of stationarity of the mean and variance of the autoregressive covariate processes (Szuwalski and Hollowed, 2016). While this assumption appeared reasonable in our dataset (by visual inspection; Fig. 4), it may not hold generally, for example, in systems with long-term trends or increasing variability. We recommend further simulations to test the behavior of SCEAM when assumptions of stationarity are violated (e.g., drawing upon the econometrics literature regarding co-integration models, Johansen, 1995), but for now analysts should be cautious with variables that visually appear nonstationary. Second, a common concern when relating natural resource and environmental variables is that a relationship might not hold through time (Myers, 1998; Haltuch *et al.*, 2009; Punt *et al.*, 2014). For example, omitted variables or shifts in system dynamics may break the assumed causal links. Although structural causal models can incorporate more variables than regression and may be more robust to such issues, we did not perform simulation testing to evaluate this in our application.

An unexpected outcome of integrating population models with environmental covariate estimation in a state-space framework is that population data can inform environmental covariate predictions. For example, in our study, although we assume environmental and ecosystem variables inform recruitment, the age composition data also indirectly inform predictions of environmental and ecosystem variables before the start of the time series and in missing years (Fig. 4, blue line) when predictions for the iid configuration revert to the mean of the time series as AR processes do (Fig.4, red line). Note that these missing year predictions remain highly imprecise for all model configurations. Similar effects have been observed in other studies and are likely to be pronounced when indicator time series contain large data gaps, high observation error, inconsistencies with population data (Miller *et al.*, 2018; Correa *et al.*, 2023), or when causal pathways are weak (e.g., small effect size) or statistically unsupported. If this behavior is considered undesirable, it can be mitigated by assuming that environmental covariates are measured without error, such that there is no leverage from biological processes upon their estimated value in years with data.

The coupled structure of SCEAM also increases code complexity, particularly for implementing diagnostics such as residual analysis or simulations. Integrating SCEAM into existing or future stock assessment platforms (e.g., SS3, WHAM, SAM) will require careful design and testing. Additionally, the flexibility of causal diagrams may increase the risk of overfitting, especially when relying on model selection tools such as marginal AIC. Most of the nonconverged repetitions of our self-testing analysis had $\sigma_{R,unexplained}$ going to 0, meaning that recruitment deviations are fully explained by environmental and ecosystem variables. These cases could be examples of overfitting and should be investigated more in the future by conducting leave-one-out cross-validation (Yates *et al.*, 2023), or using a recent generic approximation to conditional AIC (Zheng *et al.*, 2024), which is designed to measure expected predictive performance for leave-one-out cross-validation designs (Thorson, 2024). Finally, our simulation testing focused on basic statistical behavior, and we recommend future work to conduct more extensive simulations to evaluate sensitivity to misspecification (in either the causal diagram or stock assessment model), violations of stationarity, observation error, and missing data.

4.2 Ecological interpretation of causal diagram for our case study

Our study demonstrates the benefits of including environmental and ecosystem covariates into a stock assessment model which align with recent applications (Miller *et al.*, 2016; du Pontavice *et al.*, 2022; Rogers *et al.*, 2025). However environmentally-linked SSAMs usually rely on regression-based methods while SCEAM uses a causal approach. Thus, we highlight several aspects of the causal diagram interpretation that may be less familiar. In a causal diagram, a driver can influence an outcome through both direct and indirect pathways. The estimates of these direct and indirect effects can have opposite signs, potentially leading to a null total effect. However, this result would not imply the exposure variable is not important for prediction or for driving dynamics of intermediate variables. Such effects would likely be missed in a standard regression analysis. Our causal diagrams included physical drivers (wind, SST), biological indicators (pollock body

condition, diet, prey abundance), and early observations of recruitment (larvae, offshore and nearshore YOY). However, these variables differ in interpretation. For instance, early recruitment indicators do not cause recruitment, rather they are early measurements of year-class strength. Including them allows for a more realistic representation of the system's causal structure at different life stages, helping bridge mechanistic understanding with observed outcomes, and improving short-term projections. Notably, almost all estimated effect sizes resulted in the expected sign, consistent with the literature and process studies presented in Table 1. For instance, when adult pollock are in better condition prior to spawning, recruitment is estimated to be higher, which is consistent with hypothesized maternal effects on reproductive investment and offspring success (McBride *et al.*, 2015). This highlights how expert-informed hypotheses can yield models that both reflect ecological understanding and perform well statistically. One exception was the negative effect of YOY condition on recruitment across all the three diagrams (Fig. 3b), which contrasted with our expectation that higher YOY body condition would improve survival to the next life stage and increase recruitment (Siddon *et al.*, 2013). This unexpected result may point to missing causal pathways in our causal diagram, creating bias in the estimated link. But the Moderate causal diagrams fit the environmental and ecosystem time series well (Fig. 4), reinforcing its predictive strength despite some counterintuitive pathways. Together, these findings underscore the importance of distinguishing predictive performance from causal interpretation and highlight the need for further investigation into the mechanisms or model structures contributing to unexpected results.

4.3 Future directions

The distinction between prediction and causal inference is particularly important in structural causal modelling, where additional steps are required to validate causal claims. First, one must assess whether the implied independencies in the causal diagram are consistent with the observational data (i.e., causal diagram-data consistency; Arif and MacNeil, 2023). This can be tested using the d-separation rule to evaluate conditional independence among variables (Pearl, 1988; Shipley, 2000, 2016; Thorson *et al.*, 2025). The second step uses the backdoor criterion to identify which variables must be directly measured to allow a causal effect to be identified from a given dataset (Pearl, 2009). Applying this criterion helps avoid common statistical biases such as confounding, overcontrol and collider bias (Arif and MacNeil, 2023). If the causal diagram contains a known, but unmeasured, variable that may confound the results, one can use the front-door criterion instead of the backdoor one (Pearl, 1995, 2009). Tools to perform these validation steps on time series are beginning to become available (Thorson *et al.*, 2025) and were beyond the scope of our study. However, they are critical for moving beyond prediction toward inference about underlying mechanisms. Together, these tools provide a foundation for extending SCEAM beyond its initial application to recruitment and for increasing the complexity of modelled processes in a principled way. Building on this framework, future iterations of SCEAM could extend its scope to capture time-varying dynamics across multiple population processes such as growth and natural mortality (Fig. 1, far right). This would involve the development of specific

causal diagrams for each process, potentially linked through shared environmental or ecological drivers (e.g., temperature or prey abundance). We are conscious of the challenges of modelling multiple time-varying population processes known as hard to estimate and sometimes correlated/confounded with each other (Punt, 2023). But we think that explicit modelling of shared drivers could account for part of this correlation in a transparent way. Other time-varying components commonly modelled in assessments such as fishery selectivity, survey catchability, and weight-at-age may also benefit from this framework, particularly where indicators can help explain observed trends. Advancing this line of work will require sustained interdisciplinary collaboration to incorporate expert opinion and careful attention to structural assumptions.

As the scope and complexity of causal modelling in stock assessments expands, realizing its full potential in practice will require institutional support and investment in modelling capacity. Broader adoption of causal analysis in stock assessments will require new workflows, tools, and training. Encouragingly, causal modelling for observational data is gaining traction in ecology, and recent reviews provide accessible entry points for the fisheries science community (e.g., Grace and Irvine, 2020; Arif *et al.*, 2022; Arif and MacNeil, 2023; Thorson *et al.*, 2024, 2025; Byrnes and Dee, 2025; Siegel and Dee, 2025). To guide future development and the more formal integration of causal reasoning into fisheries science, we propose a conceptual framework (Fig. 6) that outlines an idealized, iterative framework for incorporating causal analysis into ecosystem-linked stock assessments. While inspired by the general principles of explanatory modelling (e.g., Grace and Irvine, 2020), this framework is tailored to the structure, constraints, and practical needs of the fishery management process. Figure 6 highlights the cyclical relationship between data collection, synthesis of existing knowledge, causal diagram development, statistical modelling, and targeted process research. Initial causal diagrams are constructed from expert knowledge and empirical evidence, implemented within the SCEAM framework, and they are used to evaluate both stock dynamics and ecosystem linkages. Model results, including those indicating weak or uncertain causal pathways, then inform the design of future ecosystem monitoring or experiments, which in turn can be used to refine causal structure. This adaptive process supports the development of models that are not only statistically robust but also mechanistically grounded and responsive to management needs. Such a framework is well aligned with the principles of EBFM (Levin *et al.*, 2009; Link, 2010), which emphasize the importance of accounting for ecological interactions, environmental variability, and broader ecosystem drivers in the management of fish stocks.

4.4 Benefits and challenges of SCEAM for management

Causal diagrams are promising tools for stakeholder communication, as they have the potential to enhance transparency, support the sharing of knowledge across disciplines, and enable co-construction of the linkages in ecosystems. SCEAM would also directly respond to current mandates for including ecosystem and socioeconomic considerations within the estimation of optimum yield (Magnuson-Stevens Fishery Act, 2007). However, important challenges remain in

implementing SCEAM for advice production. A key difficulty lies in integrating changes in stock productivity into the existing science-to-management framework, which often relies on static reference points. While many agree that strong environmental linkages should be incorporated into stock assessment models (Link *et al.*, 2021), there is less consensus on using time-varying reference points or management targets (Berger, 2019; O'Leary *et al.*, 2020; Szuwalski *et al.*, 2023; Bessell-Browne *et al.*, 2024). In addition, rigid management systems may struggle to accommodate dynamic processes, even as evidence grows that ecosystem-based approaches can reduce risk of collapse under increased climate variability (Holsman *et al.*, 2020). In the near-term, time-varying parameters could inform management by reporting stock status relative to "Dynamic B0" (the spawning biomass that would have occurred without fishing to provide context for drivers of population dynamics) and to improve short-term forecasts. In the longer-term, models based on SCEAM could form the basis for the operating models used to evaluate candidate harvest control rules (whether they are static or dynamically respond to ecosystem drivers). In the near-term, robustly constructed causal diagrams (i.e., based on expert input, process studies, and model validation) not yet approved for direct use in management, could still support management advice qualitatively. For example, they could help contextualize indicator trends in syntheses or reviews that summarize environmental and ecosystem dynamics (e.g., ecosystem overview [ICES, 2024] or risk tables [Dorn and Zador, 2020)]).

