Smoothed dynamic factor analysis for identifying trends in multivariate time series

Eric J. Ward[1], Sean C. Anderson[2], Mary E. Hunsicker[3], Mike A. Litzow[4]

[1]Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd E, Seattle WA, 98112, USA

[2]Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, BC, V6T 6N7, Canada

[3]Fish Ecology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd E, Seattle WA, 98112, USA

[4]Shellfish Assessment Program, Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 301 Research Court. Kodiak, AK, 99615, USA

Open Research Statement: Code to replicate these analyses is in our repository (https://github.com/fate-ewi/gpdfa) and as an R package on CRAN, (https://cran.r-project.org/web/packages/bayesdfa/index.html). All data analyzed are previously published, but also included in our repository.

Abstract

Ecological processes are rarely directly observable, and for most systems, variance parameters must be estimated from data. State space models have become widely used in the environmental sciences, particularly for time series data, because of their ability to simultaneously estimate multiple sources of variation (process or natural variability, and variance attributed to observations, associated with measurement and sampling errors). A common state space approach for using multivariate time series to identify underlying signals is dynamic factor analysis (DFA). The conventional DFA model is flexible in that unseen processes are modeled as random walks. Whereas this may be suitable for some situations, random walks may be too flexible for other cases. In this paper, we introduce a new class of models, where latent processes are modeled as smooth functions. We highlight two alternatives to the random walk approach, using basis splines and Gaussian predictive process models. These models are applied to two long-term datasets from the west coast of the United States: (1) a 35-year dataset of juvenile rockfishes from the west coast of the United States, and (2) a 39-year dataset of fisheries catches. Estimation and model selection is done in a Bayesian framework, with code provided as the bayesdfa R package. For both applications we find that the smooth trend models have higher predictive accuracy and yield more precise predictions, compared to the conventional approach.

Key words

Dynamic factor analysis, smooth spline, B-spline, Gaussian process, Bayesian modeling, Stan

Introduction

Ecological data can be characterized by multiple sources of variability, including stochastic natural variation, and errors associated with data collection (observation, sampling, and measurement errors). Disentangling these sources of variability is often challenging, and necessitates the use of statistical methods, such as state space models. These approaches have become ubiquitous in ecology, particularly for time-series data (Auger-Méthé et al. 2020)—in part because these models allow researchers to make inferences about ecological processes that are not directly observable. Applications of these models include estimating population change over time (Clark and Bjørnstad 2004), movement dynamics (Patterson et al. 2008), and understanding spatiotemporal variation (Anderson and Ward 2019).

Estimating multiple sources of variation in state space models is numerically complex, and can be constrained explicitly or implicitly in ecological models via model assumptions. For example, discrete-time state-space models of population trajectories generally assume latent population size n_t at time t can be approximated by an autoregressive process in log-space, $x_{t+1} = f(x_t) + \epsilon_t$, where $f()$ represents some function, $x_t = \log(n_t)$, and ϵ_t are normally distributed process deviations representing stochastic variability of the natural system (Dennis et al. 2006). The autoregressive assumption is critical here; without such a constraint, the variance of the stochastic noise ϵ_t is not estimable in the presence of an observation or data model. Separating these sources of variability is critical to generating unbiased estimates of population trends or density dependence (Knape 2008). If inference is not dependent on parameters of ecological interest (e.g., growth rates, density dependence), a wide range of alternative semi-parametric approaches exist that can be used to model the trajectory of x_t, including generalized additive models (GAMs, Wood 2011) and Gaussian process models (Roberts et al. 2013). Because these models are not autoregressive with discrete time steps, the flexibility or 'wiggliness' of the model can be adjusted as part of the model fitting. In addition to their flexibility, these semi-parametric models may be better suited for situations when data are patchily distributed in time or unequally spaced, making estimation of process and observation errors more difficult.

Challenges posed by univariate time-series models also apply to multivariate models, with the additional complexity that the number of latent time series may be variable, $k = 1, \ldots, m$, where m is the number of time series observed. At one extreme, $k = m$, and each time series corresponds to a unique latent process. Motivating questions in analyzing these models include estimating correlated latent processes or trends, or estimating effects of environmental covariates (Hovel et al. 2017). At the other extreme, $k = 1$, where each time series represents repeated measurements of the same process, with optional offsets included for each time series (e.g., offsets allowing for differing detectability). Applications focused on estimating a single trend from multivariate data include the development of ecological indicators. Models with intermediate numbers of latent states $1 < k < m$ require mapping of time series to latent trends. These may be specified a priori (Ward et al. 2010) or estimated within the modeling framework using dimension reduction techniques.

Many statistical approaches have been proposed in recent years for clustering or estimating common signals in multivariate time series (Liao 2005). Examples include clustering based on similarities among time series features (Sardá-Espinosa 2019), identifying common patterns in the frequency domain (Holan and Ravishanker 2018), and clustering based on neural networks (Cherif et al. 2011). Application of these methods to ecological data has been limited, in part because many of these approaches identify clusters from raw data and ignore observation error. An alternative approach that has been used in ecology to map collections of multivariate time series to latent processes, while accounting for observation error, is dynamic factor analysis (DFA) (Zuur et al. 2003b, 2003a). DFA is an extension of factor analysis for time series data, and estimates a small number of unobserved processes ("trends"), that can describe observed data. Mapping time series to trends is done via estimated factor loadings—these allow each time series to be modeled as a mixture of estimated latent trends, rather than assigning each time series to a single trend.

To date, applications of DFA models in ecology and other fields have assumed that underlying trends are modeled as a random walk, $x_{t+1} = x_t + \epsilon_t$. The objective of this paper is to introduce a new class of DFA models based on smooth functions, instead of autoregressive processes. Recent work has highlighted the application of hierarchical GAMs for multiple data sources (Pedersen et al. 2019). These approaches are flexible and likely to provide similar inference to DFA for a single latent trend; however, these methods have not been extended to include more than one process. We illustrate two options for modeling smooth functions for latent trends: basis splines ('B-splines') and Gaussian process models. We compare both approaches to conventional autoregressive DFA models for two datasets on marine fishes from the west coast of the USA. All data and code for replicating our analysis are available on Github (*https://github.com/fate-ewi/gpdfa*), and in our existing R package 'bayesdfa' (Ward et al. 2019).

