State-space modelling in marine science

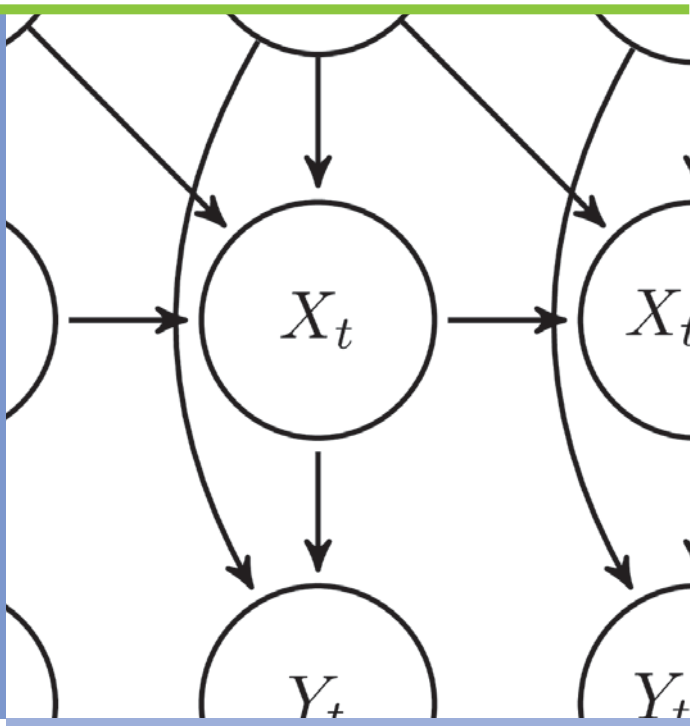
Albertsen, Christoffer Moesgaard

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

State-Space Modelling in Marine Science

PhD Thesis



Written by Christoffer Moesgaard Albertsen
Defended 9 February 2018

DTU Aqua
National Institute of Aquatic Resources

State-Space Modelling in Marine Science

PhD Thesis

Submitted November 2017

Defended 9 February 2018

Christoffer Moesgaard Albertsen

cmoe@aqua.dtu.dk

Technical University of Denmark

National Institute of Aquatic Resources

Section for Marine Living Resources

Preface

This thesis concludes my PhD studies on "State-space modelling in marine science" at the Technical University of Denmark (DTU), National Institute of Aquatic Resources (DTU Aqua), where I have been enrolled from December 1st 2014 to November 30th 2017 under supervision of Senior Researcher Anders Nielsen (DTU Aqua) and Associate professor Uffe Høgsbro Thygesen (DTU Aqua; DTU Compute since June 2017). The project was funded by the PhD School at DTU Aqua and the European Maritime Fisheries Fund.

During my studies, I have been privileged to visit Head of Stock Assessment Program Mark Maunder and colleagues at the Inter-American Tropical Tuna Commission during the first half of 2016, with financial support in the form of personal travel grants from Augustinus Fonden, Christian og Ottilia Brorsons Rejselegat, Knud Højgaards Fond, and Oticon Fonden. I have also had the pleasure of visiting Professor Joanna Mills Flemming and colleagues at Dalhousie University at several occasions for workshops, a conference, and collaboration. Experiences from which I have learned a lot.

This thesis consists of three parts. The first part is a synopsis describing the state-of-the-art in state-space modelling with relation to fisheries stock assessment and individual animal movement. My aim with the synopsis is to provide a broader context and motivation for my studies than the individual manuscript introductions can do. The next part consists of three published manuscripts and two manuscripts in the final stage of preparation. Finally, the last part gives a short description of four software packages developed during my studies.

Christoffer Moesgaard Albertsen
Kongens Lyngby, November 2017

Summary

State-space models provide a natural framework for analysing time series that cannot be observed without error. This is the case for fisheries stock assessments and movement data from marine animals. In fisheries stock assessments, the aim is to estimate the stock size; however, the only data available is the number of fish removed from the population and samples on a small fraction of the population. In marine animal movement, accurate position systems such as GPS cannot be used. Instead, inaccurate alternative must be used yielding observations with large errors.

Both assessment and individual animal movement models are important for management and conservation of marine animals. Consequently, models should be developed to be operational in a management context while adequately evaluating uncertainties in the models.

This thesis develops state-space models using the Laplace approximation within fisheries stock assessment and individual animal movement. In a fisheries stock assessment context, several observational likelihoods are implemented and evaluated using the Laplace approximation. Further, a model for multiple fish stocks is proposed. The model connects single stocks through correlation in the survival process, without requiring any data not used in the individual assessments. Both studies show improvements in evaluating the status of a stock for management.

In an individual animal movement context, the use of a normal distribution for Argos data is compared to the use of t-distributions, both implemented in a state-space model and estimated using the Laplace approximation. Using the heavy tailed t-distribution for the uncertain Argos data improves reconstruction of the true movement trajectories. Further, the commonly used first-Difference Correlated Random Walk is generalized to allow irregular time steps and drift in the movement.

Finally, an approximate filter and smoother based on sequential Laplace approximations are introduced for state-space models with Markov switching. The method can potentially be used in both animal movement and stock assessment models to account for structural changes in behaviour or environmental effects.

Resumé

Tilstandsrumsmodeller er et naturligt værktøj til at analysere tidsrækker, der ikke kan observeres uden fejl. Eksempler på dette er bestandsvurderingsmodeller for fisk og bevægelsesmodeller for individuelle havdyr. I bestandsvurderingsmodeller for fisk er målet at estimere bestandsstørrelsen. Imidlertid er det eneste tilgængelige data antallet af fisk, der er fjernet fra bestanden, samt stikprøver baseret på en lille andel af bestanden. I individuelle bevægelsesmodeller for havdyr kan præcise positionssystemer som GPS ikke benyttes. I stedet må upræcise alternativer bruges, hvilket giver observationer med store målefejl.

Både bestandsvurderings- og bevægelsesmodeller er af stor betydning for forvaltning og beskyttelse af havdyr. Derfor bør modeller udvikles, så de er operationelle i en forvaltningssammenhæng, samtidig med, at de giver realistiske vurderinger af usikkerhederne i modellerne.

Denne afhandling udvikler tilstandsrumsmodeller, som benytter Laplaces integraltilnærmelse indenfor bestandsvurderings- og bevægelsesmodellering. Inden for bestandsvurderingsmodellering implementeres og sammenlignes en række fordelinger til at beskrive observationerne givet den sande bestandstilstand. Dette gøres ved hjælp af Laplaces integraltilnærmelse. Derudover præsenteres en model til at vurdere flere bestande sammen. Modellen forbinder enkelte bestande gennem korrelationer i overlevelsesforløbet uden at kræve andre oplysninger end de allerede benyttede for hver enkelt bestand. Begge undersøgelser forbedrer evnen til at vurdere bestandens tilstand til brug i forvaltning.

Inden for bevægelsesmodeller for individuelle havdyr sammenlignes brugen af normalfordelte målefejl for Argos-observationer med brugen af t-fordelte målefejl. Begge implementeres i en tilstandsrumsmodel og estimeres ved hjælp af Laplaces integraltilnærmelse. Brugen af den tung-halede t-fordeling til at modellere de usikre Argos-observationer forbedrer genskabelsen af de sande bevægelsesbaner. Derudover generaliseres den ofte brugte selv-korrelerede Brownske bevægelse på første-ordens-differencer, således at den tillader varierende tidsskridt, samt drevet bevægelse.

Endelig præsenteres en tilnærmet filtrerings- og udglatningsmetode baseret på fortløbende Laplace-integraltilnærmelser til tilstandsrumsmodeller med Markovianske parameterskift. Metoden kan potentielt benyttes til bevægelses- såvel som bestandsvurderingsmodeller for at tage højde for strukturelle ændringer i adfærd eller miljøpåvirkninger.

Contents

Papers

Chapter 1

Introduction

State-space models are time series models where there is not only stochasticity in the process evolving over time, but also in the measurements given the process. That is, where a classical time series model describes a process by the transition densities, $X_t \mid X_{t-1}$, a state-space model has an additional layer for the measurement stochasticity described by the conditional distribution $Y_t \mid X_t$. In the state-space model, the process, X_t, is not observed. Therefore, inference can only be based on the marginal observation process, Y_t, which is similar to random effects models; however, obtaining the marginal observation process requires integrating over the entire X_t process, which can be computationally demanding. In the most general setting, X_t and Y_t can be vectors of any dimension with values in any measurable space.

The separation of processes and measurements makes state-space modelling a natural framework for marine science. In fisheries stock assessment models, the aim is to estimate the abundance of a fish stock; however, the fish stock is only measured indirectly, and with error, through catches and scientific surveys. In this setting, the abundance is the unobserved process, X_t, while catches and survey indices of abundance combine to a measurement process, Y_t. Likewise, models for individual marine animal movement must account for both the inherent stochasticity of the movement behaviour and the stochasticity of the imprecise measurements that can be obtained in marine environments. In this setting, the actual position of the animal is the unobserved process, X_t, while a location measurement from, for instance, tagging the animal are the observed measurement process, Y_t.

While seemingly very different fields, fisheries stock assessment models and individual animal movement models not only share statistical methodology, but can both be used for management and conser-

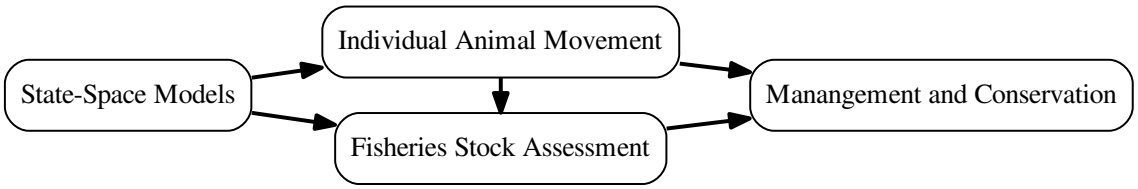


Figure 1.1: Fisheries stock assessment models and individual animal movement models are connected by the common statistical framework, and the common importance for management and conservation of marine living resources. Further, individual movement models can be included in assessment models.

vation of marine animals. To be operational in a management and conservation context, models should be able to run in a reasonable time; should require the same input as models already used; and should produce output comparable to models already used.

Recent advancements in computational statistics have facilitated the use of the Laplace approximation for maximum likelihood inference in advanced state-space models. While simulation-based methods such as Markov Chain Monte Carlo may take hours or days to produce results, software utilizing the Laplace approximation require seconds or minutes, at the cost of a less general framework.

The aim of this PhD study has been to investigate the applicability of the Laplace approximation to efficiently approximate marginal observation processes, Y_t, in state-space models in marine science, and produce operational models that can be applied in, and improve upon, management and conservation of marine animals.

Paper I and Paper II contributes to the fisheries stock assessment literature. Paper I compares several choices of observational likelihoods in an age-based assessment model and illustrates that the choice can influence subsequent management decisions. Paper II extends the same age-based assessment model from a single-stock to a multi-stock model. The paper combines stocks through correlated survival processes; a highly operational approach, albeit simple, that does not require any data sources not used in the single-stock model. Three case studies illustrate that correlations are in fact present in the data, while simulation studies show the consequences of ignoring these correlations.

Paper III and Paper IV contributes to the animal movement literature. Paper III investigates the applicability of the Laplace approximation for a movement model. The method is compared to a Kalman filter and Markov Chain Monte Carlo; both in simulations and for real data. Further, the paper extends a previous model from requiring Gaussian data to allow the more appropriate t-distribution for the error prone movement data. Likewise, Paper IV generalizes one of the most frequently used animal movement models to allow irregular time points and drift in the movement.

Finally, Paper V introduces a method for maximum likelihood-based analysis of state-space models with Markov switching parameters. The method has potential to be used in both animal movement and fisheries stock assessment models. In animal movement models, Markov switching parameters are frequently considered, but in a Bayesian context. Markov switching state-space models are not as widely used in fisheries stock assessment models; however, they have been considered recently (e.g., Fay, 2015; Baggio, 2016).

The following four chapters describe the state-of-the-art of state-space modelling in marine science. Chapter 2 describes modern methods for state-space models common for both fisheries stock assessment and animal movement modelling. Chapter 3 focuses on state-space modelling in fisheries stock assessment modelling with a focus on age-based models, while Chapter 4 provides a similar overview for individual animal movement. All three chapters also attempts to provide a broad context and motivation for the included papers. Chapter 5 concludes on the previous chapters. Following these synopsis chapters, the five papers and four short descriptions of software developed during the studies are included.

Chapter 2

State-space models

Many systems, especially in marine environments, cannot be observed directly. When modelling these systems, it is convenient to separate the stochasticity of the system process and the stochasticity of the measurements conditional on the system process. This is exactly what state-space models does. State-space models are time series models consisting of two parts: a model for the system process, and a model for the measurements given the process. That is, the state-space of the process is modelled separately from the observation-space. In a general setting, the current state of the system, X_t, is a Markov process given by the transition distribution $X_t \mid X_{t-1} \sim F_\theta$ with density f_θ, while the observations Y_t are independent given the process. In other words, the process is determined by the conditional distributions $Y_t \mid X_t \sim G_\theta$ with density g_θ. In the most general setting, both X_t and Y_t can be vectors of any dimension. Both distributions are parameterised with a parameter vector $\theta \in \Theta \subset \mathbb{R}^p$. The state-space model can conveniently be expressed by the state and measurement equations:

$$\begin{aligned} X_t &= \phi(X_{t-1}, \theta, \epsilon_t) \\ Y_t &= \gamma(X_t, \theta, \nu_t), \end{aligned} \tag{2.1}$$

where ϵ_t and ν_t are random variables. In this way, the model can be described intuitively through functions, ϕ and γ, of parameters, θ, and random variables $\big(X_{t-1}$ and ϵ_t, respectively X_t and $\nu_t\big)$ instead of through the corresponding conditional densities, f_θ and g_θ.

State-space models have been given several names in the literature such as hidden Markov models, dynamic (linear) models, partially observed Markov processes, latent variable models, and random effects models. Here, "hidden Markov model" will be used for a system developing over time where the state-space is discrete, $X_t \in \mathcal{X}^d \subset \mathbb{N}^d$, and "state-space model" will be used for systems developing over time where the state-space is continuous, $X_t \in \mathcal{X}^d \subset \mathbb{R}^d$.

State-space models have a strong tradition in engineering, in particular in control theory. The classical linear Gaussian state-space model,

$$\begin{aligned} X_t &= AX_{t-1} + Bu_t + \epsilon_t \\ Y_t &= HX_t + \nu_t \end{aligned} \tag{2.2}$$

has been thoroughly investigated. In this setting, A is a matrix giving the autoregressive properties of the process, B is a matrix of effects of the input u_t, and H is a matrix giving the expected measurements conditional on the states. Finally, ϵ_t and ν_t are independent random variables following normal distributions with covariances Q and R respectively. In control theory, the objective is to influence the system in an optimal way through the inputs, u_t; however, in the models presented in this thesis, the objective is not to control the system, but to learn from and predict the system.

2.1 Maximum likelihood inference

To learn from data generated by a system with imperfect measurements, maximum likelihood inference can be used. In a state-space model, the model description directly leads to a likelihood function for the fully observed system; the joint likelihood of data and latent variables:

$$L(\theta; X_{0:T}, Y_{0:T}) = f_\theta(X_0) g_\theta(Y_0 \mid X_0) \prod_{t=1}^{T} f_\theta(X_t \mid X_{t-1}) g_\theta(Y_t \mid X_t),$$

that is, the joint probability of observing the data at hand as a function of the parameters. From the likelihood function, the aim is to find the parameters that makes the data at hand most likely.

In state-space models, however, the state process is not observed. Therefore, only the marginal likelihood of the actual observations[1],

$$L_M(\theta; Y_{0:T}) = \int L(\theta; X_{0:T}, Y_{0:T}, Y) dX_{0:T}$$

can be used to obtain the estimates

$$\hat{\theta} = \text{argmax}_\theta \, L_M(\theta; Y_{0:T})$$

In general, finding the marginal likelihood of the data is impossible to do analytically. Several approximate methods exist to obtain the maximum likelihood estimates. In the sections below, only the methods used in the appended papers will be introduced, while other methods such as the EM algorithm (Dempster et al., 1977), Markov Chain Monte Carlo (e.g., Jacquier et al., 2007), importance sampling (e.g., Geweke, 1989; Skaug and Fournier, 2006; Kleppe and Skaug, 2012), discretization of the state-space (e.g., Pedersen et al., 2011), numerical integration, and others, are silently left out.

2.2 Filters

One approach to calculate the marginal likelihood is through recursive filters. Originating in engineering, the filters are constructed to filter the measurement noise from the signal of the state process. As a by-product the likelihood function can be calculated during the process. In a maximum likelihood framework, the viewpoint is reversed. The filtering process consists of three steps in a recursion (e.g., Kitagawa, 1987). For a model defined by the transition density of $X_t \mid X_{t-1}$, f_θ, and the measurement density of $Y_t \mid X_t$, g_θ, the first step is predicting the next state given the previous observation. Knowing the density of $X_{t-1} \mid Y_{0:(t-1)}$ from the previous step in the recursion, the density of $X_t \mid Y_{t-1}$ can be calculated through the joint distribution $X_t, X_{t-1} \mid Y_{0:(t-1)}$, using the transition density, f_θ, since $X_t \mid X_{t-1} \overset{D}{=} X_t \mid X_{t-1}, Y_{0:(t-1)}$. Hence,

$$p(X_t \mid Y_{0:(t-1)}) = \int f_\theta(X_t \mid X_{t-1}) p(X_{t-1} \mid Y_{0:(t-1)}) dX_{t-1}.$$

Here, p denotes derived densities. For instance, $p(X_t \mid Y_{0:(t-1)})$ is the density of the conditional distribution $X_t \mid Y_{0:(t-1)}$. The derived densities only depend on parameters through f_θ and g_θ. Therefore, the parameter dependendence is excluded from the notation for simplicity.

[1] Throughout the thesis, the region of integration is omitted when it is the entire space.

Second step is calculating the likelihood of the next information, Y_t, given the previous, $Y_{0:(t-1)}$. Again, this can be calculated through a joint distribution. The joint density of $Y_t, X_t \mid Y_{0:(t-1)}$ can be calculated using the density of $Y_t \mid X_t, Y_{0:(t-1)} \overset{D}{=} Y_t \mid X_t, g_\theta$, and the density of $X_t \mid Y_{0:(t-1)}$:

$$p(Y_t \mid Y_{0:(t-1)}) = \int g_\theta(Y_t \mid X_t) p(X_t \mid Y_{0:(t-1)}) dX_t.$$

Final step is updating the state estimate with information from a new observation. The density of $X_t \mid Y_{0:t}$ can be calculated by Bayes' formula,

$$p(X_t \mid Y_{0:t}) = p(X_t \mid Y_t, Y_{0:(t-1)}) = g_\theta(Y_t \mid X_t) \frac{p(X_t \mid Y_{0:(t-1)})}{p(Y_t \mid Y_{0:(t-1)})}.$$

The marginal likelihood can be calculated by

$$L_M(\theta; Y_{0:T}) = p(Y_0) \prod_{t=1}^{T} p(Y_t \mid Y_{0:(t-1)})$$

In few cases, the equations can be solved analytically. For the linear Gaussian state-space model (Equation (2.2)), the Kalman filter (Kalman, 1960) can be used. When the state space is discrete, for a hidden Markov model, the likelihood calculation reduces to a series of matrix products. When the equations cannot be solved analytically, numerical integration or simulation-based Monte Carlo methods can be used instead. In the latter case, the method is known as a particle filter or a sequential Monte Carlo method (Gordon et al., 1993). If a simulation-based method is used, the likelihood function calculated becomes a stochastic function, which requires special methods to obtain maximum likelihood estimates (e.g., Ionides et al., 2011, 2015).

2.3 Laplace approximation

A general and computationally efficient alternative to filters and simulation methods is the Laplace approximation. The Laplace approximation replaces the joint likelihood with a log-convex function that can easily be integrated. From the joint log-likelihood,

$$\ell(\theta; X_{0:T}, Y_{0:T}) = \log L(\theta; X_{0:T}, Y_{0:T}),$$

a second order Taylor series is constructed around the mode of the latent variables,

$$\hat{X}_\theta = \text{argmax}_{X_{0:T}} \, \ell(\theta; X_{0:T}, Y_{0:T}).$$

Since the gradient at the mode is zero,

$$\ell(\theta; X_{0:T}, Y_{0:T}) \approx \ell(\theta; \hat{X}_\theta, Y_{0:T}) - \frac{1}{2}(X_{0:T} - \hat{X}_\theta)^T H(\theta)(X_{0:T} - \hat{X}_\theta)$$

where $H(\theta)$ is the Hessian with respect to the latent states,

$$H(\theta) = -\nabla^2_{X_{0:T} X_{0:T}} \ell(\theta; X_{0:T}, Y_{0:T})\big|_{X_{0:T} = \hat{X}_\theta}.$$

Inserting this approximation in the formula for the marginal likelihood

$$\begin{aligned}
L_M(\theta; Y_{0:T}) &= \int L(\theta; X_{0:T}, Y_{0:T}) dX_{0:T} \\
&= \int \exp(\ell(\theta; X_{0:T}, Y_{0:T})) dX_{0:T} \\
&\approx \int \exp\left(\ell(\theta; \hat{X}_\theta, Y) - \frac{1}{2}(X_{0:T} - \hat{X}_\theta)^T H(\theta)(X_{0:T} - \hat{X}_\theta)\right) dX_{0:T} \\
&= L(\theta; \hat{X}_\theta, Y_{0:T}) \int \exp\left(-\frac{1}{2}(X_{0:T} - \hat{X}_\theta)^T H(\theta)(X_{0:T} - \hat{X}_\theta)\right) dX_{0:T} \\
&= L(\theta; \hat{X}_\theta, Y_{0:T}) \cdot (2\pi)^{n/2} \cdot \det(H(\theta))^{-1/2}
\end{aligned}$$

Taking the logarithm, we obtain the approximation

$$\ell_M(\theta; Y_{0:T}) = \ell(\theta, \hat{X}_\theta; Y_{0:T}) - \frac{1}{2}\log(\det(H(\theta))) + \frac{T+1}{2}\log(2\pi)$$

Using the Laplace approximation requires an optimization of the joint log-likelihood and obtaining the Hessian with respect to the latent states. To obtain the gradient of ℓ_M (e.g. Skaug and Fournier, 2006), note that $\hat{X}_\theta$ is an implicit function of θ. To be explicit, write

$$\ell_M(\theta, X_{0:T}; Y_{0:T}) = \ell(\theta, X_{0:T}; Y_{0:T}) - \frac{1}{2}\log(\det(H(\theta))) + \frac{T+1}{2}\log(2\pi).$$

That is, $\ell_M(\theta; Y_{0:T}) = \ell_M(\theta, \hat{X}_\theta; Y_{0:T})$ where $\hat{X}_\theta$ is a function of θ. By the chain rule,

$$\begin{aligned}
\frac{\partial \ell_M(\theta; Y_{0:T})}{\partial \theta} &= \left.\frac{\partial \ell_M(\theta, X_{0:T}; Y_{0:T})}{\partial \theta}\right|_{X_{0:T}=\hat{X}_\theta} + \frac{\partial \ell_M(\theta, \hat{X}_\theta; Y_{0:T})}{\partial \theta} \\
&= \left(\frac{\partial \hat{X}_\theta}{\partial \theta}\right)^T \left.\frac{\partial \ell_M(\theta, X_{0:T}; Y_{0:T})}{\partial X_{0:T}}\right|_{X_{0:T}=\hat{X}_\theta} + \frac{\partial \ell_M(\theta, \hat{X}_\theta; Y_{0:T})}{\partial \theta}
\end{aligned}$$

where the implicit function theorem yields

$$\frac{\partial \hat{X}_\theta}{\partial \theta} = -H(\theta)^{-1} \left.\frac{\partial \ell(\theta; X_{0:T}, Y_{0:T})}{\partial X_{0:T} \partial \theta}\right|_{X_{0:T}=\hat{X}_\theta}$$

Ignoring parameter uncertainty, the covariance of the state estimates can be approximated by

$$V(\hat{X}_{\hat{\theta}}) = H(\hat{\theta})^{-1}.$$

Including uncertainty in parameter estimates (Kass and Steffey, 1989; Skaug and Yu, 2014), the covariance approximation becomes

$$V(\hat{X}_{\hat{\theta}}) = H(\hat{\theta})^{-1} + \left.\frac{\partial \hat{X}_\theta}{\partial \theta}\right|_{\theta=\hat{\theta}} \Sigma_{\hat{\theta}} \left(\left.\frac{\partial \hat{X}_\theta}{\partial \theta}\right|_{\theta=\hat{\theta}}\right)^T, \tag{2.3}$$

where $\Sigma_{\hat{\theta}}$ is the covariance of the estimator. Using these few equations, the Laplace approximation can be used for a wide range of models with unobserved data provided there is a unique, or at least very dominant, maximum, $\hat{X}_\theta$. The generality of the Laplace approximation makes it a useful tool for software packages such as `INLA` (Rue et al., 2009)[2], `AD Model Builder` (Fournier et al., 2012), and `Template Model Builder` (Kristensen et al., 2016); however, as it is an approximation, its accuracy is a concern.

[2]Note that `INLA` does not use the Laplace approximation in the way outline in this section. Instead, `INLA` approximates $p(X_{0:T} \mid \theta, Y_{0:T})$ to calculate $p(\theta \mid Y_{0:T})$ and in turn $p(X_t \mid Y_{0:T})$, in a Bayesian framework. Bayesian statistics will not be discussed here.

2.3.1 Accuracy of the Laplace approximation

The Laplace approximation is usually justified in a setting where the number of observations grow faster than the number of random effects (e.g., Barndorff-Nielsen and Cox, 1989; Kass and Steffey, 1989; Tierney et al., 1989a,b; Kass et al., 1990; Erkanli, 1994; Pinheiro and Bates, 1995). This is the typical case in, for instance, (generalized) linear mixed models. In the literature, most theoretical results concern this asymptotic case where the number of random effects is fixed while the number of observations increases; or, at least, where the number of random effects per observation goes to zero. This is, however, not the case in state-space models. In a typical time series model, the ratio is fixed as the number of observations increases; for each new time point, additional states must be included in the model.

Ogden (2017) provides conditions for estimators based on approximate likelihoods to be consistent when the likelihood approximation does not improve with the number of observations. Instead, the error in the score function must go to zero; however, verifying the conditions is difficult in practice when the true marginal likelihood is unavailable. There are, nonetheless, general approaches that can be used: simulation studies, comparing to higher order approximations or simulation-based methods.

Effectively, the Laplace approximation approximates the joint distribution of states and observations by a multivariate Gaussian distribution. In most cases, this is a poor point-wise approximation; however, the resulting integral approximation often performs well. For example, consider the case with one observation, Y, and a latent variable, X, such that

$$\begin{aligned} X &\sim \Gamma(\alpha, \beta) \\ Y &\sim Pois(X) \end{aligned}$$

The gamma distribution is parameterised with shape parameter α and scale parameter β. Figure 2.1 shows the joint likelihood, $P(X, Y)$, for $\alpha = \beta = 2$ and observed value $Y = 3$.

While the point-wise approximation as a function of X is clearly poor, the integral approximation is fairly good. For $X < \hat{X}_\theta$, the Gaussian approximation is larger than the true density, while it is lower for $x > \hat{X}_\theta$; however, the two cases nearly cancel out. Specifically,

$$\int_{-\infty}^{\hat{X}_\theta} f(X) - \tilde{f}(X) dx = -0.0156112$$

while

$$\int_{\hat{X}_\theta}^{\infty} f(X) - \tilde{f}(X) dx = 0.0183204$$

Combined, the Laplace approximation underestimates the integral. The approximated value is 97.94% of the true value for these parameters.

2.3.2 Transformation of non-Gaussian state variables

For a fixed data set, the Laplace approximation may be improved by a change of the integration variable (Kleppe and Skaug, 2012). In the example above, the gamma distributed latent variable can be expressed as a transformation of a Gaussian random variable, thereby changing the scale at which the integral is approximated.

In the example above, let p_Γ be the density of the gamma distributed latent state with cumulative distribution function Γ and p_{Pois} be the conditional density of the observation given the latent state as a

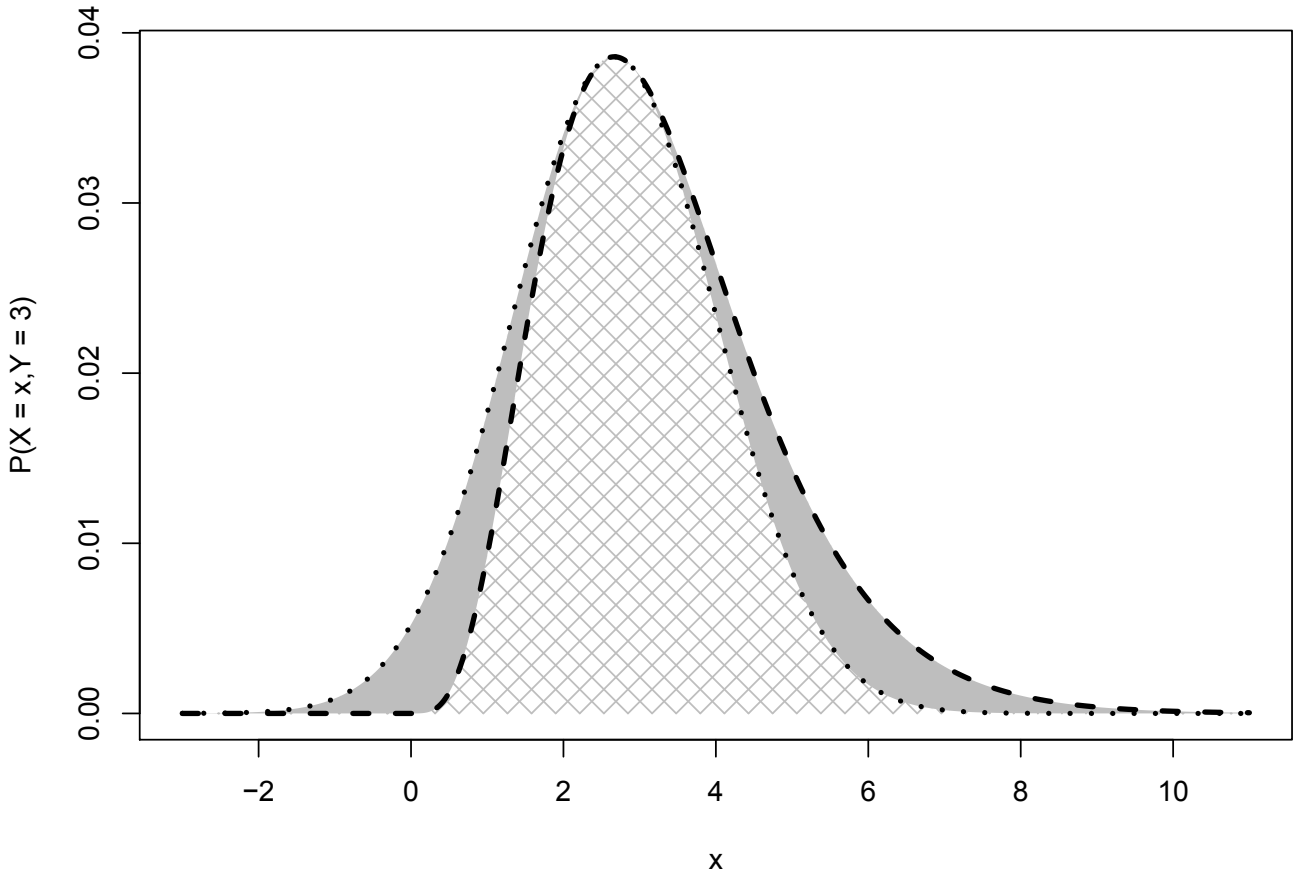


Figure 2.1: Illustration of the Laplace approximation. The dashed line shows the likelihood of the true model, while the dotted line is a second order Taylor approximation. The grey area indicates the difference between the two integrands.

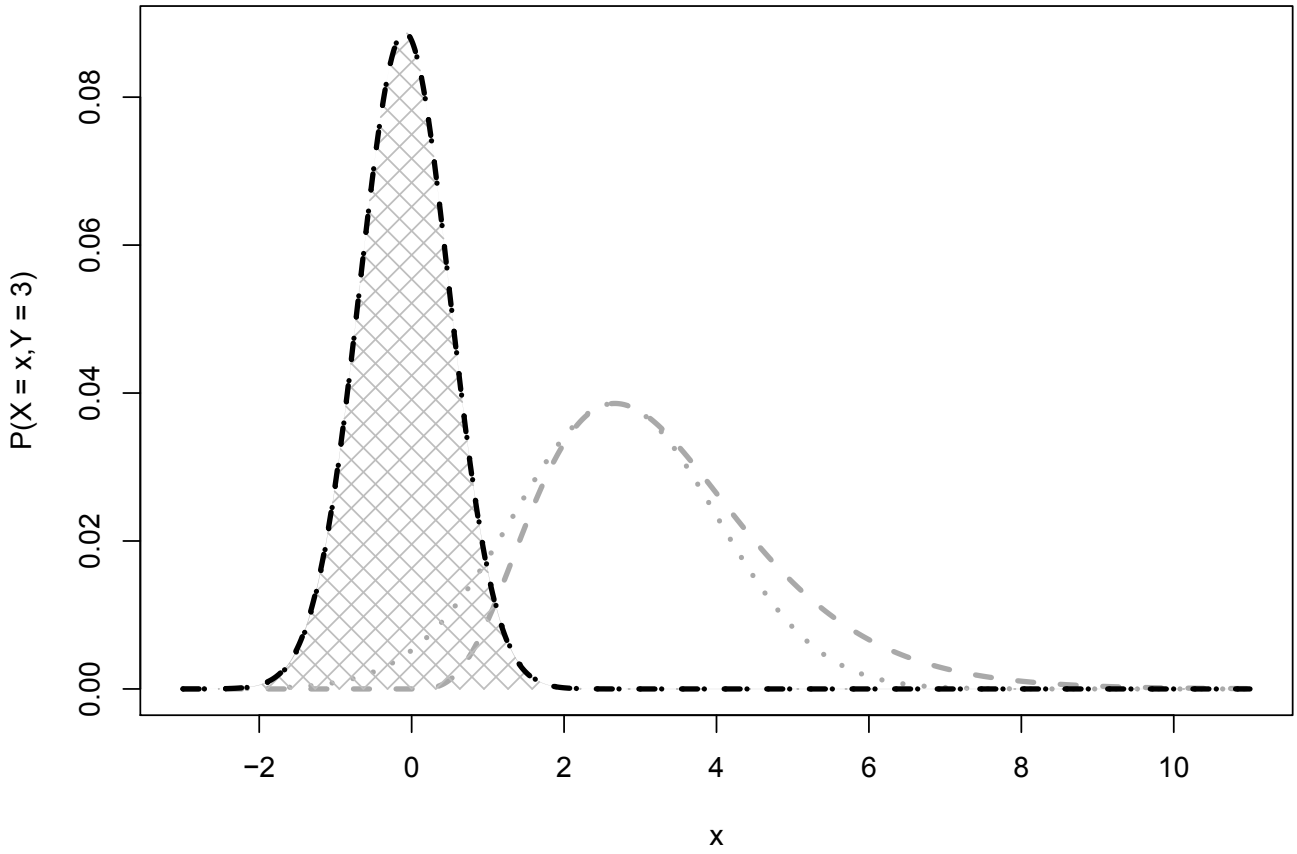


Figure 2.2: Illustration of the transformed Laplace approximation. The dashed line shows the likelihood of the true model, while the dotted line is a second order Taylor approximation. The grey area indicates the difference between the two integrands. The dashed and dotted grey lines show the corresponding values in the untransformed model.

function of the state. Employing the transformation

$$h(x) = \Phi^{-1}(\Gamma(x))$$

with inverse

$$h^{-1}(y) = \Gamma^{-1}(\Phi(y)),$$

where Φ is the standard normal cumulative density function, the integral of $p_\Gamma(x)p_{Pois}(x)$ can be transformed to an integral of $\phi(y)p_{Pois}(h^{-1}(y))$, where $\phi(y)$ is the standard normal density function, since

$$\begin{aligned}\int_{\mathbb{R}_+} p_\Gamma(x)p_{Pois}(x)dx &= \int_{\mathbb{R}_+} \phi(h(x))p_{Pois}(x)|h'(x)|dx \\ &= \int_{\mathbb{R}} \phi(h(h^{-1}(y)))p_{Pois}(h^{-1}(y))|h'(h^{-1}(y))||h^{-1\prime}(y)|dy \\ &= \int_{\mathbb{R}} \phi(y)p_{Pois}(h^{-1}(y))dy.\end{aligned}$$

In this model, applying the Laplace approximation to the last integral is more accurate than applying the Laplace approximation to the first integral. The two approaches are compared in Figure 2.2. With the transformation, the integral approximation still underestimates the integral; however, the error is much smaller. The new approximated value is 99.93% of the true value for these parameters. Further, the relative error is less dependent on the parameter values. On the original scale, the error of the Laplace approximation decreases as the scale parameter increases (Figure 2.3). In contrast, the relative error of the Laplace approximation on the transformed scale is nearly constant as a function of the scale parameter.