Regardless of their use in management, stock assessment platforms must first be built to incorporate complex, causal relationships among variables and key population processes. Our SCEAM framework provides proof of concept and demonstrates superior statistical performance than the status quo, serving as a critical first step towards progress in better managing fisheries under varying environmental conditions.

Acknowledgements

We would like to thank all the contributors of the GOA pollock ESP (Matt Callahan, Lauren Rogers, Patrick Ressler, Cole Monnahan, Ben Laurel, Kerim Aydin) for their timely response to requests and questions regarding their data, report summaries, and manuscripts. We thank Jennifer Bigman and Eric Ward for feedback on an earlier draft. Efficient calculation of GMRF likelihoods is made possible by Kasper Kristensen in the development of TMB and RTMB. This publication is funded by the Cooperative Institute for Climate, Ocean, & Ecosystem Studies (CICOES) under NOAA Cooperative Agreement NA20OAR4320271.

Data availability statement

The code underlying this manuscript may be accessed on GitHub https://github.com/jchampag/GOApollock/tree/dsem. The stock assessment data for the case study is also available in the repository, the ESP data are available upon request.

Conflict of interest statement

The authors have no conflicts of interest to declare.

References

Aeberhard, W. H., Mills Flemming, J., and Nielsen, A. 2018. Review of state-space models for fisheries science. Annual Review of Statistics and Its Application, 5: 215–235.

Arcos, D. F., Cubillos, L. A., and P. Núñez, S. 2001. The jack mackerel fishery and El Niño 1997–98 effects off Chile. Progress in Oceanography, 49: 597–617.

Arif, S., Graham, N. A. J., Wilson, S., and MacNeil, M. A. 2022. Causal drivers of climate-mediated coral reef regime shifts. Ecosphere, 13: e3956.

Arif, S., and MacNeil, M. A. 2023. Applying the structural causal model framework for observational causal inference in ecology. Ecological Monographs, 93: e1554.

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. doi.org/10.3389/fmars.2020.00703.

Berger, A. M. 2019. Character of temporal variability in stock productivity influences the utility of dynamic reference points. Fisheries Research, 217: 185–197.

Bessell-Browne, P., Punt, A. E., Tuck, G. N., Burch, P., and Penney, A. 2024. Management strategy evaluation of static and dynamic harvest control rules under long-term changes in stock productivity: A case study from the SESSF. Fisheries Research, 273: 106972.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. 2016. Time Series Analysis: Forecasting and Control. Wiley Series in Probability and Statistics. Wiley, Hoboken, New Jersey.

Boyce, D. G., Petrie, B., and Frank, K. T. 2021. Fishing, predation, and temperature drive herring decline in a large marine ecosystem. Ecology and Evolution, 11: 18136–18150.

Byrnes, J. E. K., and Dee, L. E. 2025. Causal Inference With Observational Data and Unobserved Confounding Variables. Ecology Letters, 28: e70023.

Cadigan, N. G. 2016. A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. Canadian Journal of Fisheries and Aquatic Sciences, 73.

Cadigan, N. G., Albertsen, C. M., Zheng, N., and Nielsen, A. 2024. Are state-space stock assessment model confidence intervals accurate? Case studies with SAM and Barents Sea stocks. Fisheries Research, 272: 106950.

Clark, A. T., Ye, H., Isbell, F., Deyle, E. R., Cowles, J., Tilman, G. D., and Sugihara, G. 2015. Spatial convergent cross mapping to detect causal relationships from short time series. Ecology, 96: 1174–1181.

Correa, G. M., Monnahan, C. C., Sullivan, J. Y., Thorson, J. T., and Punt, A. E. 2023. Modelling time-varying growth in state-space stock assessments. ICES Journal of Marine Science, 80: 2036–2049.

Deyle, E. R., Fogarty, M., Hsieh, C., Kaufman, L., MacCall, A. D., Munch, S. B., Perretti, C. T., *et al.* 2013. Predicting climate effects on Pacific sardine. Proceedings of the National Academy of Sciences, 110: 6430–6435.

585 Dorn, M. W., and Zador, S. G. 2020. A risk table to address concerns external to stock assessments
586 when developing fisheries harvest recommendations. Ecosystem Health and Sustainability, 6:
587 1813634.
588 du Pontavice, H., Miller, T. J., Stock, B. C., Chen, Z., and Saba, V. S. 2022. Ocean model-based
589 covariates improve a marine fish stock assessment when observations are limited. ICES Journal of
590 Marine Science, 79: 1259–1273.
591 Feng, J., Zhang, F., Sun, M., Chen, Y., and Zhu, J. 2025. Impacts of temporal growth variability
592 on fisheries stock assessment in changing oceans: a case study of Eastern Atlantic skipjack.
593 Frontiers in Marine Science, 12: 1555106.
594 Fournier, D., and Archibald, C. P. 1982. A general theory for analyzing catch at age data. Canadian
595 Journal of Fisheries and Aquatic Sciences, 39: 1195–1207.
596 Francis, R. I. C. C. 2011. Data weighting in statistical fisheries stock assessment models. Canadian
597 Journal of Fisheries and Aquatic Sciences, 68: 1124–1138.
598 Giron-Nava, A., Ezcurra, E., Brias, A., Velarde, E., Deyle, E., Cisneros-Montemayor, A. M.,
599 Munch, S. B., *et al.* 2021. Environmental variability and fishing effects on the Pacific sardine
600 fisheries in the Gulf of California. Canadian Journal of Fisheries and Aquatic Sciences, 78: 623–
601 630.
602 Grace, J. B., and Irvine, K. M. 2020. Scientist's guide to developing explanatory statistical models
603 using causal analysis principles. Ecology, 101: e02962.
604 Haltuch, M. A., Punt, A. E., and Dorn, M. W. 2009. Evaluating the estimation of fishery
605 management reference points in a variable environment. Fisheries Research, 100: 42–56.
606 Haltuch, M. A., Brooks, E. N., Brodziak, J., Devine, J. A., Johnson, K. F., Klibansky, N., Nash,
607 R. D. M., *et al.* 2019. Unraveling the recruitment problem: A review of environmentally-informed
608 forecasting and management strategy evaluation. Fisheries Research, 217: 198–216.
609 Hernán, M., and Robins, J. M. 2023. Causal inference: what if. Taylor and Francis, Boca Raton.
610 Hilborn, R., and Walters, C. J. 1992. Quantitative fisheries stock assessment: choice, dynamics,
611 and uncertainty. Chapman and Hall, New York. 570 p.
612 Holsman, K. K., Haynie, A. C., Hollowed, A. B., Reum, J. C. P., Aydin, K., Hermann, A. J.,
613 Cheng, W., *et al.* 2020. Ecosystem-based fisheries management forestalls climate-driven collapse.
614 Nature Communications, 11: 4579.
615 ICES. 2024. ICES Ecosystem Overviews Technical Guidelines. Version 4. ICES Guidelines and
616 Policies - Advice Technical Guidelines. https://doi.org/10.17895/ices.pub.22059803.
617 Johansen, S. 1995. Likelihood-based inference in cointegrated vector autoregressive models.
618 Advanced texts in econometrics. Oxford University Press, Oxford ; New York. 267 pp.
619 Karp, M. A., and Vieser, J. 2024. An Accounting of Approaches to Incorporate Environmental,
620 Climate, and Ecosystem Impacts in Fishery Stock Assessments in the U.S. from 2003–2023.
621 NOAA Tech. Memo., NMFS-F/SPO-251. http://spo.nmfs.noaa.gov/tech-memos/.
622 Kell, L. T., Pilling, G. M., and O'Brien, C. M. 2005. Implications of climate change for the
623 management of North Sea cod (*Gadus morhua*). ICES Journal of Marine Science, 62: 1483–1491.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. 2016. **TMB** : Automatic Differentiation and Laplace Approximation. Journal of Statistical Software, 70. http://www.jstatsoft.org/v70/i05/ (Accessed 1 May 2025).

Kristensen, K. 2024. RTMB: "R" Bindings for "TMB. Available from https://CRAN.R - project.org/package=RTMB.

Kuczynski, L., Chevalier, M., Laffaille, P., Legrand, M., and Grenouillet, G. 2017. Indirect effect of temperature on fish population abundances through phenological changes. PLOS ONE, 12: e0175735.

Laubach, Z. M., Murray, E. J., Hoke, K. L., Safran, R. J., and Perng, W. 2021. A biologist's guide to model selection and causal inference. Proceedings of the Royal Society B: Biological Sciences, 288: 20202815.

Laurel, B. J., Spencer, M., Iseri, P., and Copeman, L. A. 2016. Temperature-dependent growth and behavior of juvenile Arctic cod (*Boreogadus saida*) and co-occurring North Pacific gadids. Polar Biology, 39: 1127–1135.