Methods

Dynamic Factor Model

The basic DFA model can be written as a multivariate state space model, consisting of a latent process model and observation or data model. In its simplest form, the process model is expressed as a random walk, $\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{w}_t$, where $\mathbf{w}_t \sim \text{MVN}(\mathbf{0}, \mathbf{Q})$. For identifiability constraints, the covariance matrix $\mathbf{Q}$ is generally constrained to be an identity matrix (Holmes et al. 2012, Zuur et al. 2003b). Additional features may be incorporated into the process model, including autoregressive or moving-average coefficients, covariates, or deviations that are more extreme than that of the normal distribution (Ward et al. 2019). The observation model in a DFA is expressed as a linear combination of trends $\mathbf{x}_t$ and a matrix of loadings coefficients $\mathbf{Z}$, $\mathbf{y}_t = \mathbf{Z}\mathbf{x}_t + \mathbf{B}\mathbf{d}_t + \mathbf{e}_t$. In addition to the trends and loadings, time-varying covariates $\mathbf{d}_t$ may be optionally included and linked to the observations through estimated coefficients $\mathbf{B}$. The vector $\mathbf{e}_t$ represents residual observation error, which is typically modeled as a diagonal matrix, $\mathbf{e}_t \sim \text{MVN}(0, \mathbf{R})$, although off-diagonal elements may be estimated (Holmes et al. 2020). Further details of the Bayesian implementation of the DFA model and extensions are provided in Ward et al. (2019).

Modeling trends as Gaussian processes

Conventional DFA models with trends modeled as random walks are flexible, but for some datasets these models may be too complex. As a first alternative to the random walk model, we treat the trends as a Gaussian process. A discrete-time Gaussian process model of trends treats the vector representing the k^{th} trend as a stochastic process, where $\mathbf{x}_k$ is drawn from a multivariate normal distribution. As data in a DFA are generally standardized (mean 0, standard deviation 1), we can assume the mean of each trend to be 0, and all inference about the Gaussian process centers around the covariance matrix, $\mathbf{x}_k \sim \text{MVN}(\mathbf{0}, \Sigma)$. Rather than estimate each element of Σ independently, smooth covariance functions or 'kernels' are chosen to represent the covariance between points in time (typical choices include the exponential, Gaussian, and Matérn functions). For the purpose of our DFA modeling, we adopt a Gaussian kernel. With this kernel, the covariance between points i and j at times t_i and t_j on trend k can be expressed as $\text{cov}(x_{i,k}, x_{j,k}) = \sigma_k^2 \exp\left(\frac{-(t_i - t_j)^2}{2\theta_k^2}\right)$, where σ_k controls the magnitude of variation, and θ_k controls how smoothly correlation decreases as time points become further apart. We allow each trend to have its own covariance parameters (θ_k, σ_k), allowing each to have differing degrees of smoothness. Because of potential computation issues in high dimensionality problems such as spatial models (Latimer et al. 2009, Anderson and Ward 2019), we also allow this Gaussian process model to be expressed as a Gaussian predictive process model. The difference between the predictive process approach and the full Gaussian process model is that instead of modeling the $\mathbf{x}_t$ themselves as random variables, random variables are modeled at a subset of locations $\mathbf{x}_k^*$ (referred to as 'knots') and projected to the locations of the data $\mathbf{x}_k$. If we assume $\mathbf{x}_k^* \sim \text{MVN}(\mathbf{0}, \Sigma^*)$, then this projection can be done as $x_k = {\Sigma'}_{k,k^*} {\Sigma^*}^{-1} x_k^*$, where the matrix ${\Sigma'}_{k,k^*}$ is the transpose of the matrix describing the covariance between x_k and x_k^*. The location of k^* can be spaced equally or depend on data; we assume that the k^* are equally spaced within each time series (with the endpoints also acting as knots).

Modeling trends as splines

As an alternative model of latent trends in a DFA, we use a series of smoothing functions, known as basis splines ('B-splines'). These models can be thought of as a special case of Gaussian process models (Kimeldorf and Wahba 1970), and offer flexibility similar to the more familiar generalized additive models (Wood 2011). B-splines are represented as a series of piecewise polynomial functions, where higher order polynomials result in more flexible curves (Hastie 1992). A common choice of the order of these polynomials is a cubic or 3rd degree, and will be the focus of our implementation for DFA. An additional input to B-splines is the locations of the control points (knots) between polynomial segments—more knots translates into a more flexible function, but also one with more parameters to estimate. We assume knots to be uniformly distributed over the time series. Uniform knot vectors may be appropriate for data collected at regular intervals, but for observations more patchily distributed in time, defining knots based on quantiles or other metrics may be warranted. Mathematically, modeling the trends in a DFA with B-splines can be expressed as a linear combination of the B-spline weights $\mathbf{B}$ and estimated coefficients $\mathbf{a}$, $\mathbf{x}_k = \mathbf{aB}$. The matrix $\mathbf{B}$ is generated from the raw data prior to estimation. In the DFA setting, $\mathbf{B}$ is shared across trends, but for trend-specific variability, we allow the coefficients $\mathbf{a}$ to have a trend-specific variance, $\mathbf{a}_k \sim \text{Normal}\left(0, \sigma_k^2\right)$.

Application: 1-trend models of larval fish dynamics

As a first application of smooth DFA models, we analyze time series data of larval fishes collected in Southern California (USA). The California Cooperative Oceanic Fisheries Investigations (CalCOFI) survey has been collecting physical and biological samples since 1949, to monitor changes to the California Current Ecosystem (Bograd et al. 2003). The CalCOFI data have been incorporated into models used to assess population status (MacCall 2003), and numerous publications have used these time series as indicators of ecosystem state (Mcclatchie et al. 2008). These types of motivating questions also present an opportunity to apply DFA with both conventional and smoothed trends to generate ecosystem state indices. For this application, we focus on the dynamics of three species of juvenile rockfishes: aurora rockfish (*Sebastes aurora*), shortbelly rockfish (*S. jordani*), and bocaccio rockfish (*S. paucispinis*). We restrict the time series to data collected since 1985, when sampling has been consistent in space and time (Moser et al. 2001). Though CalCOFI cruises are done throughout the year, we are primarily interested in estimating interannual trends, and further restrict our analysis to considering spring cruises from 1-April to 22-May when densities of most rockfish species are highest (Mosek et al. 2000). All data were retrieved using the software R (R Core Team 2020) and the 'rerddap' package (Chamberlain 2020).