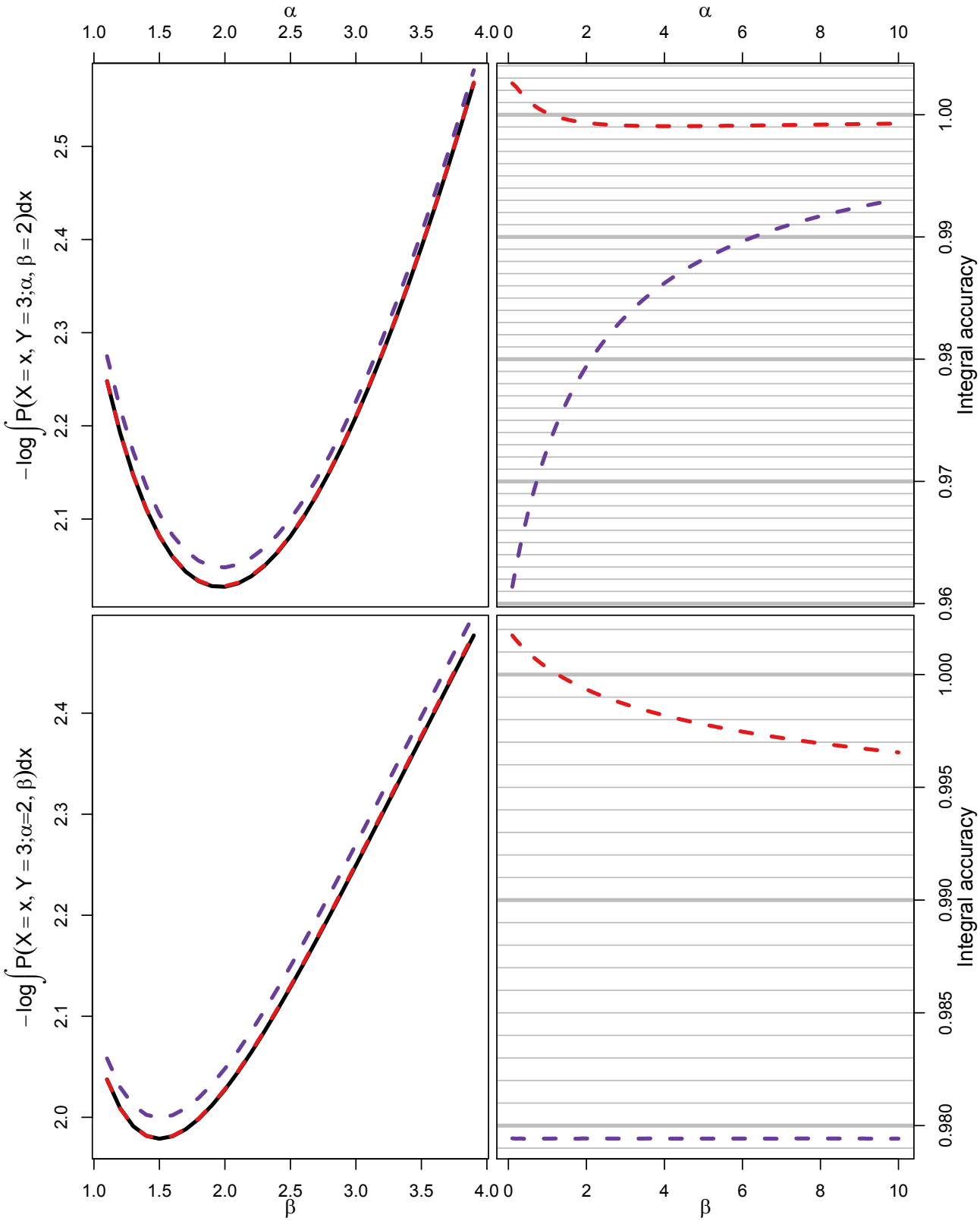


Figure 2.3: Marginal negative log-likelihood approximations (left panels) and their relative error (right panels) compared to the true likelihood (black line) with gamma distributed random effect (purple dashed line) and transformed Gaussian random effects (red dashed line) as a function of α with $\beta = 2$ (upper panels) and as a function of β with $\alpha = 2$ (lower panels).

2.3.3 Functions of latent states

The ratio between latent states and observations is not only an issue for the accuracy of the Laplace approximation, but also for calculating the uncertainty of transformations of latent states. In many applications, it is not only the parameters and latent states that are of interest. Derived quantities may also be an integral part of the analysis and interpretation of the data at hand. Extending the arguments of Kass and Steffey (1989; e.g., Kristensen et al., 2016), the covariance of latent states and parameters can be approximated by

$$V\begin{pmatrix} \hat{X}_\theta \\ \theta \end{pmatrix} = \begin{pmatrix} H(\theta)^{-1} & 0 \\ 0 & 0 \end{pmatrix} + J^T \Sigma_\theta J,$$

where J is the $p \times (T + 1 + p)$ Jacobian matrix of $(\hat{X}_\theta, \theta)^T$ with respect to θ. Note that extracting the marginals corresponding to $\hat{X}_\theta$ simplifies to Equation (2.3). From this, the delta method can be used to approximate the mean and variance of a function of the estimates, $f(\hat{\theta}, \hat{X}_\theta)$, by

$$\begin{aligned} E\left(f(\hat{X}_\theta, \hat{\theta}) \mid Y_{0:T}\right) &\approx f(\hat{X}_\theta, \hat{\theta}) \\ V\left(f(\hat{X}_\theta, \hat{\theta}) \mid Y_{0:T}\right) &\approx f'(\hat{X}_\theta, \hat{\theta})^T V \begin{pmatrix} \hat{X}_\theta \\ \theta \end{pmatrix} f'(\hat{X}_\theta, \hat{\theta}) \end{aligned}$$

This is the procedure used in both ADMB and TMB (Kristensen et al., 2016).

Using the delta method relies on asymptotic normality of the parameters. However, in a state-space model the number of observations per latent state is fixed, as previously discussed. Therefore, the usual asymptotic normality cannot be used. This can result in a bias if transformations of random effects are not corrected.

To obtain the mean of the transformed variable, as opposed to the transformed mode, Thorson and Kristensen (2016) note that

$$E(f(\hat{\theta}, X_{0:T}) \mid Y_{0:T}) = \frac{\int \exp\left(\ell(\hat{\theta}; X_{0:T}, Y_{0:T})\right) f(\hat{\theta}, X_{0:T}) dX_{0:T}}{\int \exp\left(\ell(\hat{\theta}; X_{0:T}, Y_{0:T})\right) dX_{0:T}}$$

where $\ell(\theta, X_{0:T}; Y_{0:T})$ is the joint log-likelihood of data and random effects. Defining the function

$$g(\hat{\theta}, X_{0:T}, \epsilon; Y_{0:T}) = -\log\left(\int \exp(\ell(\hat{\theta}; X_{0:T}, Y_{0:T}) - \epsilon f(\hat{\theta}, X_{0:T})) dX_{0:T}\right),$$

the gradient with respect to ϵ is

$$\frac{\partial}{\partial \epsilon} g(\hat{\theta}; X_{0:T}, \epsilon; Y_{0:T}) = \frac{\int \exp\left(\ell(\hat{\theta}; X_{0:T}, Y_{0:T}) - \epsilon f(\theta, X_{0:T})\right) f(\hat{\theta}, X_{0:T}) dX_{0:T}}{\int \exp\left(\ell(\hat{\theta}; X_{0:T}, Y_{0:T}) - \epsilon f(\theta, X_{0:T})\right) dX_{0:T}}$$

Evaluating the gradient at $\epsilon = 0$, the mean of the transformed variable is obtained:

$$\left.\frac{\partial}{\partial \epsilon} g(\hat{\theta}, X_{0:T}, \epsilon; Y_{0:T})\right|_{\epsilon=0} = E(f(\hat{\theta}, \tilde{X}_{0:T}) \mid Y_{0:T})$$

In summary, the mean of f can be obtained by appending $\epsilon f(\theta, X_{0:T})$ to the negative log-likelihood and evaluate the gradient of the new marginal likelihood of data with respect to ϵ. The integral can be evaluated by the Laplace approximation; however, then the calculated mean is also an approximation. Similar calculations can be made for the variance.

2.4 Using the Laplace approximation in filters

A clear limitation of using the Laplace approximation on the joint log-likelihood of the entire time series, $\ell(\theta; X_{0:T}, Y_{0:T})$, is the requirement of a unique mode

$$\hat{X}_\theta = \text{argmax}_{X_{0:T}} \ell(\theta; X_{0:T}, Y_{0:T}).$$

This requirement can be circumvented by combining the filtering method and the Laplace approximation (e.g., Shimada and Tsukuda, 2005; Koyama et al., 2010; Paper V Albertsen, 2017b).

Koyama et al. (2010) use the Laplace approximation to approximate the filtering distribution, $p(X_n \mid Y_n)$, while the prediction density is calculated by numerical integration or an asymptotic expansion. Shimada and Tsukuda (2005) and Paper V (Albertsen, 2017b) use a slightly different approach. Instead of the filtering distribution, they apply the Laplace approximation to the one step prediction $p(Y_n \mid Y_{n-1})$, while an approximate filtering distribution is obtained as a by-product. The filtering distribution is approximated by a Gaussian distribution where the mode and Hessian from the Laplace approximation are used as mean and precision, respectively. With this approach, it is enough to have a unique maximum of $g_\theta(Y_n \mid X_n)p(X_n \mid Y_{n-1})$ for each time step in the filter. This is an important observation for Paper V (Albertsen, 2017b) to allow regime switching in the model.

2.5 Regime switching state-space models

In many dynamical systems, it is reasonable to consider changes in the parameters. For instance, in econometrics, the return of an asset may alternate between high volatility and low volatility periods; an economy may be expanding or contracting; in fisheries stock assessment models, a fish stock may alternate between high and low production periods; and in animal movement models, an animal may alternate between migration and foraging. These changes in behaviour of the system dynamics can be linked to an abrupt change in the parameters. This change can be modelled by extending the state-space model with a latent finite state Markov process, S_t. The current state of the Markov chain determines the parameters to use.

Extending the model for the fully observed system is straight forward. The general state-space model defined in Equation (2.1) becomes

$$\begin{aligned} P(S_t = j \mid S_{t-1} = k) &= p_{tkj} \\ X_t &= \phi(X_{t-1}, \theta_{S_t}, \epsilon_t) \\ Y_t &= \gamma(X_t, \theta_{S_t}, \nu_t), \end{aligned} \tag{2.4}$$

In words, the probability of going from regime k to regime j is p_{tkj}, which may depend on time. Naturally, $\sum_j p_{tkj} = 1$ for all t and k. The process and measurement equations are changed such that the parameter vector used depends on the current state of S_t.

Although extending the full system model is simple, calculating the marginal likelihood of the observations is complicated even further than for the regular state-space model. To obtain the marginal likelihood, the discrete latent states must be summed over, while the continuous latent states must be integrated over. One approach to solve the problem is to sum over the discrete latent states and apply the Laplace approximation to the resulting joint likelihood of $X_{0:T}$ and $Y_{0:T}$; however, this will cause problems for the Laplace approximation.

To illustrate the problem, consider the simple case of a linear Gaussian model with two time steps and a regime switching mean parameter, $\mu_1 = -1$ and $\mu_2 = 1$:

$$X_0 \sim N(\mu_{S_0}, 0.5^2)$$
$$X_1 \sim N(\mu_{S_1} + 0.5(X_0 - \mu_{S_1}), 0.5^2),$$

where the regime transition densities are given by

$$P(S_t = 1 \mid S_{t-1} = 1) = P(S_t = 2 \mid S_{t-1} = 2) = 0.7,$$

and the observations are given by

$$Y_t \sim N(X_t, \sigma_Y^2).$$

Even in this simple example, summing over the discrete states, S_0 and S_1, results in a multimodal distribution of the continuous states for fixed data (Figure 2.4). Clearly, there is no unique optimum to construct a Taylor polynomial around, and if one of the modes is used, a Gaussian approximation will be poor. Increasing the information contained in the data by reducing the measurement variance will eventually lead to a unimodal distribution. Therefore, this approach may be feasible with high quality observations.

An alternative approach, taken by Paper V (Albertsen, 2017b), is to use the Laplace approximation before the summation for each time step of a filter. In the example above, without observations, the bimodal distribution of continuous latent variables, reduces to four regime-wise unimodal Gaussian distributions, that can be integrated by the Laplace approximation and combined to give the data likelihood (Figure 2.5).

In the linear Gaussian state-space model, Kim (1994) used this approach to extend the Kalman filter to a regime switching model. For each time step, the Kalman filter updates are made conditional on the current and the previous regime. Afterwards, the regime-wise likelihood contributions are weighted by the regime probabilities to give the marginal likelihood of the data at the current time step, given data from all previous time steps.

Building on the ideas of Kim (1994), Paper V (Albertsen, 2017b) replaces the Kalman filter updates with regime-wise Laplace approximations in each time step to calculate the marginal likelihood. In the special case of one regime, the procedure is similar to Shimada and Tsukuda (2005). By using sequential regime-wise Laplace approximations, the multi-modality problem is circumvented; however, it is at the cost of assuming the filtering distributions $p(X_t \mid Y_t, S_t)$ are Gaussian. In Paper V (Albertsen, 2017b), the mean and variance of $p(X_t \mid Y_t, S_t)$ is approximated directly from the Laplace approximation. It may be feasible to describe the filtering distribution using higher order moments using, for example, the corrections discussed in subsection 2.3.3; however, thais is a topic for future research.

2.6 Automatic differentiation

Applying the Laplace approximation requires optimizing the joint likelihood and calculating the Hessian at the optimum; both involves calculating derivatives of the joint likelihood. This process can be automated using automatic differentiation (AD, see e.g. Griewank, 2000; or Skaug and Fournier, 2006, combined with Software III for a simple implementation).

Given a computer program that defines a function, AD computes the derivatives by traversing the computational graph. The computer program is a long list of simple operations such as addition, multiplication, exponential, and square root. For each of these operations, the derivative is known, and the chain rule can be used to combine them:

$$(f(g(x)))' = f'(g(x))g'(x);$$

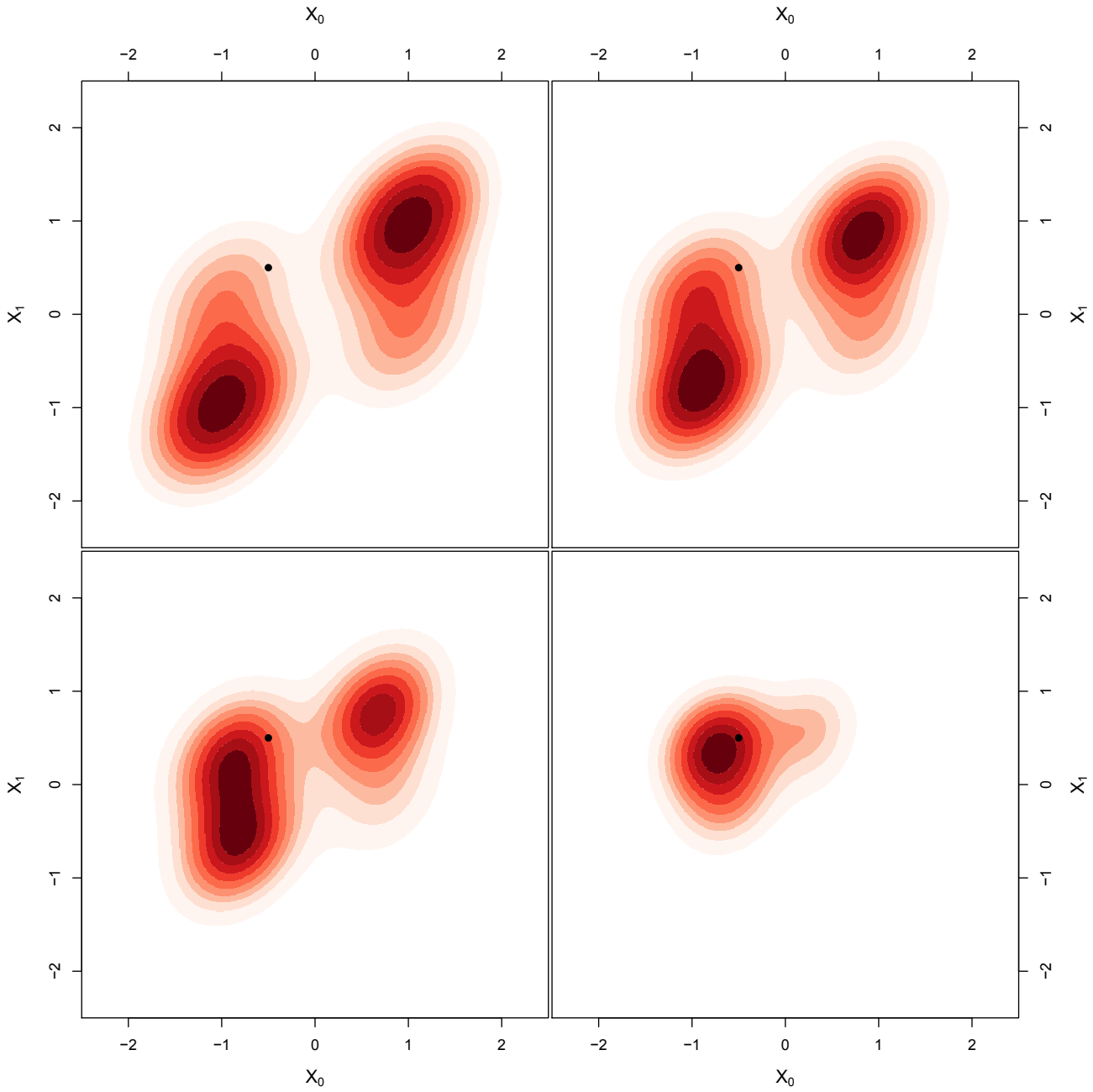


Figure 2.4: Highest density regions (Hyndman, 1996) of 10% (●), 20% (●), 30% (●), 40% (●), 50% (●), 60% (●), 70% (●), 80% (●), and 90% (●) for the density of $X_0, X_1 \mid Y_0 = -0.5, Y_1 = 0.1$ without a likelihood contribution from the observations (upper left), with $\sigma_Y = 3\sigma_X$ (upper right), with $\sigma_Y = 2\sigma_X$ (lower left), and $\sigma_Y = 1\sigma_X$ (lower right). Black points indicate the observation (Y_0, Y_1).

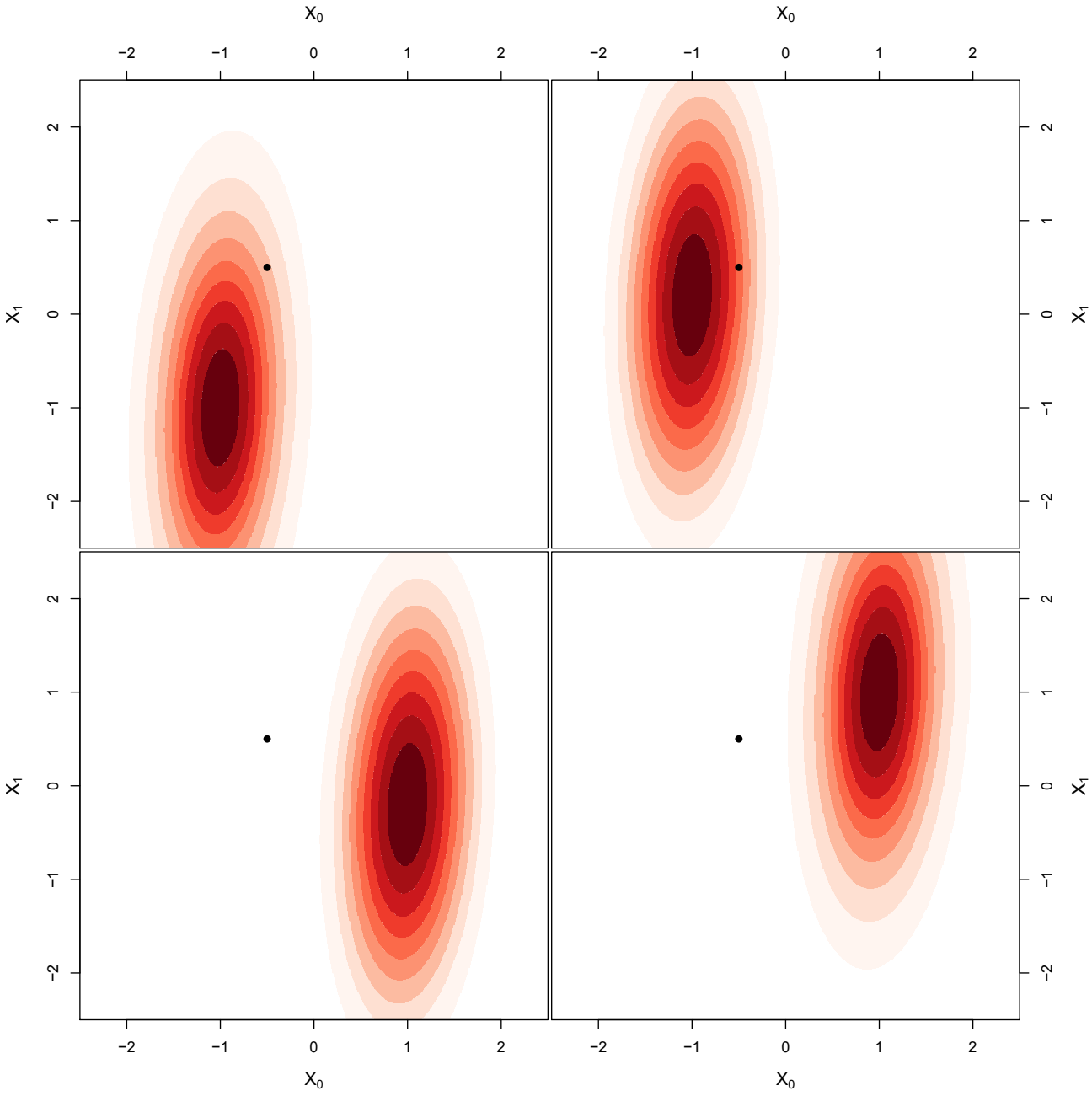


Figure 2.5: Regime-wise highest density regions (Hyndman, 1996) of 10% (●), 20% (●), 30% (●), 40% (●), 50% (●), 60% (●), 70% (●), 80% (●), and 90% (●) for the density of $X_0, X_1 \mid S_0, S_1$ without a likelihood contribution from observations for $S_0 = 1, S_1 = 1$ (upper left), $S_0 = 1, S_1 = 2$ (upper right), $S_0 = 2, S_1 = 1$ (lower left), and $S_0 = 2, S_1 = 2$ (lower right). Black points indicate the observation (Y_0, Y_1).

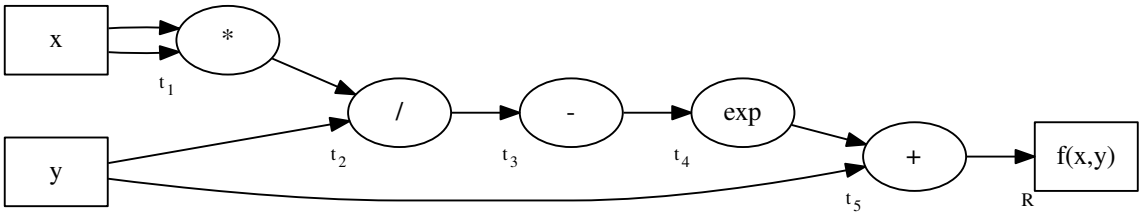


Figure 2.6: Computational graph for $f(x, y) = \exp\left(\frac{-x^2}{y}\right) + y$.

Table 2.1: Computations for forward mode AD of $f(x, y) = \exp\left(\frac{-x^2}{y}\right) + y$ with respect to x.

Node	Operation	Value	Derivative
t_1	$x \cdot x$	x^2	$\frac{\partial t_1}{\partial x} = 2x$
t_2	t_1/y	x^2/y	$\frac{\partial t_2}{\partial t_1}\frac{\partial t_1}{\partial x} = 1/y \cdot 2x$
t_3	$-t_2$	$-x^2/y$	$\frac{\partial t_3}{\partial t_2}\frac{\partial t_2}{\partial x} = -1 \cdot 2x/y$
t_4	$\exp(t_3)$	$\exp(-x^2/y)$	$\frac{\partial t_4}{\partial t_3}\frac{\partial t_3}{\partial x} = \exp(t_3) \cdot (-2x/y)$
t_5	$t_4 + y$	$\exp(-x^2/y) + y$	$\frac{\partial t_5}{\partial t_4}\frac{\partial t_4}{\partial x} = 1 \cdot (-\exp(t_3) \cdot 2x/y)$
R	t_5	$\exp(-x^2/y) + y$	$\frac{\partial R}{\partial t_5}\frac{\partial t_5}{\partial x} = -2x/y \cdot \exp(-x^2/y)$

or in a more convenient notation,

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y}\frac{\partial y}{\partial x}.$$

To calculate the derivative of the entire function, the computer only needs to keep track of all the simple operations used, use the simple derivative formulas, and combine them by the chain rule. Two variants of AD exist. Forward mode AD calculates derivatives while traversing forward through the graph, while reverse mode AD calculates derivatives while traversing backward through the graph.

2.6.1 Forward mode AD

To illustrate the calculations of forward mode AD, consider the function

$$f(x, y) = \exp\left(\frac{-x^2}{y}\right) + y \tag{2.5}$$

with the computational graph shown in Figure 2.6.

To calculate the derivative $\frac{\partial}{\partial x} f(x, y)$, start with input values x and y. The first operation, node t_1, is multiplication of x with itself. The derivative of $x \cdot x$ with respect to x is known to be $2x$. Hence, the value x^2 and the derivative $2x$ can be recorded at node t_1. The next node, t_2, is the division t_1/y. The derivative of this operation with respect to t_1 is $1/y$. By using the chain rule, the derivative of t_2 with respect to x is $1/y \cdot 2x$, which can be recorded together with the value x^2/y at note t_2. This procedure is continued until the final result is returned (Table 2.1).

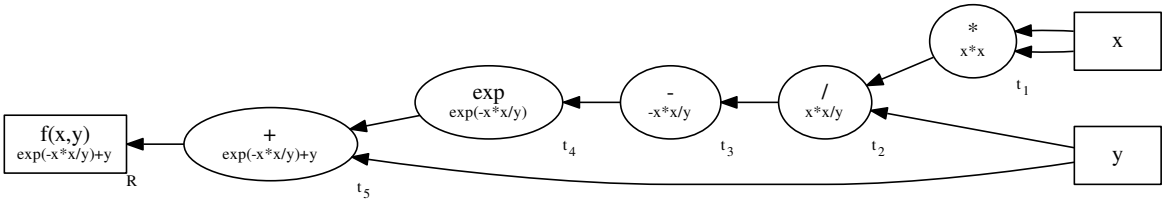


Figure 2.7: Reversed computational graph for $f(x,y) = \exp\left(\frac{-x^2}{y}\right) + y$ after a forward sweep recording the function value.

Table 2.2: Computations for reverse mode AD of $f(x,y) = \exp\left(\frac{-x^2}{y}\right) + y$.

Edge	Derivative to calculate	Calculated derivative	Combined derivative
$R \leftarrow t_5$	$\frac{\partial R}{\partial t_5}$	1	$\frac{\partial R}{\partial t_5} = 1$
$t_5 \leftarrow t_4$	$\frac{\partial t_5}{\partial t_4}$	1	$\frac{\partial R}{\partial t_4} = 1$
$t_4 \leftarrow t_3$	$\frac{\partial t_4}{\partial t_3}$	$\exp(t_3)$	$\frac{\partial R}{\partial t_3} = \exp(-x^2/y)$
$t_3 \leftarrow t_2$	$\frac{\partial t_3}{\partial t_2}$	-1	$\frac{\partial R}{\partial t_2} = -\exp(-x^2/y)$
$t_2 \leftarrow t_1$	$\frac{\partial t_2}{\partial t_1}$	$1/y$	$\frac{\partial R}{\partial t_1} = -\exp(-x^2/y)/y$
$t_1 \leftarrow x$	$\frac{\partial t_1}{\partial x}$	$2x$	$\frac{\partial R}{\partial x} = -2x\exp(-x^2/y)/y$
$t_5 \leftarrow y; t_2 \leftarrow y$	$\frac{\partial t_5}{\partial y}; \frac{\partial t_2}{\partial y}$	$1; -t_1/y^2$	$\frac{\partial R}{\partial y} = \frac{\partial R}{\partial t_5}\frac{\partial t_5}{\partial t_y} + \frac{\partial R}{\partial t_2}\frac{\partial t_2}{\partial t_y}$ $= -\exp(-x^2/y)\cdot(-x^2/y^2) + 1\cdot 1$

At the final node, R, we have the value $\exp(-x^2/y) + y$ and derivative $\frac{\partial}{\partial x} f(x,y) = -2x/y \cdot \exp(-x^2/y)$. To calculate the derivative with respect to y, the graph is traversed again.

2.6.2 Reverse mode AD

To calculate derivatives of the function (2.5) using reverse mode AD, a forward sweep through the computational graph must be made to record the function value at each node. Afterwards, the graph is traversed in reverse (Figure 2.7), starting at the final node, R.

Sweeping backwards through the graph, the derivative at each egde is calculated. The first edge in the sweep is $t_5 \rightarrow R$ corresponding to the derivative $\frac{\partial R}{\partial t_5}$. Going from t_5 to R, the result is returned. The derivative of this operation is 1. The next edge is $t_4 \rightarrow t_5$ corresponding to the derivative $\frac{\partial t_5}{\partial t_4}$. The derivative of $t_4 + y$ with respect to t_4 is 1. By using the chain rule, the combined derivative, $\frac{\partial R}{\partial t_4}$, is obtained:

$$\frac{\partial R}{\partial t_4} = \frac{\partial R}{\partial t_5}\frac{\partial t_5}{\partial t_4} = 1 \cdot 1 = 1.$$

This procedure is continued until the parameter nodes are reached (Table 2.2). Special care is needed for y. Since two edges are leaving the y node, the derivative with respect to y is calculated as the sum of the derivatives obtained from each path.

After a single reverse sweep of the computational graph, the full gradient

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right)$$

evaluated at the input parameters has been calculated. Higher order derivatives can be calculated by constructing the computational graph for the gradient and apply reverse mode AD on the new graph.

In statistical applications, the function of interest is the negative log-likelihood, $\ell : \mathbb{R}^p \mapsto \mathbb{R}$. Hence, with reverse mode AD, the score, $\nabla_\theta \ell$, can be obtained with a single forward sweep and a single backwards sweep; irrespectively of the dimension, the cheap gradient principle yields that (in theory) the gradient can be calculated with less than four[3] times the computational cost of calculating the function value (e.g., Skaug and Fournier, 2006) Likewise, for the Laplace approximation the first and second order derivatives of the joint log-likelihood is needed for finding the mode and calculating the marginal likelihood. These can be obtained automatically through AD.

2.7 Sparsity of Hessians

While the gradient of the joint likelihood of a state-space model can efficiently be calculated using AD, the Hessian (i.e., the gradient of the function $\nabla_{X_{0:T},\theta} \ell : \mathbb{R}^{T+1+p} \mapsto \mathbb{R}^{T+1+p}$) requires more work as the graph must be traversed several times. Further, finding the mode of latent states requires computations involving the large Hessian matrix.

Therefore, care must be taken in formulating the state-space model to be used with the Laplace approximation. If the Hessian matrix with respect to the latent states is dense, the computational complexity will increase rapidly with the number of states. On the other hand, if the Hessian matrix is sparse, matrix algorithms can utilize the structure to reduce the computational complexity.

For illustration, consider a random walk with measurement error. It can be formulated through the independent state increments

$$\Delta X_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_s^2)$$

where the observations are

$$Y_t = \sum_{i=0}^{t} \Delta X_i + \eta_t$$

with $\eta_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_o^2)$. Alternatively, the same model can be formulated through the state locations as

$$\begin{aligned} X_t &= X_{t-1} + \epsilon_t \\ Y_t &= X_t + \eta_t \end{aligned}$$

where $\epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_s^2)$ and $\eta_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_o^2)$.

Both formulations describe the same statistical model; however, the former is not suited for efficient inference with the Laplace approximation. With the iid model formulation, the Hessian matrix of the latent states is dense (Figure 2.8 upper left panel). As a result, using the Laplace approximation is slow, and the time to compute the marginal likelihood increases rapidly with the number of time steps. Conversely, the location-based model formulation results in a sparse Hessian matrix. The sparse Hessian results in faster

[3] or three to five depending on the assumptions (Griewank, 2000).

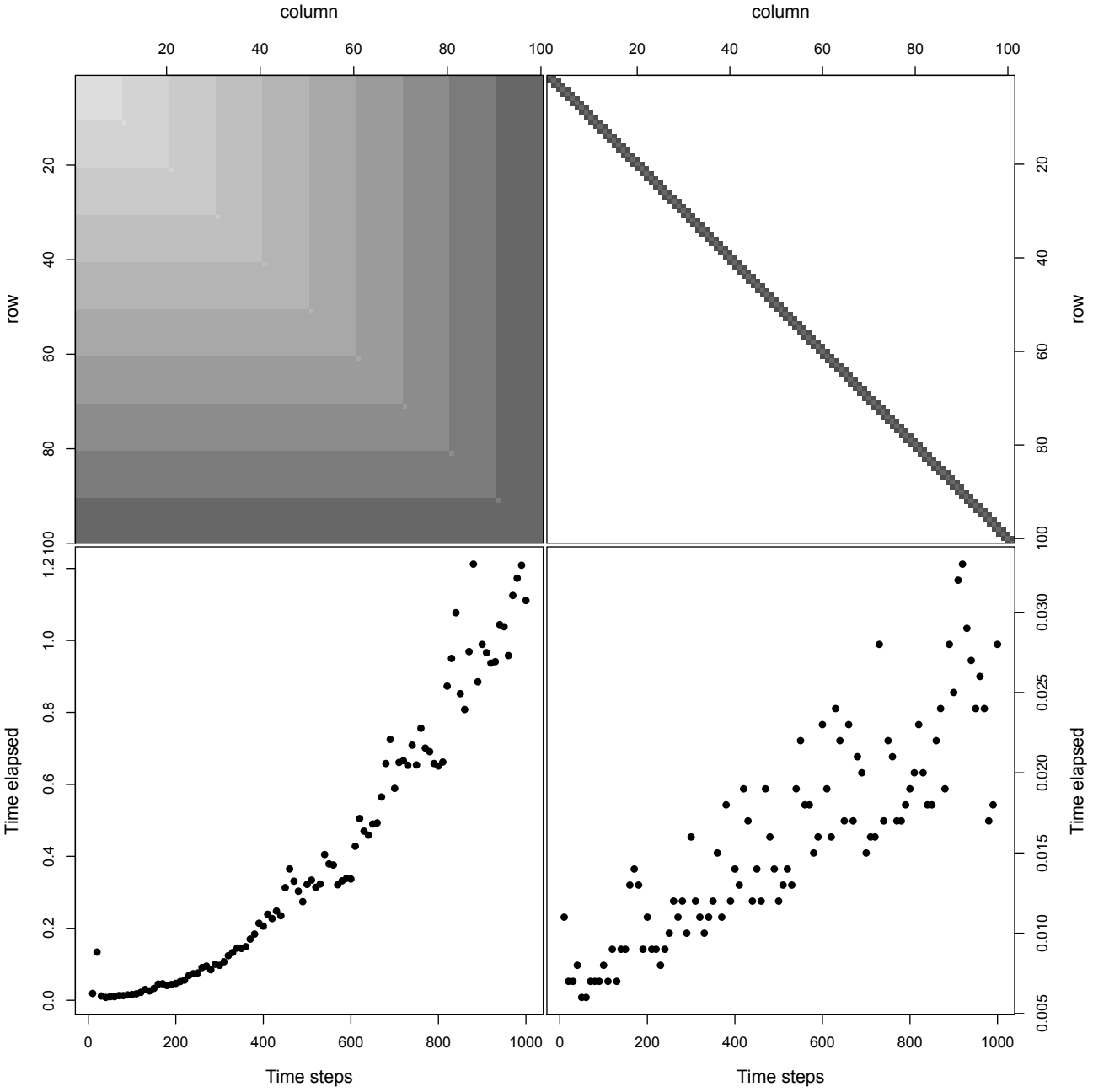


Figure 2.8: Hessian for 100 latent states (upper panels) and time to calculate the marginal likelihood as a function of the number of latent states (lower panels) for the iid (left panels) and location (right panels) based formulations of the random walk with measurement error. In the Hessian matrices, white indicates a structural zero while the remaining colour breakpoints are:[-10; 0) (●), [0; 10) (●), [10; 20) (●), [20; 30) (●), [30; 40) (●), [40; 50) (●), [50; 60) (●), [60; 70) (●), [70; 80) (●), [80; 90) (●), [90; 100) (●), and [100; 110) (●)

calculations and better scaling of the problem to longer time series. In the example above, evaluating the marginal likelihood for 1000 latent states took 1.111 seconds in the iid parameterisation, while it only took 0.028 seconds in the location based parameterisation (Figure 2.8).

2.8 Model validation and selection

Another difficulty in state-space models is validation. Using the marginal model, regular statistical concept such as forecast residuals, hypothesis test and information criteria can be used for model validation and evaluation. Further, a model may be validated by its consistency when adding data, and its predictive ability. If a system is adequately modelled, the model should be able to predict new observations generated by the system. Likewise, from a certain point, adding new observations should not drastically change parameter or state estimates.

2.8.1 Forecast residuals

Through one-step predictions, residuals are often used to investigate shortcomings of a model; autocorrelation, non-zero mean or variance heterogeneity may indicate, for instance, missing dependence, covariates or flexibility in the model. In general non-linear non-Gaussian models, simulation based quantile residuals may be used for model validation (Smith, 1985). In a time series context, the quantile residuals of continuous observations, y_t, rely on the probability

$$U_t = P(Y_t \leq y_t \mid Y_{0:t-1} = y_{0:t-1}).$$

Under the true model, the U_t's are independent and uniformly distributed. Independent standard normally distributed residuals are obtained by the transformation

$$Z_t = \Phi^{-1}(U_t),$$

where Φ is the cumulative distribution function of a standard normal distribution. For discrete observations (Smith, 1985; Früiiwirth-Schnatter, 1996), U_t defined above is not uniformly distributed. Instead, a random interpolation can be used to calculate U_t,

$$U_i = (1-\alpha_t)P(Y_i \leq y_i - 1 \mid Y_{1:i-1} = y_{1:i-1}) + \alpha_t P(Y_i \leq y_i - 1 \mid Y_{1:i-1} = y_{1:i-1}),$$

where $\alpha_t \sim Unif(0,1)$. This corrected version of U_t will be continuous uniformly distributed on $(0,1)$.