Laurel, B. J., Copeman, L. A., Spencer, M., and Iseri, P. 2018. Comparative effects of temperature on rates of development and survival of eggs and yolk-sac larvae of Arctic cod (*Boreogadus saida*) and walleye pollock (*Gadus chalcogrammus*). ICES Journal of Marine Science, 75: 2403–2412.

Lefcheck, J. S. 2016. PIECEWISESEM : Piecewise structural equation modelling in R for ecology, evolution, and systematics. Methods in Ecology and Evolution, 7: 573–579.

Levin, P. S., Fogarty, M. J., Murawski, S. A., and Fluharty, D. 2009. Integrated Ecosystem Assessments: Developing the Scientific Basis for Ecosystem-Based Management of the Ocean. PLOS Biology, 7: e1000014.

Lima, M., Canales, T. M., Wiff, R., and Montero, J. 2020. The Interaction Between Stock Dynamics, Fishing and Climate Caused the Collapse of the Jack Mackerel Stock at Humboldt Current Ecosystem. Frontiers in Marine Science, 7: 123.

Lindegren, M., Möllmann, C., Nielsen, A., and Stenseth, N. C. 2009. Preventing the collapse of the Baltic cod stock through an ecosystem-based management approach. Proceedings of the National Academy of Sciences, 106: 14722–14727.

Link, J. S. 2010. Ecosystem-Based Fisheries Management: Confronting Tradeoffs. Cambridge University Press, Cambridge ; New York. 207 p.

Link, J. S., Karp, M. A., Lynch, P., Morrison, W. E., and Peterson, J. 2021. Proposed business rules to incorporate climate-induced changes in fisheries management. ICES Journal of Marine Science, 78: 3562–3580.

Litzow, M. A., Abookire, A. A., Duffy-Anderson, J. T., Laurel, B. J., Malick, M. J., and Rogers, L. A. 2022. Predicting year class strength for climate-stressed gadid stocks in the Gulf of Alaska. Fisheries Research, 249: 106250.

Magnuson-Stevens Fishery Conservation and Management Reauthorization Act of 2006. Pub. L. No. 109-479, 120 Stat. 3575 (2007).

Maunder, M. N., and Piner, K. R. 2015. Contemporary fisheries stock assessment: many issues still remain. ICES Journal of Marine Science, 72: 7–18.

McBride, R. S., Somarakis, S., Fitzhugh, G. R., Albert, A., Yaragina, N. A., Wuenschel, M. J., Alonso-Fernández, A., *et al.* 2015. Energy acquisition and allocation to egg production in relation to fish reproductive strategies. Fish and Fisheries, 16: 23–57.

McClatchie, S., Goericke, R., Auad, G., and Hill, K. 2010. Re-assessment of the stock–recruit and temperature–recruit relationships for Pacific sardine (*Sardinops sagax*). Canadian Journal of Fisheries and Aquatic Sciences, 67: 1782–1790.

McElreath, R. 2018. Statistical Rethinking: A Bayesian Course with Examples in R and Stan. Chapman and Hall/CRC. https://www.taylorfrancis.com/books/9781315362618 (Accessed 1 May 2025).

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

Miller, T. J., Hare, J. A., and Alade, L. A. 2016. A state-space approach to incorporating environmental effects on recruitment in an age-structured assessment model with an application to southern New England yellowtail flounder. Canadian Journal of Fisheries and Aquatic Sciences, 73: 1261–1270.

Miller, T. J., and Hyun, S.-Y. 2018. Evaluating evidence for alternative natural mortality and process error assumptions using a state-space, age-structured assessment model. Canadian Journal of Fisheries and Aquatic Sciences, 75: 691–703.

Miller, T. J., O'Brien, L., and Fratantoni, P. S. 2018. Temporal and environmental variation in growth and maturity and effects on management reference points of Georges Bank Atlantic cod. Canadian Journal of Fisheries and Aquatic Sciences, 75: 2159–2171.

Monnahan, C. C., Adams, G. D., Ferris, B. E., Shotwell, S. K., McKelvey, D. R., and McGowan, D. W. 2023. Assessment of the walleye pollock stock in the Gulf of Alaska. In Stock assessment and fishery evaluation report for the Gulf of Alaska. North Pacific Fishery Management Council, 1007 West 3rd Ave., Suite 400, L92 Building, 4th floor, Anchorage, AK 99501.

Myers, R. A. 1998. When Do Environment–recruitment Correlations Work? Reviews in Fish Biology and Fisheries, 8: 285–305.

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

O'Leary, C. A., Thorson, J. T., Miller, T. J., and Nye, J. A. 2020. Comparison of multiple approaches to calculate time-varying biological reference points in climate-linked population-dynamics models. ICES Journal of Marine Science, 77: 930–941.

Pearl, J. 1988. Embracing causality in default reasoning. Artificial Intelligence, 35: 259–271.

Pearl, J. 1995. Causal inference from indirect experiments. Artificial Intelligence in Medicine, 7: 561–582.

Pearl, J. 2009. Causal inference in statistics: An overview. Statistics Surveys, 3. https://projecteuclid.org/journals/statistics-surveys/volume-3/issue-none/Causal-inference-in-statistics-An-overview/10.1214/09-SS057.full (Accessed 5 December 2024).

Pershing, A. J., Alexander, M. A., Hernandez, C. M., Kerr, L. A., Le Bris, A., Mills, K. E., Nye, J. A., *et al.* 2015. Slow adaptation in the face of rapid warming leads to collapse of the Gulf of Maine cod fishery. Science, 350: 809–812.

Punt, A. E., A'mar, T., Bond, N. A., Butterworth, D. S., De Moor, C. L., De Oliveira, J. A. A., Haltuch, M. A., *et al.* 2014. Fisheries management under climate and environmental uncertainty: control rules and performance simulation. ICES Journal of Marine Science, 71: 2208–2220.

Punt, A. E. 2023. Those who fail to learn from history are condemned to repeat it: A perspective on current stock assessment good practices and the consequences of not following them. Fisheries Research, 261: 106642.

Quinn, T. J., and Deriso, R. B. 1999. Quantitative fish dynamics. Biological resource management series. Oxford University Press, New York. 542 p.

Rogers, L. A., Wilson, M. T., Duffy-Anderson, J. T., Kimmel, D. G., and Lamb, J. F. 2021. Pollock and "the Blob": Impacts of a marine heatwave on walleye pollock early life stages. Fisheries Oceanography, 30: 142–158.

Rogers, L. A., Monnahan, C. C., Williams, K., Jones, D. T., and Dorn, M. W. 2025. Climate-driven changes in the timing of spawning and the availability of walleye pollock (*Gadus chalcogrammus*) to assessment surveys in the Gulf of Alaska. ICES Journal of Marine Science, 82: fsae005.

Rue, H., and Held, L. 2005. Gaussian Markov Random Fields. Chapman and Hall/CRC. https://www.taylorfrancis.com/books/9780203492024 (Accessed 8 April 2025).

Samhouri, J. F., Andrews, K. S., Fay, G., Harvey, C. J., Hazen, E. L., Hennessey, S. M., Holsman, K., *et al.* 2017. Defining ecosystem thresholds for human activities and environmental pressures in the California Current. Ecosphere, 8: e01860.

Schaub, M., Maunder, M. N., Kéry, M., Thorson, J. T., Jacobson, E. K., and Punt, A. E. 2024. Lessons to be learned by comparing integrated fisheries stock assessment models (SAMs) with integrated population models (IPMs). Fisheries Research, 272: 106925.

Shipley, B. 2000. A New Inferential Test for Path Models Based on Directed Acyclic Graphs. Structural Equation Modeling: A Multidisciplinary Journal, 7: 206–218.

Shipley, B. 2016. Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference with R. Cambridge University Press. doi.org/10.1017/CBO9781139979573

Shotwell, S. K., Dorn, M. W., Deary, A. L., Fissel, B., Rogers, L. A., and Zador, S. 2019. Appendix 1A. Ecosystem and Socioeconomic Profile of the Walleye Pollock stock in the Gulf of Alaska - Report Card. In Stock assessment and fishery evaluation report for the Gulf of Alaska. North Pacific Fishery Management Council, 1007 West 3rd Ave., Suite 400, L92 Building, 4th floor, Anchorage, AK 99501.

Shotwell, S. K., Ferriss, B. E., Monnahan, C. C., Oke, K., Rogers, L. A., and Zador, S. 2023a. Appendix 1A. Ecosystem and Socioeconomic Profile of the Walleye Pollock stock in the Gulf of Alaska. Report Card. https://apps-afsc.fisheries.noaa.gov/Plan_Team/2023/GOApollock_appA.pdf