With only three time series, we focus on DFA models with one trend and single observation error variance, shared across species. Other types of models, including hierarchical GAMs (Pedersen et al. 2019) or models allowing estimated offsets may also be useful in this type of application. Where the DFA model differs is that unlike models with random intercepts or additive terms, the DFA factor loadings $\mathbf{Z}$ are multiplicative and may be close to zero. These cases may arise when a particular time series has a low signal-to-noise ratio, or if there is low correspondence with the latent trends estimated among all other time series. In addition to estimating a conventional 1-trend DFA model with a latent autoregressive process, we evaluate 1-trend B-spline and Gaussian process models. Because we have no a priori hypotheses about the complexity of these smoothed factor models, we evaluated a range of models for each (Table 1), using equally spaced knots.

Application: 2-trend models of commercial fisheries catches

As a slightly more complex example of the smooth factor analysis model, we examine the performance of 2-trend models, using a dataset of commercial fisheries catches (landings) from the west coast of the USA. This dataset consists of 13 species or groups reported annually over a 39-year period (1981–2019)(PFMC 2020). Landings on the US West Coast are dominated by Pacific hake (also Pacific whiting, *Merluccius productus*), but also include substantial catches of rockfishes (*Sebastes spp.*) and flatfishes (e.g., Dover sole, *Solea solea*). Over the course of the last 4 decades, these species have experienced variability associated with population dynamics and the environment, but the patterns of landings also reflects a dynamic fisheries management process. Examples of changes include temporarily closing areas to fishing to protect species of conservation concern, and implementing catch share programs. These processes, combined with environmental conditions that have been positive for many species, have resulted in many increasing populations (Warlick et al. 2018). Given these various management and ecological changes, it is important to summarize patterns of landings, and identify common trends as indicators for management and ecosystem status (Harvey et al. 2018).

As with our previous example, we compared conventional DFA models to those modeling the trends with smooth functions. Preliminary model fitting suggested that 2-trend models were most supported by the data, and thus will be the focus of our analysis. In addition to modeling the 2-trend model with conventional DFA, we evaluated B-spline and Gaussian process models with equally spaced knots (Table 1). All models included a single observation error variance, shared across time series.

Estimation and model selection

We developed our smooth trend DFA model by extending an existing approach that implements conventional DFA in a Bayesian framework (Ward et al. 2019). For the spline models, we assigned priors on the weights $\mathbf{a} \sim N(0,1)$. Similarly, we assigned standard half-normal priors for the Gaussian Process variances $\sigma_k \sim N(0,1)$, and inverse Gamma priors for the scale $\theta_k \sim IG(3,1)$. Estimation in the 'bayesdfa' package is done using Stan and the package R package 'rstan' (Stan Development Team 2016), which implements Markov chain Monte Carlo (MCMC) using the No-U Turn Sampling (NUTS) algorithm (Hoffman and Gelman 2014, Carpenter et al. 2017). For each model considered, we ran 3 parallel MCMC chains for 4000 iterations each, discarding the first 50% of the samples. We assessed convergence using split-$\hat{R}$ and effective samples size (Gelman et al. 2013) along with trace plots. Following previous approaches, we used the Leave One Out Information Criterion (LOOIC, Vehtari et al. 2017, 2020) as a model selection tool (Ward et al. 2019), which approximates leave-one-out cross-validation. Preliminary model checks using LOOIC for the models included in our analysis indicated that many models had 1–4 data points that had high Pareto-k statistics (possibly because of model-misspecification or model flexibity, Vehtari et al. (2017)). To avoid re-fitting these models, we implemented moment matching in the loo package (Vehtari et al. 2020, Paananen et al. 2021).

Results

For our application of smooth dynamic factor models to the CalCOFI juvenile rockfish dataset, we found that the full rank Gaussian process DFA model had slightly lower LOOIC values compared to alternative models (Table 1), with most performing better than the conventional DFA model. Varying the number of knots for the B-spline and Gaussian process models resulted in similar predictions and data support between the two smoothed trend approaches (Table 1). Varying the number of knots did allow for greater flexibility, however, allowing for more complex models to better capture recent variability in rockfish densities (Fig. 1). Trend 1 can be seen as largely capturing the variability in the timeseries of aurora rockfish, which had the loading that was largest in magnitude (0.6, 90% credible interval = 0.04–2.14). Bocaccio rockfish also loaded positively on trend 1, though the effect was weaker (0.53, 90% credible interval = -1.07–2.31). The loading for shortbelly rockfish was smallest in magnitude (0.25, 90% credible interval = -1.44–1.66).

When smooth-trend B-spline and Gaussian predictive process models were applied to commercial fisheries landings data, the model with the lowest LOOIC was the B-spline model with 6 knots. The first trend exhibited nearly linear change from 1981–2001 and was relatively stationary from 2001–2019 (Fig. 2). The second trend represented change from the early 1990s, with the strongest change occuring 2010–present. Estimates of the loadings from this B-spline model indicated many species or species groups loaded negatively on trend 1 (lingcod, sablefish, rockfishes), but Arrowtooth flounder and Pacific whiting had opposite loadings (Fig. 2). Trend 2 from this model appeared to contrast species with relatively stationary catches before declining in 2010 (e.g., Arrowtooth flounder, *Atheresthes stomas*) versus Petrale sole (*Eopsetta jordani*)—one of the only non-whiting species that has experienced positive catches since 2010. Predictions across all models appeared to characterize the trends of most species, and trends from the B-spline model generated more precise predictions relative to the random walk, although neither model was able to capture the variability in Pacific whiting catches since 2000 (Fig. 3).

While low dimensional Gaussian process and B-spline models perform similarly (Table 1), comparing higher order models highlighted an interesting contrast between these two smooth approaches. As more knots were added to the B-spline model of fisheries landings, the wiggliness of the estimated trends generally increased (Fig. 4). The opposite is generally true of the Gaussian process model for this application, with trends becoming smoother as more knots were added (Fig. 4). Estimates of θ_k for this Gaussian process model were relatively large (8.32, 4.4), allowing correlation between neighboring points to decrease slowly and neighboring points further away to have a larger effect. In contrast, the full rank Gaussian process model was most supported for the CalCOFI data — this model had a relatively small value of $\theta_k = 1.12$, allowing correlation between adjacent points to decrease rapidly, translating into greater flexibility.