In filters, quantile residuals can be calculated during the recursive loop over time steps (Früiiwirth-Schnatter, 1996). During the recursion, $p(Y_t \mid Y_{t-1})$ is readily available and the residuals can be calculated through

$$P(Y_t \leq y_t \mid Y_{0:t-1} = y_{0:t-1}) \int_{-\infty}^{y_t} p(Y_t = y \mid Y_{t-1}) dy,$$

provided that the Markov property can be invoked. The integral can be solved by, for example, numerical methods.

For models marginalized by the Laplace approximation, the one-step data predictions are not readily available. Instead, Thygesen et al. (2017) provides a procedure for calculating the residuals; also when the Markov property does not hold. Thygesen et al. (2017) recognize the uniform residuals, U_t, as

$$U_t = \frac{P(Y_t \leq y_t, Y_{0:t-1} = y_{0:t-1})}{P(Y_{0:t-1} = y_{0:t-1})}.$$

Defining

$$\begin{aligned} f_t(y) &= P(Y_t = y, Y_{0:t-1} = y_{0:t-1}) \\ &= \int f_{X_{0:T}Y_{0:t}}(X_{0:T}, [y_{1:i-1}, y]; \theta) dX_{0:T} \end{aligned}$$

where $f_{X_{0:T}Y_{0:t}}$ is the joint density of all states and all observations up to time i, and $[y_{1:i-1}, y]$ is the reduced data set up to time i where y_i is replaced by the integration variable y. Using f_t, the uniform residuals can be calculated by

$$\begin{aligned} U_t &= \frac{\int_{(-\infty, y_t]} f_t(y) dy}{\int f_i(y) dy} \\ &= \frac{\int_{(-\infty, y_i]} f_i(y) dy}{\int_{(-\infty, y_i]} f_i(y) dy + \int_{(y_i, \infty)} f_i(y) dy} \end{aligned}$$

Note that if the observations are discrete, the integrals reduce to a sum. In this expression, $f_i(y)$ can be calculated by the Laplace approximation, and integrated by numerical integration.

It is important to note that using the Laplace approximation, Pearson residuals based on the state estimates, $Y_{0:T} - \hat{X}_\theta$, will not be independent. A dependence is introduced because the state estimates naturally provided by the Laplace approximation are smoothed in the sence that they depend on the entire data set, $Y_{0:T}$.

2.8.2 Akaike information criterion

While residuals are used for validating a single model, information criteria can be used for comparing several models. Several information criteria exist; however, the Akaike information criterion (AIC; Akaike, 1974) has a convenient interpretation. The AIC is calculated by

$$AIC = -2\ell(\hat{\theta}) + 2p$$

where $\ell(\hat{\theta})$ is the log-likelihood function evaluated at the maximum likelihood estimates, and p is the number of parameters. Differences in the AIC between models asymptotically estimates the expected difference in Kullback-Liebler divergences to the true data generating system; a measure of the information lost by using the candidate model instead of the true model.

Since the $2p$ term only holds asymptotically, several variants of the AIC exist where the term has been replaced by other approximations derived for specific settings. Likewise, other information criteria exist that does not estimate Kullback-Liebler divergence. Paper I (Albertsen et al., 2017a) and Paper II (Albertsen et al., 2017b) use the AIC to compare models; however, another variant could be used instead.

While Paper II (Albertsen et al., 2017b) uses AIC to compare several nested models, Paper I (Albertsen et al., 2017a) compares different measurement distributions in a fisheries stock assessment model. As the data is multivariate, each measurement distribution defines a family of models where the largest model estimates all parameters freely, and the smallest model estimates all parameters as equal, $\theta_1 = \theta_2 = \ldots = \theta_p$. To give a fair comparison between measurement distributions, the model attaining the minimal AIC from each family should be used. Calculating the AIC for each sub-model in each family would require extensive model fitting. Instead, Paper I (Albertsen et al., 2017a) notes that a lower and an upper bound for each family can be found. The upper bound for the minimal AIC within a family is simply the lowest AIC of a model fitted from the family. A lower bound can be found by noting that the minimal negative log-likelihood within the family is obtained for the full model, while the minimal

number of parameters is found for the smallest model. Combining these two lower bounds for each term in the AIC, a lower bound is obtained for the minimal AIC within the family. In other words, for n model specifications where an i subscript indicates a value for the ith model,

$$\min_{i\in\{1,\ldots,n\}}(AIC_i) \geq \min_{j\in\{1,\ldots,n\}}(\ell(\hat{\theta})_j) + \min_{k\in\{1,\ldots,n\}}(2p_k)$$

Comparing the resulting intervals, poor models can be excluded.

For any model selection tool, care must be taken in subsequent use of parameter estimates and confidence intervals, as these do not take model selection uncertainty into account (see e.g., Chatfield, 1995, for a discussion). In marine ecological applications, model selection is not the only source of uncertainty not taken into account. Often, data is not actually observed data, but aggregated or results from other models; only a small subset of relevant predictors are available; measuring a system can influence it unpredictably; and the data generating system is too complex to adequately describe by any model.

2.9 Using state-space models in marine science

There are several advantages to using state-space models over a regular time series model in marine science. In many marine systems, it is often reasonable to believe, that there is a good understanding of either the system dynamics, at least at some level, or the measurement errors.

Subject knowledge of physical or biological processes can provide the deterministic part of the state dynamics in many cases. Likewise, when the measuring equipment can be tested, independent studies with known fixed states can provide knowledge of the distribution of measurement errors. Extending the deterministic system model to a marginal model of observations including both system and measurement noise is, in general, not easy. However, using the state-space model frame-work, the procedure is obvious. Similarly, extending a marginal time series model with subject knowledge of the measurement errors is not straightforward without the separation of process and measurements the state-space model provides.

Further, state-space model can easily handle missing data and data from different sources. If a system is observed in more than one way, both measurement equations can be included to utilize all available information about the system. Conversely, if data is missing, the corresponding latent state variable can still be included, while it is not influenced directly by an observation; corresponding to including an observation censored on the entire observation space. In a marginal time series model, an observed value is often needed to calculate the likelihood of the next observations.

Naturally, there are also limitations of state-space models. Because the data is often highly correlated, some parameters can be weakly unidentifiable, unless the time series is long (e.g., Auger-Méthé et al., 2016). Likewise, the ratio between measurement and process stochasticity can result in models emulating either independent observations or observations without error. Further, since the process is unobservable it is impossible to validate directly. It can only be validated by its internal consistency, through observations, or by comparing with data from the same system collected differently. Finally, methods for estimating state-space models are computationally demanding, even when based on approximations such as the Laplace approximation. However, in systems that cannot be observed directly or perfectly, state-space models are a necessity for statistical analysis.

Chapter 3

Fisheries stock assessment

The first application of state-space models in this thesis is in fisheries stock assessment models. In fisheries stock assessment models, the aim is to infer the abundance of a fish stock over time from catches and scientific surveys. The true abundance can never be observed directly, which makes state-space models a natural framework to use.

Contemporary fisheries stock assessment models depend on various biological and fisheries data for the analysis. Further, several approaches exist to analyse these data for use in management and conservation. Paper I (Albertsen et al., 2017a) and Paper II (Albertsen et al., 2017b) are focused on extending current models for management and conservation in an operational way. For a model to be operational, it must use the data sources currently used, be estimated quickly, and provide the output required by managers. To limit the scope of this chapter, data and models will be described in a European management context for data rich species with a focus on state-space models. Occasionally, the descriptions will branch out to other regions and methods.

3.1 European fisheries management

In Europe, the International Council for the Exploration of the Sea (ICES) provides advice on sustainable fisheries for policymakers. The advice builds on fisheries stock assessment models.

The production of the advice can be separated into four phases (ICES, 2016b). For advice in 2018, the first phase is collecting data up to December 31st 2017. For some stocks, the data collection may continue into 2018. In the second phase, the stock status by January 1st 2018 is assessed. Here, a stock assessment model is used to estimate the size of the stock and the mortality induced by fishing. The next phase is to make assumptions about 2018, as the year is not over yet, to use in the final phase. In the final phase, the stock status is forecasted for 2019 based on the possible catch options to be taken in the year.

The advice production process poses some constraints on stock assessment models to be operational for management. Firstly, the stock status should be assessed by January 1st. Secondly, the model must be able to forecast the stock status given assumptions. Finally, European stocks are managed with a stock by stock system, which requires that models deliver stock specific statuses (ICES, 2016a).

For each stock, reference points are set for the fisheries-induced mortality and the stock size, which the management strategy should ensure the stock is below respectively over, to ensure a sustainable fishery.

To be operational in fisheries management, models should either estimate the reference points or produce output from which they can be derived.

3.2 Models

Several models have been developed for fisheries stock assessments. The models differ in complexity, required data, and methodology; some models estimate the total biomass of a population while others estimate a length or age distribution, and some models are space aggregated while others are spatially explicit. Historically, state-space models have received limited attention in fisheries stock assessment modelling (see e.g., Quinn, 2003; and Aeberhard et al., 2018, for historical accounts); however, the framework has received renewed attention due to the technological advances in computational statistics (Aeberhard et al., 2018). This section will focus on biomass models, which illustrates many of the current management concepts, and age-based assessment models which are used in Paper I (Albertsen et al., 2017a) and Paper II (Albertsen et al., 2017b).

3.2.1 Biomass models and reference points

The change in biomass of a stock can be expressed through the production, $R(B_t)$, the natural mortality rate, M, and the fisheries induced mortality rate, F (Schaefer, 1954):

$$\frac{dB_t}{dt} = R(B_t) - MB_t - FB_t$$

The production is a combination of growth and recruitment; both are functions of the current biomass. Combining $R(B_t) - MB_t$ to a single term, $P(B_t)$, the change in biomass can be expressed as the difference between the surplus production of the stock and the fisheries catch.

In its essence, fisheries management is built on the observation that a stable population is maintained when $\frac{dB_t}{dt} = 0$; that is, when the catch equals the surplus production of the stock. If the catch is higher than the surplus production, the stock size will decrease; on the other hand, if less than the surplus production is caught, the stock size will increase, and a stable population could have been achieved with a higher catch. For each stock size, a different fishing mortality will produce a stable population. The aim is to select the level at which the production is maximized[1], yielding the highest sustainable catch (or yield; illustrated in Figure 3.1).

State-space models have been used to fit stochastic versions of the biomass model above and extensions (e.g, Freeman and Kirkwood, 1995; Millar and Meyer, 2000; Montenegro and Branco, 2016; Pedersen and Berg, 2017). While the deterministic model can easily be extended to a stochastic differential equation, the resulting reference points must be corrected for stochasticity (e.g., Pedersen and Berg, 2017).

3.2.2 Stock recruitment relationship

When age-wise information is available, modelling mortality and recruitment separately is often convenient, as age-wise mortality and maturity may be available. Instead of the total biomass, the

[1] Because of data and model uncertainty, a lower catch is usually used as the management target.

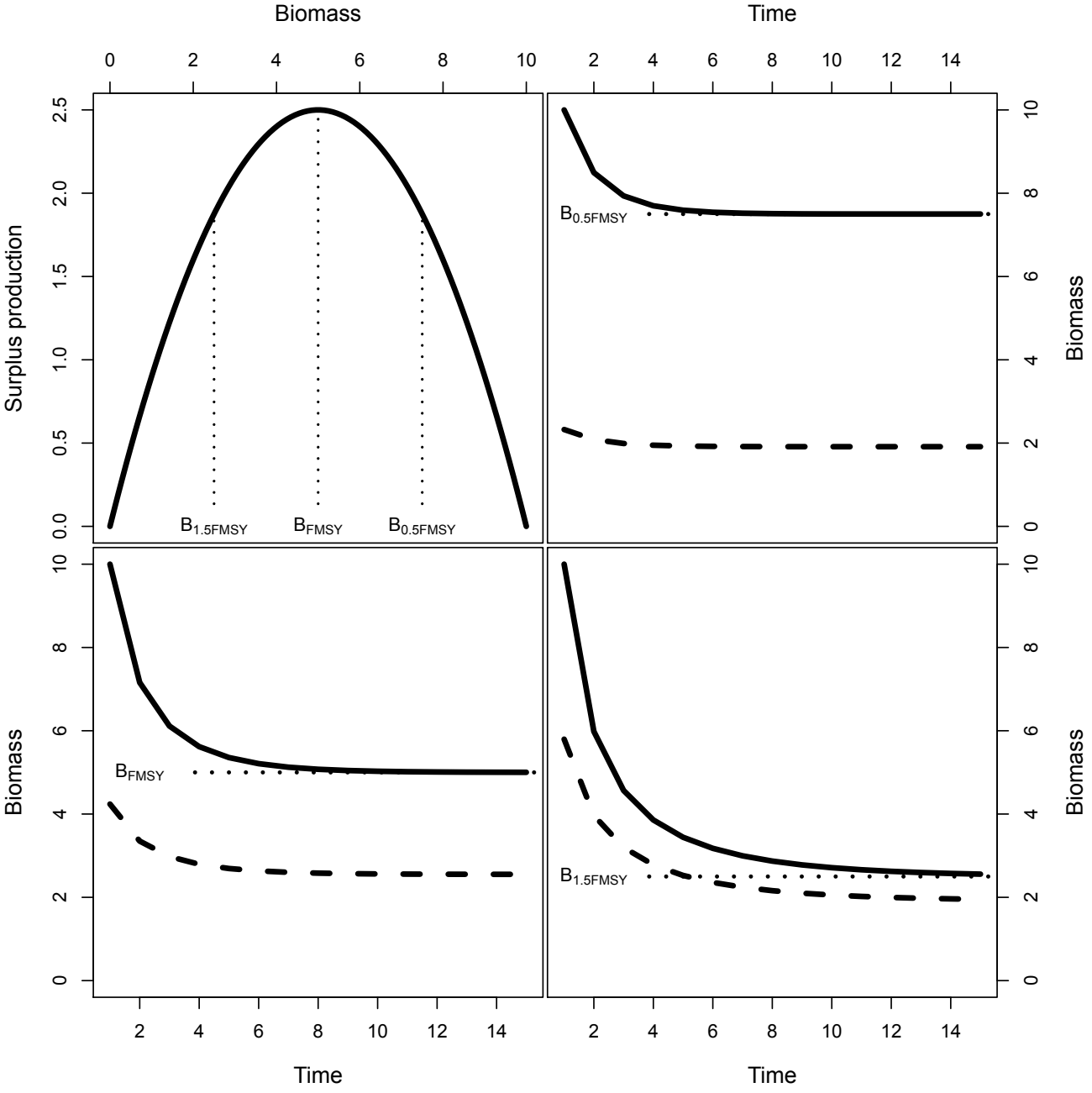


Figure 3.1: Schaefer surplus production $P(B_t) = rB_t\left(1 - \frac{B_t}{K}\right)$ (upper left panel; full line) and deterministic biomass trajectories when fishing at the rate giving the maximum sustainable yield (lower left panel; full line), when fishing at half the rate giving the maximum sustainable yield (upper right panel; full line), and when fishing at 1.5 times the rate giving the maximum sustainable yield (lower right panel; full line). Corresponding catch biomasses are indicated by dashed lines, while equilibrium biomass is indicated by dotted lines.

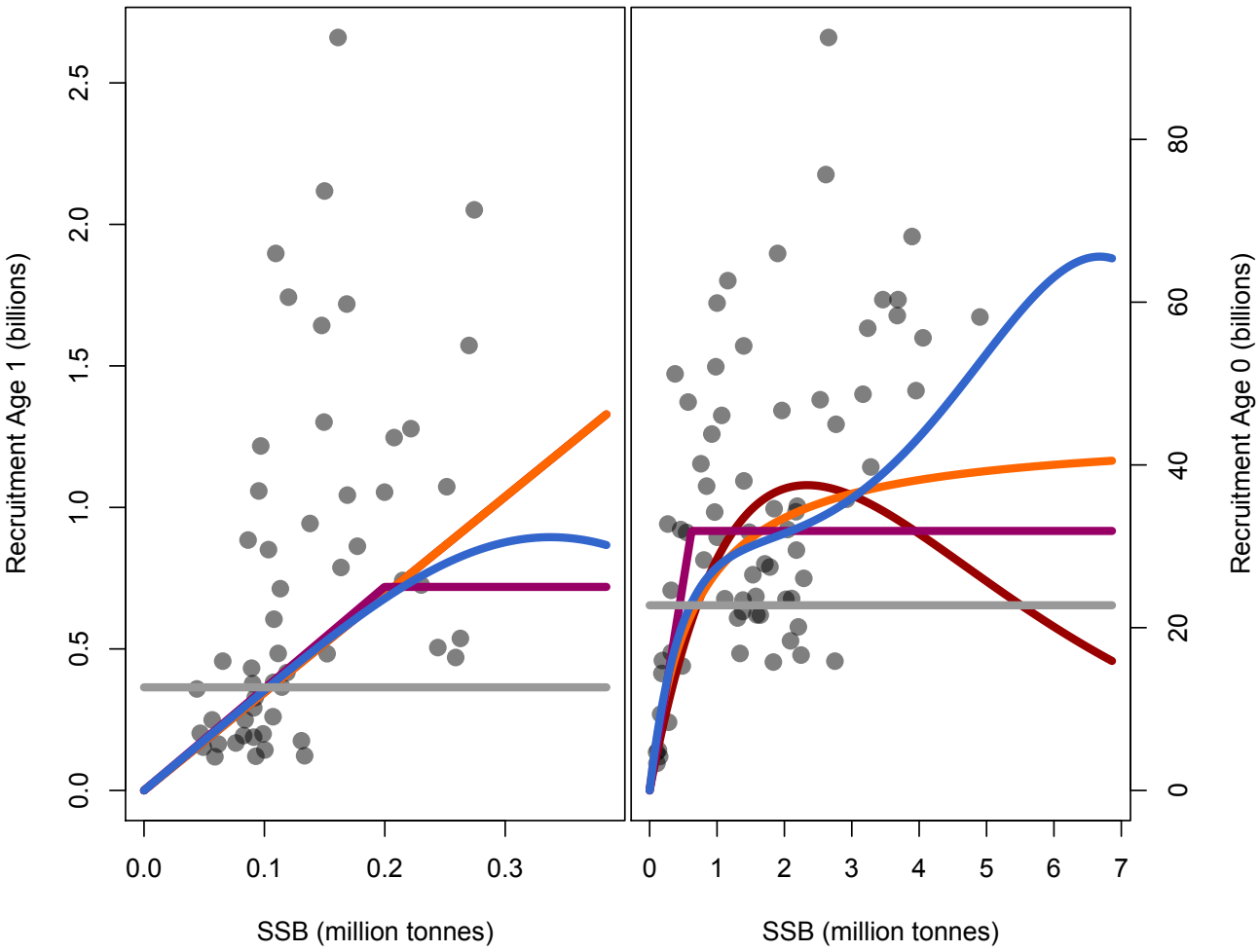


Figure 3.2: Stock-recruitment relationship for North Sea cod (left panel; data from http://standardgraphs.ices.dk/ViewCharts.aspx?key=9251 accessed Nov. 25 2017) and North Sea herring (right panel; data from http://standardgraphs.ices.dk/ViewCharts.aspx?key=8742 accessed Nov. 26 2017). Points indicate estimated recruitment and spawning stock biomass from the assessment model used for the 2017 advice. Lines are post-processed fitted Ricker (red line), Beverton-Holt (orange line), Hockey-stick (purple line), semi-parametric (blue line), and constant (grey line) stock-recruitment relationships. All relationships were fitted with log-normal errors. Note: For North Sea Cod, the Ricker and Beverton-Holt curves are coinciding.

spawning-stock biomass is often used. To calculate the spawning-stock biomass, the age-wise biomasses are weighted by the proportion of spawning individuals and added. The relationship between the stock size and recruitment, $R(S)$, is used to determine reference points for management (ICES, 2017). Therefore, it is a key output of any assessment model not directly calculating reference points such as the surplus production model.

The relationship between the stock size and the number of recruits, the production, can be modelled in several ways. Two of the most common models are the Ricker and the Beverton-Holt models. The Ricker model assumes the relationship is of the form

$$R(S) = \alpha S \exp(-\beta S)$$

with $\alpha, \beta > 0$, while the Beverton-Holt stock-recruitment relationship is of the form

$$R(S) = \frac{\alpha S}{\beta + S} \quad , \alpha, \beta > 0.$$

Another convenient form of the stock-recruitment relationship is the hockey stick:

$$R(S) \begin{cases} \alpha \frac{S}{\beta} & S \leq \beta \\ \alpha & S > \beta \end{cases}.$$

In Europe, a precautionary biomass reference point is defined from the biomass at which the stock is considered to have reduced reproductive capacity (ICES, 2017). In the hockey stick stock-recruitment relationship, this is exactly α. This level is the biomass limit, while the precautionary limit is corrected for uncertainty in biomass estimates. While the models above have few parameters, Cadigan (2013) proposes a flexible semi-parametric monotone decreasing B-spline for $\log\left(\frac{R}{S}\right)$:

$$R(S) = \exp(X(S)\gamma)S$$

where $X(S)$ is a B-spline basis evaluated at S and the spline knot-parameters are, for instance,

$$\gamma_1 = \beta_1, \quad \gamma_i = \gamma_{i-1} - \exp(\beta_j)$$

with $(\beta_1, \ldots, \beta_j)^T \in \mathbb{R}^j$, which ensures a decreasing spline. For this model, the number of knots can either be selected as a fixed number or numerous knots can be used with regularization on the parameters. The models above are illustrated for North Sea cod and herring in Figure 3.2. Several extensions, generalizations and other models have been proposed (see e.g., Needle, 2001), but will not be covered here.

3.2.3 Age-based models

For many stocks, data on the age distribution is available. Instead of modelling the total biomass, the additional information can be utilized to model the numbers-at-age for a stock. In a closed spatially uniform stock, the size of an age cohort over time can be described by the exponential decay model

$$N_{a,y} = \exp\left(-(F_{a-1,y-1} + M_{a-1,y-1}) + \epsilon_{N,a,y}\right) N_{a-1,y-1}.$$

Each year the cohort ages by one year, some die of natural causes while others die from fisheries; modelled by $M_{a-1,y-1}$ and $F_{a-1,y-1}$ respectively. Often the natural mortality rate is assumed to be known. In this case, the process variability can be thought of as a combination of inaccuracies in the known mortality and the closed stock assumption. In the cohort equation, and the equations below, ϵs will usually be independent Gaussian random variables. A new cohort is modelled by a stock-recruitment relationship above:

$$N_{0,y} = R\left(N_{0,y-1}, \ldots, N_{A,y}\right) \exp\left(\epsilon_{N,0,y}\right).$$

Besides the stock size, fishing mortality, $F_{a,y}$, is a quantity of interest for management. In non-state-space models, it is often assumed that the fishing mortality can be separated into a yearly contribution and an age contribution, $F_{a,y} = S_a f_y$, such that the parameters can be estimated. The age contribution describes the selectivity of the fishery, while the year contribution describes the fishing intensity.

In a state-space framework, Gudmundsson (1994) models fishing mortality with the separable structure extended with independent random deviations. To obtain a more flexible structure with time-varying selectivity, Nielsen and Berg (2014) models the fishing mortality on log-scale by a random walk with correlated increments,

$$\log F_{a,y} = \log F_{a,y-1} + \epsilon_{F,a,y}.$$

The stock size and fishing mortality are only observed indirectly through the catch and scientific surveys. The catch-at-age is often assumed to follow the Baranov catch equation,

$$C_{a,y} = \frac{F_{a,y}}{F_{a,y} + M_{a,y}} \left(1 - \exp\left(-(F_{a,y} + M_{a,y})\right)\right) N_{a,y} \exp\left(\epsilon_{C,a,y}\right).$$

In this equation, the catch during year y is calculated as the proportion that die from fishing, $\frac{F_{a,y}}{F_{a,y}+M_{a,y}}$, of the proportion that die $\left(1 - \exp\left(-(F_{a,y} + M_{a,y})\right)\right)$ of the number of fish at the beginning of the year, $N_{a,y}$. The Baranov catch equation assumes constant mortality rates throughout the year.

Scientific surveys are used to produce indices of abundance. It is assumed that the catch from scientific surveys are small enough to be ignored, while the resulting indices are proportional to the true stock size at the time of the survey:

$$I_{a,y} = q \exp\left(-(F + M)d/365\right) N_{a,y} \exp\left(\epsilon_{I,a,y}\right),$$

where d is the day of the survey.

While state-space modelling is not the predominant framework for analysing catch-at-age data, several examples exists of its use. For example, Gudmundsson (1994) and Schnute (1994)[2] use Kalman filters to estimate parameters and stock status in age-based state-space models. Brinch et al. (2011) investigated the Laplace approximation and importance sampling based on Laplace approximations, and found that both provided accurate estimates for a catch-at-age model. Likewise, Gudmundsson and Gunnlaugsson (2012) compared the Kalman filter to the Laplace approximation with `AD Model Builder` and found that both approaches provide accurate parameter and stock status estimates. The authors note two limitations of the Laplace approximation compared to the Kalman filter: it does not produce residuals, and it is slower. The former is now resolved by Thygesen et al. (2017), while the latter was correctly predicted to be a vanishing problem as technology and methodology advances (see e.g., Kristensen et al., 2016). Nielsen and Berg (2014) use the Laplace approximation to compare correlation structures for the fishing mortality rate, modelled as random walks. This model is extended in Berg and Nielsen (2016) and Paper I (Albertsen et al., 2017a) to investigate different observational models, and in Paper II (Albertsen et al., 2017b) to estimate correlations between stocks (see also Software IV). As the final examples, Cadigan (2016) and Van Beveren et al. (2017) use the Laplace approximation for models accounting for under-reporting of catches through censored likelihoods.

3.3 Stock assessment data

As mentioned above, assessment models mainly rely on data from the fisheries and data from scientific surveys. New data from European commercial fisheries are not readily available, however, the data used in previous assessments can be found[3]. The commercial fisheries data available are age-wise yearly aggregates over a management area.

In contrast, data from scientific surveys are available at a high resolution; however, for use in assessment models, they are aggregated to abundance indices that match the spatio-temporal resolution of the catch data. In both commercial and survey data, age-wise data is obtained through a series of sub-sampling procedures. From the total catch (either commercial or survey), samples are weighed and

[2] Schnute (1994) also considers the "Error in Variables" method, where the joint likelihood is used for estimation, and latent states are estimated together with parameters. This approach is also used in Schnute and Richards (1995).

[3] e.g., at https://stockassessment.org

measured to estimate the length distribution and average weight for each length. The measured samples are then sub-sampled to estimate age-length keys used to divide the total catch into catch-at-age.

3.3.1 Distribution of data

For such a sampling procedure, it is difficult to find the true distribution of the aggregated data a priori. It is further complicated by the fact that the sampled population - or even the sampling protocol - changes over time.

Figure 3.3 shows the spatial distribution of samples with age readings from a scientific survey in four different years. Clearly, the number of spatial samples have increased from 1987 to 2017; however, the spatial distribution of the samples has also changed slightly. This affects the distribution of the aggregated data. Similarly, changes in the spatial distribution of the population, or targeted fisheries, can influence the distribution of measurements conditional on the true population.

Likewise, Figure 3.4 illustrates the distribution of age-length samples in the same survey. The number of samples that could not be aged (group -9) has increased drastically. In 1987, 4 of 7,598 (0.053 %) samples were classified with an unknown age. In 2017, the number of samples classified as unknown had increased to 327 of 5,158 corresponding to 6.340 %. Like the spatial sampling distribution, the distribution of missing age readings can influence the result of an assessment, and the distribution of the aggregated data.

Paper I (Albertsen et al., 2017a) recognizes this challenge and compares different likelihoods already used for assessment data along with close extensions. Incorporating flexible correlation structures across ages improved all four case study assessments, as measured by AIC. In particular the multivariate log-normal and additive logistic normals performed well. The multivariate log-normal and additive logistic normal symbolise a difference in tradition in stock assessment modelling; traditions of either modelling catch numbers or catch proportions per age.

3.3.2 Numbers or proportions

In an age-based state-space model it is convenient to model age-wise catches, as each directly relate to a single latent state. However, it may be more natural to divide the age-wise catches into a total catch (in numbers or weight) and age-wise proportions of the catch, since the catch data is collected this way. In general, total catch is accurately monitored to enforce quotas, while age-wise proportions are collected through a series of subsamples; first to sample the length distribution, then to sample the age-at-length distribution.

In many proportions-at-age models, the multinomial distribution has been a popular choice. Based on the ageing procedure described above, it is also a natural choice; however, the covariance structure does not match the actual data. In particular, the variance is often to small when using the true sample size. As an ad hoc solution, the total sample size has been estimated as a parameter. That is, instead of the proper distribution for a vector of integers summing to a fixed number, the density has been used for the product of the observed proportions and a tuning parameter. In the latter case, it is not a proper distribution. Therefore, the new tuning parameter cannot be estimated, but iterative schemes must be used until the variance is adequately matched to data. Instead, a proper distribution for proportions can be used such as the Dirichlet or transformations of a normal distribution.

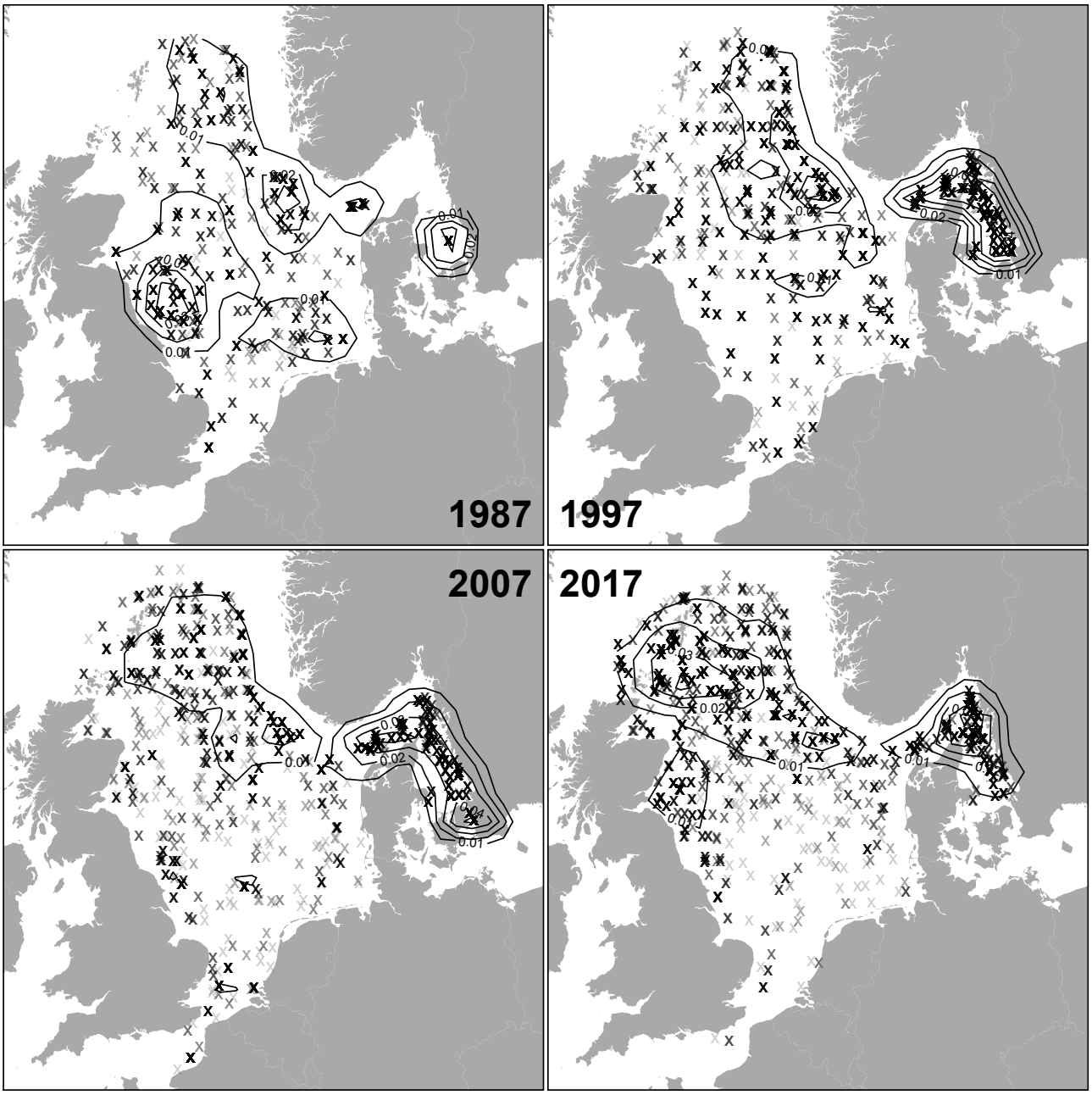


Figure 3.3: Map of survey hauls in 1987, 1997, 2007, and 2017 with age classifications in the North Sea International Bottom Trawl Survey along with kernel density estimates (ignoring land).

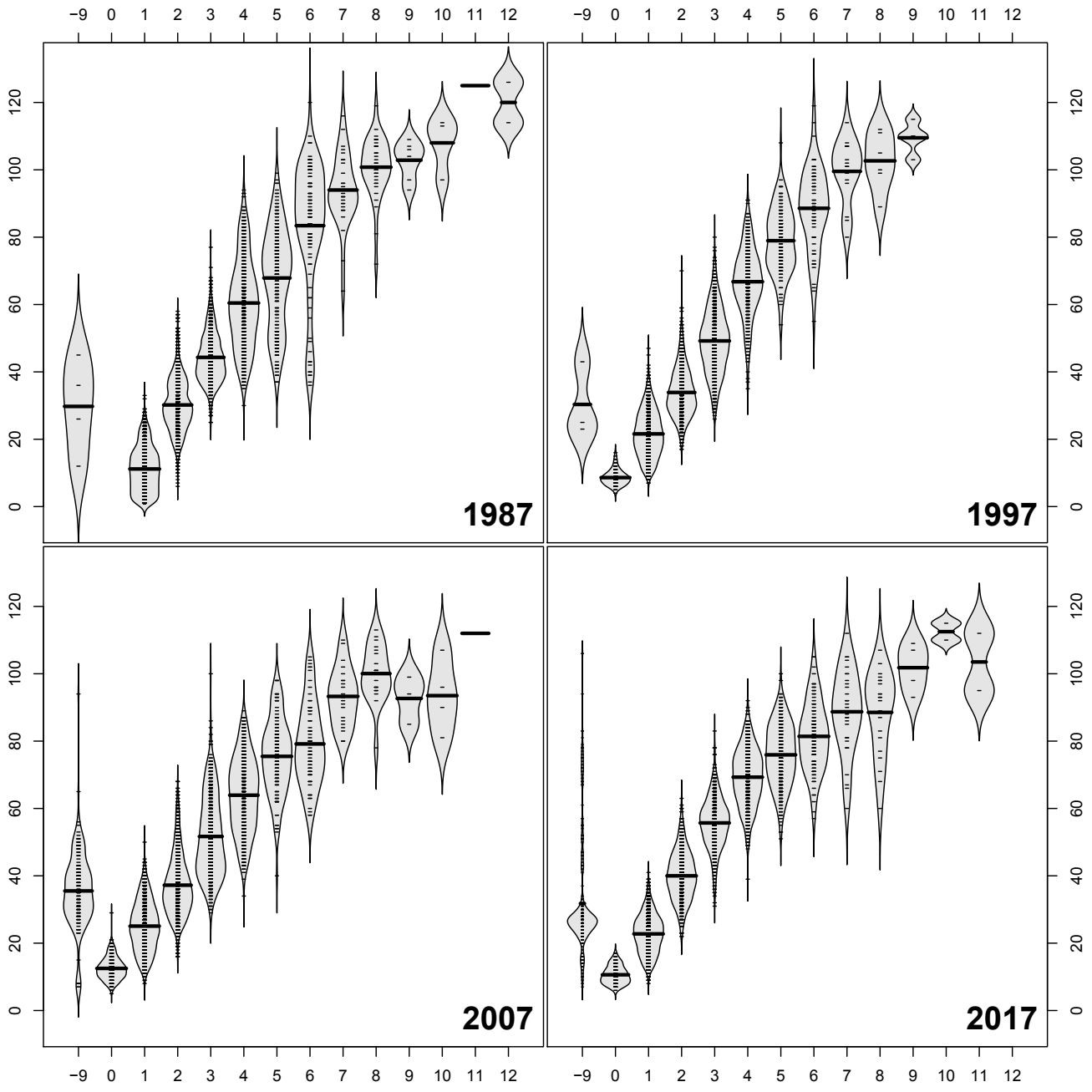


Figure 3.4: Bean plot (Kampstra, 2008) of length distributions per age group in the North Sea International Bottom Trawl Surveys 1987, 1997, 2007, and 2017. Unknown age is indicated by -9.

Paper I (Albertsen et al., 2017a) compares numbers-at-age with proportions-at-age distributions in age-based stock assessments. For models adequately including correlations, it is concluded that both formulations can be appropriate for real assessment data. Which of the alternatives is the most appropriate is determined by several factors; however, the following example indicates that the relationship between sampling error in the totals and sampling error in the proportions influences the optimal choice.

3.3.2.1 Comparing the multivariate log-normal to the additive logistic normal

Consider the following overly simplistic example for a stock with three ages where the relative stock status over time is known without error (and therefore left out of this example). To emulate the sampling scheme above, the reported catch, C_i, is simulated from a log-normal distribution with log-mean $\mu = \log(1000)$. Catch proportions-at-age, $Q_{i,a}$, are sampled from a Dirichlet distribution with concentration parameters $\alpha_a = P_a\alpha$, where P_a is the true age-distribution, $P_a = \frac{\exp(-(a-1))}{\sum_{j=1}^{3}\exp(-(j-1))}$ for $a = 1, 2, 3$. The reported catch-at-age is calculated as $C_{i,a} = Q_{i,a}C_i$. Each simulated data set consists of 100 replications of $(C_{i,1}, C_{i,2}, C_{i,3})$ $(i = 1, \ldots, 100)$. The simulation procedure has two parameters that determine the sampling error: the standard deviation on log-scale for the totals (σ), and the concentration parameter for the proportions (α). A lower value of σ indicates a more precise sample of total catches, while a higher value of α indicates a more precise sample of the proportions.