741 Shotwell, S. K., Blackhart, K., Cunningham, C., Fedewa, E., Hanselman, D., Aydin, K., Doyle,
742 M., *et al.* 2023b. Introducing the Ecosystem and Socioeconomic Profile, a Proving Ground for
743 Next Generation Stock Assessments. Coastal Management, 51: 319–352.
744 Siddon, E. C., Heintz, R. A., and Mueter, F. J. 2013. Conceptual model of energy allocation in
745 walleye pollock (*Theragra chalcogramma*) from age-0 to age-1 in the southeastern Bering Sea.
746 Deep Sea Research Part II: Topical Studies in Oceanography, 94: 140–149.
747 Siegel, K., and Dee, L. E. 2025. Foundations and Future Directions for Causal Inference in
748 Ecological Research. Ecology Letters, 28: e70053.
749 Skaug, H. J., and Fournier, D. A. 2006. Automatic approximation of the marginal likelihood in
750 non-Gaussian hierarchical models. Computational Statistics & Data Analysis, 51: 699–709.
751 Stock, B. C., and Miller, T. J. 2021. The Woods Hole Assessment Model (WHAM): A general
752 state-space assessment framework that incorporates time- and age-varying processes via random
753 effects and links to environmental covariates. Fisheries Research, 240: 105967.
754 Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., and Munch, S. 2012. Detecting
755 Causality in Complex Ecosystems. Science, 338: 496–500.
756 Szuwalski, C. S., and Hollowed, A. B. 2016. Climate change and non-stationary population
757 processes in fisheries management. ICES Journal of Marine Science, 73: 1297–1305.
758 Szuwalski, C. S., Hollowed, A. B., Holsman, K. K., Ianelli, J. N., Legault, C. M., Melnychuk, M.
759 C., Ovando, D., *et al.* 2023. Unintended consequences of climate-adaptive fisheries management
760 targets. Fish and Fisheries, 24: 439–453.
761 Thorson, J., Monnahan, C., and Rogers, L. 2025, March 17. Validating causal inference in time
762 series models with conditional-independence tests. https://ecoevorxiv.org/repository/view/8762/
763 (Accessed 20 June 2025).
764 Thorson, J. T., Monnahan, C. C., and Cope, J. M. 2015. The potential impact of time-variation in
765 vital rates on fisheries management targets for marine fishes. Fisheries Research, 169: 8–17.
766 Thorson, J. T., Hermann, A. J., Siwicke, K., and Zimmermann, M. 2021. Grand challenge for
767 habitat science: stage-structured responses, nonlocal drivers, and mechanistic associations among
768 habitat variables affecting fishery productivity. ICES Journal of Marine Science, 78: 1956–1968.
769 Thorson, J. T., Andrews, A. G., Essington, T. E., and Large, S. I. 2024. Dynamic structural
770 equation models synthesize ecosystem dynamics constrained by ecological mechanisms. Methods
771 in Ecology and Evolution, 15: 744–755.
772 Thorson, J. T. 2024. Measuring complexity for hierarchical models using effective degrees of
773 freedom. Ecology, 105: e4327.
774 Tolimieri, N., and Haltuch, M. A. 2023. Sea-level index of recruitment variability improves
775 assessment model performance for sablefish *Anoplopoma fimbria*. Canadian Journal of Fisheries
776 and Aquatic Sciences: cjfas-2022-0238.
777 Tsai, C.-H., Munch, S. B., Masi, M. D., and Stevens, M. H. 2024. Empirical dynamic modeling
778 for sustainable benchmarks of short-lived species. ICES Journal of Marine Science, 81: 1209–
779 1220.

Vert-pre, K. A., Amoroso, R. O., Jensen, O. P., and Hilborn, R. 2013. Frequency and intensity of productivity regime shifts in marine fish stocks. Proceedings of the National Academy of Sciences, 110: 1779–1784.

Ward, E. J., Hunsicker, M. E., Marshall, K. N., Oken, K. L., Semmens, B. X., Field, J. C., Haltuch, M. A., *et al.* 2024. Leveraging ecological indicators to improve short term forecasts of fish recruitment. Fish and Fisheries: faf.12850.

Wilson, M. T., Mier, K. L., and Jump, C. M. 2013. Effect of region on the food-related benefits to age-0 walleye pollock (*Theragra chalcogramma*) in association with midwater habitat characteristics in the Gulf of Alaska. ICES Journal of Marine Science, 70: 1396–1407.

Wilson, M. T., and Laman, N. 2021. Interannual variation in the coastal distribution of a juvenile gadid in the northeast Pacific Ocean: The relevance of wind and effect on recruitment. Fisheries Oceanography, 30: 3–22.

Xu, H., Miller, T. J., Hameed, S., Alade, L. A., and Nye, J. A. 2018. Evaluating the utility of the Gulf Stream Index for predicting recruitment of Southern New England-Mid Atlantic yellowtail flounder. Fisheries Oceanography, 27: 85–95.

Yates, L. A., Aandahl, Z., Richards, S. A., and Brook, B. W. 2023. Cross validation for model selection: A review with examples from ecology. Ecological Monographs, 93: e1557.

Ye, H., Beamish, R. J., Glaser, S. M., Grant, S. C. H., Hsieh, C., Richards, L. J., Schnute, J. T., *et al.* 2015. Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. Proceedings of the National Academy of Sciences, 112. https://pnas.org/doi/full/10.1073/pnas.1417063112 (Accessed 23 January 2025).

Zheng, N., Cadigan, N., and Thorson, J. T. 2024. A note on numerical evaluation of conditional Akaike information for nonlinear mixed-effects models. arXiv. https://arxiv.org/abs/2411.14185 (Accessed 1 May 2025).

Zuur, A. F., Tuck, I. D., and Bailey, N. 2003. Dynamic factor analysis to estimate common trends in fisheries time series. Canadian Journal of Fisheries and Aquatic Sciences, 60: 542–552.

Tables

Table 1: Environmental and ecosystem variables used in our modelling framework. This is a subset of the Ecosystem and Socioeconomic Profile (ESP) presented in Shotwell *et al.* (2023a) relevant to our work. In Shotwell *et al.* (2023a), each indicator is associated with a contact name.

Variable name	Description of data and proposed mechanism relating to recruitment
Spring SST	Spring (April-May) daily sea surface temperatures (SST) for the western and central (combined) GOA from the NOAA Coral Reef Watch Program (https://coralreefwatch.noaa.gov/). Temperature directly affects metabolic rates for pollock early life stages and indirectly affects pollock through ecosystem processes (Shotwell *et al.*, 2019).
Wind	Mean springtime (April-May) north/south surface wind strength from the National Data Buoy Center for site B-AMAA2 located in the NE Kodiak Archipelago (https://www.ndbc.noaa.gov/station_history.php?station=amaa2). Northerly surface winds are hypothesized to retain larvae and juveniles in areas that favor survival (Wilson and Laman, 2021).
Euphausiids	Summer euphausiid abundance from the Alaska Fishery Science Center (AFSC) acoustic survey for the Kodiak core survey area. Euphausiids are an energy-rich prey source for juvenile pollock (Wilson *et al.*, 2013).
Fall adult condition	Fall body condition for adults from the pollock fishery sampled by observers. Body condition of adults prior to spawning may affect reproductive output and success (McBride *et al.*, 2015).
Larvae	Spring pollock larvae catch-per-unit-of-effort (CPUE) from the EcoFOCI spring survey. Larval CPUE reflects reproductive output and early life stage mortality and can be an early indicator of year-class strength (Rogers *et al.*, 2021).
Offshore YOY	Summer young-of-the-year (YOY) pollock catch-per-unit-of-effort (CPUE) from the EcoFOCI summer survey. Relative abundance of YOY pollock in pelagic habitat over the shelf reflects reproductive output and cumulative mortality through the first summer (Litzow *et al.*, 2022).
Nearshore YOY	Summer catch-per-unit-of-effort (CPUE) of young-of-the-year (YOY) pollock from the AFSC beach seine survey in the Kodiak region. Relative abundance of YOY pollock in nearshore habitat has been associated with year-class strength (Litzow *et al.*, 2022).
YOY Condition	Summer body condition for young-of-the-year (YOY) pollock from EcoFOCI summer survey. Fish with better body condition have greater energetic reserves to survive the first winter (Siddon *et al.*, 2013).
Euph Diet	Proportion-by-weight of euphausiids in the diets of juvenile (age 1+) GOA pollock from summer bottom-trawl surveys. Increased euphausiids in juvenile diets may reflect higher availability of euphausiids as prey, including for YOY pollock.
Copepod	Summer large copepods from the EcoFOCI summer survey. Large copepods are an important prey source for YOY pollock (Wilson *et al.*, 2013).

Table 2: Models results across configurations, as shown in Fig. 3b. Configurations are described in terms of numbers of parameters, value of the joint negative log-likelihood, marginal AIC, value of the recruitment standard deviation ($\sigma_{R,unexplained}$, indicating the unexplained part of recruitment variability) and RMSE. RMSE is the Root Mean Square Error of projected recruitment (in billions) one year ahead compared to estimated recruitment, computed and averaged for 5 or 10 peeled years. Skill-prediction analysis was not performed for the DFA configuration.

Configuration name	**No. parameters**	**-log-likelihood**	ΔAIC	$\sigma_{R,unexplained}$	**RMSE 5y**	**RMSE 10y**
iid	313	639	35	1.01	3.7	3.6
AR1	314	636	31	0.94	4.1	4.3
Regression	314	632	22	0.687	3.1	9.0
DFA	314	649	58	0.90	--	--
Simple causal diagram	318	622	12	0.587	2.7	3.0
Moderate causal diagram	322	612	0	0.558	1.8	2.5
Complex causal diagram	324	617	13	0.588	1.7	2.7