Discussion

Dynamic factor analysis represents a flexible approach for using state space models to capture latent processes in multivariate time series (Zuur et al. 2003b, 2003a). For some ecological processes — particularly those with high variability — random walks may be too constraining, while for others, using a random walk may be overly complex. Examples of cases where random walks may overfit trends may exist when there are large temporal gaps between observations, or data are collected from systems with high signal to noise ratios. As alternatives to the conventional random walk, we illustrate how DFA trends may be modeled using Gaussian process models or B-splines. Both of these alternatives are flexible in that their smoothness may be specified a priori by the user, and compared via model selection. As the variability of latent trends is nearly always fixed in a conventional DFA for identifiability (Holmes et al. 2012, Zuur et al. 2003b), adopting an alternative model of the trend does not limit inference or change the meaning of other parameters (e.g., loadings).

In both of our case studies comparing smooth DFA models to conventional ones, we found that using smooth functions to model DFA trends resulted in models with higher predictive ability (as measured with LOOIC). Our two case studies contrast two datasets with different degrees of variability. The CalCOFI dataset on juvenile rockfish abundance represents data with relatively high variability — both because of the sampling process, and because the nature of fish recruitment is stochastic. In comparing conventional 1-trend DFA versus smooth trend DFA models to the CalCOFI data, the conventional random walk had difficulty in capturing extremes (Fig. 1), while the B-spline and Gaussian process models generally did better (Table 1). Our second example consisted of applying DFA models to time series of fisheries catches; these data are generally less variable than the CalCOFI data because catches are aggregated across space and individual vessels. Like the CalCOFI example, we found that smooth trend DFA models were better supported over the conventional random walk, however, the models receiving the most support were lower dimension models (e.g., B-spline with 6 knots; Table 1). For both of our case studies, knot locations were assigned uniformly, and these results would be expected to change slightly if the knot locations were adjusted. For models with missing data, or datasets with unevenly distributed replicate samples, it may be important to consider non-uniform knot locations.

Because of their flexibility, applications of LOOIC or related model selection tools to state-space models, including the DFA models in our analysis, may result in poor diagnostics (e.g., high Pareto-k statistics). Though not explored here, alternative approaches for evaluating predictive performance may be used, including the expected log posterior density (ELPD) (Vehtari et al. 2020, 2017). Rather than performing parameter estimation once per model, as was done in our analysis, calculating ELPD is more computationally challenging because with cross-validation, a model must be fit once per fold. With this added cost comes new opportunities, in that cross validation methods specific to time series data may be more easily applied. Commonly used approximations like LOOIC represent an approximation to leave-one-out cross validation where each data point is held out in turn. An alternative approach for time series data is that the observations in each time step can be treated as a fold, and held out in turn. Extensions of this time series approach include leave-future-out cross-validation, where data points are only used to predict future observations, not historical ones (Bürkner et al. 2020).

There are a number of possible extensions to the smooth-function DFA models described in this paper. One extension would be to penalize the wiggliness of the B-spline basis functions, resulting in P-splines (Eilers and Marx 1996, Wood 2017). This would reduce the impact of the number of B-spline basis functions on model fit (Wood 2017). Another extension would be to further constrain the wiggliness defined by the Gaussian process rate of correlation decay (θ) via a prior such as the penalized complexity (PC) prior (Simpson et al. 2017). Such a prior which would allow one to more easily impart prior beliefs about the parameter scale. Third, the smooth trends could themselves be hierarchical: the trends could share their wiggliness, draw wiggliness parameters from a shared distribution, or share a global smoother combined with group-specific smoothers (Pedersen et al. 2019). While these and other developments would further enhance the flexibility of the approach, the trend models included in our analysis already represent a robust approach for DFA that may also be considered in hindcasting or forecasting scenarios.

Tables

Table 1. Leave One Out Information Criterion (LOOIC, with standard errors in parentheses) for each of the models applied to our cases studies (CalCOFI time series of juvenile rockfishes, and the time series of commercial groundfish landings from the west coast of the USA). The B-spline models are generated with basis splines, and the Gaussian predictive process models are generated using a Gaussian covariance function. For each model, knots (or locations of control points) are assumed to be uniformly spaced over the time series. To aid in interpretation, the minimum LOOIC value across models has been subtracted from each case study.

Trend.model	Knots	CalCOFI	Landings
Random walk	NA	16.46 (12.37)	27.56 (49.21)
B-spline	6	13.61 (12.14)	0 (54.64)
B-spline	12	16.73 (12.28)	2.46 (50.72)
B-spline	18	14.56 (11.99)	17.01 (49.72)
B-spline	24	11.34 (12)	30.74 (48.22)
B-spline	30	10.06 (12.07)	49.94 (47.68)
Gaussian process	6	13.95 (12.41)	3.3 (53.64)
Gaussian process	12	15.13 (12.33)	3.34 (53.81)
Gaussian process	18	14.53 (12.29)	6.24 (53.31)
Gaussian process	24	13.79 (12.39)	6.26 (53.38)
Gaussian process	30	13.32 (12.68)	6.75 (53.26)
Gaussian process	Full rank	0 (13.59)	4.48 (53.42)