For each simulated data set, a multivariate log-normal and an additive logistic normal with AR(1) covariance, (Paper I, Albertsen et al., 2017a) were fitted. The AR(1) correlation structure is determined by the age difference between catch groups. For the multivariate log-normal, the correlation between log-catches are $Cor(\log C_{i,a}, \log C_{i,a'}) = \rho^{|a-a'|}$. For the additive logistic normal, the AR(1) covariance is not for the log-catch but for logit catch proportions. The log-normal model estimated the true catch-at-age while the additive logistic normal was combined with a log-normal for the totals to estimate the true total catch and the true catch proportions. For each simulation, the difference in AIC was calculated. Negative values indicated that the multivariate log-normal had a better fit to data than the additive logistic normal. Figure 3.5 shows the AIC differences for $\alpha = 0.5$ as a function of σ. For each value of σ, 100 replications were simulated. For low values of σ (i.e., for accurate samples of the total), the additive logistic normal provided a better fit than the log-normal as measured by AIC. As σ increases, the log-normal provides better data fits compared to the logistic normal.

Similarly, Figure 3.6 shows the AIC differences as a function of both σ and α. Again, for a fixed value of α the log-normal provides a better fit for large values of σ, while the additive logistic normal provides a better fit for small values of σ. Likewise, for a fixed value of σ, the log-normal provides a better fit for large values of α, while the additive logistic normal provides the best fit for small values of α.

The simplistic simulation study indicates that modelling catch-at-age is appropriate when the catch proportions are sampled with high precision relative to the precision of the total catch. Likewise, it is indicated that modelling proportions-at-age is appropriate when the total catch is sampled with high precision relative to the precision of the proportions.

3.4 Extending the single stock model

One of the main assumptions of the age-based assessment model presented above is that the stock is closed. There is no migration out of, or into, the management area, and interactions with other species are completely determined through the known natural mortality. In many cases, this assumption is too strong.

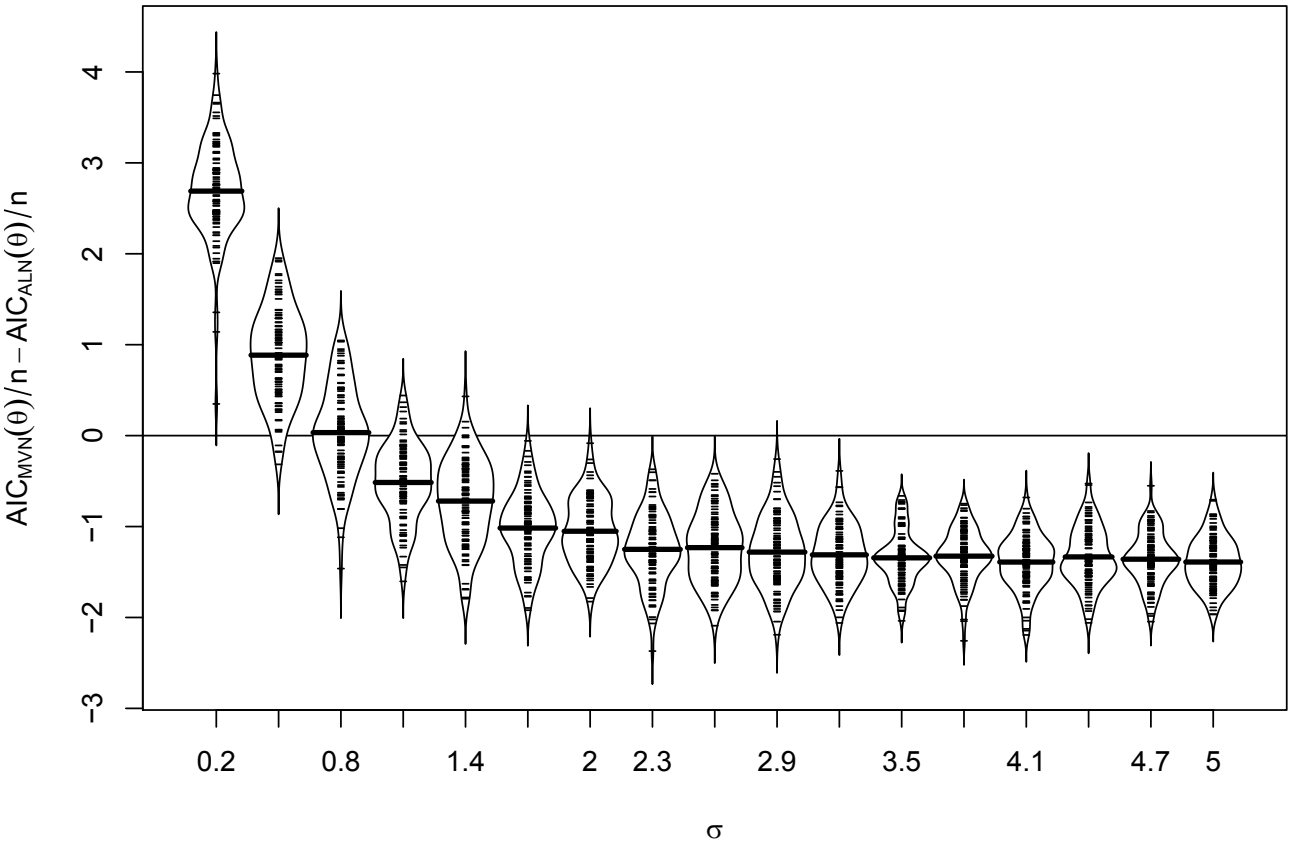


Figure 3.5: AIC difference per data point between the multivariate log-normal and the additive logistic normal for $\alpha = 0.5$.

Stocks may mix at the boundaries of management areas, and the data have to be split between the stocks before single-stock models are used (e.g., Hüssy et al., 2016).

To account for connections between stocks, several models ranging in complexity have been developed. However, most ecosystem or multi-stock models require additional data on interactions between stocks, or stocks and their surroundings, such as stomach content, migration data, or environmental data. Collecting new data sources is time-consuming and costly. Before a data set can be incorporated in an assessment, a time-series of several years is required. Further, complex ecosystem models often have a vast amount of settings and can be difficult to calibrate and run.

The aim of Paper II (Albertsen et al., 2017b) is to provide an operational compromise between the simple single-stock model, and the complex ecosystem models. Without requiring any new data sources, the single-stock model (described in subsection 3.2.3) is extended to include several stocks. Instead of deterministic interactions, the model includes correlations in the survival between the stock. The correlations are parameterized through partial correlations between the stocks. Conditional on the other stocks, ages within a stock are assumed to be uncorrelated; this formulation provides a natural extension of the single-stock models, where survival is typically uncorrelated. The proposed model improves the independent individual assessment, as measured by AIC, in all three case studies considered.

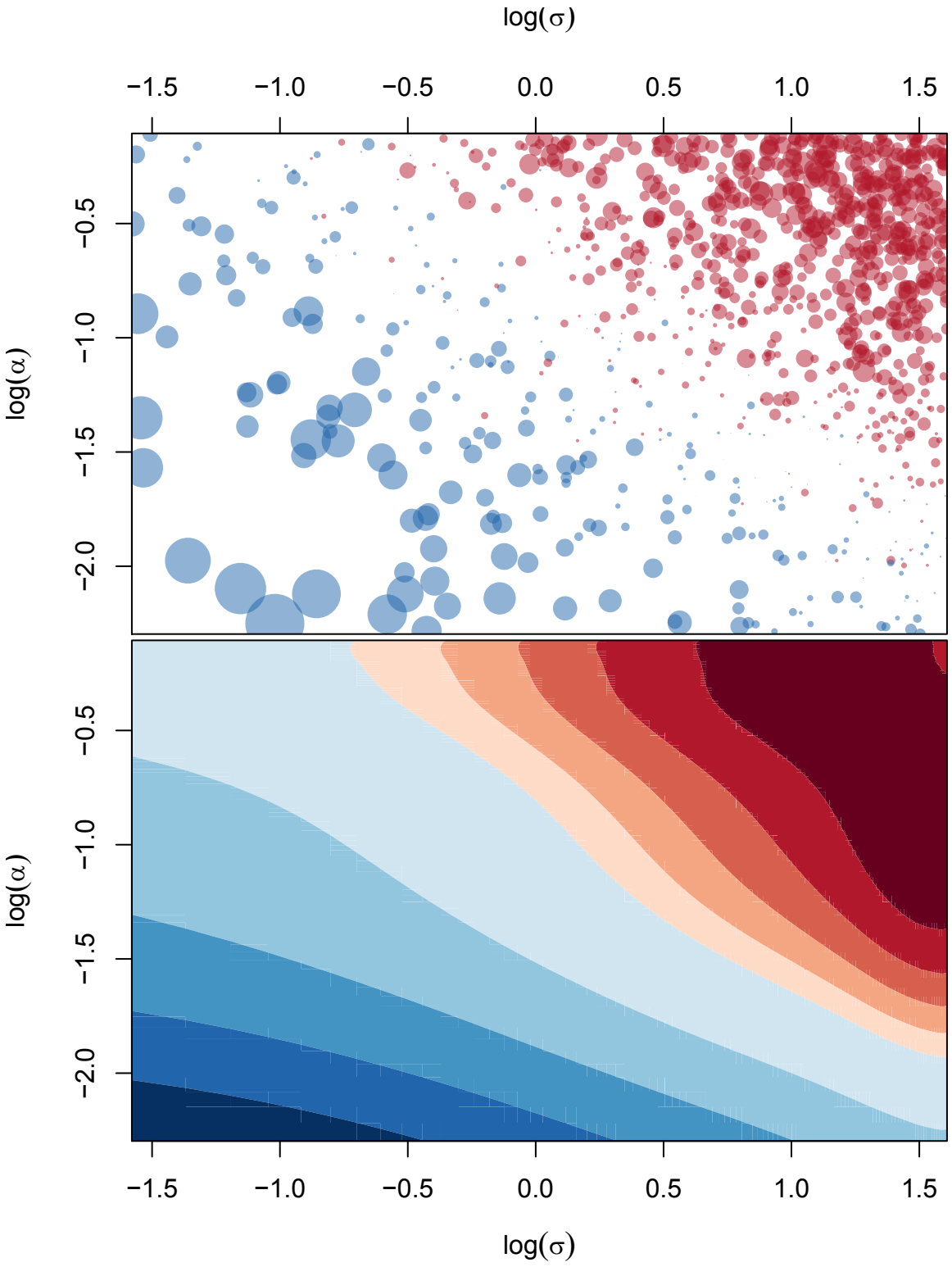


Figure 3.6: AIC difference between the multivariate log-normal and the additive logistic normal for various values of σ and α. Upper panel: AIC difference at uniformly distributed points; larger points indicate a larger difference - blue points indicate a positive difference, red points indicate a negative difference. Lower panel: Smoothed surface calculated by local polynomial regression; colors show the estimated differences (-1.566; -1.253) (●), (-1.253; -0.94) (●), (-0.94; -0.626) (●), (-0.626; -0.313) (●), (-0.313; 0) (●), (0; 1.119) (●), (1.119; 2.239) (●), (2.239; 3.358) (●), (3.358; 4.478) (●), and (4.478; 5.597) (●).

3.5 Using state-space models for fisheries stock assessment models

State-space models provide a natural framework for fisheries stock assessment models; both age-based, length-based, and biomass models. Population processes naturally evolves stochastically over time, which requires a time-series model for statistical analysis. Further, the true quantities of interest cannot be measured directly without uncertainty in marine environments. This is conveniently handled in state-space models. The separation of population process and measurements allows a cohesive approach to combining different data sources measuring the population; similar to the popular approach integrated analysis (e.g., Maunder and Punt, 2013).

Further, state-space models naturally allow variability from the deterministic population model. Alternatives to state-space models can include additional parameters to allow time-varying deviations from the deterministic model. However, to be identifiable a parametric form must be chosen, such as constant blocks or overall trends. Finally, state-space models for fisheries stock assessments inherits the advantages and disadvantages of general state-space models. The framework can provide flexible and fast inference; however, the structural form the latent states can be difficult to verify, and stock assessment models often rely on short time-series, which can make inference difficult.

Chapter 4

Individual animal movement

Understanding the movement behaviour of marine animals is crucial for evaluating the impact on their conservation from humans and environmental changes. In individual animal movement modelling, the aim is to reconstruct a movement path from uncertain data to learn about the behaviour causing the specific path. Combining several paths, even more can be learned. Modelling several animals together, additional information such as population parameters, diversity, migration patterns and spatial distributions can be estimated; all of which are important for evaluating strategies for managing species. Nonetheless, for the models to be applicable outside purely academic applications, they must be able to handle animals observed at different time scales, and the inference methods must be efficient.

In particular in marine environments, it is difficult and expensive to gather observations from a moving animal. Therefore, it is natural to separate the stochasticity in the data into two parts: the stochasticity from the animal's movement, and the stochasticity from the observation method. Therefore, like fisheries stock assessment models, individual animal movement is a natural application of state-space models.

4.1 Movement data

Several methods exist to track moving animals. For animals that surface (and birds), a common observation method is using Argos transmitters. Argos transmitters transmits a signal which is received by the Argos satellites. From the difference between the signal frequency emitted and the signal frequency received, the Doppler effect can be used to calculate a position. An approximate position, with potentially very large error, can be obtained even if the animal only surfaces shortly. The longer the animal surfaces, the more satellites can receive the signal and, in turn, the more precise the calculated position will be. As a by-product of the calculations, the Argos system provides a classification of the amount of error in the positions. From best to worst, the classes are 3, 2, 1, 0, A, B, Z. In the last class, the position is considered completely unknown. The error distribution is known to be non-normal with heavy tails for the worst classes. Therefore, Argos data is often modelled by a Student's t-distribution; either independent in the coordinates (Jonsen et al., 2005) or a multivariate distribution (Paper III, Albertsen et al., 2015). Alternatively, the data can be pre-processed to remove the most error-prone locations and in turn be modelled by a Gaussian distribution in the state-space model. Finally, several semi-parametric methods have been used to model the measurement error of Argos data.

In marine environments, GPS can typically not be used because many marine animals surface too briefly for the tag to communicate with the GPS satellites. Instead, the TrackTag or Fastloc-GPS system

can use the GPS satellites to transmit signals, while requiring a shorter surface time; less than 60 ms (Rutz and Hays, 2009). The accuracy of Fastloc-GPS positions is higher than the accuracy of Argos positions (e.g., Hazel, 2009; Witt et al., 2010); however, only the lighter-weight Argos tags can be used for small animals. With the high accuracy of GPS and Fastloc-GPS, the data is adequately described by a normal distribution; in some cases, the measurement error can even be ignored to greatly simplify the modelling framework.

For large animals that do not surface, a pop-up satellite tag can be used instead. These tags record, amongst other variables, the light level and depth. Combined with a time stamp, these variables can be used to infer the position. The resulting positions are often modelled by a normal distribution with a constant variance in the longitudinal coordinate. In the latitudinal coordinate, however, the variance depends on the time of year. Close to an equinox, the tags produce highly inaccurate positions in the latitude coordinate. Therefore, an appropriate model must increase the measurement variance for observations on days near an equinox (e.g., Sibert et al., 2003).

Alternatively, acoustic tags or passive integrated transponder (PIT) tags can be used to monitor animals in bounded regions. Both tags require a receiver to be close to the tag to get an observation. Acoustic tags emit a sound signal, that can be caught by a receiver station. Knowing the positions of the receiver stations and whether they receive a signal, the movement of an animal can be estimated (e.g., Pedersen and Weng, 2013; Baktoft et al., 2017).

As the name suggests, PIT tags are passive until they are close to the receiver. An electro-magnetic field from the receiver lets the tag transmit a signal containing a unique identifier to be recorded. The tag itself does not indicate a position; however, since the tag must be within centimetres of the receiver, the position can be inferred from the position of the receiver.

This thesis uses Argos data to model animal movement. Nevertheless, the separability of movement and measurement in state-space models, allows the same movement models to be used with other data sources. Using a new data source only requires a suitable modification of the distribution of an observation given the true position. In the `argosTrack` package (Software I), the implemented movement models can currently be freely combined with a multivariate normal, multivariate Student's t-distribution, or normals where the latitudinal variance depends on the day of year through a spline or a parametric form. The package links observations to the process through the formula (e.g., Jonsen et al., 2005)

$$Y_{s_j} = (1-q)X_{t_i} + qX_{t_{i+1}} + \eta_j$$

such that $t_i \leq s_j < t_{i+1}$, $q = \frac{s_j - t_i}{\Delta_{i+1}}$, and η_i is a zero-mean bivariate random variable for discrete time models, and

$$Y_{s_j} = X_{s_j} + \eta_j$$

for continuous time models

4.2 Models

The movement models for individual animal movement are more diverse than for the measurements; however, most models can be grouped into one of seven categories: Discrete time models of turning angle and step length; discrete time models of location or velocity; continuous time location models; continuous time velocity models; step-selection models; and functional models.

One of the simplest ways to describe animal movement is the random walk with drift in discrete time:

$$X_i = X_{i-1} + c_i + \epsilon_i$$

This model, which only includes dispersion parameters, is often used in settings with long periods between observations (e.g., Anderson-Sprecher and Ledolter, 1991; Sibert et al., 2003; Nielsen et al., 2006; Nielsen and Sibert, 2007; Lam et al., 2008; Nielsen et al., 2009; Thygesen and Nielsen, 2009). Data processed from pop-up satellite tags is typically average locations based on light-levels throughout a time period such as 12 or 24 hours. In this case, little information about small scale movement remains in the data at hand, making the random walk a sufficient model to retrace the true trajectory.

4.2.1 Discrete time angle and step length model

When the random walk does not suffice, a natural way to think of animal movement is through a series of step lengths and directions (e.g., Turchin, 1998; Morales et al., 2004; Gurarie et al., 2009; Tracey et al., 2010; McClintock et al., 2012, 2014; Michelot et al., 2016). This way of thinking is inherently discrete in time. The models are known as correlated random walks. Step lengths are often modelled by a distribution on the positive numbers, such as a log-normal, Weibull, or gamma distribution,

$$S_i \sim \Gamma(\alpha, \beta),$$

while the bearing is modelled by a circular distribution, such as a von Mises or wrapped Cauchy distribution where the mean angle depends on the last bearing,

$$\rho_i \sim WCauchy(\rho_{i-1}, \sigma)$$

From the step lengths and bearings, the coordinate position is known by

$$\begin{aligned} X_{1,i} &= \sum_{i=0}^{N} S_i \cos(\phi_i) \\ X_{2,i} &= \sum_{i=0}^{N} S_i \sin(\phi_i) \end{aligned} \tag{4.1}$$

For these models, the Laplace approximation is typically not an appropriate method for inference. Firstly, the model structure is not sparse if the observations are modelled as coordinates, which will lead to computational problems when the number of time points increases (section 2.7). For each time step, the true location must be calculated using all previous locations. To circumvent this problem, the true locations can be used as random effects instead. However, this leads to the second problem. If the variance, modelled by σ, of the circular distribution is large, and the step length distribution is located far from zero, the resulting one-step displacement distribution will be far from normal; it will wrap around zero as a doughnut.

Consider the two cases where the turning angle follows a wrapped Cauchy distribution with location parameter 0, and concentration parameter $\frac{1}{4}$ and $\frac{3}{2}$, respectively (Figure 4.1). In both cases, the step length is gamma distributed with shape parameter 90 and scale parameter 0.1; that is, mean 9 and variance 0.9.

Simulating 5,000 step lengths and turning angles, and calculating the corresponding displacements by Equation (4.1), we get the corresponding displacement distribution Figure 4.2. In both cases, there is a non-zero probability of the animal moving in any direction; however, when the wrapped Cauchy

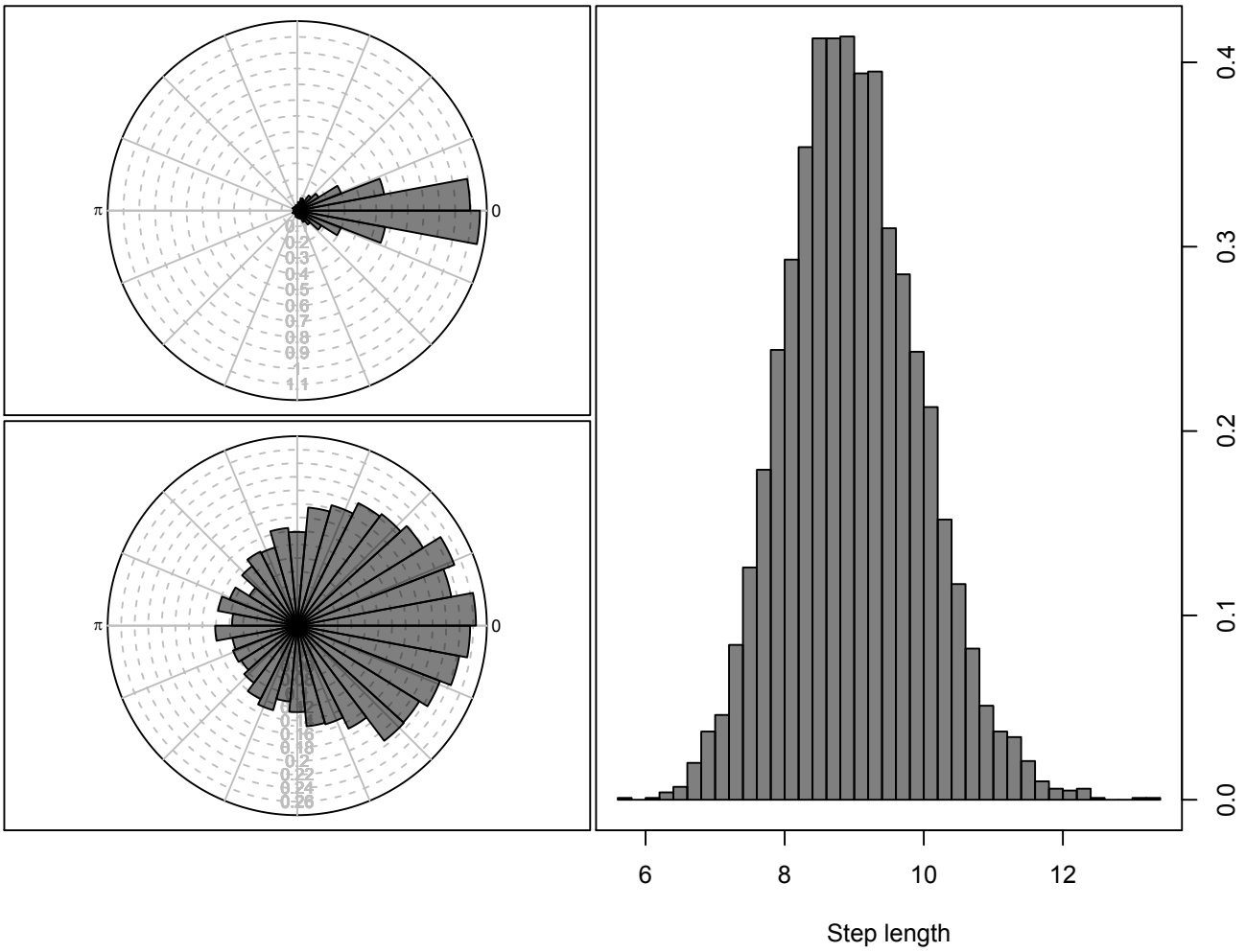


Figure 4.1: Rose plot of wrapped Cauchy distributions and histogram of step lengths.

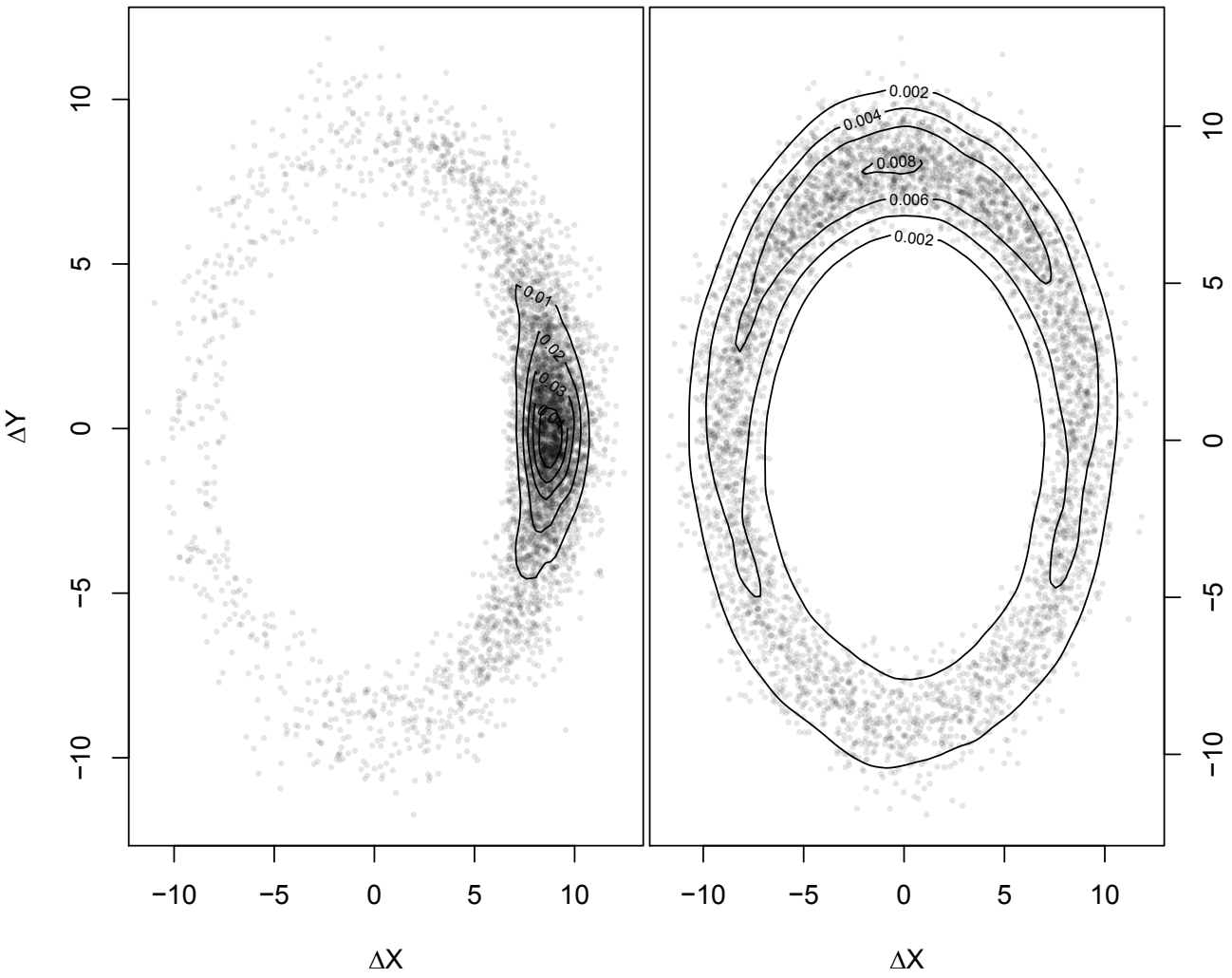


Figure 4.2: Resulting one-step displacement distributions.

distribution is concentrated, the majority of the probability mass is centred around one point. In this case, the displacement distribution can reasonably be approximated by the Laplace approximation. On the contrary, when the turning angle distribution has a larger variance, a large part of the displacement distribution wraps around $(0, 0)$. This can be a problem for the Laplace approximation; both for finding the mode, and for the accuracy of the Gaussian approximation.

4.2.2 Discrete time models

An alternative to the step lengths and direction parameterisation is to model locations or velocities directly. The first-Difference Correlated Random Walk (DCRW; Jonsen et al., 2005) was introduced as a random walk on the differences in locations with a rotational transition matrix to mimic the step length-direction formulations:

$$X_i = X_{i-1} + \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} (X_{i-1} - X_{i-2})\,\gamma + \epsilon_i$$

With this formulation, the location process, X_i, is a second order Markov process with Gaussian increments. As a limiting case for $\gamma \to 0$, the DCRW includes the random walk model without drift, $c_i = 0$. The model is frequently used for analysing movement data.

4.2.3 Time irregular models

Recently, models extending the DCRW to data observed at irregular time intervals have been proposed. Auger-Méthé et al. (2017) proposed a modified DCRW without rotation where the autocorrelation movement displacements are scaled by the time intervals,

$$X_{t_i} = X_{t_{i-1}} + \frac{\Delta_i}{\Delta_{i-1}} \left(X_{t_{i-1}} - X_{t_{i-2}} \right) \gamma_i + \epsilon_i.$$

In this model,

$$\epsilon_i \sim \mathcal{N}\left(0, \Delta_i^2 \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right)$$

By examining the continuous time velocity model

$$\begin{aligned} dV_t &= -\begin{pmatrix} -\log\gamma_1 & \theta \\ -\theta & -\log\gamma_2 \end{pmatrix} (V_t - \mu)dt + SdB_t \\ &= -\Theta(V_t - \mu)dt + SdB_t \end{aligned}$$

where the resulting locations are given by

$$dX_t = V_t dt,$$

further steps can be taken in the direction of a time irregular DCRW model. By solving the velocity model analytically and discretising the location model, Paper IV (Albertsen, 2017a) proposes the Generalized first-Difference Correlated Random Walk (GDCWR):

$$X_{t_i} = X_{t_{i-1}} + \Delta_i \exp(-\Theta\Delta_{i-1})(X_{t_{i-1}} - X_{t_{i-2}})/\Delta_{i-1} + \Delta_i \left(I - \exp(-\Theta\Delta_{i-1})\right) \mu + \Delta_i \epsilon_{t_i}$$

The error terms, ϵ_{t_i}, follow a zero-mean normal distribution with variance

$$Var(\epsilon_{t_i}) = C - \exp(-\Theta\Delta_i) C \exp(-\Theta^T \Delta_i)$$

where

$$\text{vec}(C) = (\Theta \oplus \Theta)^{-1} \text{vec}(\Sigma),$$

and $\Sigma = SS^T$ is a covariance matrix. The GDCRW generalizes the DCRW in several ways. As the modified DCRW, the model allows irregularly observed data. Further, the model allows drift in the movement and different autocorrelation parameters in the two coordinates. Finally, Paper IV (Albertsen, 2017a) provides time-scale invariant parameters, and illustrates how the parameters in the DCRW can be transformed to be time-scale invariant. By decreasing the time steps, Δ_i, the GDCRW can be interpreted as a continuous time model.

4.2.4 Continuous time models

Both the discrete time models with equidistant time steps and irregular time steps can be constructed from a continuous time model. The true animal behaviour most likely occurs in continuous time, making continuous time models more realistic. However, they are also more difficult to work with. Continuous time models can be specified through stochastic differential equations (SDE). SDEs describe the instantaneous change in the process; however, an analytical solution describing the actual process of interest can only be obtained in few specific cases.

For instance, a continuous time random walk with drift can be described by

$$dX_t = cdt + \sigma dB_t.$$

The transition densities of the solution to the SDE are

$$X_t \mid X_s \sim \mathcal{N}\left(X_s + c \cdot (t-s), \sigma^2(t-s)\right), \quad t > s$$

Another model often used for animal movement is the Ornstein-Uhlenbeck process (e.g., Dunn and Gipson, 1977; Blackwell, 1997, 2003; Pedersen and Weng, 2013; Breed et al., 2016; Niu et al., 2016). In a simplified form, the model is described by

$$dX_t = -\beta\left(X_t - \mu\right)dt + \sigma dB_t$$

with transition densities

$$X_t \mid X_s \sim \mathcal{N}\left(\mu + \exp(\beta(t-s)) \cdot (X_s - \mu), \frac{\sigma^2}{2\beta} \cdot (1 - \exp(-2\beta(t-s)))\, I_{2\times 2}\right)$$

where $I_{2\times 2}$ is a two-by-two identity matrix, β is a scalar autocorrelation parameter, and σ is a scalar variance parameter. A solution can also be found in the case where the autocorrelation and variance are modelled by matrices instead (see e.g., Paper IV (Albertsen, 2017a) or several of the references above). Unlike the random walk, the Ornstein-Uhlenbeck process is stationary. The process reverts around the mean, μ. An animal modelled by this process is assumed to stay close to the mean locations, which may be the case for territorial and homing animals. As a result, the Ornstein-Uhlenbeck process can be used to estimate a home range for an animal. A clear disadvantage of the Ornstein-Uhlenbeck process for animal movement is that the attraction towards the mean increases linearly with the distance. That is, the further an animal is from home, the faster it will move.

For animals that does not have a strong attraction to a small area, the Ornstein-Uhlenbeck process can be used to model the velocity instead (Johnson et al., 2008).

$$dV_t = -\beta\left(V_t - \mu\right)dt + \sigma_X dB_t$$

Locations are obtained as the integrated velocity process.

$$dX_t = V_t dt$$

This model was introduced as a continuous time version of the DCRW; however, the DCRW and the CTCRW are only the same in the specific case where the turning angle of the DCRW is zero, and the drift of the CTCRW is zero. Rather, the continuous time model used to construct the GDCRW encompasses both the DCRW and the CTCRW (Paper IV, Albertsen, 2017a).

A common difficulty for all continuous-time models, is the ability to obtain the transition densities $p(X_{t_i} \mid X_{t_{i-1}})$. The transition densities can only be obtained analytically for few models, such as the Ornstein-Uhlenbeck and integrated Ornstein-Uhlenbeck models. In general, the transition densities cannot be described in closed form. Instead, the process can be discretised to obtain approximate solutions. As for the GDCRW, the Euler-Maruyama discretization can be used to approximate the process by Gaussian increments in discrete time (e.g. Pedersen, 1995; Elerian et al., 2001; Cleur, 2001; Shoji, 1997; Preisler et al., 2004; Prathumwan et al., 2012; Paper V Albertsen, 2017b; Russell et al., 2017). The approximation can be improved by adding auxiliary time steps, such that the discretization steps are shorter (Pedersen, 1995). Integrating the auxiliary process observations out of the likelihood, the approximate likelihood of the actual observations is obtained. In a state-space context, the auxiliary time points are added for an

unobserved process which is already integrated out of the likelihood. If the Euler-Maruyama discretization is not accurate enough, higher order approximations can be used (e.g., Hanks et al., 2017). Alternatively, the change of the transition density in time can be described by the Fokker-Planck partial differential equation, which can be solved numerically (e.g. Hurn and Lindsay, 1999; Lux, 2012; Albertsen et al., 2014). Finally, methods not directly approximating the likelihood function can be used, such as the method of moments or estimating functions.

4.2.5 Functional models

Recently, functional data analysis (e.g., Ramsay and Silverman, 2005) has been used in the context of animal movement. Instead of modelling the movement through a time-series of steps, a functional model considers the entire trajectory as an outcome from a multivariate normal where the mean is a smooth curve determined by a weighted sum of basis functions, for instance a B-spline (Buderman et al., 2016) but other basis functions can be used (Hooten and Johnson, 2017). While functional data methods provide smooth continuous time trajectories, the models can easily become detached from the time series structure of the actual process. Further, when model parameters are spline weights they are not directly interpretable in a movement context. However, in an attempt to bridge this gap, Hooten and Johnson (2017) provides a class of velocity based functional movement models containing the CTCRW.

4.2.6 Models including available resources

All the models above solely utilize observed locations in the analysis. However, in many cases animal movement and behaviour may be influenced by the environment and resources available.

To accommodate this, step selection models incorporate a regression on available resources in the movement model. Two approaches exist to combine movement and resource selection. In the first approach (e.g., Rhodes et al., 2005; Forester et al., 2009; Potts et al., 2014; Avgar et al., 2016), the movement density is weighted by a resource (or habitat) selection function such that the combined transition density of moving from X_t to X_{t+1} is

$$f(X_{t+1} \mid X_t, Z) \propto \phi(X_{t+1} \mid X_t; \theta)\omega(Z(X_{t+1}); \beta)$$

where $Z(X_{t+1})$ are covariates at location X_{t+1}, θ and β are parameter vectors, ϕ is a resource independent movement density and ω is, for instance, a log-linear function

$$\omega(Z(X_{t+1}); \beta) = \exp\left(Z(X_{t+1})\beta\right).$$

In the second approach (e.g., Fortin et al., 2005; Duchesne et al., 2015; Nicosia et al., 2017), the resource independent movement model, ϕ, is omitted. Instead, movement quantities such as the turning angle and step length can be included in the log-linear function, ω. In both cases, the likelihood function can be calculated by simulating J control points for each move. The probability of an observed move, conditional on the control points $X^*_{1:J}$, can be calculated by

$$P(X_{t+1} \mid X_t, Z, X^*_{1:J}) = \frac{f(X_{t+1} \mid X_t, Z)}{\sum_{i=1}^{J} f(X^*_i \mid X_t, Z)}$$

where the normalization constants cancel out. If the control points are not sampled uniformly, the sampling must be accounted for.

In a maximum likelihood framework, step selection models cannot easily be used for data with measurement error, since incorporating measurement error would require integrating $P(X_{t+1} \mid X_t, Z, X^*_{1:J})$ over X_t and X_{t+1}, remembering that the control points depend on X_t.

4.2.6.1 Ornstein-Uhlenbeck with spatial constraints

In movement models based on stochastic differential equations with measurement error, environmental and resource influence can be incorporated in the drift (e.g., Preisler et al., 2004). For instance, the case study of Paper IV (Albertsen, 2017a) incorporates a location dependent velocity through a semi-parametric surface (Software II, Albertsen, 2017c). Similarly, environmental covariates such as sea surface temperature or bathymetry could have been incorporated; however, the covariate must be available for any value of the latent state.

As an example, consider the Ornstein-Uhlenbeck process

$$dX_t = -\beta(X_t - \mu)dt + \sigma dB_t.$$

To include spatial constraints, the drift term can be extended to include repulsion from the coast. The new process is

$$dX_t = -\nabla P(X_t) - \beta(X_t - \mu)dt + \sigma dB_t.$$

where $P(X_t) = -\alpha \min(D(X_t), d_0)$ and $D(X_t)$ is the Euclidean distance to the coast. The movement of an animal following this process will be a compromise of attraction towards μ and repulsion from the coast. After a certain distance, d_0, the movement will be a pure Ornstein-Uhlenbeck process. The process is illustrated in Figure 4.3.