Figure legends

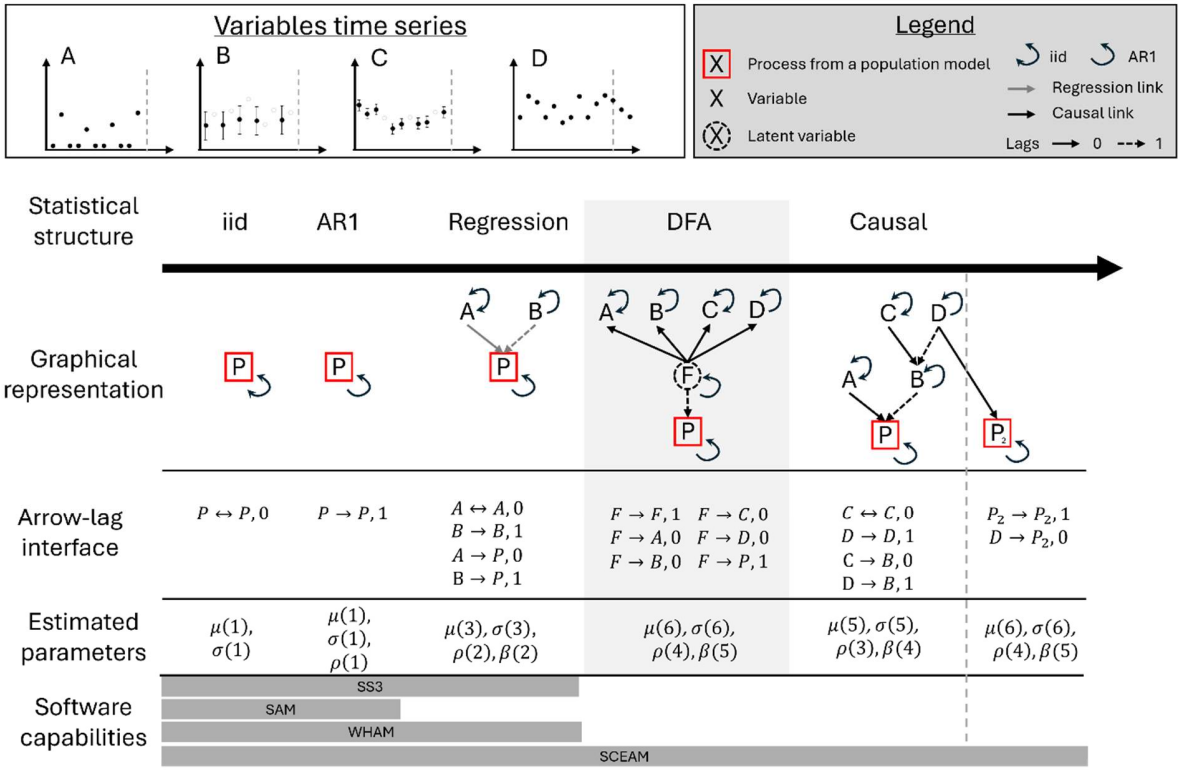


Figure 1. Overview of the various options (columns) to represent time-variation in a population process (P in the red square): random deviation around a mean (iid), autoregressive structure of order one (AR1), regression on environmental variables, Dynamic Factor Analysis (DFA) and causal. Each statistical structure has a graphical representation, the number of estimated parameters, and existing software able to estimate it: Stock Synthesis (SS3, Methot and Wetzel, 2013); SAM (Nielsen and Berg, 2014), WHAM (Stock and Miller, 2021), and Structural Causal Enhanced Assessment Model (SCEAM, this study). Some options include the use of latent states (F) and/or covariates (A, B, C, D) and the top insert illustrates the variety of data types that SCEAM can handle (A-C: observations with or without NAs and observation error; D: output from a model with forecast). The arrow-lag interface row highlights the additional code needed to move from one option to another within the SCEAM framework (except for the causal column where the code adds on the code of regression not DFA, as indicated by the gray box). The estimated parameters row details the type of fixed effects: mean (μ), standard deviation (σ), correlation coefficient (ρ), effect size (β) and their number in parenthesis.

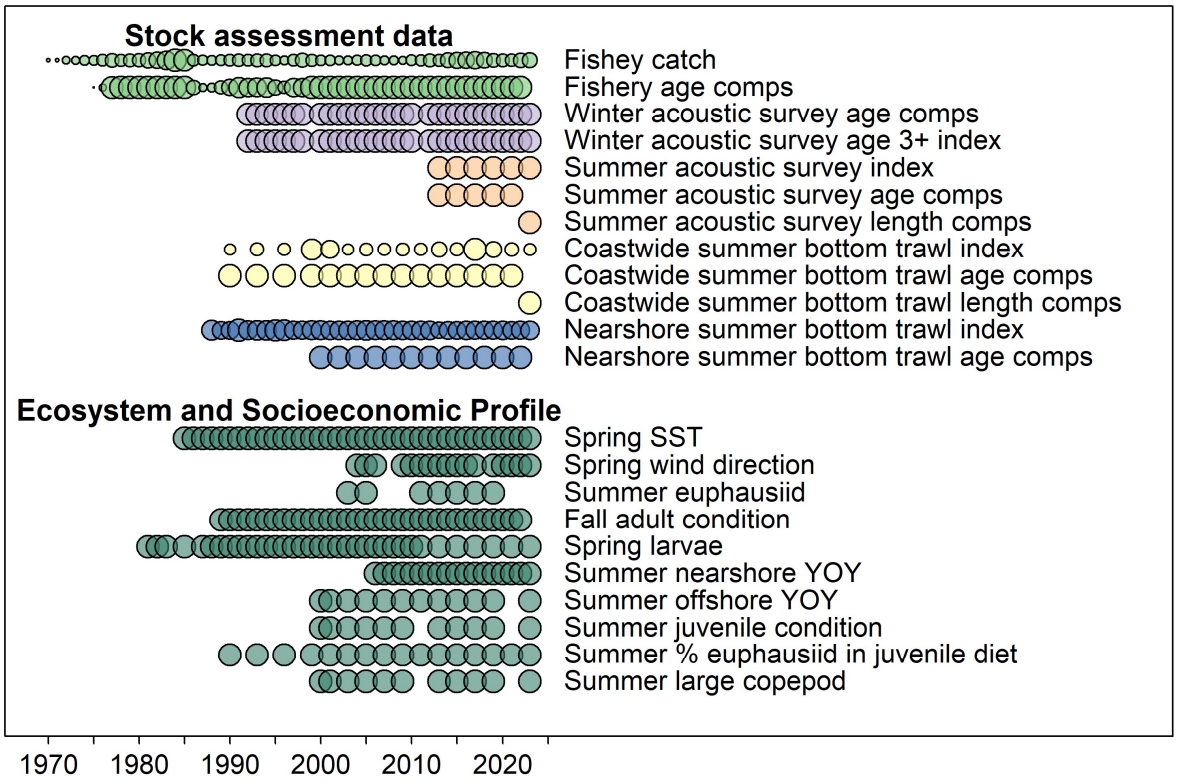
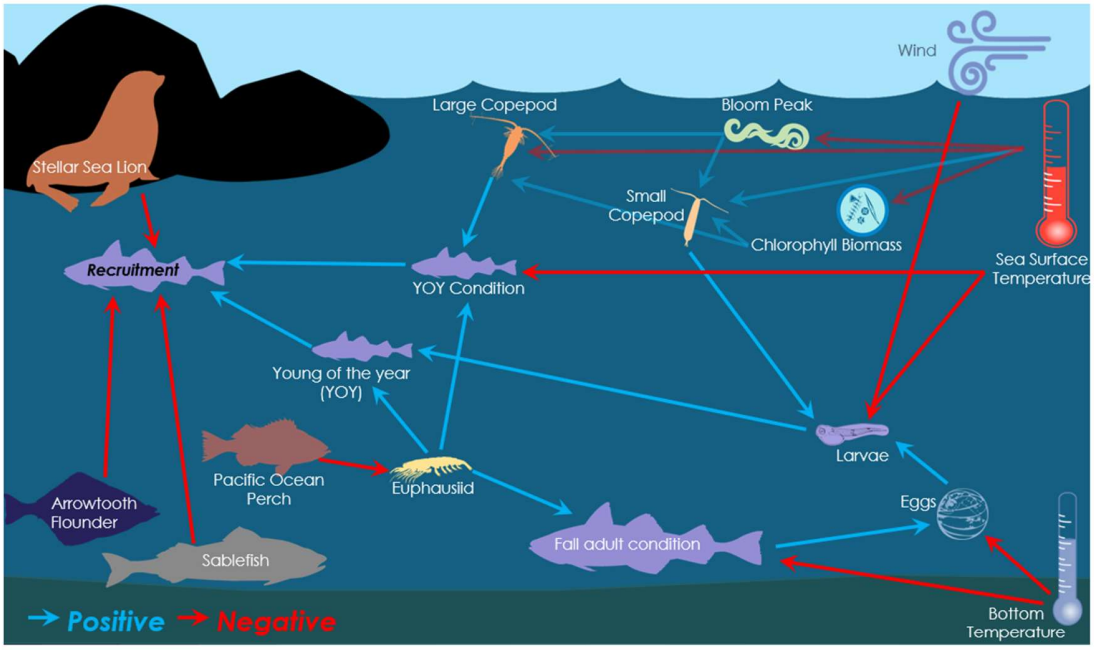


Figure 2. Overview of data sources for the GOA pollock case study including stock assessment data and environmental and ecosystem variables used in the Ecosystem and Socioeconomic Profile (ESP). Colors indicate different data type. Circle sizes are relative to values within a row, with larger circles indicating larger catches, smaller coefficients of variation for abundance indices and ESP time series, and larger effective sample sizes for compositions (comps). YOY is young of the year, SST is sea surface temperature.