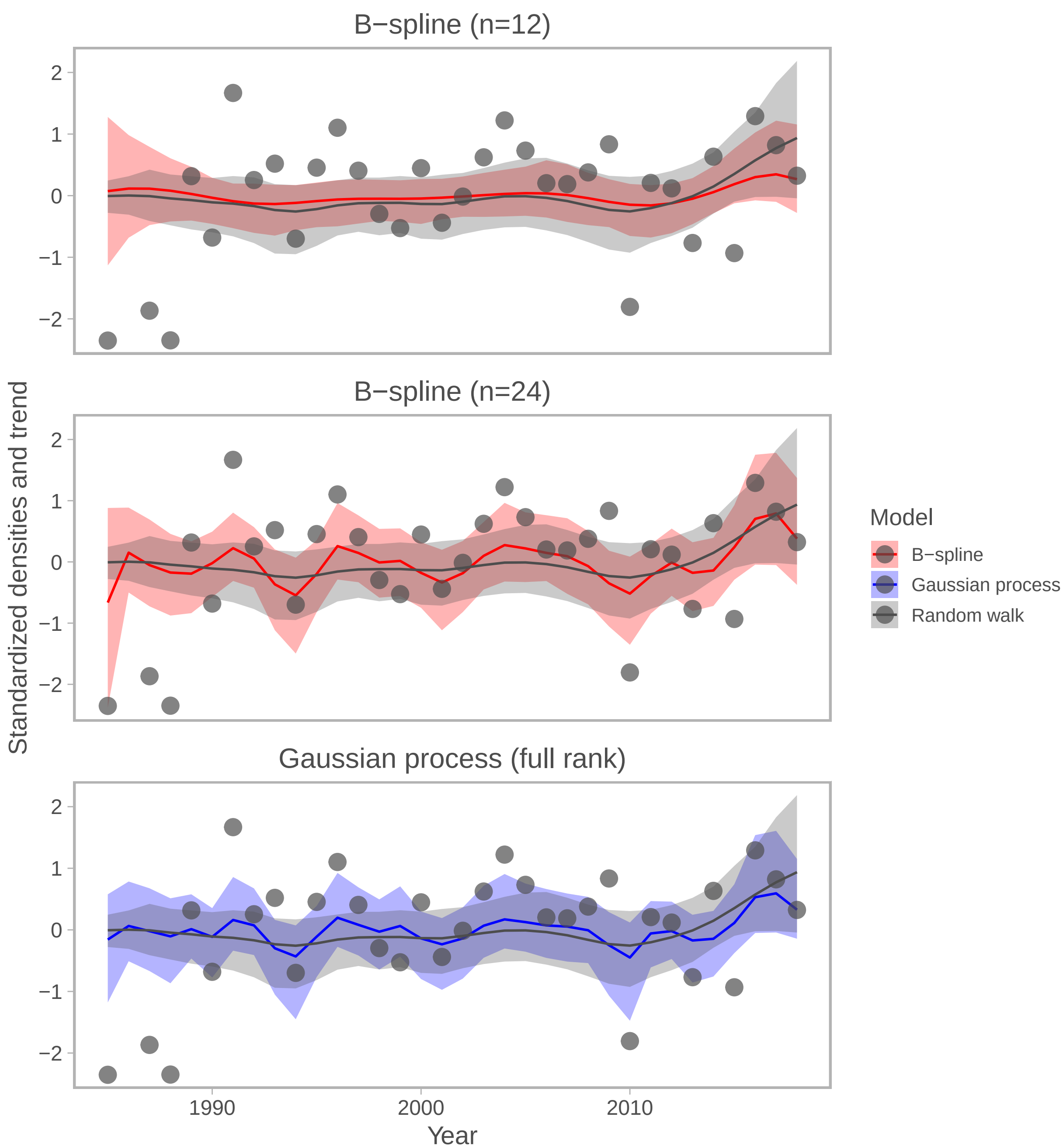
Figure Captions

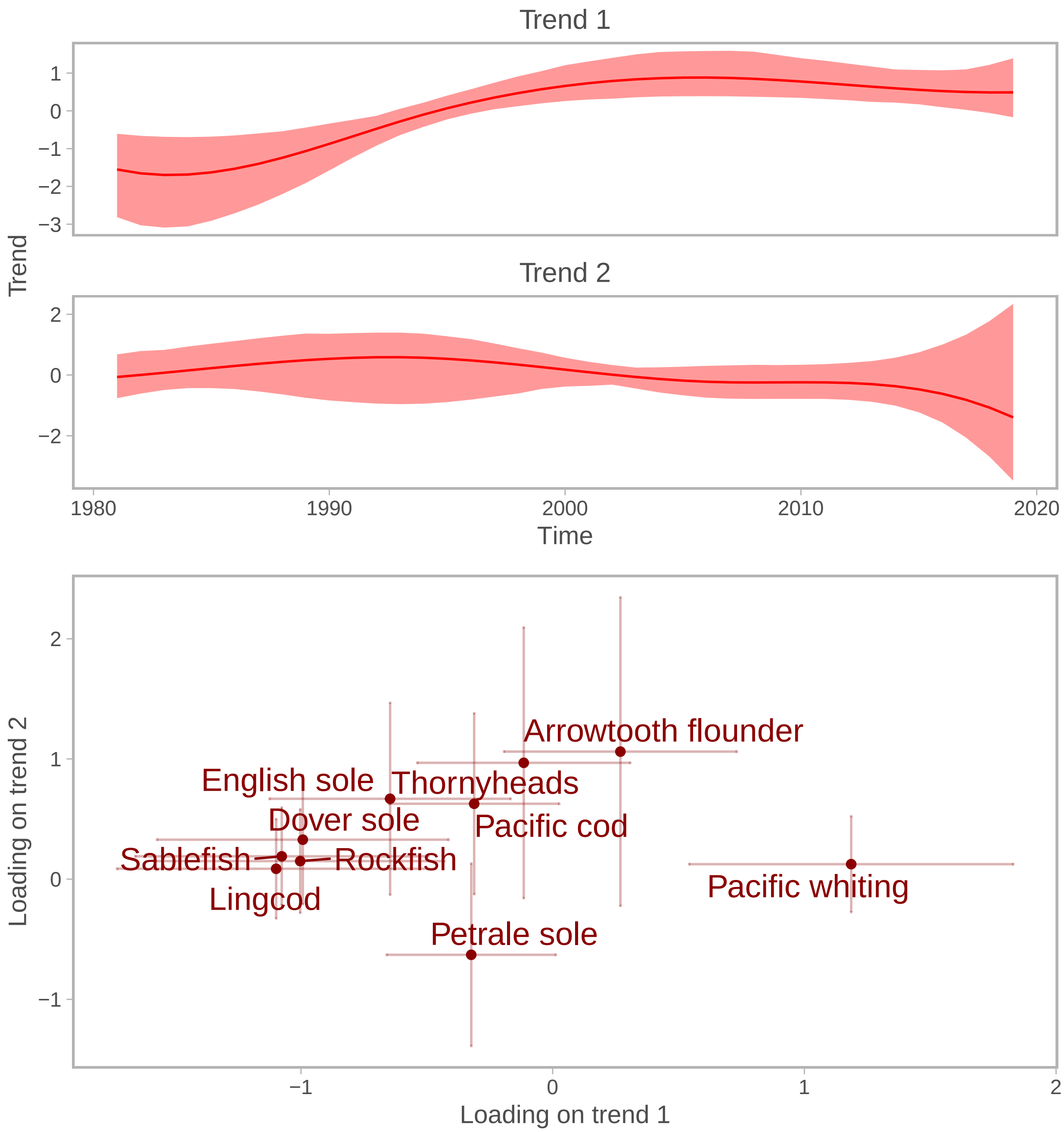
Figure 1. Standardized densities of juvenile shortbelly rockfish (*Sebastes jordani*) collected in the CalCOFI survey, and estimates of latent trends for three candidate models, representing a range of flexibility in splines compared to the conventional random walk. In addition to the conventional DFA model with a latent random walk (included in all panels for reference), predictions from a full rank Gaussian process model, and B-spline model with 12 knots and 24 knots are shown. The posterior mean from each model is shown as a solid line, and 90% credible intervals are shown with ribbons.