4.3 Using the Laplace approximation for animal movement data

The Laplace approximation can readily be used for the discrete, irregular, and continuous time location and velocity-based models above. Paper III (Albertsen et al., 2015) implements the CTCRW with multivariate t-distributed measurement errors using the Laplace approximation. The model is compared to the CTCRW with Gaussian measurement errors, the CTCRW with Gaussian errors and pre-processed observations, and the DCRW fitted with Bayesian methods. The paper illustrates that the Laplace approximation can adequately be used for movement data. Further, in comparing the DCRW and CTCRW with t-distributed measurement errors, the paper illustrates that more accurate trajectories are reconstructed by having latent states at the same time as the observations.

The latter result is also found by Paper IV (Albertsen, 2017a). In Paper IV (Albertsen, 2017a), the DCRW and GDCRW were compared in a simulation study, simulating from the underlying continuous time model. The parameters were chosen such that the only difference between the DCRW and GDCRW were the time steps. In the DCRW, the time steps were equidistant, while the GDCRW had latent states at the time of the observations. The GDCRW provided more accurate reconstructed trajectories when the measurement error was low, while there was no difference for large measurement errors.

Using the Laplace approximation for movement data allows rapid and accurate estimation and reconstruction of movement trajectories (Paper III Albertsen et al., 2015; Paper IV Albertsen, 2017a; Auger-

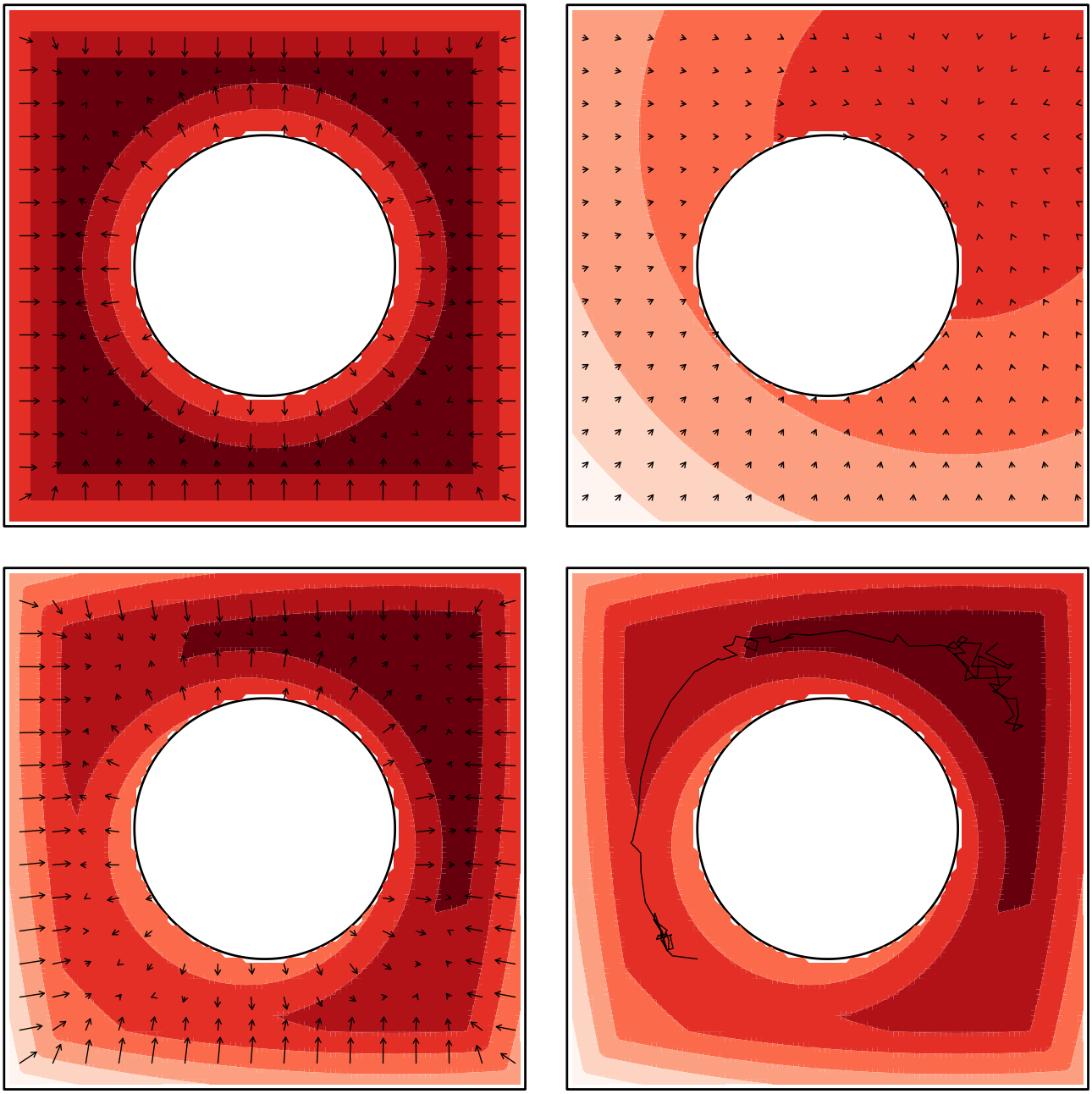


Figure 4.3: Potential functions and their gradients for the Ornstein-Uhlenbeck with spatial constraints. Upper left panel: potential function for the coast repulsive part with gradient at different points (arrows); upper right panel: potential function for the attraction part with gradient at different points (arrows); lower left panel: Combined potential function with gradient at different points (arrows); lower right panel: Combined potential function with a simulated trajectory.

Méthé et al., 2017). The use of the Laplace approximation is, however, limited to models that fit within the framework.

4.4 Incorporating behavioural switching

In the discrete and continuous time models described above, the animal is assumed to behave homogeneously throughout the sampled trajectory. This assumption can be relaxed in two ways; by incorporating time-varying parameters or behavioural switching.

Time-varying parameters can be included in models fitted by the Laplace approximation. Auger-Méthé et al. (2016) allows the autocorrelation parameter of the modified DCRW to vary over time. The parameter is assumed to follow a random walk where the variance parameter is estimated. A similar approach can be used in continuous time models; however, closed form transition densities will not be available.

Including behavioural switching is, as previously discussed (section 2.5), conceptually straight forward, but estimation is difficult in a maximum likelihood framework. In discretised models, the state-space model can be extended to include regime-switching parameters by including a Markov process determining which parameter value to use (e.g., Morales et al., 2004; Jonsen et al., 2005; Whoriskey et al., 2017). Jonsen et al. (2005) presents the DCRW in a regime-switching state-space model in a Bayesian context. In a maximum likelihood framework, Whoriskey et al. (2017) fits the DCRW to data without measurement error, while Paper IV (Albertsen, 2017a) outlines how the switching model can be extended to irregular data. Further research will determine if the method proposed in Paper V (Albertsen, 2017b) can be used to provide reliable estimates for movement models.

In continuous time models, behavioural switching can be included by extending the model with a continuous time Markov chain determining the parameter values. Continuous switching can be estimated by simulation-based methods (e.g., Blackwell, 1997; Blackwell et al., 2016; Parton and Blackwell, 2017). Alternatively, continuous time switching can be estimated by approximate methods. In the GDCRW, continuous time switching can be approximated by including auxiliary latent states with small time steps. The filtering method in Paper V (Albertsen, 2017b) can easily incorporate the missing observations, however, further research is needed to determine how the filtering method will behave in such a situation.

Chapter 5

Conclusion

This PhD study has contributed to the field of fisheries stock assessment modelling and the field of individual animal movement modelling. In both fields, the contributions relate to the use of the Laplace approximation for accurate and efficient maximum likelihood inference in complex state-space models.

With an increasing data collection, the demands for efficient data use and analysis is warranted. Applying the results of movement data analysis in dynamic and near real time management and conservation of marine resources is not feasible if fitting single models of moderate amounts of data require hours or days.

Providing rapid and accurate results is pivotal for management and conservation of large marine animals. As increasing amounts of data are used in near real time for dynamic ocean management, data analysis must be efficient, both in time and data use (e.g., Lewison et al., 2015). This makes state-space models fitted with the Laplace approximation a promising framework for both animal movement and fisheries stock assessment models. Further, the rapid inference using the Laplace approximation makes it feasible to extend current models. For instance, it is within reach to extend fisheries stock assessment models to include a spatial component, and to model raw survey data more directly than the aggregates currently used.

Both fisheries stock assessment and individual animal movement modelling face a common challenge in adequately accounting for the distribution of data given the true process; in terms of the choice of distribution family and in terms of including flexible correlation structures that matches the data. Paper I (Albertsen et al., 2017a) compares two common, but usually completely separated, approaches to modelling fisheries catch data; as numbers-at-age or proportions-at-age. The paper illuminates that both approaches can be suitable for assessment data; however, further studies are needed to investigate when one approach should be preferred over another. Likewise, Paper III (Albertsen et al., 2015) compares the use of the normal distribution and a t-distribution for Argos data. The paper illustrates that using an appropriate measurement distribution improves inference on the latent states. Together, Paper I (Albertsen et al., 2017a) and Paper III (Albertsen et al., 2015) illustrates that the Laplace approximation is applicable for a wide range of measurement distributions, and that the choice of distribution can be important for subsequent management.

Paper II (Albertsen et al., 2017b) builds on Paper I (Albertsen et al., 2017a) by extending the model to multiple stocks. The connection made is simplistic, but improves the individual independent assessments by including correlations in the survival process. A similar approach might be used to account for correlations in fishing mortality between fleets or stocks; for instance, for species where the catch of one influences the catch of another.

Paper IV (Albertsen, 2017a) generalizes the commonly used DCRW movement model. Combined with the `covafillr` R package, the model allows including environmental and spatial effects to better understand the drivers of animal movement. Drivers that needs to be understood for efficient management and protecting of large marine animals. Finally, Paper V (Albertsen, 2017b) introduces a filter and smoother for regime-switching state-space models that allows maximum likelihood inference of a wider range of movement and assessment models while still being operational. Combined, the manuscripts contribute models and methods in marine science that can improve the basis for management and conservation of marine animals. The models are directly applicable in current management frameworks.

Bibliography

Aeberhard, W. H., Flemming, J. M., & Nielsen, A. (2018). Review of state-space models for fisheries science. *Annual Review of Statistics and Its Application*, *5*(1). In Press.

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, *19*(6), 716–723. doi:10.1109/TAC.1974.1100705

Albertsen, C. M. (2017a). A generalized first-difference correlated random walk for animal movement data. *Advance draft*. In Preparation.

Albertsen, C. M. (2017b). An approximate filter and smoother for general regime-switching state-space models. *Advance draft*. In Preparation.

Albertsen, C. M. (2017c). *Covafillr: local polynomial regression of state dependent covariates in state-space models*. R package version 0.4.1. https://CRAN.R-project.org/package=covafillr. Retrieved from https://CRAN.R-project.org/package=covafillr

Albertsen, C. M., Nielsen, A., & Thygesen, U. H. (2017a). Choosing the observational likelihood in state-space stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, *74*(5), 779–789. doi:10.1139/cjfas-2015-0532

Albertsen, C. M., Nielsen, A., & Thygesen, U. H. (2017b). Connecting single-stock assessment models through correlated survival. *ICES Journal of Marine Science*. In Press. doi:10.1093/icesjms/fsx114

Albertsen, C. M., Støttrup, J. G., Nielsen, A., & Christoffersen, M. O. (2014). Depth preference in released juvenile turbot *Psetta maxima*. *Journal of Experimental Marine Biology and Ecology*, *461*(0), 179–184. doi:10.1016/j.jembe.2014.07.013

Albertsen, C. M., Whoriskey, K., Yurkowski, D., Nielsen, A., & Flemming, J. M. (2015). Fast fitting of non-gaussian state-space models to animal movement data via template model builder. *Ecology*, *96*(10), 2598–2604. doi:10.1890/14-2101.1

Anderson-Sprecher, R. & Ledolter, J. (1991). State-space analysis of wildlife telemetry data. *Journal of the American Statistical Association*, *86*(415), 596–602. doi:10.1080/01621459.1991.10475084

Auger-Méthé, M., Albertsen, C. M., Jonsen, I. D., Derocher, A. E., Lidgard, D. C., Studholme, K. R., … Flemming, J. M. (2017). Spatiotemporal modelling of marine movement data using template model builder. *Marine Ecology Progress Series*, *565*, 237–249. doi:10.3354/meps12019

Auger-Méthé, M., Field, C., Albertsen, C. M., Derocher, A., Lewis, M. A., Jonsen, I. D., & Flemming, J. M. (2016). State-space models' dirty little secrets: even simple linear gaussian models can have parameter and state estimation problems. *Scientific Reports*, *6*, 26677. doi:10.1038/srep26677

Avgar, T., Potts, J. R., Lewis, M. A., & Boyce, M. S. (2016). Integrated step selection analysis: bridging the gap between resource selection and animal movement. *Methods in Ecology and Evolution*, *7*(5), 619–630. doi:10.1111/2041-210x.12528

Baggio, M. (2016). Optimal fishery management with regime shifts: an assessment of harvesting strategies. *Environmental and Resource Economics*, *64*(3), 465–492. doi:10.1007/s10640-015-9906-0

Baktoft, H., Gjelland, K. Ø., Økland, F., & Thygesen, U. H. (2017). Positioning of aquatic animals based on time-of-arrival and random walk models using YAPS (yet another positioning solver). *Scientific Reports*, *7*(1). doi:10.1038/s41598-017-14278-z

Barndorff-Nielsen, O. & Cox, D. (1989). *Asymptotic techniques for use in statistics*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Berg, C. W. & Nielsen, A. (2016). Accounting for correlated observations in an age-based state-space stock assessment model. *ICES Journal of Marine Science: Journal du Conseil*, *73*(7), 1788–1797. doi:10.1093/icesjms/fsw046

Blackwell, P. G. (1997). Random diffusion models for animal movement. *Ecological Modelling*, *100*(1-3), 87–102. doi:10.1016/s0304-3800(97)00153-1

Blackwell, P. G. (2003). Bayesian inference for markov processes with diffusion and discrete components. *Biometrika*, *90*(3), 613–627. doi:10.1093/biomet/90.3.613

Blackwell, P. G., Niu, M., Lambert, M. S., & LaPoint, S. D. (2016). Exact bayesian inference for animal movement in continuous time. *Methods in Ecology and Evolution*, *7*(2), 184–195. doi:10.1111/2041-210x.12460

Breed, G. A., Golson, E. A., & Tinker, M. T. (2016). Predicting animal home-range structure and transitions using a multistate ornstein-uhlenbeck biased random walk. *Ecology*, *98*(1), 32–47. doi:10.1002/ecy.1615

Brinch, C. N., Eikeset, A. M., & Stenseth, N. C. (2011). Maximum likelihood estimation in nonlinear structured fisheries models using survey and catch-at-age data. *Canadian Journal of Fisheries and Aquatic Sciences*, *68*(10), 1717–1731. doi:10.1139/f2011-085

Buderman, F. E., Hooten, M. B., Ivan, J. S., & Shenk, T. M. (2016). A functional model for characterizing long-distance movement behaviour. *Methods in Ecology and Evolution*, *7*(3), 264–273. doi:10.1111/2041-210x.12465

Cadigan, N. G. (2013). Fitting a non-parametric stock-recruitment model in r that is useful for deriving MSY reference points and accounting for model uncertainty. *ICES Journal of Marine Science*, *70*(1), 56–67. doi:10.1093/icesjms/fss183

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*(2), 296–308. doi:10.1139/cjfas-2015-0047

Chatfield, C. (1995). Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *158*(3), 419. doi:10.2307/2983440

Cleur, E. M. (2001). Maximum likelihood estimates of a class of one□dimensional stochastic differential equation models from discrete data. *Journal of Time Series Analysis*, *22*(5), 505–515. doi:10.1111/1467-9892.00238

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1–38.

Duchesne, T., Fortin, D., & Rivest, L.-P. (2015). Equivalence between step selection functions and biased correlated random walks for statistical inference on animal movement. *PLOS ONE*, *10*(4), e0122947. doi:10.1371/journal.pone.0122947

Dunn, J. E. & Gipson, P. S. (1977). Analysis of radio telemetry data in studies of home range. *Biometrics*, *33*(1), 85. doi:10.2307/2529305

Elerian, O., Chib, S., & Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, *69*(4), 959–993.

Erkanli, A. (1994). Laplace approximations for posterior expectations when the mode occurs at the boundary of the parameter space. *Journal of the American Statistical Association*, *89*(425), 250–258. doi:10.1080/01621459.1994.10476466

Fay, G. (2015). Including regime shifts in fisheries assessments and harvest recommendations using hidden markov models. ICES CM 2015/A:17.

Forester, J. D., Im, H. K., & Rathouz, P. J. (2009). Accounting for animal movement in estimation of resource selection functions: sampling and data analysis. *Ecology*, *90*(12), 3554–3565. doi:10.1890/08-0874.1

Fortin, D., Beyer, H. L., Boyce, M. S., Smith, D. W., Duchesne, T., & Mao, J. S. (2005). Wolves influence elk movements: behaviour shapes a trophic cascade in yellowstone national park. *Ecology*, *86*(5), 1320–1330. doi:10.1890/04-0953

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., … Sibert, J. (2012). AD model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, *27*(2), 233–249. doi:10.1080/10556788.2011.597854

Freeman, S. & Kirkwood, G. (1995). On a structural time series method for estimating stock biomass and recruitment from catch and effort data. *Fisheries Research*, *22*(1-2), 77–98. doi:10.1016/0165-7836(94)00304-f

Früiiwirth-Schnatter, S. (1996). Recursive residuals and model diagnostics for normal and non-normal state space models. *Environmental and Ecological Statistics*, *3*(4), 291–309. doi:10 . 1007 / bf00539368

Geweke, J. (1989). Bayesian inference in econometric models using monte carlo integration. *Econometrica*, *57*(6), 1317. doi:10.2307/1913710

Gordon, N., Salmond, D., & Smith, A. (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings F Radar and Signal Processing*, *140*(2), 107–113. doi:10.1049/ip-f-2.1993.0015

Griewank, A. (2000). *Evaluating derivatives: principles and techniques of algorithmic differentiation*. Frontiers in Applied Mathematics. Philadelphia, PA, U.S.A: Society for Industrial and Applied Mathematics.

Gudmundsson, G. (1994). Time series analysis of catch-at-age observations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *43*(1), 117–126. doi:10.2307/2986116

Gudmundsson, G. & Gunnlaugsson, T. (2012). Selection and estimation of sequential catch-at-age models. *Canadian Journal of Fisheries and Aquatic Sciences*, *69*(11), 1760–1772. doi:10.1139/f2012-095

Gurarie, E., Andrews, R. D., & Laidre, K. L. (2009). A novel method for identifying behavioural changes in animal movement data. *Ecology Letters*, *12*(5), 395–408. doi:10.1111/j.1461-0248.2009.01293.x

Hanks, E. M., Johnson, D. S., & Hooten, M. B. (2017). Reflected stochastic differential equation models for constrained animal movement. *Journal of Agricultural, Biological and Environmental Statistics*, *22*(3), 353–372. doi:10.1007/s13253-017-0291-8

Hazel, J. (2009). Evaluation of fast-acquisition GPS in stationary tests and fine-scale tracking of green turtles. *Journal of Experimental Marine Biology and Ecology*, *374*(1), 58–68. doi:10.1016/j.jembe.2009.04.009

Hooten, M. B. & Johnson, D. S. (2017). Basis function models for animal movement. *Journal of the American Statistical Association*, *112*(518), 578–589. doi:10.1080/01621459.2016.1246250

Hurn, A. & Lindsay, K. (1999). Estimating the parameters of stochastic differential equations. *Mathematics and Computers in Simulation*, *48*(4-6), 373–384. doi:10.1016/s0378-4754(99)00017-8

Hüssy, K., Mosegaard, H., Albertsen, C. M., Hemmer-Hansen, J., Nielsen, E. E., & Eero, M. (2016). Evaluation of otolith shape as a tool for stock discrimination in marine fishes using baltic sea cod as a case study. *Fisheries Research*, *174*, 210–218. doi:10.1016/j.fishres.2015.10.010

Hyndman, R. J. (1996). Computing and graphing highest density regions. *The American Statistician*, *50*(2), 120–126. doi:10.1080/00031305.1996.10474359

ICES. (2016a). Advice basis. In *Report of the ICES advisory committee* (Chap. 1.2). ICES Advice 2016.

ICES. (2016b). Timeline of ICES advice. In *Technical guidelines - report of the ICES advisory committee* (Chap. 12.4.4). ICES Advice 2016.

ICES. (2017). Technical guidelines - ICES fisheries management reference points for category 1 and 2 stocks. In *Report of the ICES advisory committee* (Chap. 12.4.3.1). ICES Advice 2017. doi:10.17895/ices.pub.3036

Ionides, E. L., Bhadra, A., Atchadé, Y., & King, A. (2011). Iterated filtering. *The Annals of Statistics*, *39*(3), 1776–1802. doi:10.1214/11-aos886

Ionides, E. L., Nguyen, D., Atchadé, Y., Stoev, S., & King, A. A. (2015). Inference for dynamic and latent variable models via iterated, perturbed bayes maps. *Proceedings of the National Academy of Sciences*, *112*(3), 719–724. doi:10.1073/pnas.1410597112

Jacquier, E., Johannes, M., & Polson, N. (2007). MCMC maximum likelihood for latent state models. *Journal of Econometrics*, *137*(2), 615–640. doi:10.1016/j.jeconom.2005.11.017

Johnson, D. S., London, J. M., Lea, M.-A., & Durban, J. W. (2008). Continuous-time correlated random walk model for animal telemetry data. *Ecology*, *89*(5), 1208–1215. doi:10.1890/07-1032.1

Jonsen, I. D., Flemming, J. M., & Myers, R. A. (2005). Robust state–space modeling of animal movement data. *Ecology*, *86*(11), 2874–2880. doi:10.1890/04-1852

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, *82*(1), 35. doi:10.1115/1.3662552

Kampstra, P. (2008). Beanplot: a boxplot alternative for visual comparison of distributions. *Journal of Statistical Software*, *28*(Code Snippet 1). doi:10.18637/jss.v028.c01

Kass, R. E. & Steffey, D. (1989). Approximate bayesian inference in conditionally independent hierarchical models (parametric empirical bayes models). *Journal of the American Statistical Association*, *84*(407), 717–726. doi:10.1080/01621459.1989.10478825

Kass, R. E., Tierney, L., & Kadane, J. B. (1990). The validity of posterior expansions based on laplace's method. In S. Geisser, J. Hodges, S. Press, & A. Zellner (Eds.), *Bayesian and likelihood methods in statistics and econometrics: essays in honor of george a. barnard* (pp. 473–488). Amsterdam, North Holland: Elsevier Science Publishers B.V.

Kim, C.-J. (1994). Dynamic linear models with markov-switching. *Journal of Econometrics*, *60*(1-2), 1–22. doi:10.1016/0304-4076(94)90036-1

Kitagawa, G. (1987). Non-gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association*, *82*(400), 1032. doi:10.2307/2289375

Kleppe, T. S. & Skaug, H. J. (2012). Fitting general stochastic volatility models using laplace accelerated sequential importance sampling. *Computational Statistics & Data Analysis*, *56*(11), 3105–3119. doi:10.1016/j.csda.2011.05.007

Koyama, S., Pérez-Bolde, L. C., Shalizi, C. R., & Kass, R. E. (2010). Approximate methods for state-space models. *Journal of the American Statistical Association*, *105*(489), 170–180. doi:10.1198/jasa.2009.tm08326

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: automatic differentiation and laplace approximation. *Journal of Statistical Software*, *70*(5). doi:10.18637/jss.v070.i05

Lam, C. H., Nielsen, A., & Sibert, J. R. (2008). Improving light and temperature based geolocation by unscented kalman filtering. *Fisheries Research*, *91*(1), 15–25. doi:10.1016/j.fishres.2007.11.002

Lewison, R., Hobday, A. J., Maxwell, S., Hazen, E., Hartog, J. R., Dunn, D. C., … Crowder, L. B. (2015). Dynamic ocean management: identifying the critical ingredients of dynamic approaches to ocean resource management. *BioScience*, *65*(5), 486–498. doi:10.1093/biosci/biv018

Lux, T. (2012). Inference for systems of stochastic differential equations from discretely sampled data: a numerical maximum likelihood approach. *Annals of Finance*, *9*(2), 217–248. doi:10.1007/s10436-012-0219-9

Maunder, M. N. & Punt, A. E. (2013). A review of integrated analysis in fisheries stock assessment. *Fisheries Research*, *142*, 61–74. doi:10.1016/j.fishres.2012.07.025

McClintock, B. T., Johnson, D. S., Hooten, M. B., Ver Hoef, J. M., & Morales, J. M. (2014). When to be discrete: the importance of time formulation in understanding animal movement. *Movement Ecology*, *2*(1), 21. doi:10.1186/s40462-014-0021-6

McClintock, B. T., King, R., Thomas, L., Matthiopoulos, J., McConnell, B. J., & Morales, J. M. (2012). A general discrete-time modeling framework for animal movement using multistate random walks. *Ecological Monographs*, *82*(3), 335–349.

Michelot, T., Langrock, R., & Patterson, T. A. (2016). moveHMM: an r package for the statistical modelling of animal movement data using hidden markov models. *Methods in Ecology and Evolution*, *7*(11), 1308–1315. doi:10.1111/2041-210x.12578

Millar, R. B. & Meyer, R. (2000). Non-linear state space modelling of fisheries biomass dynamics by using metropolis-hastings within-gibbs sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *49*(3), 327–342.

Montenegro, C. & Branco, M. (2016). Bayesian state-space approach to biomass dynamic models with skewed and heavy-tailed error distributions. *Fisheries Research*, *181*, 48–62. doi:10.1016/j.fishres.2016.03.021

Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., & Fryxell, J. M. (2004). Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology*, *85*(9), 2436–2445. doi:10.1890/03-0269

Needle, C. L. (2001). Recruitment models: diagnosis and prognosis. *Reviews in Fish Biology and Fisheries*, *11*(2), 95–111. doi:10.1023/A:1015208017674

Nicosia, A., Duchesne, T., Rivest, L.-P., & Fortin, D. (2017). A multi-state conditional logistic regression model for the analysis of animal movement. *The Annals of Applied Statistics*, *11*(3), 1537–1560. doi:10.1214/17-aoas1045. arXiv: 1611.02690v1 [stat.ME]

Nielsen, A. & Berg, C. W. (2014). Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research*, *158*, 96–101. doi:10.1016/j.fishres.2014.01.014

Nielsen, A., Bigelow, K. A., Musyl, M. K., & Sibert, J. R. (2006). Improving light-based geolocation by including sea surface temperature. *Fisheries Oceanography*, *15*(4), 314–325. doi:10.1111/j.1365-2419.2005.00401.x

Nielsen, A. & Sibert, J. R. (2007). State–space model for light-based tracking of marine animals. *Canadian Journal of Fisheries and Aquatic Sciences*, *64*(8), 1055–1068. doi:10.1139/f07-064

Nielsen, A., Sibert, J. R., Kohin, S., & Musyl, M. K. (2009). State space model for light based tracking of marine animals: validation on swimming and diving creatures. In J. L. Nielsen, H. Arrizabalaga, N. Fragoso, A. Hobday, M. Lutcavage, & J. Sibert (Eds.), *Tagging and tracking of marine animals with electronic devices* (pp. 295–309). Dordrecht: Springer Netherlands. doi:10.1007/978-1-4020-9640-2_18

Niu, M., Blackwell, P. G., & Skarin, A. (2016). Modeling interdependent animal movement in continuous time. *Biometrics*, *72*(2), 315–324. doi:10.1111/biom.12454

Ogden, H. E. (2017). On asymptotic validity of naive inference with an approximate likelihood. *Biometrika*, *104*(1), 153–164. doi:10.1093/biomet/asx002

Parton, A. & Blackwell, P. G. (2017). Bayesian inference for multistate 'step and turn' animal movement in continuous time. *Journal of Agricultural, Biological and Environmental Statistics*, *22*(3), 373–392. doi:10.1007/s13253-017-0286-5

Pedersen, A. R. (1995). Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli*, *1*(3), 257. doi:10.2307/3318480

Pedersen, M. W., Berg, C. W., Thygesen, U. H., Nielsen, A., & Madsen, H. (2011). Estimation methods for nonlinear state-space models in ecology. *Ecological Modelling*, *222*(8), 1394–1400. doi:10.1016/j.ecolmodel.2011.01.007

Pedersen, M. W. & Berg, C. W. (2017). A stochastic surplus production model in continuous time. *Fish and Fisheries*, *18*(2), 226–243. doi:10.1111/faf.12174

Pedersen, M. W. & Weng, K. C. (2013). Estimating individual animal movement from observation networks. *Methods in Ecology and Evolution*, *4*(10), 920–929. doi:10.1111/2041-210X.12086

Pinheiro, J. C. & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, *4*(1), 12–35. doi:10.1080/10618600.1995.10474663

Potts, J. R., Bastille-Rousseau, G., Murray, D. L., Schaefer, J. A., & Lewis, M. A. (2014). Predicting local and non-local effects of resources on animal space use using a mechanistic step selection model. *Methods in Ecology and Evolution*, *5*(3), 253–262. doi:10.1111/2041-210x.12150

Prathumwan, D., Lenbury, Y., Satiracoo, P., & Rattanakul, C. (2012). Euler-maruyama approximation and maximum likelihood estimator for a stochastic differential equation model of the signal transduction process. *International Journal of Mathematical Models and Methods in Applied Sciences*, *6*(2), 323–331.

Preisler, H. K., Ager, A. A., Johnson, B. K., & Kie, J. G. (2004). Modeling animal movements using stochastic differential equations. *Environmetrics*, *15*(7), 643–657. doi:10.1002/env.636

Quinn, T. J. (2003). Ruminations on the development and future of population dynamics models in fisheries. *Natural Resource Modeling*, *16*(4), 341–392. doi:10.1111/j.1939-7445.2003.tb00119.x

Ramsay, J. & Silverman, B. (2005). *Functional data analysis*. Springer Series in Statistics. New York, NY, U.S.A: Springer.

Rhodes, J. R., McAlpine, C. A., Lunney, D., & Possingham, H. P. (2005). A spatially explicit habitat selection model incorporating home range behaviour. *Ecology*, *86*(5), 1199–1205. doi:10.1890/04-0912

Rue, H., Martino, S., & Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319–392. doi:10.1111/j.1467-9868.2008.00700.x

Russell, J. C., Hanks, E. M., Haran, M., & Hughes, D. P. (2017). A spatially-varying stochastic differential equation model for animal movement. Preprint. arXiv: 1603.07630v2 [stat.AP]

Rutz, C. & Hays, G. C. (2009). New frontiers in biologging science. *Biology Letters*, *5*(3), 289–292. doi:10.1098/rsbl.2009.0089

Schaefer, M. (1954). Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. *Bulletin of the Inter-American Tropical Tuna Commission*, *1*(2), 27–56. Reprinted in Bulletin of Mathematical Biology 53(1/2) 253-279 1991. doi:10.1016/s0092-8240(05)80049-7

Schnute, J. T. (1994). A general framework for developing sequential fisheries models. *Canadian Journal of Fisheries and Aquatic Sciences*, *51*(8), 1676–1688. doi:10.1139/f94-168

Schnute, J. T. & Richards, L. J. (1995). The influence of error on population estimates from catch-age models. *Canadian Journal of Fisheries and Aquatic Sciences*, *52*(10), 2063–2077. doi:10.1139/f95-800

Shimada, J. & Tsukuda, Y. (2005). Estimation of stochastic volatility models: an approximation to the nonlinear state space representation. *Communications in Statistics - Simulation and Computation*, *34*(2), 429–450. doi:10.1081/sac-200055729

Shoji, I. (1997). A note on asymptotic properties of the estimator derived from the euler method for diffusion processes at discrete times. *Statistics & Probability Letters*, *36*(2), 153–159. doi:10.1016/s0167-7152(97)00058-8

Sibert, J. R., Musyl, M. K., & Brill, R. W. (2003). Horizontal movements of bigeye tuna (thunnus obesus) near hawaii determined by kalman filter analysis of archival tagging data. *Fisheries Oceanography*, *12*(3), 141–151. doi:10.1046/j.1365-2419.2003.00228.x

Skaug, H. J. & Fournier, D. A. (2006). Automatic approximation of the marginal likelihood in non-gaussian hierarchical models. *Computational Statistics & Data Analysis*, *51*(2), 699–709. doi:10.1016/j.csda.2006.03.005

Skaug, H. J. & Yu, J. (2014). A flexible and automated likelihood based framework for inference in stochastic volatility models. *Computational Statistics & Data Analysis*, *76*, 642–654. doi:10.1016/j.csda.2013.10.005

Smith, J. Q. (1985). Diagnostic checks of non-standard time series models. *Journal of Forecasting*, *4*(3), 283–291. doi:10.1002/for.3980040305

Thorson, J. T. & Kristensen, K. (2016). Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. *Fisheries Research*, *175*, 66–74. doi:10.1016/j.fishres.2015.11.016

Thygesen, U. H., Albertsen, C. M., Berg, C. W., Kristensen, K., & Nielsen, A. (2017). Validation of ecological state space models using the laplace approximation. *Environmental and Ecological Statistics*, *24*(2), 317–339. doi:10.1007/s10651-017-0372-4

Thygesen, U. H. & Nielsen, A. (2009). Lessons from a prototype geolocation problem. In J. L. Nielsen, H. Arrizabalaga, N. Fragoso, A. Hobday, M. Lutcavage, & J. Sibert (Eds.), *Tagging and tracking of marine animals with electronic devices* (pp. 257–276). Dordrecht: Springer Netherlands. doi:10.1007/978-1-4020-9640-2_16

Tierney, L., Kass, R. E., & Kadane, J. B. (1989a). Approximate marginal densities of nonlinear functions. *Biometrika*, *76*(3), 425–433. doi:10.1093/biomet/76.3.425

Tierney, L., Kass, R. E., & Kadane, J. B. (1989b). Fully exponential laplace approximations to expectations and variances of nonpositive functions. *Journal of the American Statistical Association*, *84*(407), 710–716. doi:10.1080/01621459.1989.10478824

Tracey, J. A., Zhu, J., & Crooks, K. R. (2010). Modeling and inference of animal movement using artificial neural networks. *Environmental and Ecological Statistics*, *18*(3), 393–410. doi:10.1007/s10651-010-0138-8

Turchin, P. (1998). *Quantitative analysis of movement: measuring and modeling population redistribution in animals and plants*. Sunderland, MA, U.S.A: Sinauer.

Van Beveren, E., Duplisea, D., Castonguay, M., Doniol-Valcroze, T., Plourde, S., & Cadigan, N. (2017). How catch underreporting can bias stock assessment of and advice for northwest atlantic mackerel and a possible resolution using censored catch. *Fisheries Research*, *194*, 146–154. doi:10.1016/j.fishres.2017.05.015

Whoriskey, K., Auger-Méthé, M., Albertsen, C. M., Whoriskey, F. G., Binder, T. R., Krueger, C. C., & Flemming, J. M. (2017). A hidden markov movement model for rapidly identifying behavioral states from animal tracks. *Ecology and Evolution*, *7*(7), 2112–2121. doi:10.1002/ece3.2795

Witt, M., Åkesson, S., Broderick, A., Coyne, M., Ellick, J., Formia, A., … Godley, B. (2010). Assessing accuracy and utility of satellite-tracking data using argos-linked fastloc-GPS. *Animal Behaviour*, *80*(3), 571–581. doi:10.1016/j.anbehav.2010.05.022

Paper I

Choosing the observational likelihood in state-space stock assessment models

Albertsen, C. M.*, Nielsen, A., Thygesen, U. H.

National Institute of Aquatic Resources, Technical University of Denmark, Charlottenlund, Denmark

* **Corresponding author. Email: cmoe@aqua.dtu.dk**

Post-print. Published in Canadian Journal of Fisheries and Aquatic Sciences

Full citation:

Albertsen, C. M., Nielsen, A., and Thygesen, U. H. (2017). Choosing the observational likelihood in state-space stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, *74*(5), 779–789. doi:10.1139/cjfas-2015-0532

Abstract

Data used in stock assessment models result from combinations of biological, ecological, fishery, and sampling processes. Since different types of errors propagate through these processes it can be difficult to identify a particular family of distributions for modelling errors on observations a priori. By implementing several observational likelihoods, modelling both numbers- and proportions-at-age, in an age based state-space stock assessment model, we compare the model fit for each choice of likelihood along with the implications for spawning stock biomass and average fishing mortality. We propose using AIC intervals based on fitting the full observational model for comparing different observational likelihoods. Using data from four stocks, we show that the model fit is improved by modelling the correlation of observations within years. However, the best choice of observational likelihood differs for different stocks, and the choice is important for the short-term conclusions drawn from the assessment model; in particular, the choice can influence total allowable catch advise based on reference points.

Keywords:

Numbers-at-age, proportions-at-age, data weighting, state-space model, stock assessment model

I.1 Introduction

Stock assessment models often use aggregated and uncertain data such as surveys and landings-at-age which rely on age classification of effectively few individuals (Aanes and Pennington, 2003). Commercial fishing and scientific surveys sample from populations that vary according to, for example, season, sex, age and region. From this catch, samples are weighed and measured to estimate the length distribution, weight-at-length, and total catch in numbers. Additional sub-samples are taken to age classify individuals for estimating proportions-at-age; either directly or through an age-length key. The samples consist of many individuals from few hauls (Aanes and Pennington, 2003), which may lead to underestimated uncertainties of estimates if ignored. Finally, all this information is aggregated to numbers-at-age for each year. This aggregation may be via models including e.g. spatial location, season, gear and length effects. Even though the stock population growth process at this level of aggregation is well described (Each year the fish age by one year, some die of natural causes, and others die from fishing) aggregating the different sources of uncertainty makes it difficult to find the optimal (or true) distributions of the observations a priori.