(a) Initial conceptual model



(b) Alternative simplified models

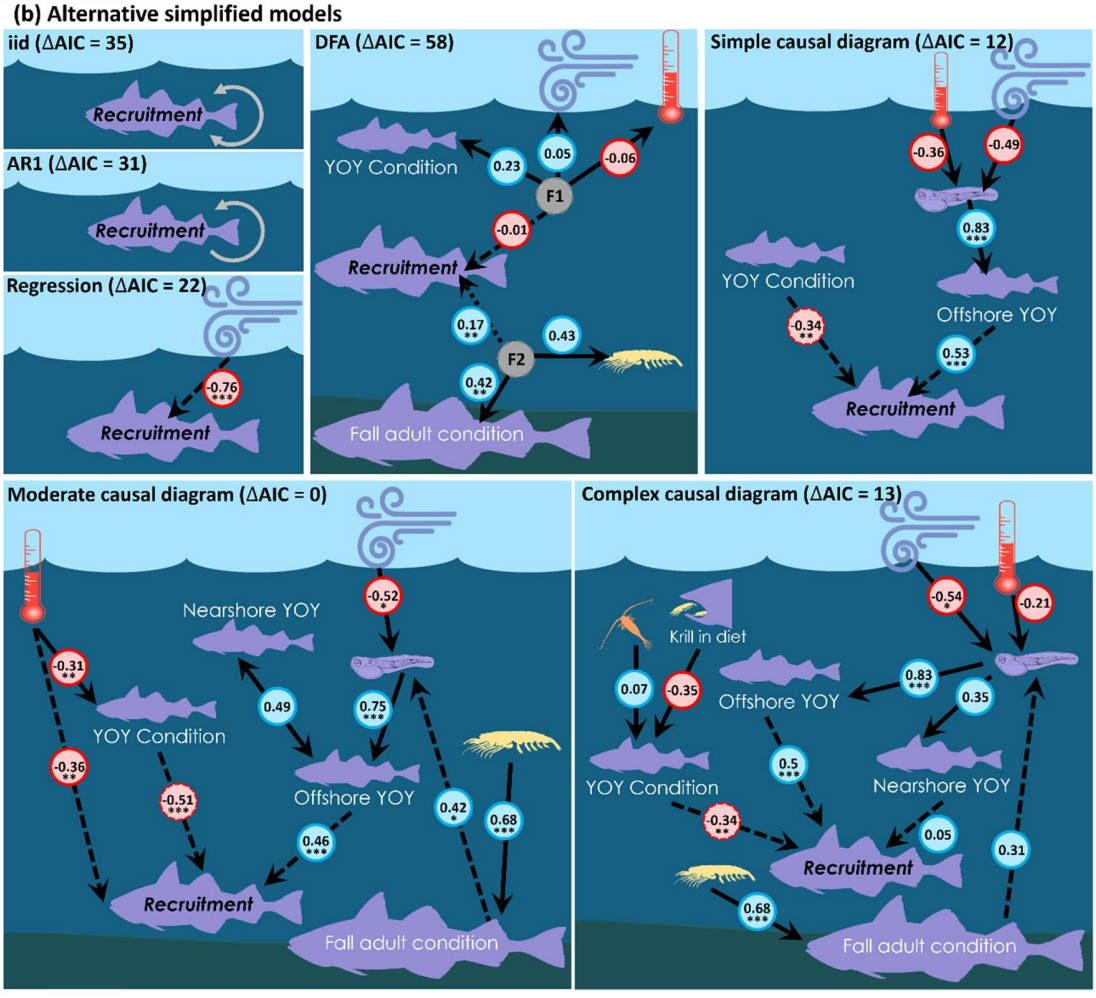
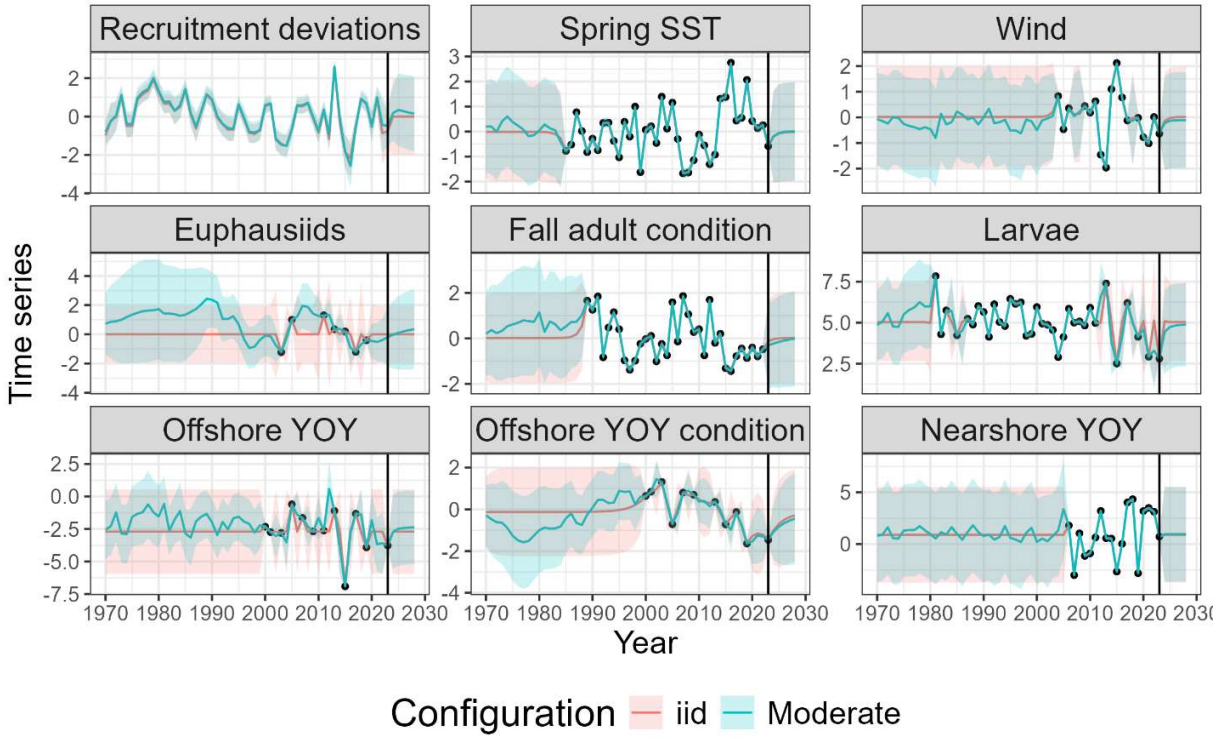


Figure 3: (a) Initial conceptual model summarizing ecosystem processes affecting the early life stages of GOA pollock until recruitment (age 1). Each arrow represents a causal link; for example, the relationship between two variables where a change in the exposure variable implies a change in the outcome variable being pointed to. Arrows color indicate hypothesized signs for the relationship: blue for positive, red for negative. Transparent arrows represent links for which mechanism is less established and/or known. (b) Alternative simplified models explored with SCEAM. The top left text indicates the model's name and the difference in the marginal Akaike Information Criterion (*AIC)* between a given model and the model with the lowest *AIC* (ΔAIC, Table 2) is reported in the parentheses. Black arrows indicate the hypothesized causal links, with plain arrows being simultaneous links and dashed ones lagged by one year (note that the small dash arrow in the Dynamic Factor Analysis (DFA) panel is a lag of 2 years, see text). These causal hypotheses are required as a model input. Red and blue circled numbers indicate the estimated values of causal links and stars the significance of a *p*-value from a two-sided Wald test where * is <0.05 , ** is <0.03 and *** is <0.01. The sign of the causal relationship for circles with dashed lines is opposite from our *a priori* expectation. In the DFA panel, the grey circles represent latent variables.

858



859
860 **Figure 4**: Estimated values for recruitment deviations and eight covariates (a subset of the indicators used in the
861 Moderate causal diagram) between two configurations (colors), shown as estimates (lines) and 95% confidence
862 intervals (ribbons). The data are shown as black points and the black vertical line marks the start of the projection
863 period. Most variables are scaled to have a mean of zero and variance of one, except Larvae, Offshore and Nearshore
864 young of the year which are in log scale. Configuration names correspond to Fig. 3b and Table 2.
865