Figure 2. Estimated trends and loadings from the 2-trend DFA model applied to commercial groundfish landings off the west coast of the United States. The model results with highest LOOIC is shown, a model that allows trends to be approximated with B-spines (6 knots). The posterior mean for each trend is shown, with ribbons representing 90% credible intervals. The loadings of each species on each trend are shown as points, with lines representing 90% credible intervals.

Figure 3. Estimated landings for 2 species included in our analysis, with contrasting trends (lingcod, Pacific whiting). Posterior means and 90% credible intervals (ribbons) for two candidate models are shown: a B-spline trend model with 6 knots, and a random walk model representing the conventional DFA.

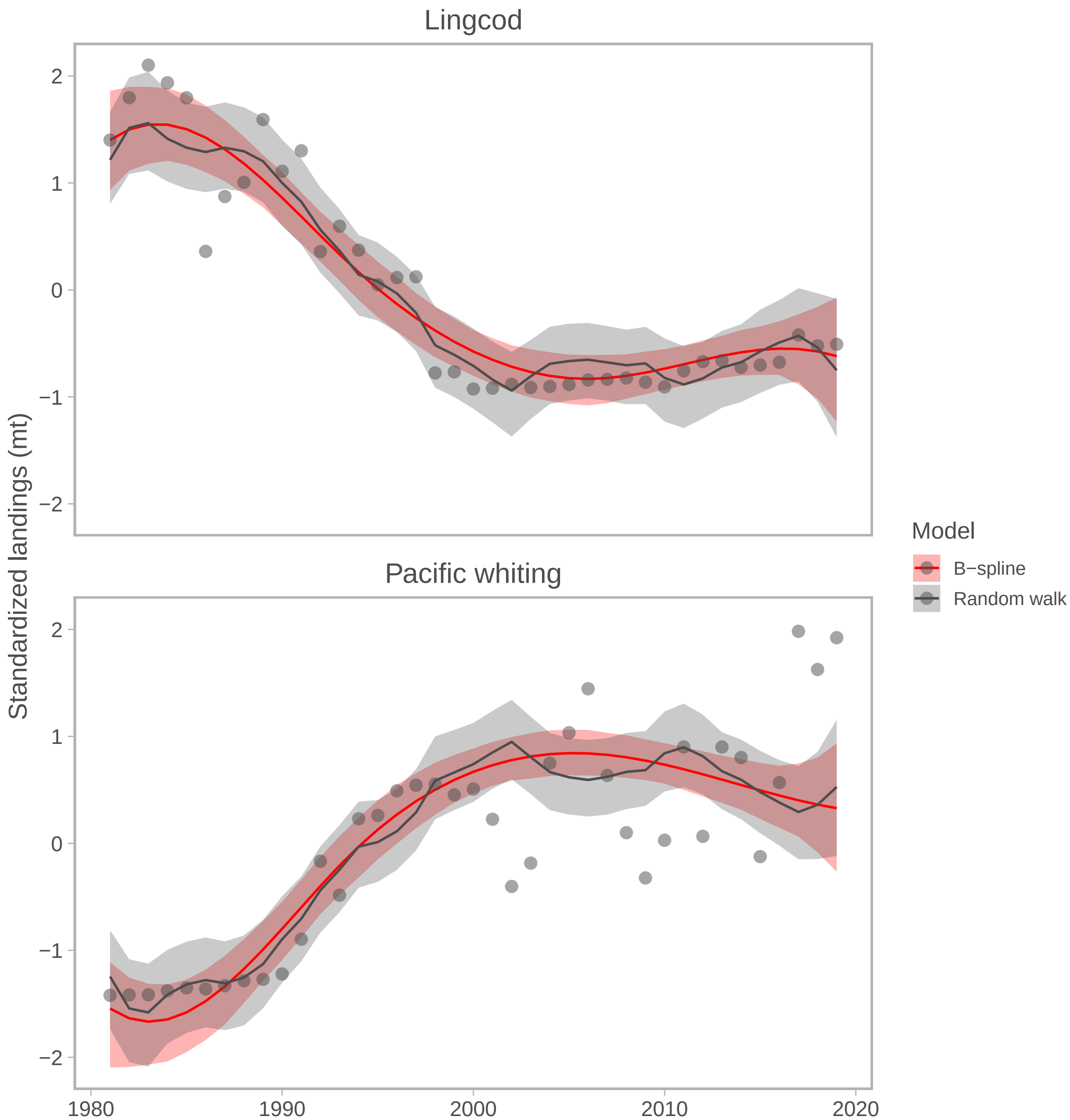
Figure 4. Estimated trends for the 2-trend model of fisheries landings on the west coast of the USA. Shown are results for the B-spline and Gaussian process models with 6 and 18 knots (or control points). Solid lines represent the posterior means and 90% credible intervals are shown as ribbons.