Age-based stock assessment models can be divided into two classes depending on the way they utilize the data. Either the data can be modelled as numbers-at-age or as proportions-at-age along with total weight or numbers. Most currently used age-based stock assessment models exclusively consider either numbers- or proportions-at-age and only one or few observational likelihoods (ICES, 2010b). When modelling numbers-at-age, the normal distribution, parameterized to avoid too much probability on negative observations, has been used (Gudmundsson, 1994; Fryer, 2002) along with the log-normal distribution (Cook, 2013; Nielsen and Berg, 2014) and its multivariate extension (Myers and Cadigan, 1995). Although recommended over the log-normal by Cadigan and Myers (2001), the gamma distribution is infrequently used to model numbers-at-age in assessment models (ICES, 2010b).

The multinomial distribution has been popular when modelling proportions-at-age (Fournier and

Archibald, 1982; Methot Jr. and Wetzel, 2013; Williams and Shertzer, 2015). Based on the age classification sampling, it is an intuitive choice; however, when using the true number of data generating samples, the variances of the modelled proportions are often too small, and the correlation structure too restrictive (Crone and Sampson, 1998; Aanes and Pennington, 2003; Francis, 2014). Efforts have been made to increase the variance by estimating an effective sample size (McAllister and Ianelli, 1997; Francis, 2011; Hulson et al., 2011, 2012). Nonetheless, the effective sample size must be estimated by iterative optimization (McAllister and Ianelli, 1997; Francis, 2011; Maunder, 2011) since the multinomial distribution is improper when used for continuous data (Francis, 2014). Hence, the multinomial distribution will not be considered here. To avoid iterative estimation of the effective sample size, it has been suggested to replace the multinomial with the Dirichlet distribution (Williams and Quinn, 1998; Francis, 2014) in which the variance is only determined by parameters.

While the Dirichlet distribution is an improvement over the multinomial distribution, they both have a very restrictive variance-covariance structure that only allows negative correlations, which may not be appropriate (Francis, 2014). Therefore distributions based on transformations of multivariate normals, such as the additive logistic normal (Francis, 2014) and the multiplicative logistic normal (Cadigan, 2015), have recently been proposed for proportions-at-age in stock assessment models.

Although several authors have compared different proportions-at-age models (Maunder, 2011; Francis, 2014), not much effort has been given to compare different observational likelihoods for numbers-at-age data (Cadigan and Myers, 2001), and even less has been given to compare between the proportions- and numbers-at-age. Using the R-package Template Model Builder (Kristensen et al., 2016), we implement 13 observational likelihoods, including both numbers- and proportions-at-age models, in an age-based state-space stock assessment model. Using assessment data from four European stocks, we compare the model fit for each choice of likelihood along with the implications for key outputs such as spawning stock biomass (SSB) and average fishing mortality ($\bar{F}$).

I.2 Methods

We implemented age-based state-space stock assessment models (Nielsen and Berg, 2014) with 13 different observational likelihoods (Table I.1) for four different European stocks. For simplicity the same observational likelihood was used for both commercial catch data and survey indices. While the process model was kept unchanged for each stock, we compared the goodness-of-fit of the observational likelihoods by AIC. We considered models for numbers-at-age and proportions-at-age combined with total catch. We considered seven different distributions for numbers-at-age. When using data in the form of a total and proportions-at-age we followed the wide-spread convention of modelling total catch as univariate log-normal, but considered two alternatives where the total was either in numbers or biomass. These two alternatives for total catch were crossed with three alternative distributions for the proportions. The observational likelihoods implemented cover frequently used distributions in fisheries stock assessments and close extensions.

I.2.1 Process model

The processes described in the state-space model involved the true unobserved numbers-at-age in the stock, and the true unobserved fishing mortality (See Nielsen and Berg, 2014, or Appendix I.A for details). Following Nielsen and Berg (2014), the logarithm of the fishing mortality was assumed to follow

Table I.1: Overview of the observational models used in the case studies and some properties: if zero observations are allowed; whether the Baranov catch equation determines the mean, median or location; the number of estimated observational parameters per age (a) and fleet (f); and whether a correlation parameter is estimated. The models are divided in to model classes: Univariate numbers-at-age (UN@A), multivariate numbers-at-age (MN@A), proportions-at-age with log-normal total numbers (P@AwN), and proportions-at-age with log-normal total weight (P@AwW).

Model	Distribution	Class	Allows 0	Baranov	Est. par.s	Est. cor.
M_1	log-Normal	UN@A	No	Median	1 a $f^\dagger$	No
M_2	Gamma	UN@A	Some	Mean	1 a f	No
M_3	Generalized Gamma	UN@A	Some	Location	2 a f	No
M_4	Normal	UN@A	Yes	Mean	1 a f	No
M_5	Left Truncated Normal	UN@A	Yes	Location	1 a f	No
M_6	log-Student's t	UN@A	No	Location	2 a f	No
M_7	Multivariate log-Normal	MN@A	No	Median	1 a f+1 $f^\ddagger$	Yes
M_8	Additive Logistic Normal	P@AwN	No	Location	1 a f+1 f	Yes
M_9	Multiplicative Logistic Normal	P@AwN	No	Location	1 a f + 1 f	Yes
M_{10}	Dirichlet	P@AwN	No	Mean	1 f	No
M_{11}	Additive Logisitc Normal	P@AwW	No	Location	1 a f+1 f	Yes
M_{12}	Multiplicative Logistic Normal	P@AwW	No	Location	1 a f + 1 f	Yes
M_{13}	Dirichlet	P@AwW	No	Mean	1 f	No

† Should be read: One per age per fleet.

‡ Should be read: One per age per fleet and one additional per fleet.

a multivariate Gaussian random walk, where the correlation had an AR(1) structure (model D in Nielsen and Berg, 2014). The true population numbers-at-age were assumed to follow an exponential decay model where the natural mortality is known. The model included recruitment to the first age group. The error-terms for the true numbers-at-age were assumed to follow a log-normal distribution without correlation. All variance, correlation and stock-recruitment parameters were estimated. Quantities such as weight-at-age and maturity were assumed to be known. The process model was related to the observations through the Baranov catch equation for catch data, and through an assumption of proportionality to abundance-at-age for surveys. The proportionality constants were estimated. We denoted the calculated catch (or survey index) by $\tilde{C}_{a,f,y}$.

I.2.2 Observational models

Our model M_1 was the log-normal distribution with its usual parameterization. The median was determined by $\tilde{C}_{a,f,y}$, while a scale parameter was estimated for each age and fleet. The model M_2 was the gamma distribution parameterized to have constant coefficient of variation (Cadigan and Myers, 2001). The mean was determined by $\tilde{C}_{a,f,y}$, while a coefficient of variation (CV) was estimated for each age and fleet. The generalized gamma distribution was included as model M_3 with the parameterization of Prentice (1974). This parameterization was preferred over the Stacy (1962) parameterization as it both extends it, and is numerically more stable when reducing to the log-normal distribution (Prentice, 1974; Farewell and Prentice, 1977). The log-location parameter was determined by $\log(\tilde{C}_{a,f,y})$ while a shape and scale parameter was estimated for each age and fleet. The models M_4 and M_5 were the normal and truncated normal (with left truncation at zero). Both were parameterized based on the mean determined by $\tilde{C}_{a,f,y}$ and separate CV parameters for each age and fleet (which applied to the un-truncated values

for the truncated normal). The Student's t-distribution on log-scale was our model M_6. The distribution was parameterized with a log-location parameter determined by $\log(\tilde{C}_{a,f,y})$ along with log-scale and log-degrees-of-freedom parameters estimated separately for each age and fleet.

Model M_7 was the multivariate log-normal with its usual parameterization. The marginal medians were determined by $\tilde{C}_{a,f,y}$, while a one-parameter AR(1) structure was used for the correlation between ages on logarithmic scale (Pinheiro and Bates, 2000; Francis, 2014). Separate correlation parameters were estimated for each fleet along with scale parameters estimated for each age and fleet. Our models M_8 and M_9 were the additive logistic normal and the multiplicative logistic normal (Aitchison, 2003) with log-normal total numbers. For both models, the location parameters were determined by the $\tilde{C}_{a,f,y}$s while the scale matrices were parameterized as M_7. For the log-normal distributions, the medians were determined by $\sum_a \tilde{C}_{a,f,y}$, while a separate scale parameter was estimated for each fleet. The Dirichlet distribution with log-normal total numbers was our model M_{10}. The Dirichlet distribution was parameterized with concentration parameters proportional to $(\tilde{C}_{1,f,y}, \ldots, \tilde{C}_{A,f,y})^T$. A proportionality parameter was estimated for each fleet. The log-normal distributions for total numbers were parameterized as M_8 and M_9. Finally, the models M_{11}, M_{12}, and M_{13} were the additive logistic normal, multiplicative logistic normal, and Dirichlet with log-normal total weight parameterized as M_8, M_9, and M_{10} respectively. All estimated observational parameters were assumed to be constant over years. The densities and further details can be seen in Appendix I.A.

I.2.3 Comparing by AIC

To compare the different observational models, we employed the Akaike Information Criterion (Akaike, 1974). However, the AIC applies to comparison between specific models, whereas each observational model represents an entire family of models, differing in assumed relationships between parameters for different age groups. These families include "full models" where each age group and fleet are assigned independent parameters, a "minimal model" where all age groups share common parameters, as well as a range of models between these two extremes. A standard application of the AIC would require that the optimal model in each family is identified, a task that would involve estimation of parameters in thousands of models.

To avoid this step, which is computationally very demanding and tangential to our purpose, we chose to identify an AIC interval which characterized each model family. This AIC interval gave an upper and a lower bound on the optimal AIC within that family. The upper bound of the interval was attained by the AIC for the full model. The lower bound of the interval was calculated as the AIC that would hypothetically be obtained by the smallest possible nested sub-model if the negative log-likelihood would not increase compared to the full model. The difference between the upper and lower bound is thus twice the difference in the number of parameters between the full model and the minimal model.

A model family was considered clearly superior to another if the upper bound of its AIC interval was below the lower bound of the other model family's interval (i.e., the other model family had a higher interval). Clearly inferior model families could be discarded. To compare the remaining model families, it would be possible to narrow the AIC intervals through testing within each model family, but for simplicity we base the comparison on the full model in each family.

Using AIC to compare the models required that the models were defined on the same data, which was not the case when we compared between numbers-at-age models and proportions-at-age models. The proportions-at-age data were, however, a one-to-one transformation of the numbers-at-age data. Thus, using a standard transformation of densities we derived the log-likelihood for the numbers-at-age data

Table I.2: Overview of the data sources used in the case study. Q1-Q4 indicates at which quarter of the year the survey is conducted.

Fleet	First year	Last year	First age	Last age	Years with missing	Process parameters
Blue Whiting						7
Commercial	1981	2013	1	10		
Survey Q1	2004	2014	3	8	2010	
North-East Arctic Haddock						30
Commercial	1950	2013	3	11		
Survey Q4	1991	2013	3	7		
Survey Q1	1992	2013	3	7	1992-1995	
					2000-2002,2004	
Survey Q1	1992	2013	3	8	1994,1995,2001	
Survey Q3	2004	2013	3	8	2005	
North Sea Cod						11
Commercial	1950	2011	1	7		
Survey Q1	1983	2012	1	5		
Northern Shelf Haddock						33
Commercial	1963	2011	0	8		
Survey Q3	1977	1991	0	6		
Survey Q3	1992	2011	0	6		
Survey Q3	1982	1997	0	6		
Survey Q3	1998	2011	0	6		
Survey Q1	1982	2011	0	4		

that is consistent with our specified distributions based on proportions and totals (Appendix I.B). Using the transformed likelihood in the AIC calculation allowed for valid comparisons of models using numbers-at-age directly versus those using totals and proportions-at-age. Note that a similar transformation was required so that models that used total weight could be compared to models that used total numbers with the proportions.

I.2.4 Case study

We implemented the models for four different data sets used for assessments (Table I.2): The Blue Whiting data set was the basis of the 2014 ICES advice (ICES, 2014a) for Subareas I-IX and XIV; the North-East Arctic Haddock data was used for the 2014 ICES advice (ICES, 2014b) for Subarea IV (North Sea) and Division IIIa West (Skagerrak); The North Sea Cod data was obtained from the 2012 ICES advice (ICES, 2012) for subarea IV (North Sea) and Divisions VII (Eastern Channel) and IIIa West (Skagerrak); and the Northern Shelf Haddock data from the ICES advice for Subarea IV (North Sea) and Division IIIa West (Skagerrak) in 2012 (ICES, 2012). A Beverton-Holt curve was assumed for the relationship between stock and recruitment for North Sea Cod, whereas a random walk was assumed for the other stocks. Two of the data sets had missing data. For the Blue Whiting data set a whole year was missing for the survey, whereas for the North-East Arctic Haddock data set, values were missing for at most three ages per year. The proportions-at-age models above could not easily handle years with missing age observations. Hence, to give a fair comparison of the univariate and multivariate models, we treated years with missing age observations as if it was missing all ages. Process model parameters were assumed equal between ages in the same way as they were for the advice.

I.3 Results

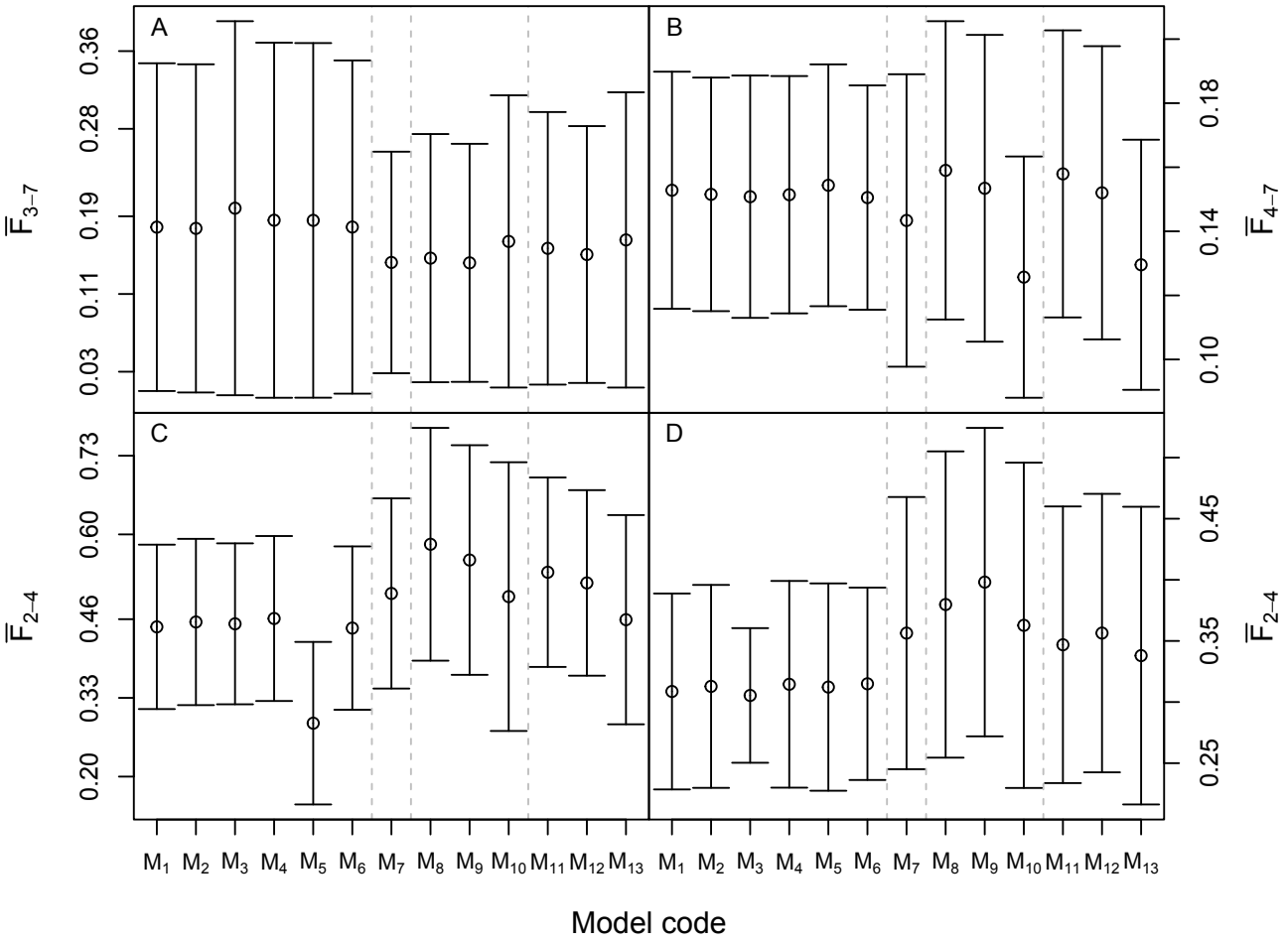


Figure I.1: Last year fishing mortalities with 95% confidence intervals for models M_1 to M_{13} (Table 1) in the case studies: Blue Whiting (A), North-East Arctic Haddock (B), North Sea Cod (C), and Northern Shelf Haddock (D). Vertical dashed grey lines separates the models in model classes (Table 1). Subscripts to $\bar{F}$ indicates the ages the average is over. All ages are weighed equally in the average.

In all four case studies we found that the estimated average fishing mortality (Figure I.1), spawning stock log-biomass (Figure I.2), and their standard errors in the final year differed between models. In particular, we found that for North Sea Cod, the highest fishing mortality was 2 times the lowest fishing mortality, and the widest confidence interval was 1.7 times the narrowest. For Northern Shelf Haddock, the confidence interval of the estimated final year spawning stock biomass for model M_4, which had the highest estimate, did not overlap with the confidence interval for model M_9, which had the lowest estimate.

We found that the models including correlation parameters obtained better fit to the data for the full models than models without correlation parameters within each model class (Figure I.3); the AIC for the full model (upper bound of the interval) was lower for the multivariate log-normal than for the univariate numbers-at-age, and similarly the logistic normals had better model fits than the Dirichlet distribution, which provided one of the highest AIC intervals of all models in all case studies. In the North Sea Cod and Northern Shelf Haddock cases, the lower bounds of the AIC intervals for the Dirichlet distribution were clearly separated from the upper AIC bounds of all other models with an AIC difference of more than 190.

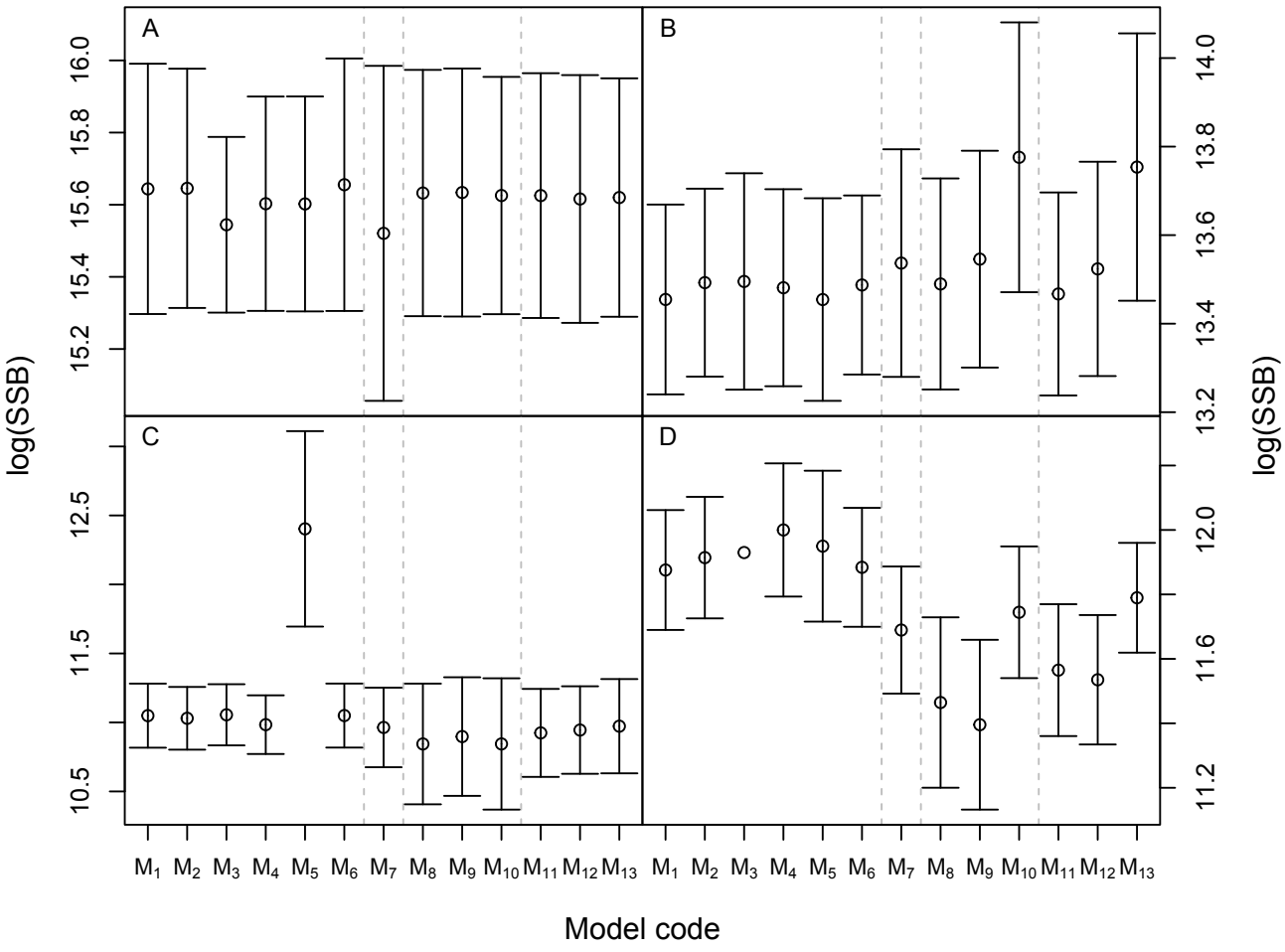


Figure I.2: Last year spawning stock biomass logarithmic (log(SSB)) for models M_1 to M_{13} (Table 1) in the case studies: Blue Whiting (A), North-East Arctic Haddock (B), North Sea Cod (C), and Northern Shelf Haddock (D). Vertical dashed grey lines separates the models in model classes (Table 1).

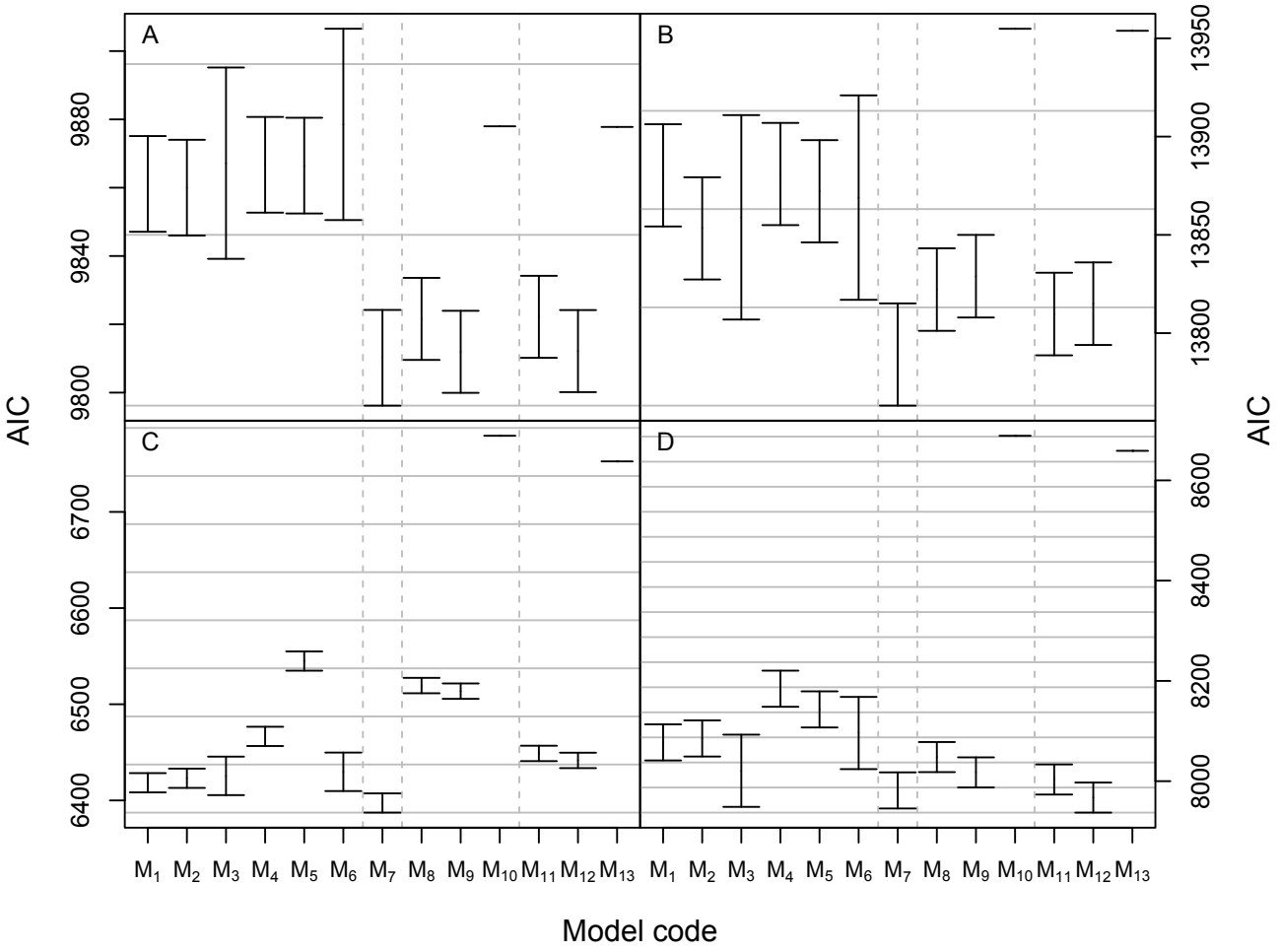


Figure I.3: AIC intervals for models M_1 to M_{13} (Table 1) in the case studies: Blue Whiting (A), North-East Arctic Haddock (B), North Sea Cod (C), and Northern Shelf Haddock (D). The horizontal grey lines indicate AIC differences of 50 starting at the lowest lower bound of the models. Vertical dashed grey lines separates the models in model classes (Table 1).

For North Sea Cod, the multivariate log-normal achieved the lowest AIC for the full model. The AIC interval for this distribution (upper bound: 6407.32) only barely overlapped with the intervals for the generalized gamma (lower bound: 6405.52). Hence among the observational likelihoods we considered, the multivariate log-normal was the most appropriate for the North Sea Cod data and this particular process model. The multivariate log-normal also had the lowest AIC for the full model for North-East Arctic Haddock, whereas it was the multiplicative logistic normal with total weight for Northern Shelf Haddock and the additive logistic normal with total numbers for Blue Whiting. However, in these cases the AIC intervals of the multivariate log-normal, the additive logistic normals, and the multiplicative logistic normals all overlapped. For the two haddock cases, the AIC interval of the generalized gamma also overlapped with the AIC interval of the model with the lowest AIC interval upper bound. We further found that the AIC intervals for proportions-at-age models using total weight overlapped with the corresponding model using total numbers-at-age, except for the North Sea Cod where the total weight models had lower AIC intervals. In addition, the intervals for the additive and multiplicative logistic normals overlapped for all four data sets.

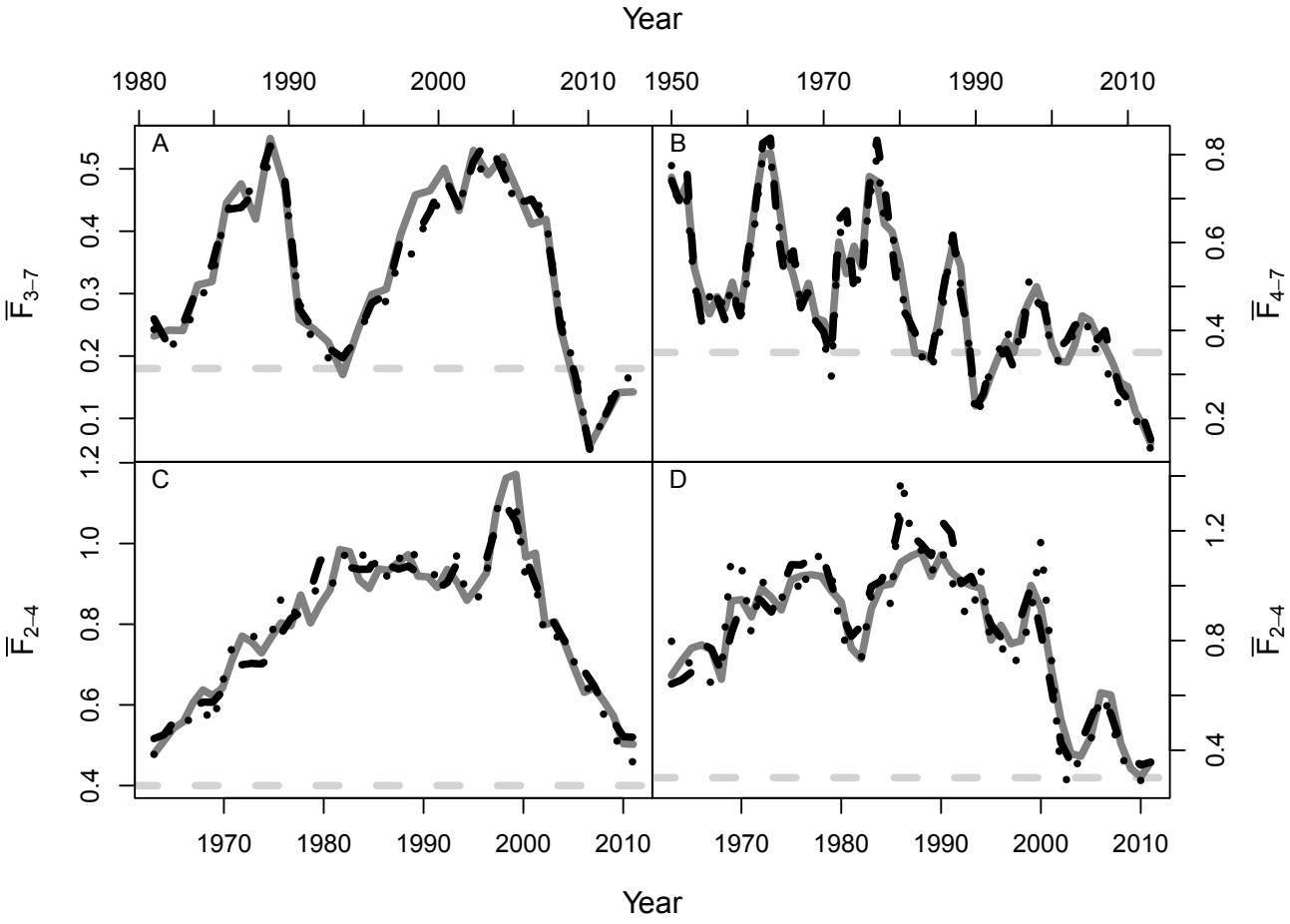


Figure I.4: Estimated average fishing mortality, $\bar{F}$, for Multivariate log-Normal (dark grey line), Multiplicative Logistic Normal with log-Normal Weight (dashed black line), and Dirichlet with log-Normal Weight (dotted black line) in the case studies: Blue Whiting (A), North-East Arctic Haddock (B), North Sea Cod (C), and Northern Shelf Haddock (D). Horizontal dashed grey lines show the management plan reference point. Subscripts to $\bar{F}$ indicates the ages the average is over. All ages are weighed equally in the average.

Overall, the trends in estimated fishing mortality (Figure I.4) and spawning stock biomass (Figure I.5) were similar between the models (See also supplementary material). However, there were noticeable differences in single years. For Blue Whiting, the estimated fishing mortality and spawning stock biomass for multivariate log-normal, multiplicative logistic-normal, and Dirichlet distribution followed each other

closely. The largest difference in average fishing mortality between the multiplicative logistic-normal and the multivariate log-normal was 0.07 (16%), and the difference in spawning stock biomass was up to 12%. In the North Sea Cod case, the multivariate log-normal and the logistic normals had larger differences in fishing mortality and spawning stock biomass. The largest difference in mortality was 0.12 (11.2%), while the spawning stock biomass differed as much as 23%. The resulting confidence intervals also differed between the models. For North Sea Cod the standard errors of the estimated average fishing mortality were up to 76.6% larger for the Dirichlet model, which had the highest AIC, compared to the multivariate log-normal, which had the lowest AIC. Although the trends were similar to the other models, the spawning stock biomass was estimated to be 3.8 to 14.2 times higher for the left truncated normal than for the other models [1]. Likewise, the average fishing mortality was estimated to be lower. For both North Sea Cod[2] and Northern Shelf Haddock[3], the logistic normals provided less volatile estimated time series of fishing mortality and spawning stock biomass than other models. For these models the CVs for commercial catch were estimated to be higher than for the other data sets[4].

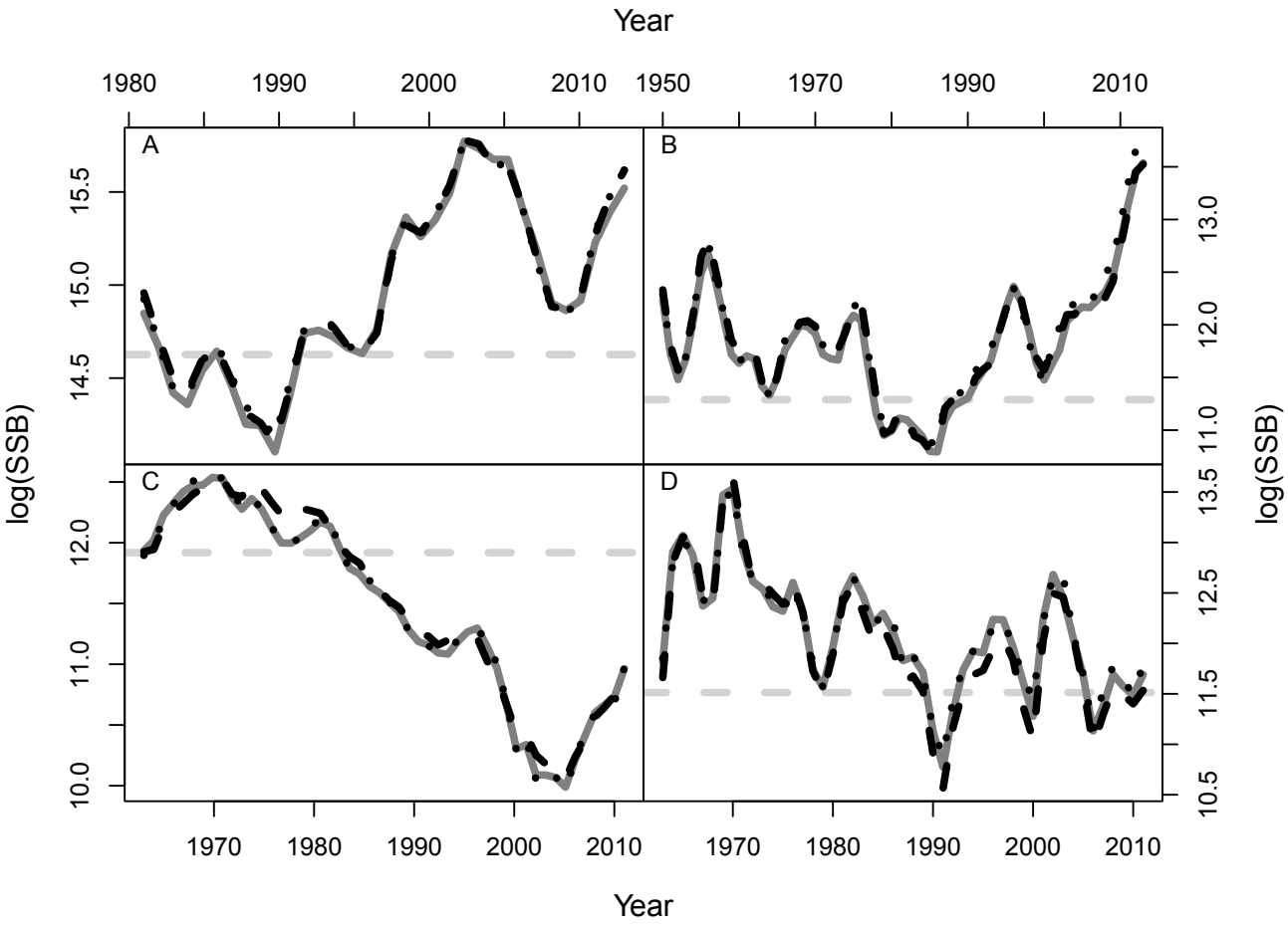


Figure I.5: Natural logarithm of estimated spawning stock biomass, $\log(SSB)$, for Multivariate log-Normal (dark grey line), Multiplicative Logistic Normal with log-Normal Weight (dashed black line), and Dirichlet with log-Normal Weight (dotted black line) in the case studies: Blue Whiting (A), North-East Arctic Haddock (B), North Sea Cod (C), and Northern Shelf Haddock (D). Horizontal dashed grey lines show the management plan reference point.

For Blue Whiting, North-East Arctic Haddock, and Northern Shelf Haddock, estimated CVs were

[1] Figure S31
[2] Figures S34, S35, S37, S38
[3] Figures S47, S48, S50, S51
[4] Table S5

similar for the two logistic normals[5]. The CVs were estimated to be between 0.09 and 0.37. For North Sea Cod, the logistic normals with total numbers had higher CVs (0.32 for commercial catch; 0.27 for survey) than the logistic normals with total weight (0.15 for commercial catch; 0.20 for survey). For the multivariate log-normal, the estimated marginal CVs were estimated to be between 0.01 and 1.86[6]. The CVs were typically estimated to be higher for the first and last ages.