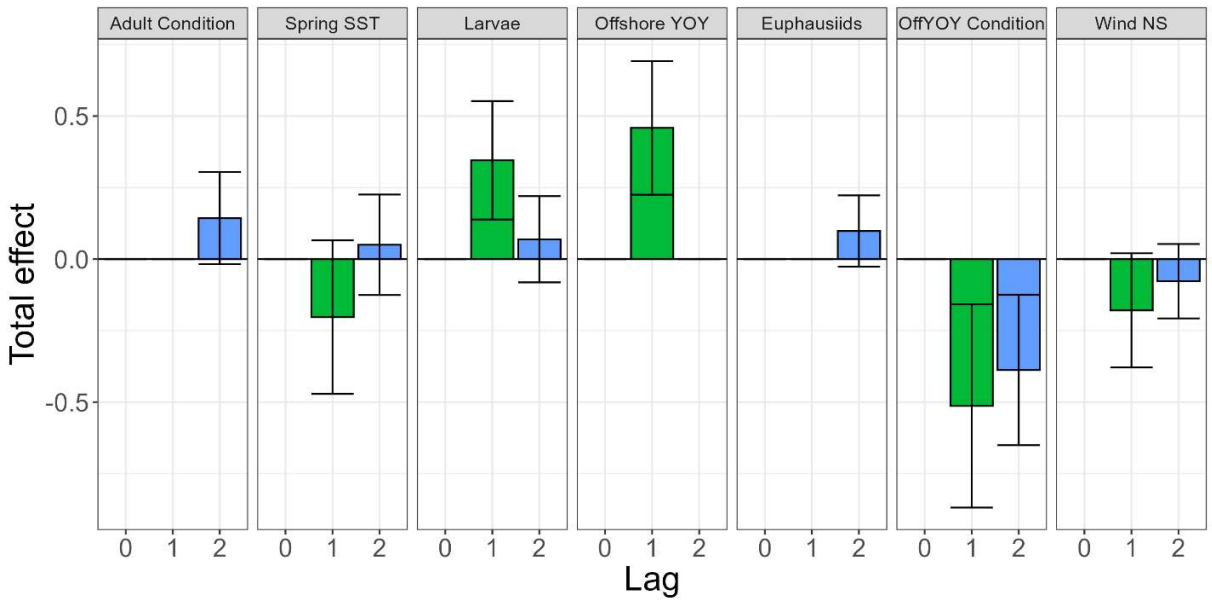
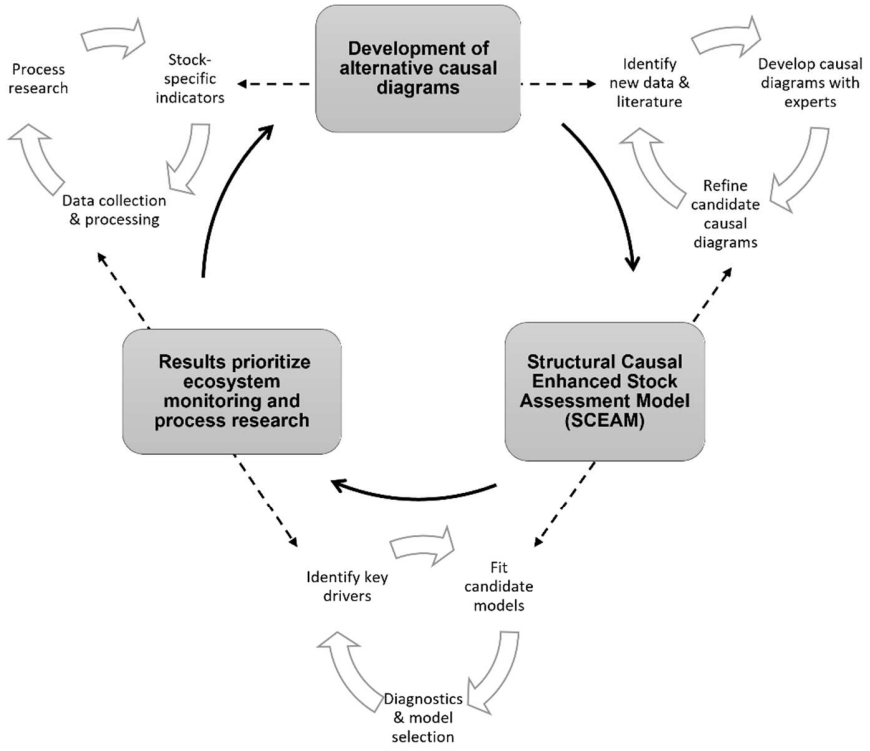


Figure 5: Total effects of the environmental variables on GOA pollock recruitment for various time lags (in year, colored by time-lag) for the selected Moderate model configuration. The total effect value accounts for the direct and indirect effects of causal links. 95% confidence intervals are shown as line range. A positive total effect indicates an increase in the covariate would lead to an increase in recruitment.



873
874 **Figure 6.** Representation of the ideal development of causal diagram within the stock assessment framework and
875 associated ecosystem monitoring and process research.
876

Supplementary Information

Supplementary Information 1: SCEAM details

S1.1 Equations of the GOA pollock stock assessment model

The full details of the model structure can be found in Appendix 1C of Monnahan *et al.* (2023). Several notable changes were made from the 2023 assessment model for this study. First, vaguely informative priors were placed on selectivity parameters to stabilize estimation in the simulation studies. Specifically, $N(0,2)$ priors were put on descending slope parameters in log space, $N(0,3)$ priors on initial inflection points, and $N(10,3)$ on descending inflection points. Further details can be found in the online model file in the link given in the main text. Second, the age-1 and age-2 indices from the Shelikof index were removed completely from the model presented here. This was done because subsequent to the 2023 model they were deemed to have strong and unreliable impacts on estimates of recruits. This was also done in the 2024 operational model (Monnahan *et al.*, 2024).

S1.2. Pseudo-code block for DSEM integration into stock assessment model

Table S1.1: Pseudo-code blocks required to integrate DSEM into future stock and ecosystem models

Code type	Purpose	Example
Logical code (R)	User creates the covariates table covering stock assessment + projection years	y_tj =data.frame(P_dev= NA, A, B, C, D)

User interface (R)	User creates his causal structure using expressive arrow-lag notation	my_sem = “
		# time series structure
		A<->A,0,
		B->B,1,
		C<->C,0,
		D->D,1,
		P<->P,1
		
		#causal relationships
		A->P,0,
		C->B,0,
		D->B,1,
		B->P,1”

User interface (R)	User builds DSEM-related object through the dsem() function	obj_dsem=fit_dsem(tsdata=ts(y_tj), family=rep(‘fixed’, ncol(y_tj)), run_model=FALSE, use_REML=FALSE, sem =my_sem)
Logical code (R)	User constructs data objects for coupled model	data = c(input_assess$data, obj_dsem$tmb_inputs$dat) pars= c(input_assess$pars, obj_dsem$tmb_inputs$parameters) map= c(input_assessmap, obj_dsemtmb_inputs$map)

Logical code (R)	User specifies parameters to be estimated or pre-specified	pars$sigmaP <- NULL pars$meanP <- NULL
Statistical code (TMB)	User combines DSEM and assessment TMB code and defines new parameters to be estimated	vector<Type> P_dev=x_tj.col(0);
Statistical code (TMB)	Model calculates joint negative log-likelihood for causal variables given GMRF	jnll_dsem -= loglik_tj.sum() jnll_dsem += GMRF(Q_kk)(x_tj - xhat_tj - delta_k);
Statistical code (TMB)	Model calculates log-likelihood for assessment (loglik elements depending on model structure) and creates a global objective function to be optimized by combining assessment and dsem likelihood.	objfun = -sum(loglik) + jnll_dsem;

Supplementary information 2: Additional model results and validation

S2.1 Model results

Table S2.1 Estimated fixed effect and standard error (SE) for assessment parameters for the Moderate causal diagram contrasted with the iid configuration results.

Parameter	Type	Name	Moderate		iid	
			Estimate	SE	Estimate	SE
Mean log age 1 recruitment (billions)	Recruitment	log_mean_recruit	1.248	0.215	1.306	0.16
Descending slope of fishery selectivity	Selectivity	log_slp2_fsh_mean	0.808	0.395	0.833	0.42
Descending inflection point of fishery selectivity	Selectivity	inf2_fsh_mean	9.733	0.144	9.758	0.14
Descending slope of Shelikof selectivity	Selectivity	log_slp2_srv1	0.622	0.54	0.608	0.54
Descending inflection point of Shelikof selectivity	Selectivity	inf2_srv1	9.897	0.372	9.905	0.38
Ascending slope of NMFS BT selectivity	Selectivity	log_slp1_srv2	-0.387	0.214	-0.405	0.21
Ascending inflection point of NMFS BT selectivity	Selectivity	inf1_srv2	3.749	0.581	3.829	0.6
Ascending slope of ADF&G selectivity	Selectivity	log_slp1_srv3	0.533	0.112	0.528	0.11
Ascending inflection point of ADF&G selectivity	Selectivity	inf1_srv3	4.23	0.228	4.235	0.23
Ascending slope of summer AT selectivity	Selectivity	log_slp2_srv6	0.316	0.473	0.32	0.47
Ascending inflection point of summer AT selectivity	Selectivity	inf2_srv6	7.76	0.745	7.75	0.74
Catchability for NMFS BT	Catchability	log_q2_mean	-0.234	0.095	-0.234	0.1
Catchability for summer AT	Catchability	log_q6	-0.355	0.1548	-0.367	0.15