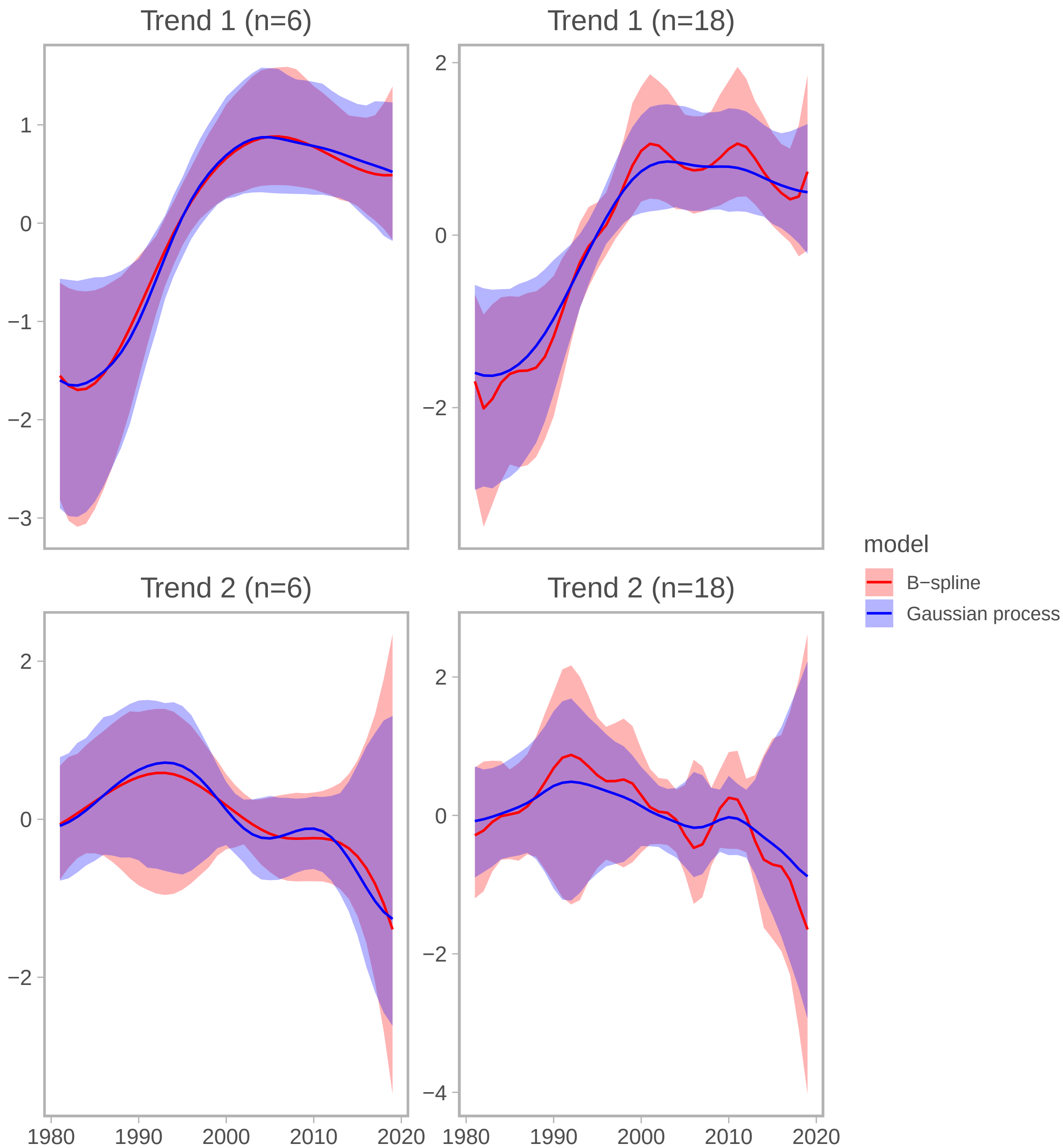

B−spline (n=12)
B−spline (n=24)
Gaussian process (full rank)
Standardized densities and trend
Year
1990
2000
2010
2
1
0
−1
−2
Model
B−spline
Gaussian process
Random walk


Trend 1
1
0
−1
−2
−3
Trend
Trend 2
2
0
−2
1980
1990
2000
2010
2020
Time
2
1
0
−1
Loading on trend 2
Arrowtooth flounder
English sole
Thornyheads
Dover sole
Pacific cod
Sablefish
Rockfish
Lingcod
Pacific whiting
Petrale sole
−1
0
1
2
Loading on trend 1

305

306


Lingcod
Pacific whiting
Standardized landings (mt)
2
1
0
−1
−2
1980
1990
2000
2010
2020
Model
B−spline
Random walk


Trend 1 (n=6)
Trend 1 (n=18)
Trend 2 (n=6)
Trend 2 (n=18)
model
B−spline
Gaussian process
1
0
−1
−2
−3
2
0
−2
2
0
−2
2
0
−2
−4
1980
1990
2000
2010
2020
1980
1990
2000
2010
2020

307

References

Anderson, S. C., and E. J. Ward. 2019. Black swans in space: Modeling spatiotemporal processes with extremes. Ecology 100:e02403.

Auger-Méthé, M., K. Newman, D. Cole, F. Empacher, R. Gryba, A. A. King, V. Leos-Barajas, J. M. Flemming, A. Nielsen, G. Petris, and L. Thomas. 2020. A guide to state-space modeling of ecological time series. arXiv:2002.02001.

Bograd, S. J., D. A. Checkley, and W. S. Wooster. 2003. CalCOFI: A half century of physical, chemical, and biological research in the California Current System. Deep Sea Research Part II: Topical Studies in Oceanography 50:2349–2353.

Bürkner, P.-C., J. Gabry, and A. Vehtari. 2020. Approximate leave-future-out cross-validation for Bayesian time series models. Journal of Statistical Computation and Simulation 90:2499–2523.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. 2017. Stan: A probabilistic programming language. Journal of Statistical Software 76:1–32.

Chamberlain, S. 2020. Rerddap: General purpose client for "erddap" servers.

Cherif, A., H. Cardot, and R. Boné. 2011. SOM time series clustering and prediction with recurrent neural networks. Neurocomputing 74:1936–1944.

Clark, J. S., and O. N. Bjørnstad. 2004. Population Time Series: Process Variability, Observation Errors, Missing Values, Lags, and Hidden States. Ecology 85:3140–3150.

Dennis, B., J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples. 2006. Estimating Density Dependence, Process Noise, and Observation Error. Ecological Monographs 76:323–341.

Eilers, P. H. C., and B. D. Marx. 1996. Flexible smoothing with B-splines and penalties. Statist. Sci. 11:89–102.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. Bayesian data analysis, third edition. CRC press.