I.4 Discussion

When modelling highly aggregated stock assessment data, the optimal observational likelihood to use can not be known a priori. We provide an objective method for limiting the number of candidate models. By fitting each model with the largest possible number of parameters, upper and lower bounds for the lowest attainable AIC can be calculated. After discarding models with AIC intervals that does not intersect with the lowest interval, the remaining AIC intervals can be narrowed by combining parameters until a single observational likelihood is left. Once a final stock assessment model is found, model validation tools such as residuals and retrospective analysis should be used. AIC was used to determined the optimal observational likelihood, since the AIC estimates the Kullback-Liebler divergence between the candidate model and the true data generating system (Akaike, 1974), i.e. the information lost by using the candidate model instead of the true data generating system (Burnham and Anderson, 2002). The Kullback-Liebler divergence can not be used directly because the calculation requires full knowledge of the true data generating system. Although the AIC was used here, other criterias allowing multivariate data could be used instead.

Choosing the best possible observational likelihood for the data is vital for the short-term management and conservation of fish stocks (Figure I.1, Figure I.2). In 2012, the ICES advice for Northern Shelf Haddock (ICES, 2012), based on XSA (Shepherd, 1999), suggested a 15% increase in the total allowable catch. The suggested increase was based on the difference between the estimated average fishing mortality in the last year and the management reference point. XSA can be seen as a special case of the univariate log-normal (ICES, 2010a) by fixing parameters in a suitable way. Here we saw that the multivariate log-normal and multiplicative logistic normal were more suitable for the Northern Shelf Haddock data than the univariate log-normal, which in turn provides a better fit than the XSA. The estimated average fishing mortality from both the multivariate log-normal and the multiplicative logistic normal are above the reference point. Hence, using one of these models would have suggested a 19% decrease of the total allowable catch to get the fishing mortality at a sustainable level (Figure I.4; panel D). Thus, the choice of observational likelihood in an assessment model can have a substantial effect on the advice brought forward to the fisheries management system.

All models implemented in this study fall in to one of two categories; either they are formulated for numbers-at-age or proportions-at-age with total catch. Most assessment tools only consider one of these categories. We have shown how to compare models between these two ways of using the stock assessment data, and that models from both categories can be suitable, depending on the specific stock. Besides the choice between modelling numbers or proportions there are other differences between the models in, e.g., tail probabilities and skewness, but there are also more subtle differences. When we choose between the log-normal distribution and the gamma distribution, we also choose between whether the Baranov catch equation should model the median or the mean observed catch. Some likelihoods can be re-parameterized to link the Baranov catch equation to either the mean, median or modal observed catch. Although subtle, this difference is important for prediction and interpretation of the results. These choices can be compared

[5]Table S5

[6]Table S1-S4

objectively by including them in the analysis.

Accounting for the correlations in the data is also important for more reliable stock assessment models. In all four case studies, the distributions with correlation parameters, in particular the multivariate log-normal, performed well. However, the correlation structure must be flexible enough to mimic the data, unlike the Dirichlet distribution, which performed poorly, even compared to the models assuming independence between ages. The Dirichlet distribution arises as the distribution of proportions of gamma distributed numbers, where the gamma distributions have equal scale parameters. Stock assessment data is often believed to have constant CV (Cadigan and Myers, 2001), which leads to a parameterization of the gamma distribution where the scale parameters are not equal. Therefore the covariance structure of the Dirichlet distribution does not match the appropriate structure for the gamma distributions. This corresponds with previous findings, that the additive logistic normal generally is more suitable than the Dirichlet distribution in describing the correlation structure in stock assessment data (Francis, 2014). For simplicity we restricted ourselves to the simple AR(1) structure for correlations in this study. The correlation structure can easily be exchanged for other structures such as a linear, AR(2), ARMA(p,q), compound symmetric, or unstructured covariance matrix (Pinheiro and Bates, 2000; Francis, 2014), and some of these may be even more suitable than the AR(1) structure (Berg and Nielsen, 2016).

We only compared frequently used distributions in fisheries stock assessments and close extensions of them, yet any conceivable distribution can be used in the framework we presented, as long as they are all comparable. Correlations between age classes could be introduced in the univariate models through copulas or multivariate extensions such as the multivariate t-distribution or multivariate gamma distributions. Likewise, different ways of handling zero observations, such as zero inflating the models, could be included. For simplicity, years with missing data were removed to compare between univariate and multivariate models. For multivariate numbers-at-age models, missing data may be handled by finding the marginal distribution of the remaining ages. We also restricted the study to use the same likelihood for both commercial catches and surveys. This can be relaxed by including combinations of models (such as Cadigan, 2015), or if the survey index generation is well-understood, a suitable observational likelihood may be derived a priori. Further, the analyses were made conditional on the process models specified in the assessments from which the data was collected. A similar analysis could be made to choose the most appropriate process model conditional on the observational model, or the analyses could be combined. This may influence the specific choice of observational likelihood, as the observational model can, somewhat, compensate for misspecification in the process model and vice versa. Finally, these methods could just as well be used for, e.g., length-based models.

Statistical assessment modelling involves a choice of observational likelihood, and current practice is often to make this choice arbitrarily and subjectively. This applies particularly to the choice of whether the data inputs should be numbers-at-age or proportions-at-age along with total catch in numbers. Here, we have provided methods for an objective choice, by correcting the AICs so that they can be compared between these two families, and we have outlined a computationally efficient method for choosing between families of distributions, by bounding the AICs, which avoids elaborate hypothesis testing within each family, and demonstrated that the best fitting family depends on the particular case. These results will allow stock assessment modellers to choose objectively between these representations of uncertainty on observations, thereby improving model fit and ultimately allowing more accurate assessments.

Acknowledgements

The authors wish to thank Casper W. Berg and three anonymous reviewers for their valuable comments to improve the presentation of this manuscript.

Supplemental Material

Supplemental material is available from the publishers online version of the paper: https://dx.doi.org/10.1139/cjfas-2015-0532

References

Aanes, S. & Pennington, M. (2003). On estimating the age composition of the commercial catch of northeast arctic cod from a sample of clusters. *ICES Journal of Marine Science: Journal du Conseil*, *60*(2), 297–303. doi:10.1016/S1054-3139(03)00008-0

Aitchison, J. (2003). *The statistical analysis of compositional data*. Caldwell, N.J.: The Blackburn Press.

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, *19*(6), 716–723. doi:10.1109/TAC.1974.1100705

Berg, C. W. & Nielsen, A. (2016). Accounting for correlated observations in an age-based state-space stock assessment model. *ICES Journal of Marine Science: Journal du Conseil*, *73*(7), 1788–1797. In Press. doi:10.1093/icesjms/fsw046

Burnham, K. P. & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach* (2nd ed.). Springer.

Cadigan, N. G. (2015). 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*. In press. doi:10.1139/cjfas-2015-0047

Cadigan, N. G. & Myers, R. A. (2001). A comparison of gamma and lognormal maximum likelihood estimators in a sequential population analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, *58*(3), 560–567. doi:10.1139/f01-003

Cook, R. M. (2013). A fish stock assessment model using survey data when estimates of catch are unreliable. *Fisheries Research*, *143*, 1–11. doi:http://dx.doi.org/10.1016/j.fishres.2013.01.003

Crone, P. R. & Sampson, D. B. (1998). Evaluation of assumed error structure in stock assessment models that use sample estimates of age composition. In F. Funk, T. J. Quinn, J. Heifetz, J. N. Ianelli, J. E. Powers, J. F. Schweigert, …C.-I. Zhang (Eds.), *Fishery stock assessment models* (pp. 355–370). Alaska Sea Grant College Program Report No. AK-SG-98-01, University of Alaska Fairbanks.

Farewell, V. T. & Prentice, R. L. (1977). A study of distributional shape in life testing. *Technometrics*, *19*(1), 69–75.

Fournier, D. & Archibald, C. P. (1982). A general theory for analyzing catch at age data. *Canadian Journal of Fisheries and Aquatic Sciences*, *39*(8), 1195–1207. doi:10.1139/f82-157

Francis, R. I. C. C. (2011). Data weighting in statistical fisheries stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, *68*(6), 1124–1138. doi:10.1139/f2011-025

Francis, R. I. C. C. (2014). Replacing the multinomial in stock assessment models: a first step. *Fisheries Research*, *151*, 70–84. doi:10.1016/j.fishres.2013.12.015

REFERENCES

Fryer, R. (2002). Tsa: is it the way? In *Report of working group on methods of fish stock assessment, dec. 2001* (Chap. Appendix D, pp. 86–93).

Gudmundsson, G. (1994). Time series analysis of catch-at-age observations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *43*(1), 117–126.

Hulson, P.-J. F., Hanselman, D. H., & Quinn, T. J. (2011). Effects of process and observation errors on effective sample size of fishery and survey age and length composition using variance ratio and likelihood methods. *ICES Journal of Marine Science: Journal du Conseil*, *68*(7), 1548–1557. doi:10.1093/icesjms/fsr102

Hulson, P.-J. F., Hanselman, D. H., & Quinn, T. J. (2012). Determining effective sample size in integrated age-structured assessment models. *ICES Journal of Marine Science: Journal du Conseil*, *69*(2), 281–292. doi:10.1093/icesjms/fsr189

ICES. (2010a). *Report of the working group on methods of fish stock assessment (WGMG), 20–29 October 2009, Nantes, France*. ICES CM 2009/RMC:12.

ICES. (2010b). *Report of the workshop on reviews of recent advances in stock assessment models worldwide: "Around the World in AD Models" (WKADSAM), 27 September - 1 October 2010, Nantes, France*. ICES CM 2010/SSGSUE:10.

ICES. (2012). *Report of the ICES advisory committee 2012*. ICES Advice 2012, Book 6.

ICES. (2014a). *Report of the ICES advisory committee 2014*. ICES Advice 2014, Book 9.

ICES. (2014b). *Report of the ICES advisory committee 2014*. ICES Advice 2014, Book 3.

Kristensen, K., Nielsen, A., Berg, C. W., & Bell, H. S. B. (2016). TMB: automatic differentiation and laplace approximation. *Journal of Statistical Software*, *70*(1), 1–21. doi:10.18637/jss.v070.i05

Maunder, M. N. (2011). Review and evaluation of likelihood functions for composition data in stock-assessment models: estimating the effective sample size. *Fisheries Research*, *109*(2–3), 311–319. doi:10.1016/j.fishres.2011.02.018

McAllister, M. K. & Ianelli, J. N. (1997). Bayesian stock assessment using catch-age data and the sampling - importance resampling algorithm. *Canadian Journal of Fisheries and Aquatic Sciences*, *54*(2), 284–300. doi:10.1139/f96-285

Methot Jr., R. D. & Wetzel, C. R. (2013). Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research*, *142*, 86–99. doi:http://dx.doi.org/10.1016/j.fishres.2012.10.012

Myers, R. A. & Cadigan, N. G. (1995). Statistical analysis of catch-at-age data with correlated errors. *Canadian Journal of Fisheries and Aquatic Sciences*, *52*(6), 1265–1273. doi:10.1139/f95-123

Nielsen, A. & Berg, C. W. (2014). Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research*, *158*, 96–101. doi:10.1016/j.fishres.2014.01.014

Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-effects models in s and s-plus*. New York, USA: Springer.

Prentice, R. L. (1974). A log gamma model and its maximum likelihood estimation. *Biometrika*, *61*(3), 539–544.

Shepherd, J. G. (1999). Extended survivors analysis: an improved method for the analysis of catch-at-age data and abundance indices. *ICES Journal of Marine Science: Journal du Conseil*, *56*(5), 584–591. doi:10.1006/jmsc.1999.0498

Stacy, E. W. (1962). A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, *33*(3), 1187–1192.

Williams, E. H. & Quinn, T. J. (1998). A parametric bootstrap of catch-age compositions using the dirichlet distribution. In F. Funk, T. J. Quinn, J. Heifetz, J. N. Ianelli, J. E. Powers, J. F. Schweigert, …C.-I. Zhang (Eds.), *Fishery stock assessment models* (pp. 371–384). Alaska Sea Grant College Program Report No. AK-SG-98-01, University of Alaska Fairbanks.

Williams, E. H. & Shertzer, K. W. (2015). *Technical documentation of the beaufort assessment model (bam)* (NOAA Technical Memorandum No. NMFS-SEFSC-671). U.S. Department of Commerce. doi:10.7289/V57M05W6

I.A Likelihoods

I.A.1 Process model

The process model is identical to Nielsen and Berg (2014). For a model including age groups from 1 to A^+ (where age group A^+ contains all ages from A and up), the fishing mortality is modelled by a multivariate random walk, where the oldest modelled ages may be grouped together, indicated by using A^* rather than A^+. Let $F_y = (F_{1,y}, F_{2,y}, \ldots, F_{A^*,y})^T$ be a vector of age specific fishing mortalities in year y. Then

$$\log F_y = \log F_{y-1} + \epsilon_y,$$

where $\epsilon_y \sim N(0, \Sigma)$. The covariance matrix, Σ, is parameterized by and AR(1) structure, $\Sigma_{i,j} = \rho^{|i-j|}\sigma_i\sigma_j$.

Given the fishing mortalities, the population is modelled by an exponential decay model,

$$\begin{aligned}
\log N_{1,y} &= \log\left(R\left(w_{1,y-1}, \ldots, w_{A^+,y-1}, p_{1,y-1}, \ldots, p_{A^+,y-1}, N_{1,y-1}, \ldots, \ldots, N_{A^+,y-1}\right)\right) + \eta_{1,y}, \\
\log N_{a,y} &= \log N_{a-1,y-1} - F_{a-1,y-1} - M_{a-1,y-1} + \eta_{a,y}, 2 \le a < A^+, \\
\log N_{A^+,y} &= \log\left(N_{A^+-1,y-1}e^{-F_{A^+-1,y-1}-M_{A^+-1,y-1}} + N_{A^+,y-1}e^{-F_{A^+,y-1}-M_{A^+,y-1}}\right) + \eta_{A^+,y},
\end{aligned}$$

where all error terms are assumed independent normal distributed. The natural mortalities ($M_{a,y}$), the age specific weight in stock ($w_{a,y}$), and the proportion mature ($p_{a,y}$) are all assumed to be known. The function R describes the relationship between recruitment and spawning population. For North Sea Cod, R was modelled by a Beverton-Holt curve,

$$R\left(w_{1,y-1}, \ldots, w_{A^+,y-1}, p_{1,y-1}, \ldots, p_{A^+,y-1}, N_{1,y-1}, \ldots, N_{A^+,y-1}\right) = \frac{a \cdot SSB_{y-1}}{1 + b \cdot SSB_{y-1}}, \quad a, b > 0$$

with $SSB_y = \sum_{a=1}^{A^+} p_{a,y}w_{a,y}N_{a,y}$, whereas for the other stocks, R was modelled by a random walk

$$R\left(w_{1,y-1}, \ldots, w_{A^+,y-1}, p_{1,y-1}, \ldots, p_{A^+,y-1}, N_{1,y-1}, \ldots, N_{A^+,y-1}\right) = N_{1,y-1}.$$

I.A.2 Observational models

I.A.1 Univariate numbers-at-age models

We consider six univariate observational models for numbers-at-age. Their densities are listed below for each age each year. The joint density for the vector of catches-at-age each year is the product of the age-wise densities. Unless otherwise noted, $x > 0$ is the observed catch (or survey index) at a given age, $\mu > 0$ is the calculated catch at that age (or survey index) based on the Baranov catch equation (or proportional to the total abundance), $\sigma > 0$ is a scale matrix, and $\tau \in \mathbb{R}$ is a shape parameter.

I.A.1.1 Log-normal distribution

$$f_1(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\log(x) - \log(\mu))^2}{2\sigma^2}\right) x^{-1}$$

The mean of the log-normal distribution is $\mu \exp(\sigma^2/2)$, the median is μ and the variance is $(\exp(\sigma^2) - 1)\exp(2\log(\mu) + \sigma^2)$.

I.A.1.2 Gamma distribution

$$f_2(x;\mu,\sigma) = \frac{1}{\Gamma(\sigma)\left(\frac{\mu}{\sigma}\right)^{\sigma}} x^{\sigma-1}\exp\left(-x\sigma/\mu\right)$$

The gamma distribution has mean μ and variance $\frac{\mu^2}{\sigma}$.

I.A.1.3 Generalized gamma distribution

$$f_3(x;\mu,\sigma,\tau) = \begin{cases} |\tau|(\tau^{-2})^{\tau^{-2}}\exp\left(\tau^{-2}\left(\tau\frac{\log(x)-\log(\mu)}{\sigma} - \exp\left(\tau\frac{\log(x)-\log(\mu)}{\sigma}\right)\right)\right)/(\sigma x\Gamma(\tau^{-2})) & \tau\neq 0 \\ (2\pi)^{-1/2}\exp\left(-\frac{(\log(x)-\log(\mu))^2}{2\sigma^2}\right)(\sigma x)^{-1} & \tau = 0 \end{cases}$$

Note that $f_3(x;\mu,\sigma,0) = f_1(x;\mu,\sigma)$, and $f_3(x;\mu,\sigma,\sigma) = f_2(x;\mu,\sigma^{-2})$ for $\sigma > 0$ (Cox et al., 2007). The mean of the generalized gamma is

$$\begin{cases} \frac{\mu\left(\tau^{-2}\right)^{\tau^{-2}}}{\Gamma(\tau^{-2})}\left(\tau^2\right)^{\frac{\sigma\tau+1}{\tau^2}}\Gamma\left(\frac{\sigma\tau+1}{\tau^2}\right) & \tau<0 \\ \mu\exp(\sigma^2/2) & \tau=0 \\ \frac{\mu}{\Gamma(\tau^{-2})}\tau^{2\frac{\sigma}{\tau}}\Gamma\left(\frac{\sigma\tau+1}{\tau^2}\right) & \tau>0 \end{cases}$$

and the variance is

$$\begin{cases} -\frac{\mu^2}{(\Gamma(\tau^{-2}))^2}\left(\left(\tau^{-2}\right)^{2\tau^{-2}}\left(\tau^2\right)^{2\frac{\sigma\tau+1}{\tau^2}}\left(\Gamma\left(\frac{\sigma\tau+1}{\tau^2}\right)\right)^2 - \left(\tau^2\right)^{\frac{2\sigma\tau+1}{\tau^2}}\left(\tau^{-2}\right)^{\tau^{-2}}\Gamma\left(\frac{2\sigma\tau+1}{\tau^2}\right)\Gamma\left(\tau^{-2}\right)\right) & \tau<0 \\ \left(\exp\left(\sigma^2\right)-1\right)\exp\left(2\log(\mu)+\sigma^2\right) & \tau=0 \\ \frac{\mu^2}{(\Gamma(\tau^{-2}))^2}\tau^{4\frac{\sigma}{\tau}}\left(\Gamma\left(\frac{2\sigma\tau+1}{\tau^2}\right)\Gamma\left(\tau^{-2}\right) - \left(\Gamma\left(\frac{\sigma\tau+1}{\tau^2}\right)\right)^2\right) & \tau>0 \end{cases}$$

I.A.1.4 Normal distribution

$$f_4(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi(\mu\sigma)^2}}\exp\left(\frac{-(x-\mu)^2}{2(\mu\sigma)^2}\right)$$

For the normal distribution $x\in\mathbb{R}$ and $\mu\in\mathbb{R}$. The mean is μ and the variance is $\mu^2\sigma^2$

I.A.1.5 Truncated normal distribution

$$f_5(x;\mu,\sigma) = \frac{f_4(x;\mu,\sigma)1_{x\geq 0}(x)}{1-\int_{-\infty}^{0} f_4(y;\mu,\sigma)dy}$$

For the truncated normal distribution $x\geq 0$ and $\mu\geq 0$. The mean of the distribution is

$$\mu + \mu\sigma\frac{f_4(0;\mu,\sigma)}{1-\int_{-\infty}^{0} f_4(y;\mu,\sigma)dy}$$

and the variance is

$$\mu^2\sigma^2 - \mu^2\sigma^2\frac{f_4(0;\mu,\sigma)^2}{\left(1-\int_{-\infty}^{0} f_4(y;\mu,\sigma)dy\right)^2}$$

I.A.1.6 Student's t-distribution on log-scale

$$f_6(x;\mu,\sigma,\tau) = \frac{\Gamma\left(\frac{\tau+1}{2}\right)}{\sigma x\sqrt{\tau\pi}\Gamma\left(\frac{\tau}{2}\right)}\left(1+\frac{((\log(x)-\log(\mu))/\sigma)^2}{\tau}\right)^{-(\tau+1)/2}$$

For the Student's t-distribution on log-scale, $\tau > 0$ is the degrees of freedom. As $\tau \to \infty$ the distribution converges to a log-normal distribution. The distribution does not have mean and variance.

I.A.2 Multivariate numbers-at-age models

We consider one multivariate observational model for numbers-at-age. In the density listed below, A is the number of ages, $x > 0$ is the observed vector of catches (or survey indices), $\mu > 0$ is the calculated catches (or survey indices) based on the Baranov catch equation (or proportional to the total abundance), and Σ is an $A \times A$ symmetric positive-definite scale matrix.

I.A.2.1 Multivariate log-normal distribution

$$f_7(x;\mu,\Sigma) = (2\pi)^{-A/2}|\Sigma|^{-1/2}\exp\left(-\frac{1}{2}(\log(x)-\log(\mu))^T\Sigma^{-1}(\log(x)-\log(\mu))\right)(x_1\cdots x_k)^{-1}$$

For the multivariate log-normal distribution, the scale matrix is the covariance matrix of the logarithm of the observations. When the scale matrix is diagonal, the distribution reduces to univariate log-normals. The marginal means are $\mu_i \exp\left(\frac{1}{2}\Sigma_{ii}\right)$, and the variance/covariance is $\mu_i\mu_j \exp\left(\frac{1}{2}(\Sigma_{ii}+\Sigma_{jj})\right)(\exp(\Sigma_{ij})-1)$.

I.A.3 Proportions-at-age models

We consider three multivariate observational model for proportions-at-age. In the density listed below, A is the number of ages, $x > 0$ with $\sum_{i=1}^{A} x_i = 1$ is a vector of A observed catch proportions (or survey proportions), $\mu > 0$ with $\sum_{i=1}^{A} \mu_i = 1$ is a vector of A calculated catch proportions (or survey proportions) based on the Baranov catch equation (or proportional to the total abundance), and Σ is an $A-1 \times A-1$ symmetric positive-definite scale matrix. A vector with subscript $-A$ denotes the vector without the Ath element.

I.A.3.1 Additive logistic-normal distribution

$$f_8(x;\mu,\Sigma) = (2\pi)^{-(A-1)/2}|\Sigma|^{-1/2}\exp\left(-\frac{1}{2}\left(\alpha(x)-\alpha(\mu)\right)^T\Sigma^{-1}\left(\alpha(x)-\alpha(\mu)\right)\right)(x_1\cdots x_A)^{-1}$$

Here, α is the additive logratio transformation $\alpha(x) = \log\left(\frac{x_{-A}}{x_A}\right)$. The scale matrix is the covariance of the additive logratio transformed observations. Note that if a numbers-at-age vector y follows a multivariate log-normal, then the proportions $y/\sum_i y_i$ follows an additive logistic-normal distribution (Aitchison, 2003). The mean and variance does not have simple forms (Aitchison, 2003).

I.A.3.2 Multiplicative logistic-normal distribution

$$f_9(x;\mu,\Sigma) = (2\pi)^{-(A-1)/2}|\Sigma|^{-1/2}\exp\left(-\frac{1}{2}\left(\mathrm{m}(x)-\mathrm{m}(\mu)\right)^T\Sigma^{-1}\left(\mathrm{m}(x)-\mathrm{m}(\mu)\right)\right)(x_1\cdots x_A)^{-1}$$

Here, m is the multiplicative logratio transformation

$$\mathrm{m}(x) = \left(\log\left(\frac{x_1}{1-x_1}\right),\ldots,\log\left(\frac{x_{A-1}}{1-x_1-\cdots-x_{A-1}}\right)\right).$$

The scale matrix is the covariance of the multiplicative logratio transformed observations. The mean and variance does not have simple forms (Aitchison, 2003).

I.A.3.3 Dirichlet distribution

$$f_{10}(x;\mu,\sigma) = \frac{\Gamma\left(\sum_{i=1}^{A}\sigma\mu_i\right)}{\prod_{i=1}^{A}\Gamma(\sigma\mu_i)}\prod_{i=1}^{A}x_i^{\sigma\mu_i-1}$$

For the Dirichlet distribution, $\sigma > 0$ is a shape parameter. The marginal means of the Dirichlet distribution are μ_i, the variances are $\frac{\mu_i-\mu_i^2}{\sigma+1}$, and the covariances are $\frac{-\mu_i\mu_j}{\sigma+1}$. The Dirichlet distribution is related to the gamma distribution, since if each element of a numbers-at-age vector follows a gamma distribution where the scale parameters in the usual parameterization are equal, then the vector of proportions $y/\sum_i y_i$ follow a Dirichlet distribution. Note that f_2 does not have the same scale parameters for all ages, as they depend on the mean value.

References

Aitchison, J. (2003). *The statistical analysis of compositional data*. Caldwell, N.J.: The Blackburn Press.

Cox, C., Chu, H., Schneider, M. F., & Muñoz, A. (2007). Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statistics in Medicine*, *26*(23), 4352–4374. doi:10.1002/sim.2836

Nielsen, A. & Berg, C. W. (2014). Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research*, *158*, 96–101. doi:10.1016/j.fishres.2014.01.014

I.B Transformation of densities for proportions-at-age models

To compare the AIC of models for numbers-at-age with models for proportions-at-age, the data must be on the same scale. We note that we can transform the numbers-at-age data to proportions-at-age with total catch in numbers by the function

$$g\left((x_1,\ldots,x_A)^T\right) = \left(\frac{x_1}{\sum_{i=1}^A x_i},\ldots,\frac{x_{A-1}}{\sum_{i=1}^A x_i},\sum_{i=1}^A x_i\right)^T,$$

with inverse function

$$h\left((y_1,\ldots,y_{A-1},y_{total})^T\right) = \left(y_1\cdot y_{total},\ldots,y_{A-1}\cdot y_{total},(1-\sum_{i=1}^{A-1} y_i)\cdot y_{total}\right)^T.$$

Hence, if Y is a random vector of proportions-at-age with total catch in numbers with density f, then the corresponding numbers-at-age is $X = h(Y)$. By a change of variable,

$$\begin{aligned} P(X\in B) &= P(Y\in g(B)) \\ &= \int_{g(B)} f(y)dy \\ &= \int_B f(g(x))|\det(Dg)|dx, \end{aligned}$$

the density for the numbers-at-age data is $f(g(y))|\det(Dg)|$. Hence, to compare the AIC between the (natively) numbers-at-age and proportions-at-age models, the log-likelihoods of the proportions-at-age models must be corrected by the logarithm of the absolute determinant of the Jacobian of g, $\log|\det(Dg)|$. The entries of (Dg) are $(Dg)_{A,j} = 1$ for all j, $(Dg)_{i,i} = \frac{1}{\sum_{k=1}^A x_k} - \frac{x_i}{\left(\sum_{k=1}^A x_k\right)^2}$ for all $i < A$ and $(Dg)_{i,j} = \frac{-x_i}{\left(\sum_{k=1}^A x_k\right)^2}$ otherwise. If the total is in weight, where the weight-at-age is assumed to be known, then g is adjusted to

$$g\left((x_1,\ldots,x_A)^T\right) = \left(\frac{x_1}{\sum_{i=1}^A x_i},\ldots,\frac{x_{A-1}}{\sum_{i=1}^A x_i},\sum_{i=1}^A w_i x_i\right)^T$$

Paper II

Connecting single-stock assessment models through correlated survival

Albertsen, C. M.*, Nielsen, A., Thygesen, U. H.

National Institute of Aquatic Resources, Technical University of Denmark, Lyngby, Denmark

* **Corresponding author. Email: cmoe@aqua.dtu.dk**

Post-print. Accepted for publication in ICES Journal of Marine Science

Full citation:

Albertsen, C. M., Nielsen, A., and Thygesen, U. H. (2017). Connecting single-stock assessment models through correlated survival. *ICES Journal of Marine Science*. In Press. doi:10.1093/icesjms/fsx114

Abstract

Fisheries management is mainly conducted via single-stock assessment models assuming that fish stocks do not interact, except through assumed natural mortalities. Currently, the main alternative is complex ecosystem models which require extensive data, are difficult to calibrate, and have long run times. We propose a simple alternative. In three case studies each with two stocks, we improve the single-stock models, as measured by AIC, by adding correlation in the cohort survival. To limit the number of parameters, the correlations are parameterized through the corresponding partial correlations. We consider six models where the partial correlation matrix between stocks follows a band structure ranging from independent assessments to complex correlation structures. Further, a simulation study illustrates the importance of handling correlated data sufficiently by investigating the coverage of confidence intervals for estimated fishing mortality. The results presented will allow managers to evaluate stock statuses based on a more accurate evaluation of model output uncertainty. The methods are directly implementable for stocks with an analytical assessment and does not require any new data sources.

Keywords:

fisheries stock assessment, correlated survival, state-space model, data weighting, multi-stock model

II.1 Introduction

Fisheries management is often based on single-species models for single management areas where the stocks are assumed to have no interaction with surrounding stocks or other species, except via the predetermined natural mortality. Such a simplification is often unrealistic (Eero et al., 2014), and ignoring stock interactions can affect model output (Holsman et al., 2016). These assumptions in single-stock models are, however, convenient as they reduce the data needed on interactions such as migration and predator-prey relations.

Collecting fisheries data is costly and time-consuming: The current surveys require expensive scientific cruises and lab processing; tagging studies require catching and manually tagging many individuals with either low probability of recapture, or expensive electronic tags for gathering telemetry data; food-web data such as stomach samples require manual sampling of stomach content and possibly subsequent DNA analysis; spatially explicit catch data require extensive monitoring systems. Further, incorporating such new data sources in stock assessment models requires historic data that covers a long period for reliable statistical results.

Several methods ranging in complexity have been developed to improve single-stock assessments by accounting for interactions between stocks (e.g. Gislason and Helgason, 1985; Christensen and Pauly, 1992; Begley and Howel, 2004; Fulton et al., 2004; Lewy and Vinther, 2004; Senina et al., 2008; Methot and Wetzel, 2013). Migration between areas can be explicitly modelled by a transportation matrix that moves, for example, a proportion of the population to another area (Methot and Wetzel, 2013). These proportions can be informed by tagging studies such as mark-recapture (Quinn et al., 1990; Hannesson et al., 2008) or, for simple models, be estimated from catch and survey data; however, when estimated from catch and survey data, the results may be influenced by other factors such as inaccuracies in the assumed natural mortality or model misspecification. Likewise, predation can be modelled by a rate-of-consumption ma-

trix (Hollowed, 2000b) consisting of factors which describe the predator induced mortality. Estimating predator induced mortality requires knowledge of complex food webs (Yodzis, 1998; Tarnecki et al., 2016) along with consumption data from stomachs (e.g. Gislason and Helgason, 1985; Daan, 1987; Hollowed, 2000b; Frøysa et al., 2002; Lewy and Vinther, 2004; Richards and Jacobson, 2016).

Complex ecosystem or multi-species models often depend on extensive data from different sources such as tagging studies, DNA analysis, environment data or stomach content data (e.g. Christensen and Walters, 2004). Such data sources are not readily available from single-species assessments, are expensive to produce, and may take several years to collect. Data requirements can also include human activity, biochemistry, physiology, geology and oceanography (Link et al., 2010) raising the complexity and data collection cost. Further, ecosystem models often have a vast amount of settings and can be difficult to calibrate and run (Hollowed, 2000a; Chen et al., 2009; Ainsworth and Walters, 2015). Currently, there is a gap between single-stock assessment models and the complex ecosystem models (Collie et al., 2016), with a lack of operational models that combine multiple stocks while only utilizing the data readily available for single-stock assessments, making it directly implementable for fisheries management.

The objective of this study is to introduce a method for combining single-stock state-space age-based assessments models to provide a simple alternative to complex ecosystem models. The method is illustrated in three case studies including a total of six European stocks combined two by two. In each case study, we investigate whether combining single-stock models improves the model fit to data compared to individual assessments.

We extend the single-stock age-based state-space stock assessment model to allow simultaneous estimation for several stocks through correlated survival processes. Correlations are included in the model through the corresponding partial correlations. Assuming Gaussianity, two random variables from a collection of variables are partially uncorrelated correcting for the remaining variables in the collection if and only if they are conditionally independent given the remaining variables in the collection. Hence, modelling partial correlations allows a clear interpretation of reducing model parameters (i.e. by fixing partial correlations at zero), while still allowing a flexible correlation structure. The multi-stock model introduced does not require further data than the single-stock model.

II.2 Methods

II.2.1 Stock Assessment Model

We implemented age-based state-space stock assessment models for six European stocks (Table II.1). The models were implemented with the R-package Template Model Builder (Kristensen et al., 2016).

II.2.1 Process Model

The process model for an individual stock, s, was identical to Nielsen and Berg (2014) model D. The population process followed an exponential decay model,

$$\log N_{s,a,y} = \log N_{s,a-1,y-1} - F_{s,a-1,y-1} - M_{s,a-1,y-1} + \eta_{s,a,y}, a \in \{2, 3, \ldots, A_s^+ - 2, A_s^+ - 1\}$$

with special conditions for the last age group A_s^+ (See Nielsen and Berg, 2014; Paper I, for details). In the model, $F_{s,a,y}$ is the fishing mortality, $M_{s,a,y}$ is the natural mortality, and $\eta_{s,a,y}$ is a random Gaussian zero

mean increment, which allow inaccuracies in the deterministic part of the model. These inaccuracies can be thought of as random deviations in the natural mortality, which is assumed to be known. While Nielsen and Berg (2014) and Albertsen et al. (Paper I) model $\eta_{s,a,y}$ as independent, we will allow correlations across stocks and ages (see subsection II.2.2), while maintaining independence across years. Recruitment, $\log N_{s,1,y}$, was modelled by a random walk. Configuration of model parameters with respect to coupling across ages was done in the same way as for the official advice.

II.2.2 Observational Model

Following Albertsen et al. (Paper I), the observed log-catch conditional on the process model were modelled by a multivariate normal where the covariance $\Sigma_s^{(O)}$ had an AR(1) covariance structure, $(\Sigma_s^{(O)})_{i,j} = \rho_{o,s}^{|i-j|}\tau_{o,s,i}\tau_{o,s,j}$ with $\rho_{o,s} \in (-1,1)$ and $0 < \tau_{o,s,i}$ for all indices i. The Catch vector, $C_{s,y} = (C_{s,1,y}, \ldots, C_{s,A_s^+,y})$, for stock s in year y followed the Baranov catch equation

$$\log C_{s,y} = \log F_{s,y} - \log(F_{s,y} + M_{s,y}) + \log N_{s,y} + \log\left(1 - \exp(-F_{s,y} - M_{s,y})\right) + \epsilon_{s,y}^{(O)}$$

were $\epsilon_{s,y}^{(O)} \sim N(0, \Sigma_s^{(O)})$, $N_{s,y} = (N_{s,1,y}, \ldots, N_{s,A_s^+,y})$, $F_{s,y} = (F_{s,1,y}, \ldots, F_{s,A_s^+,y})$, $M_{s,y} = (M_{s,1,y}, \ldots, M_{s,A_s^+,y})$, and all operations are element-wise. Likewise, survey indices were modelled by

$$\log I_{s,y}^{(i)} = \log q_s^{(i)} + \log N_{s,y} - (F_{s,y} + M_{s,y}) d_s^{(i)}/365 + \epsilon_{s,y}^{(I,i)}$$

where $\epsilon_{s,y}^{(I,i)} \sim N(0, \Sigma_s^{(I,i)})$, $I_{s,y}^{(i)}$ is a vector of age observations for the ith survey in stock s year y, $q^{(i)} > 0$, $d_S^{(i)}$ is the number of days into the year the survey is conducted, and $\Sigma_s^{(I,i)}$ has an AR(1) structure. For simplicity, years with missing data for single ages were treated as if they were missing all ages for that fleet. Alternatively, missing data could be handled by, for instance, imputing random effects. All observational parameters were estimated; however, they were configured with respect to coupling across ages in the same way as for the official advice.

II.2.2 Connecting stocks

For many stocks, assuming independence between the error terms, $\eta_{s,a,y}$, is unrealistic. Inaccuracies in the assumed natural mortality will be correlated by connections not included in the model such as environmental changes, migration, or predation. All other things held constant, environmental effects will be seen as positive correlation; migration will be seen as negative correlation; while predation will be seen as negative correlation in a predator-prey relation or as positive correlation if two stocks are influenced by the same predator.

To connect the S individual stocks, we temporarily extended the stock log-population processes to have the same number of ages, A, for convenience. As an example, consider a case with two stocks (such as the Baltic cod case study presented in subsection II.2.3) where one stock includes ages two to eight, while the other stock includes ages zero to seven. To construct the correlation matrix, both log-population processes have to be temporarily extended to include ages zero to eight to combine the stocks.

The individual processes were stacked to one process,

$$N_y = (N_{1,1,y}, \ldots, N_{1,A,y}, \ldots, N_{S,1,y}, \ldots, N_{S,A,y})^T,$$

for which we modelled the error terms, $\eta_y = (\eta_{1,1,y}, \ldots, \eta_{1,A,y}, \ldots, \eta_{S,1,y}, \ldots, \eta_{S,A,y})^T$, by a multivariate normal with covariance Σ_N.