896

897

S2.2 Self-test simulations

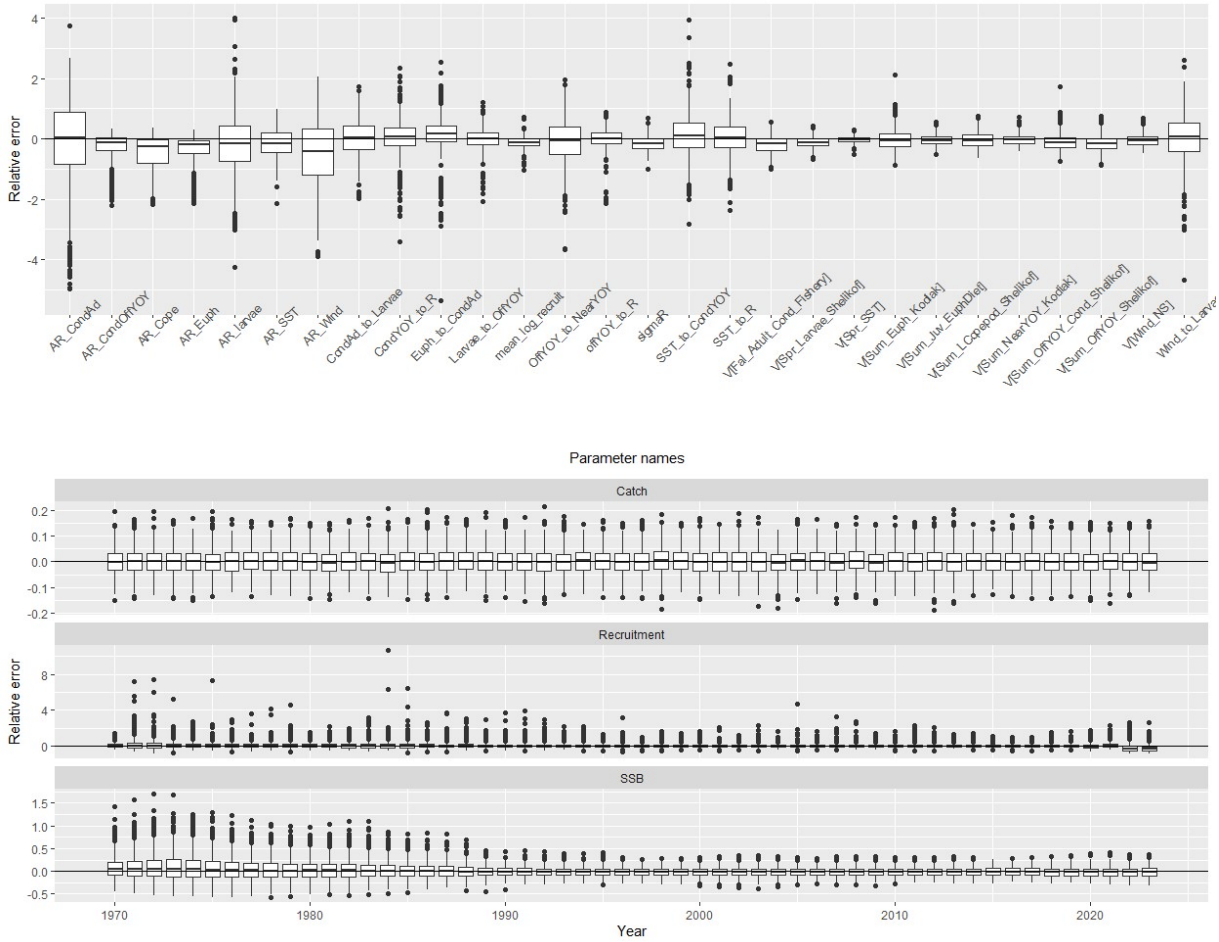


Figure S2.1: Self-test simulation results for the Moderate model configuration (600 simulations with NA in the input data). Top panel is relative error on DSEM and mean log recruitment parameters, bottom panel on the assessment quantities.

The ESP time series used as input in our model contained a lot of missing values which can make convergence difficult to reach. To evaluate if it was happening, we also tested a self-test simulation design where the simulated ESP time series have no NAs. We refer to this as an ‘idealistic” case and perform identical analysis of self-test results. For this idealistic case, 100% of the repetition converged and parameters and quantities show no bias (Fig. S2.2).

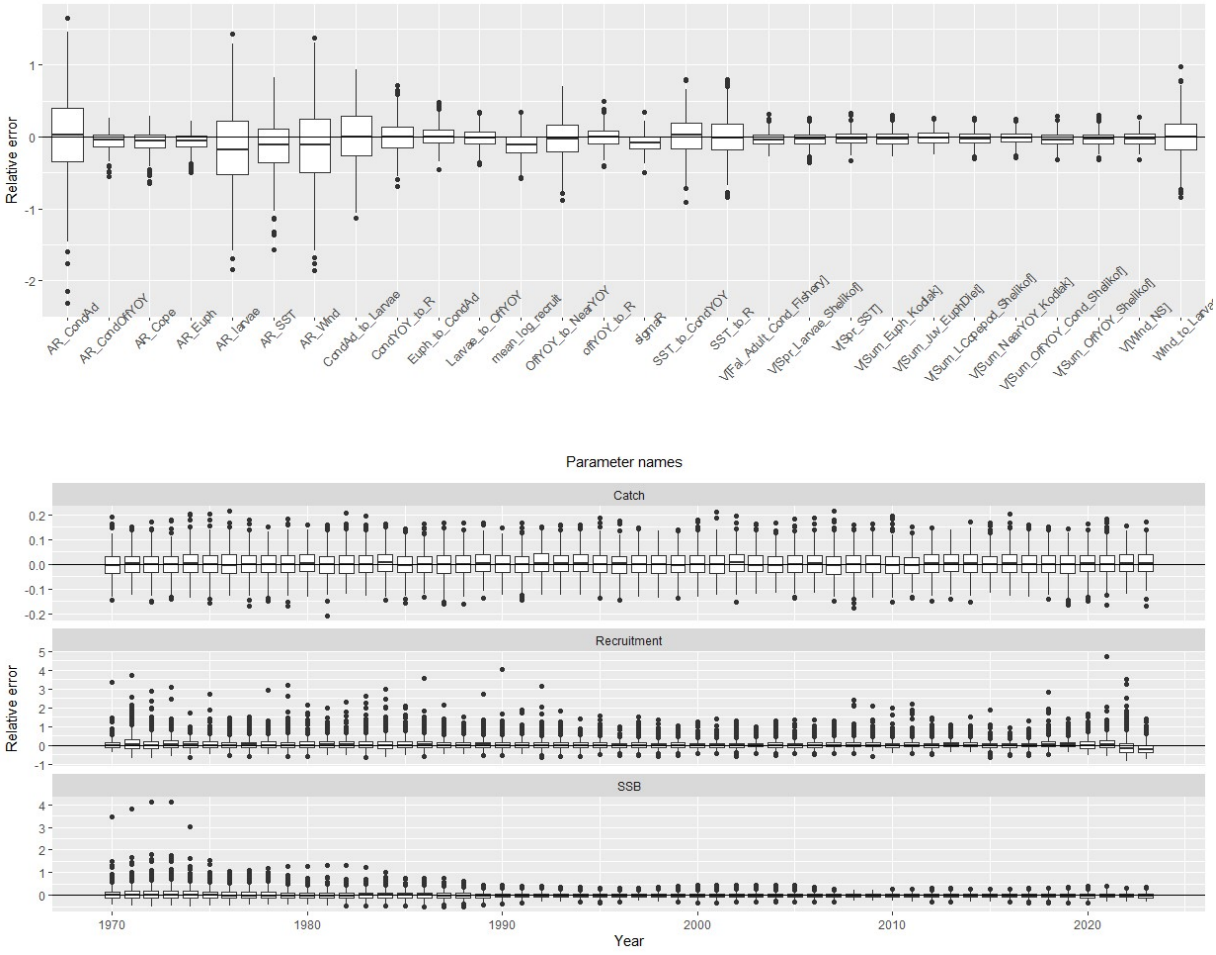


Figure S2.2: self-test simulation results for the Moderate model configuration (600 simulations with no NA in the input ESP time series - idealistic case). Top panel is relative error on DSEM and mean log recruitment parameters, bottom panel on the assessment quantities.

S2.3 Predictive skill-testing

Predictive skill testing is described in section 2.4. The Root Mean Square Error (RMSE) of recruitment was computed as follows (equation S2.1):

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\widehat{N}_{1,t+1} - N_{1,t+1})^2}$$

where $\widehat{N}_{1,t+1}$ is the recruitment forecast in year $t+1$ with a model peeled to year t and $N_{1,t+1}$ the recruitment forecast in year $t+1$ with a complete model (no peeled data).

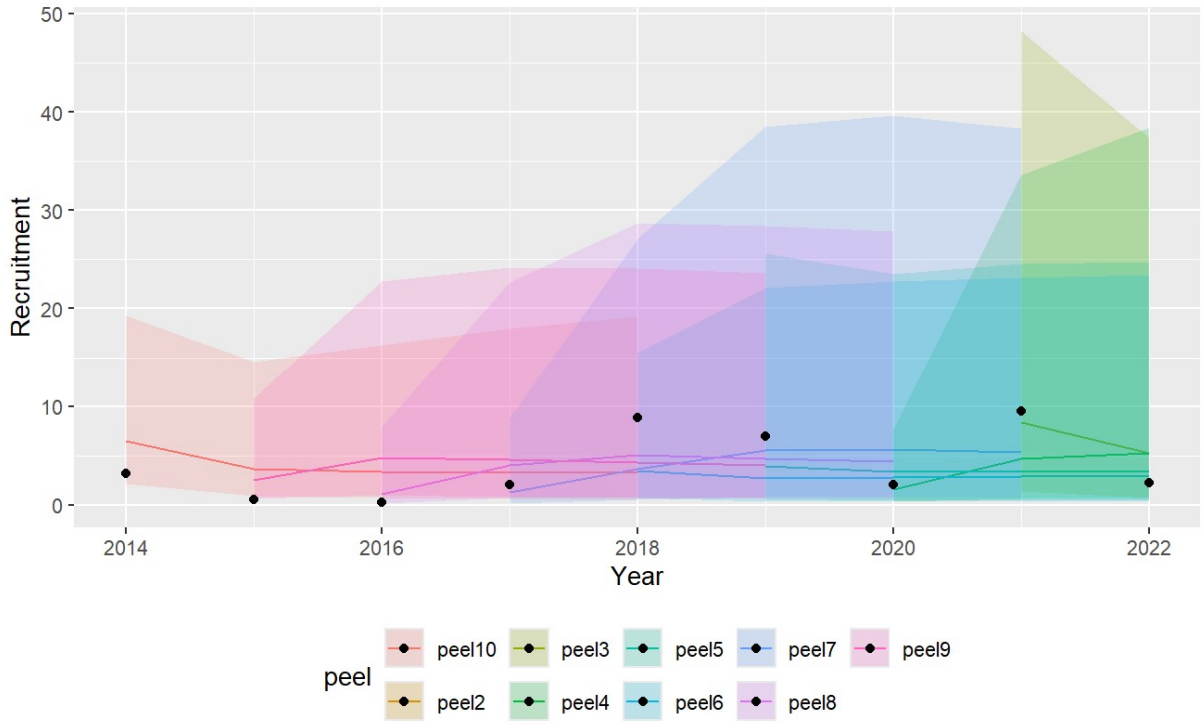


Figure S2.3: Illustration of the predictive skill-testing. For each peel (color) model predictions (lines) are compared to recruitment estimated by the complete model (i.e., model with no peeled data, black points). Ribbons indicate the 95% confidence intervals.