Harvey, C., N. Garfield, G. Williams, N. Tolimieri, I. Schroeder, E. Hazen, K. Andrews, K. Barnas, S. Bograd, R. Brodeur, B. Burke, J. Cope, L. deWitt, J. Field, J. Fisher, T. Good, C. Greene, D. Holland, M. Hunsicker, and S. Zador. 2018. Ecosystem Status Report of the California Current for 2018: A Summary of Ecosystem Indicators Compiled by the California Current Integrated Ecosystem Assessment Team (CCIEA).

Hastie, T. J. 1992. Statistical Models in S. *in* J. M. Chambers and T. J. Hastie, editors. Wadsworth & Brooks/Cole.

Hoffman, M. D., and A. Gelman. 2014. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research 15:1593–1623.

Holan, S. H., and N. Ravishanker. 2018. Time series clustering and classification via frequency domain methods. WIREs Computational Statistics 10:e1444.

Holmes, E. E., E. J. Ward, and M. D. Scheuerell. 2020. Analysis of multivariate time-series using the MARSS package. Https://cran.r-project.org/web/packages/marss/vignettes/userguide.pdf.

Holmes, E. E., E. J. Ward, and K. Wills. 2012. MARSS: Multivariate autoregressive state-space models for analyzing time-series data. The R Journal 4:11–19.

Hovel, R. A., S. M. Carlson, and T. P. Quinn. 2017. Climate change alters the reproductive phenology and investment of a lacustrine fish, the three-spine stickleback. Global Change Biology 23:2308–2320.

Kimeldorf, G. S., and G. Wahba. 1970. A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines. The Annals of Mathematical Statistics 41:495–502.

Knape, J. 2008. Estimability of Density Dependence in Models of Time Series Data. Ecology 89:2994–3000.

Latimer, A. M., S. Banerjee, H. S. Jr, E. S. Mosher, and J. A. S. Jr. 2009. Hierarchical models facilitate spatial analysis of large data sets: A case study on invasive plant species in the northeastern United States. Ecology Letters 12:144–154.

Liao, T. W. 2005. Clustering of time series data—a survey. Pattern Recognition 38:1857–1874.

MacCall, A. D. 2003. Status of Bocaccio off California in 2003. In Appendix to the status of the Pacific coast groundfish fishery through 2003: Stock assessment and fishery evaluation. Pacific Fishery Management Council, Portland, OR.

Mcclatchie, S., R. Goericke, J. Koslow, F. Schwing, S. Bograd, R. Charter, W. Watson, N. Lo, K. Hill, J. Gottschalck, M. L'Heureux, Y. Xue, W. Peterson, R. T. Emmett, C. Collins, G. Gaxiola-Castro, R. Durazo, M. Kahru, B. Mitchell, and E. Bjorkstedt. 2008. The state of the California Current, 2007-2008: La Niña conditions and their effects on the ecosystem. California Cooperative Oceanic Fisheries Investigations Reports 49:39–76.

Mosek, H., L. Charter, W. Watson, I. Ambrose, N. Shakon, K. Charter, E. Saniiknoi, S. Fisheiies,

S. Center, M. Fi, and H. Service. 2000. Abundance and distribution of rockfish (Sebastes) larvae in the Southern California Bight in relation to environmental conditions and fishery exploitation. Abundance and distribution of rockfish larvae CalCOFl Rep. 41.

Moser, H. G., R. L. Charter, W. Watson, A. Amurose, P. E. Smith, E. M. Sani, and S. R. Charter. 2001. The CalCOFI Ichthyoplankton time series: Potential contributions to the management of rocky-shore fishes 42:17.

Paananen, T., J. Piironen, P.-C. Bürkner, and A. Vehtari. 2021. Implicitly adaptive importance sampling. Statistics and Computing 31:16.

Patterson, T. A., L. Thomas, C. Wilcox, O. Ovaskainen, and J. Matthiopoulos. 2008. State–space models of individual animal movement. Trends in Ecology & Evolution 23:87–94.

Pedersen, E. J., D. L. Miller, G. L. Simpson, and N. Ross. 2019. Hierarchical generalized additive models in ecology: An introduction with mgcv. PeerJ 7:e6876.

PFMC. 2020. Status of the Pacific Coast Groundfish Fishery: Stock Assessment and Fishery Evaluation. The Pacific Fishery Management Council, Portland, OR.

R Core Team. 2020. R: A language and environment for statistical computing. manual, Vienna, Austria.

Roberts, S., M. Osborne, M. Ebden, S. Reece, N. Gibson, and S. Aigrain. 2013. Gaussian processes for time-series modelling. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371:20110550.

Sardá-Espinosa, A. 2019. Time-Series Clustering in R Using the dtwclust Package. The R Journal 11:22–43.

Simpson, D., H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye. 2017. Penalising model component complexity: A principled, practical approach to constructing priors. Statist. Sci. 32:1–28.

Stan Development Team. 2016. RStan: The R interface to Stan.

Vehtari, A., J. Gabry, M. Magnusson, Y. Yao, P.-C. Bürkner, T. Paananen, and A. Gelman. 2020. Loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models.

Vehtari, A., A. Gelman, and J. Gabry. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing 27:1413–1432.

Ward, E., J., S. Anderson C., L. Damiano A., M. Hunsicker E., and M. Litzow A. 2019. Modeling regimes with extremes: The bayesdfa package for identifying and forecasting common trends and anomalies in multivariate time-series data. The R Journal 11:46.

Ward, E. J., H. Chirakkal, M. González-Suárez, D. Aurioles-Gamboa, E. E. Holmes, and L. Gerber. 2010. Inferring spatial structure from time-series data: Using multivariate state-space models to detect metapopulation structure of California sea lions in the Gulf of California, Mexico. Journal of Applied Ecology 47:47–56.

Warlick, A., E. Steiner, and M. Guldin. 2018. History of the West Coast groundfish trawl fishery Tracking socioeconomic characteristics across different management policies in a multispecies fishery. Marine Policy 93:9–21.

Wood, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73:3–36.

Wood, S. N. 2017. Generalized additive models: An introduction with R. Second editions. Chapman; Hall/CRC.

Zuur, A. F., R. J. Fryer, I. T. Jolliffe, R. Dekker, and J. J. Beukema. 2003a. Estimating common trends in multivariate time series using dynamic factor analysis. Environmetrics 14:665–685.

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