The covariance was parameterized through the Cholesky decomposition of the precision matrix P_N. Defining L to be a lower triangular matrix of parameters with diagonal 1, a positive definite precision matrix was obtained by

$$P_N = LL^T$$

If the precision matrix was scaled to have diagonal 1, the off-diagonal elements were the partial correlations with opposite sign. With this construction, the diagonal elements of P could not be estimated freely. To obtain the covariance matrix as the inverse precision, we scaled P to have diagonal 1 and added freely estimated variance parameters. Including variance parameters at this stage maintained the interpretation of the estimated parameters from the single-stock model, where the parameter is directly related to the variance of the specific age. The scaling was done by defining a diagonal scale matrix, S, by

$$S_{ij} = \begin{cases} \frac{\sigma_i}{\sqrt{(P^{-1})_{ii}}} & i = j \\ 0 & i \neq j \end{cases}$$

where $\sigma_i > 0$ were variance parameters. Then the covariance matrix for the population process was obtained by

$$\Sigma_N = SP^{-1}S^T$$

Hence, any covariance matrix, Σ_N, can be constructed from a suitable choice of L and σ_i, $i = 1, \ldots, S \cdot A$. The marginal covariance matrix of the ages included in the assessments was found by using the corresponding indices of Σ_N. In the example above, Σ_N would be an 18×18 matrix where the indices three to nine corresponds to ages two to eight for the first stock, while indices 10 to 17 corresponds to ages zero to seven for the second stock. We considered the models $\mathcal{M}_\alpha$, $\alpha = 0, \ldots, 5$, defined such that the off-diagonal partial correlation in the error terms of $N_{s,a,t}$ and $N_{s',a',t}$ was zero if $s = s'$ or $|a - a'| \geq \alpha$ (See Appendix II.A for details). Hence, $\mathcal{M}_0$ corresponded to independent assessments while the between-stock correlation has a band structure for $\mathcal{M}_1$ to $\mathcal{M}_5$. The width of the band increases with the model subscript. When the process is Gaussian as here, a zero partial correlation implies conditional independence; hence, by setting all within-stock partial correlations to zero, any correlation within each stock is a result of partial correlations between the stocks. Therefore, the individual stock models are only changed as an effect of the between-stock connections introduced.

II.2.3 Case studies

We considered three case studies with two stocks each (Table II.1). Two cases represent stocks that are known to migrate between management areas, while the last case represent stocks with intersecting management areas. The Baltic cod (*Gadus morhua*) case consists of the Eastern and Western Baltic cod stocks; the data was the basis of the 2014 ICES advice (ICES, 2014) for sub divisions 25-32 and 23-24, respectively (Figure II.1). The Baltic herring (*Clupea harengus membras*) case comprised data for the Bothnian Sea and Central Baltic herring stocks used in the 2015 ICES advice (ICES, 2015a) for sub divisions 30 and 25-27,28.2,29,32, respectively (Figure II.1). Data for the Baltic mix case was obtained from the 2015 ICES advice for the Baltic sprat (*Sprattus sprattus*; sub divisions 22-32; ICES, 2015a) and Western Baltic herring (*Clupea harengus membras*; sub divisions 22-24; ICES, 2015b) stocks. Three of the stocks considered had missing data (Table II.1). The Central Baltic Herring and Baltic Sprat data had missing data for all ages in surveys; the Western Baltic Herring was missing data for one age in 1991 for the Q3 survey, while the remaining years with missing data were missing all ages. Process model parameters were coupled between ages in the same way as they were for the advice.

Table II.1: Overview of the data sources used in the case studies. Q1-Q4 indicates at which quarter of the year the survey is conducted.

Fleet	First year	Last year	First age	Last age	Years with missing
Baltic Cod Case Study					
Eastern Baltic Cod (Sub divisions: 25-32)					
Commercial	1966	2013	2	8	
Survey Q1	1991	2000	2	6	
Survey Q1	2001	2013	2	6	
Survey Q4	2002	2013	2	5	
CPUE tuning index	1997	2013	3	7	
Western Baltic Cod (Sub divisions: 23-24)					
Commercial	1970	2013	1	7	
CPUE tuning index	1997	2013	3	6	
Survey Q1	1992	2013	1	6	
Survey Q4	1992	2013	0	5	
Baltic Herring Case Study					
Bothnian Sea Herring (Sub division: 30)					
Commercial	1973	2014	1	9	
Survey Q4	2007	2014	2	8	
Survey Q2	1990	2014	3	9	
Central Baltic Herring (Sub divisions: 25-27,28.2,29,32)					
Commercial	1974	2014	1	8	
Survey Q4	1991	2014	1	8	1993,1995,1997
Baltic Mix Case Study					
Baltic Sprat (Sub divisions: 22-32)					
Commercial	1991	2014	1	8	
Survey Q4	1991	2014	1	8	1993,1995,1997
Survey Q2	2001	2014	1	8	
Survey Q1	1992	2014	1	1	1994,1996,1998
Western Baltic Herring (Sub divisions: 22-24)					
Commercial	1991	2014	0	8	
Survey Q3	1991	2014	1	8	1991,1999
Survey Q4	1991	2014	0	8	1991-1993,2001
Survey Q2	1992	2014	0	0	
Survey Q1	1991	2014	1	4	
Survey Q3	1991	2014	1	4	

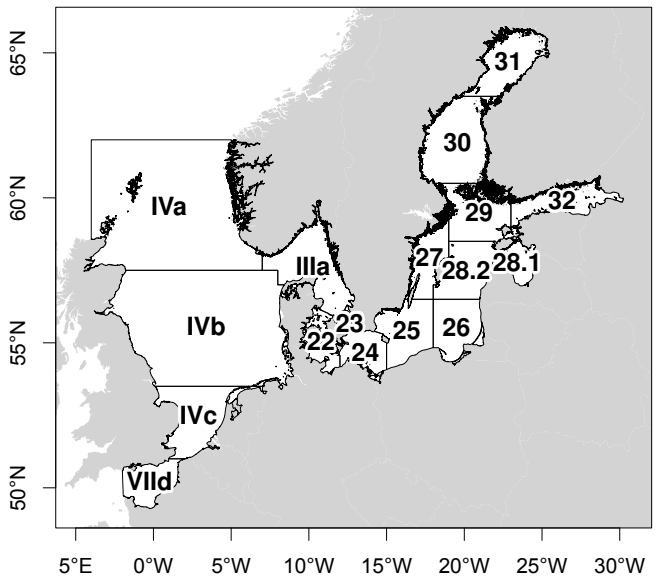


Figure II.1: ICES sub divisions in the North Sea and Baltic Sea (http://geo.ices.dk; accessed 08-02-2017).

Table II.2: AIC difference from the best model fit for each case study, with the number of free parameters needed to obtain the correlation structure in parentheses.

Model	Baltic Cod	Baltic Herring	Baltic Mix
$\mathcal{M}_0$	35.06 (0)	32.99 (0)	14.26 (0)
$\mathcal{M}_1$	37.52 (6)	**0.00** (8)	24.90 (8)
$\mathcal{M}_2$	**0.00** (18)	10.96 (23)	6.59 (23)
$\mathcal{M}_3$	3.04 (29)	7.48 (36)	**0.00** (36)
$\mathcal{M}_4$	15.91 (38)	15.24 (47)	16.76 (47)
$\mathcal{M}_5$	24.94 (45)	38.27 (56)	13.07 (56)

II.2.4 Simulation study

We simulated data sets with 50 years from the stock assessment model above with Gaussian observations (See Appendix II.B for details). The data sets were simulated with two stocks combined by partial correlation models $\mathcal{M}_0$ to $\mathcal{M}_5$. For each model we simulated 150 data sets. To investigate whether the true model could be identified by Akaike information criterion (AIC; Akaike, 1974), and in particular whether AIC would favour overfitting, the models $\mathcal{M}_0$ to $\mathcal{M}_5$ were fitted to the data. Further, the coverage of confidence intervals were compared between the models.

II.3 Results

II.3.1 Case studies

Including correlations in the survival process between stocks improved the model fit, as measured by AIC, in all three case studies (Table II.2). For Baltic Cod, $\mathcal{M}_2$ had the lowest AIC followed by $\mathcal{M}_3$ with an AIC that was 3.04 higher. Model $\mathcal{M}_3$ achieved the lowest AIC for Baltic Mix. Here, $\mathcal{M}_2$ followed with

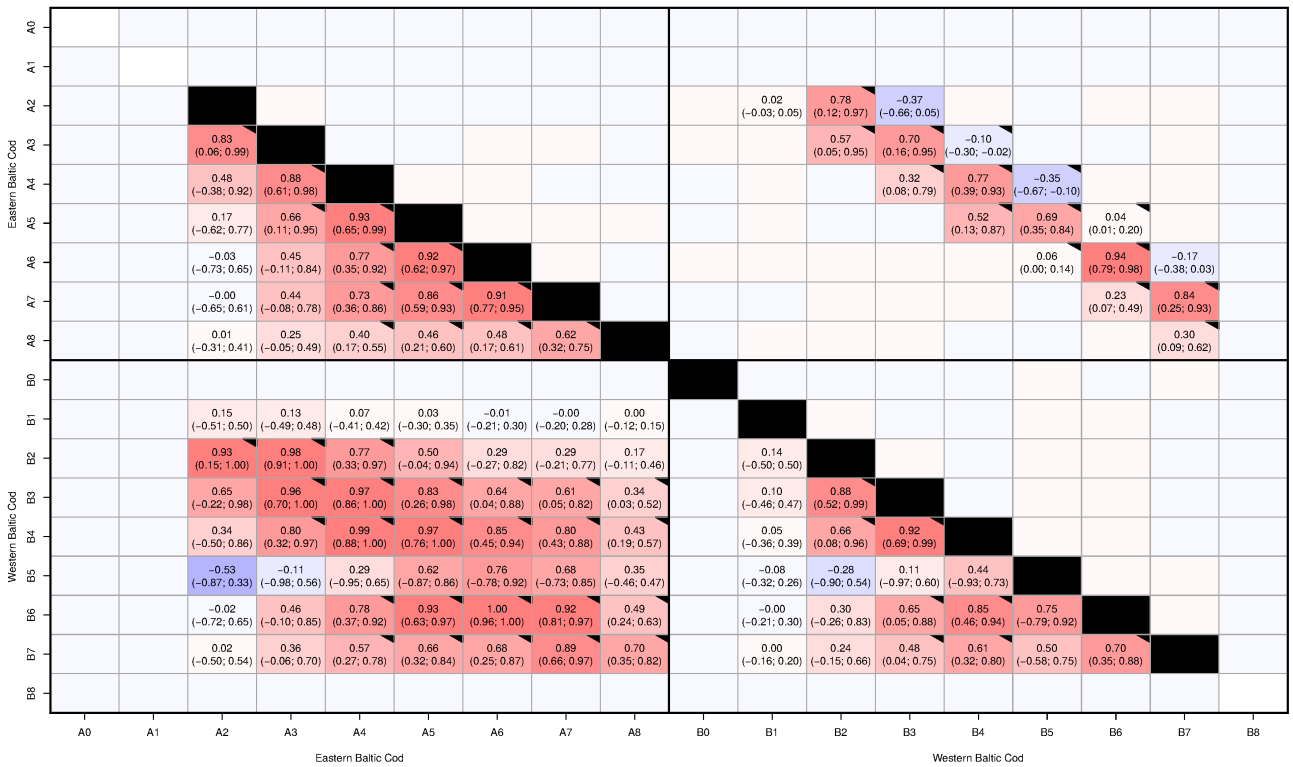


Figure II.2: Estimated partial correlation (upper triangle) and corresponding correlation (lower triangle) in non-fishery survival, $\hat{\Sigma}_N$, for model $\mathcal{M}_2$ in the Baltic Cod case study between all stocks and ages. Dark blue indicates a high negative correlation, white is no correlation, while dark red indicates a high positive correlations. By construction, both correlation and partial correlation is 1 in the diagonal. The matrix is extended to have the same ages in both stocks. Ages present in the data are represented by a black rectangle in the diagonal. 95% Confidence intervals are included in parentheses (see Appendix II.C for details). A black triangle in the upper right corner indicates that the (partial) correlation is significant. Axis labels indicate the age group, e.g., A3 is age three in the first stock, while B2 is age two in the second stock.

an AIC difference of 6.59. For Baltic Herring, $\mathcal{M}_1$ had the lowest AIC, followed by $\mathcal{M}_3$ with a difference of 7.48.

For the Baltic Cod stocks (Figure II.2), the age difference in recruitment was two years. With the recruitments two years apart, the partial correlation was zero by construction in $\mathcal{M}_3$. Further, the recruitment to Eastern Baltic was estimated to be almost partially uncorrelated with Western Baltic age one (Estimate: 0.02; lower: -0.03; upper: 0.05). The resulting correlation in recruitment was also low (Estimate: 0.15; lower: -0.51; upper: 0.50). However, both the partial correlation (Estimate: 0.78; lower: 0.12; upper: 0.97) and correlation (Estimate: 0.93; lower: 0.15; upper: 1.00) between ages two were estimated to be high for the two stocks. For Eastern Baltic Cod, close age groups were estimated to be highly correlated as a result of the partial correlations with Western Baltic Cod. These within-stock correlations between age groups no more than two years apart were estimated to be above 0.6, except for ages six and eight where the correlation was 0.48 (0.17; 0.61) and ages two and four where the correlation was 0.48 (-0.38; 0.92). A majority of the correlations between the stocks, 89%, were estimated to be positive. Of these, 81% were above 0.25, and 53% were above 0.6. With respectively seven and eight age groups in the two Baltic Cod stocks, 18 partial correlations were estimated for $\mathcal{M}_3$ based on the same number of free parameters in the Choleskey decomposition. Of these estimates, 22% were estimated to be negative, 17% had an absolute value below 0.1, and 28% had an absolute value above 0.7.

For Baltic Herring, the recruitments were estimated to be highly partially correlated in $\mathcal{M}_1$ (Fig-

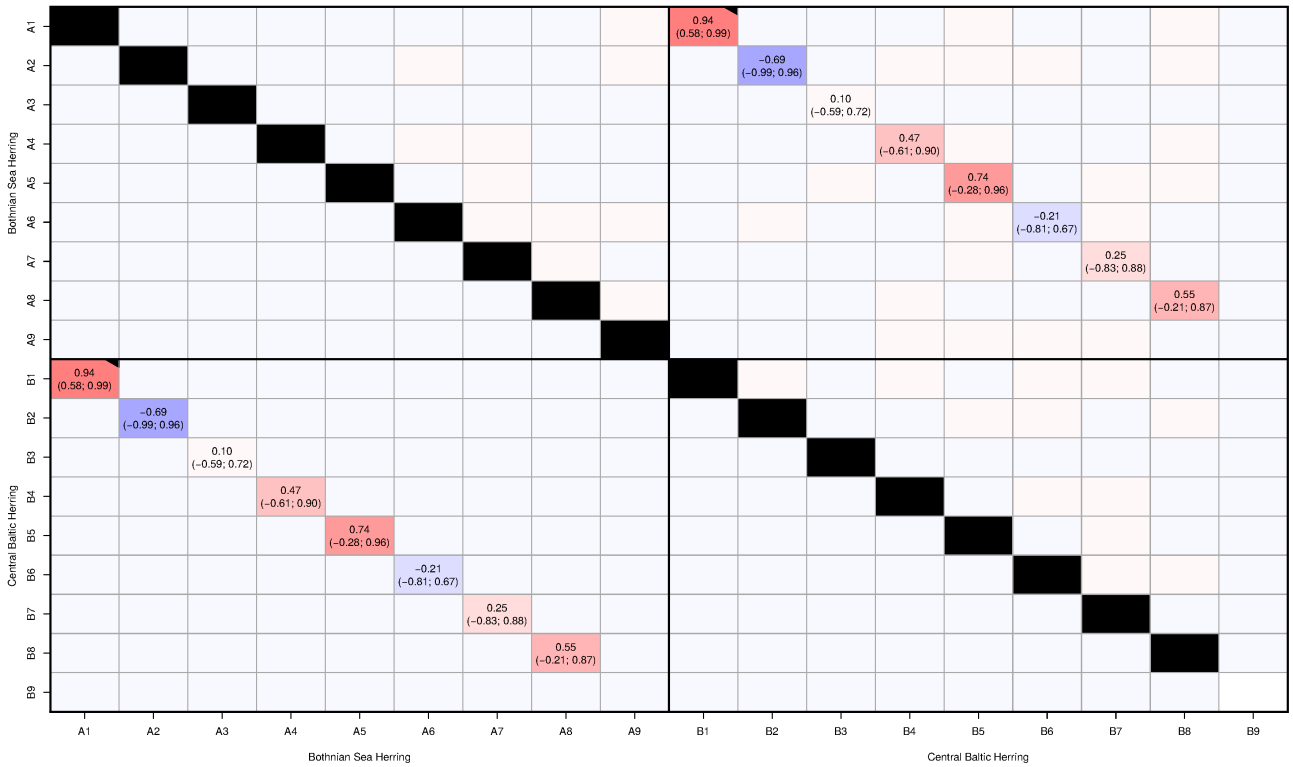


Figure II.3: Estimated partial correlation (upper triangle) and corresponding correlation (lower triangle) in non-fishery survival, $\hat{\Sigma}_N$, for model $\mathcal{M}_1$ in the Baltic Herring case study between all stocks and ages. Dark blue indicates a high negative correlation, white is no correlation, while dark red indicates a high positive correlations. By construction, both correlation and partial correlation is 1 in the diagonal. The matrix is extended to have the same ages in both stocks. Ages present in the data are represented by a black rectangle in the diagonal. 95% Confidence intervals are included in parentheses (see Appendix II.C for details). A black triangle in the upper right corner indicates that the (partial) correlation is significant. Axis labels indicate the age group, e.g., A3 is age three in the first stock, while B2 is age two in the second stock.

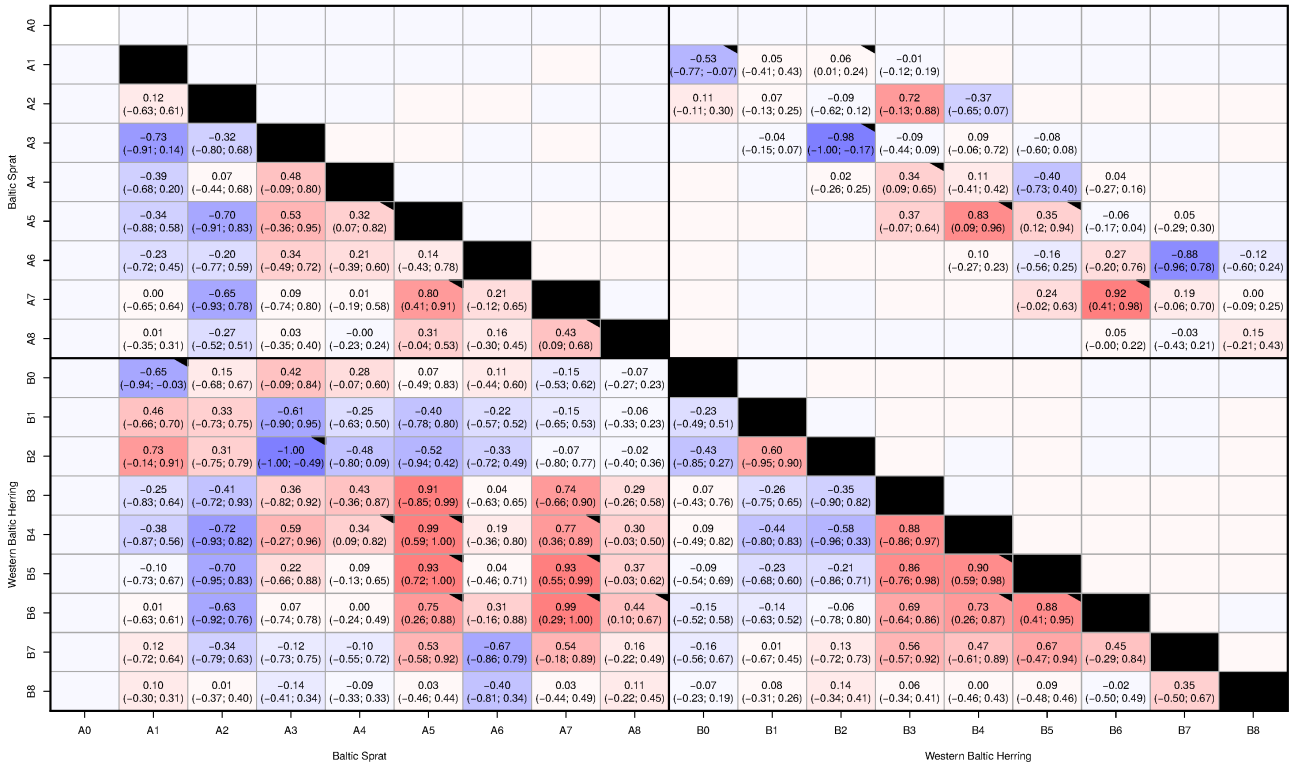


Figure II.4: Estimated partial correlation (upper triangle) and corresponding correlation (lower triangle) in non-fishery survival, $\hat{\Sigma}_N$, for model $\mathcal{M}_3$ in the Baltic Mix case study between all stocks and ages. Dark blue indicates a high negative correlation, white is no correlation, while dark red indicates a high positive correlations. By construction, both correlation and partial correlation is 1 in the diagonal. The matrix is extended to have the same ages in both stocks. Ages present in the data are represented by a black rectangle in the diagonal. 95% Confidence intervals are included in parentheses (see Appendix II.C for details). A black triangle in the upper right corner indicates that the (partial) correlation is significant. Axis labels indicate the age group, e.g., A3 is age three in the first stock, while B2 is age two in the second stock.

ure II.3) with an estimated coefficient of 0.94 (0.58; 0.99). For model $\mathcal{M}_1$, the partial correlations are equal to the correlations by construction. As with the recruitment, ages two had a high correlation. This correlation was estimated to be highly negative (Estimate: -0.69) albeit with an uncertain estimate (95% Confidence interval: -0.99; 0.96). Ages five and ages eight were both estimated to be positively correlated with coefficients of 0.74 and 0.55 respectively. However, both 95% confidence intervals included zero (Figure II.3). Likewise, the remaining estimated correlations had confidence intervals which included zero. With respectively eight and nine ages in the two Baltic Herring stocks, eight partial correlations were estimated for $\mathcal{M}_1$ using the same number of free parameters in the Choleskey decomposition.

In the Baltic Mix case (Figure II.4), recruitments were estimated (95% confidence interval) to have a negative partial correlation in $\mathcal{M}_3$ of -0.53 (-0.77; -0.07). The resulting correlation in recruitment was also highly negative with an estimated coefficient of -0.65 (-0.94; -0.03). Several age classes were estimated to be highly correlated. Baltic Spart age 3 and Western Baltic Herring age 2 had an estimated correlation of -1.00 (-1.00; -0.49); Sprat age 5 had an estimated correlation with Herring ages 3 to 5 of 0.91 (-0.85; 0.99), 0.99 (0.59; 1.00), and 0.93 (0.72; 1.00) respectively; and Sprat age 7 was estimated to be correlated with Herring ages five and six with coefficients of 0.93 (0.55; 0.99) and 0.99 (0.29; 1.00). With eight and nine ages respectively in the two Baltic Mix stocks, 36 partial correlations were estimated for $\mathcal{M}_3$ using the same number of free parameters in the Choleskey decomposition. Of these partial correlation estimates, 39% were estimated to be negative, 47% had an absolute value below 0.1, and 14% had an

Table II.3: Number of simulation studies that resulted in the lowest AIC for models $\mathcal{M}_0$ to $\mathcal{M}_5$.

	Best Model					
True Model	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$
$\mathcal{M}_0$	136	12	1	1	0	0
$\mathcal{M}_1$	0	140	6	1	2	1
$\mathcal{M}_2$	0	0	120	14	8	8
$\mathcal{M}_3$	0	0	5	105	32	8
$\mathcal{M}_4$	0	0	5	21	107	17
$\mathcal{M}_5$	0	0	0	0	33	117

Table II.4: Average (over time) coverage of Hessian based 95% confidence intervals of the logarithm of the predicted average (over ages) fishing mortality in the simulation study.

	Estimation Model					
Simulation Model	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$
$\mathcal{M}_0$	0.945	0.943	0.941	0.940	0.940	0.940
$\mathcal{M}_1$	0.942	0.940	0.939	0.936	0.935	0.934
$\mathcal{M}_2$	0.916	0.933	0.939	0.936	0.937	0.936
$\mathcal{M}_3$	0.897	0.930	0.939	0.934	0.933	0.931
$\mathcal{M}_4$	0.885	0.921	0.929	0.928	0.929	0.931
$\mathcal{M}_5$	0.919	0.917	0.897	0.902	0.934	0.932

absolute value above 0.7.

II.3.2 Simulation study

The simulation study consisted of a total of 900 simulations from the six models $\mathcal{M}_0$ to $\mathcal{M}_5$. Of these simulations, 81% of the re-estimations identified the correct model from the candidates $\mathcal{M}_0$ to $\mathcal{M}_5$ (Table II.3). Simulations from $\mathcal{M}_3$ were the most difficult to identify. Of these, 70% were correctly identified as generated by $\mathcal{M}_3$ whereas 21% were identified as $\mathcal{M}_4$, 5% as $\mathcal{M}_5$, and 3% were identified as $\mathcal{M}_2$. Only 7.1% of all simulations obtained the lowest AIC with a smaller model than the true simulation model.

Although the AIC was able to identify the true model in most cases, spurious correlations were found by the complex models when the true model did not include correlations. When simulating from $\mathcal{M}_0$, all the models $\mathcal{M}_1$ to $\mathcal{M}_5$ had high estimated correlations. For $\mathcal{M}_5$, 67% of the simulations had one or more estimated correlations with an absolute value above 0.5 while 10% had an absolute value above 0.75. The number was reduced with model complexity. The models $\mathcal{M}_1$ to $\mathcal{M}_4$, had estimated correlations with absolute values above 0.5 for 44, 57, 65, and 67% of the simulations, respectively. Likewise, respectively 3, 9, 11, and 10% of the simulations had estimated correlations with absolute values above 0.75.

Besides identifying the true model, the simulations were used to calculate the coverage of Hessian based 95% confidence intervals of the logarithm of the average fishing mortality (Table II.4). When simulating from the models $\mathcal{M}_0$ and $\mathcal{M}_1$, all four estimation models had an average coverage close to 95% based on 150 simulations. This was not the case in the simulations where substantial correlations were included in the true model. When simulating from $\mathcal{M}_2$ to $\mathcal{M}_5$, estimation model $\mathcal{M}_0$ only had a coverage of 91.6, 89.7, 88.5, and 91.9%, respectively. Likewise, when simulating from $\mathcal{M}_5$, estimation

models $\mathcal{M}_1$ to $\mathcal{M}_3$ had coverages of 91.7, 89.7, 90.2%, respectively, while models $\mathcal{M}_4$ and $\mathcal{M}_5$ had higher coverages: 93.4%, and 93.2%, respectively.

II.4 Discussion

European fisheries management mainly relies on single-stock assessment models, which assumes that stocks are not interconnected. For many stocks, this is not a reasonable assumption, as they are connected by several interactions, such as predation, migration and influences from the environment. Currently, the main alternative to single-stock models is complex ecosystem models, which are data intensive and require extensive tuning and long run times. We introduced a simple and fast alternative: connecting stocks by correlated survival variability. In our case studies, the models were fitted in 1.1 to 4.5 minutes. The models only require the data already used in single-stock models, which makes it readily usable for a wide range of stocks. Further, the model can be reduced to independent single-stock models, which allows objective testing of whether connecting stocks is an improvement, through likelihood ratio tests, AIC, or the like. We modelled complicated correlation patterns by a simple structure through the partial correlations; however, other correlation structures could be used instead.

Unlike multi-species models and ecosystem models, which try to directly model interactions or common influences of stocks, the present approach estimates whether a good year for one cohort is also a good year for another cohort it may be linked to. While simplistic and less informative about the ecosystem, this approach is highly operational and easily implemented for any group of stocks with an analytical assessment, since the model does not require any new data sources to be introduced. Further, the approach can be used to model complex correlation structures in other parts of the model such as the fishing mortality process.

The case studies presented were chosen to illustrate stocks that are believed to migrate and stocks that share environment. In general, connecting stocks should be considered when: biological knowledge suggests interactions, the management areas intersect, or patterns in model validation tools such as residuals (e.g. Thygesen et al., 2017) or retrospective analysis (e.g. Hurtado-Ferro et al., 2015) are present. If biological knowledge indicates migration or predator-prey relations, it should be tested whether assessments can be improved by correlating the abundance processes. Likewise, when the management areas intersects, the stocks likely share environmental conditions or compete for resources which can correlate the abundance processes. Finally, common or opposite patterns in the residuals or retrospective analysis can indicate missing biological knowledge which can lead to correlated abundance processes.

Although modelling partial correlations provides a simple interpretation of parameters, understanding why they arise is not as simple in the present framework. Correlations among stocks or ages can be the result of several processes such as migration, predation or environmental impact. While further research is needed to evaluate the possibility of extending the framework to directly include any of these processes, it is straight forward to evaluate whether stocks are similar in the sense that they could share parameters in the assessment model. Testing whether parameters are shared among stocks can give a basis for evaluating if stocks should be managed together or separately (Montenegro et al., 2009)

For all three case studies, the proposed model improved on the single-stock models as measured by AIC. In general, model $\mathcal{M}_3$ performed well for all stocks, although smaller models were preferred for two of the cases. The model allows a flexible correlation structure to be estimated with relatively few parameters. By modelling partial correlations with a band structure between stocks, we obtain very general correlation structures, even within the individual stocks. With this model, correlations within each stock

are interpreted as a result of a connection between the stocks. Hence, the model only changes the single-stock model as an effect of the connection between stocks.

In the simulation study we showed that the true model can be identified by AIC, although with a risk of choosing a model including too many partial correlations. It is well-known that AIC can favour overfitting, when the true model has a finite parameter space, and is among the candidate models (Burnham and Anderson, 2003). However, for stock assessment models the true data generating system is too complex to be described by a simple model; observed data will depend on, amongst other things, the hydrological systems, individual fish behaviour, individual fisherman behaviour, and a wide array of data accumulation procedures. Any practically applicable candidate model will require strong simplifications; therefore, the AIC does not necessarily select the true model in the case studies, but simply the model that has the best fit to the data, while penalizing for the additional parameters used.

While the selected model provides the best fit to data, interpretation of individual correlation parameters must be done with caution. Estimated correlations in the abundance process will be a result of several effects such as environmental conditions, predation and migration, which can not be separated in the presented model. Further, if the correlation between, for example, observations are not adequately modelled, the state-space model can somewhat compensate to increase the model fit using the correlations in the abundance process (Thygesen et al., 2017). Finally, the typically short time series available for stock assessment models can result in imprecise individual correlation estimates, which clutters the overall pattern seen.

Ignoring correlations in data results in unreliable confidence intervals for model output. In a simulation study, we investigated the confidence intervals of the estimated fishing mortality; both when correlations were present and absent. We found that when severe correlations are present in the data, (i.e. simulating from models $\mathcal{M}_3$ to $\mathcal{M}_5$), ignoring the correlations results in confidence intervals that are too narrow. When no correlations were estimated, Hessian based 95% confidence intervals only had a coverage of 89.7%, 88.5%, and 91.9%, respectively,, when simulating from models $\mathcal{M}_3$ to $\mathcal{M}_5$, while the more flexible correlation structures provided confidence intervals close to the correct coverage.

Accurate assessments of model output uncertainty is important for sustainable fisheries management (Chaput, 2004; Cadrin and Dickey-Collas, 2014). Underestimating the uncertainty of model output used for management will increase the probability of overfishing (Punt et al., 2012). Hence, if correlations in data are ignored, a fishery perceived as sustainable may in fact have a non-sustainable risk of depletion.

These results will allow managers and policymakers to make informed decisions based on a more accurate assessment of model output uncertainty. As a simple extension of single-species assessments, the model output can be evaluated per stock, while giving an accurate evaluation of uncertainty. Likewise, connecting stocks may increase the accuracy of forecasts. Together with the estimated correlations, these improvements can provide guidance on managing single stocks and the potential influence on surrounding stocks. The procedure outlined is directly applicable without additional data sources.

Acknowledgements

The authors thank handling editor Dr. Shijie Zhou and two anonymous reviewers for their valuable comments to improve this manuscript. Funding for this study was provided by a grant from the European Fisheries and Maritime Fund (33113-B-15-003, Ministry of Environment and Food in Denmark).

References

Ainsworth, C. & Walters, C. (2015). Ten common mistakes made in ecopath with ecosim modelling. *Ecological Modelling*, *308*, 14–17. doi:10.1016/j.ecolmodel.2015.03.019

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. doi:10.1109/tac.1974.1100705

Albertsen, C. M., Nielsen, A., & Thygesen, U. H. (2017). Choosing the observational likelihood in state-space stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, *74*(5), 779–789. doi:10.1139/cjfas-2015-0532

Begley, J. & Howel, D. (2004). *An overview of gadget, the globally applicable area-disaggregated general ecosystem toolbox*. ICES document CM/FF:13.

Burnham, K. & Anderson, D. (2003). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer New York.

Cadrin, S. X. & Dickey-Collas, M. (2014). Stock assessment methods for sustainable fisheries. *ICES Journal of Marine Science*, *72*(1), 1–6. doi:10.1093/icesjms/fsu228

Chaput, G. (2004). Considerations for using spawner reference levels for managing single- and mixed-stock fisheries of atlantic salmon. *ICES Journal of Marine Science*, *61*(8), 1379–1388. doi:10.1016/j.icesjms.2004.08.015

Chen, Z., Xu, S., Qiu, Y., Lin, Z., & Jia, X. (2009). Modeling the effects of fishery management and marine protected areas on the beibu gulf using spatial ecosystem simulation. *Fisheries Research*, *100*(3), 222–229. doi:10.1016/j.fishres.2009.08.001

Christensen, V. & Pauly, D. (1992). ECOPATH II — a software for balancing steady-state ecosystem models and calculating network characteristics. *Ecological Modelling*, *61*(3-4), 169–185. doi:10.1016/0304-3800(92)90016-8

Christensen, V. & Walters, C. J. (2004). Ecopath with ecosim: methods, capabilities and limitations. *Ecological Modelling*, *172*(2-4), 109–139. doi:10.1016/j.ecolmodel.2003.09.003

Collie, J. S., Botsford, L. W., Hastings, A., Kaplan, I. C., Largier, J. L., Livingston, P. A., … Werner, F. E. (2016). Ecosystem models for fisheries management: finding the sweet spot. *Fish and Fisheries*, *17*(1), 101–125. doi:10.1111/faf.12093

Daan, N. (1987). Multispecies versus single-species assessment of north sea fish stocks. *Canadian Journal of Fisheries and Aquatic Sciences*, *44*(S2), s360–s370. doi:10.1139/f87-337

Eero, M., Hemmer-Hansen, J., & Hussy, K. (2014). Implications of stock recovery for a neighbouring management unit: experience from the baltic cod. *ICES Journal of Marine Science*, *71*(6), 1458–1466. doi:10.1093/icesjms/fsu060

Frøysa, K. G., Bogstad, B., & Skagen, D. W. (2002). Fleksibest—an age–length structured fish stock assessment model. *Fisheries Research*, *55*(1-3), 87–101. doi:10.1016/s0165-7836(01)00307-1

Fulton, E. A., Smith, A. D. M., & Punt, A. E. (2004). *Ecological indicators of the ecosystem effects of fishing: final report. report no. r99/1546*. Australian Fisheries Management Authority, Canberra.

Gislason, H. & Helgason, T. (1985). Species interaction in assessment of fish stocks with special application to the north sea. *Dana*, *5*, 1–44.

Hannesson, S., Jakobsdottir, A., Begley, J., Taylor, L., & Stefansson, G. (2008). On the use of tagging data in statistical multispecies multi-area models of marine populations. *ICES Journal of Marine Science*, *65*(9), 1762–1772. doi:10.1093/icesjms/fsn132

Hollowed, A. (2000a). Are multispecies models an improvement on single-species models for measuring fishing impacts on marine ecosystems? *ICES Journal of Marine Science*, *57*(3), 707–719. doi:10.1006/jmsc.2000.0734

Hollowed, A. (2000b). Including predation mortality in stock assessments: a case study for gulf of alaska walleye pollock. *ICES Journal of Marine Science*, *57*(2), 279–293. doi:10.1006/jmsc.1999.0637

REFERENCES

Holsman, K. K., Ianelli, J., Aydin, K., Punt, A. E., & Moffitt, E. A. (2016). A comparison of fisheries biological reference points estimated from temperature-specific multi-species and single-species climate-enhanced stock assessment models. *Deep Sea Research Part II: Topical Studies in Oceanography*, *134*, 360–378. doi:10.1016/j.dsr2.2015.08.001

Hurtado-Ferro, F., Szuwalski, C. S., Valero, J. L., Anderson, S. C., Cunningham, C. J., Johnson, K. F., … Punt, A. E. (2015). Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. *ICES Journal of Marine Science*, *72*(1), 99–110. doi:10.1093/icesjms/fsu198

ICES. (2014). *Report of the baltic fisheries assessment working group (WGBFAS), 3-10 april 2014, ICES HQ, Copenhagen, Denmark.* ICES CM 2014/ACOM:10.

ICES. (2015a). *Report of the baltic fisheries assessment working group (WGBFAS), 14-21 april 2014, ICES HQ, Copenhagen, Denmark.* ICES CM 2015/ACOM:10.

ICES. (2015b). *Report of the herring assessment working group for the area south of 62°N (HAWG), 10-19 march 2015, ICES HQ, Copenhagen, Denmark.* ICES CM 2015/ACOM:06.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. (2016). Tmb: automatic differentiation and laplace approximation. *Journal of Statistical Software*, *70*(1), 1–21. doi:10.18637/jss.v070.i05

Lewy, P. & Vinther, M. (2004). *A stochastic age-length-structured multispecies model applied to north sea stocks*. ICES document CM/FF:20.

Link, J. S., Fulton, E. A., & Gamble, R. J. (2010). The northeast US application of ATLANTIS: a full system model exploring marine ecosystem dynamics in a living marine resource management context. *Progress in Oceanography*, *87*(1-4), 214–234. doi:10.1016/j.pocean.2010.09.020

Methot, R. D. & Wetzel, C. R. (2013). Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research*, *142*, 86–99. doi:10.1016/j.fishres.2012.10.012

Montenegro, C., Maunder, M. N., & Zilleruelo, M. (2009). Improving management advice through spatially explicit models and sharing information. *Fisheries Research*, *100*(3), 191–199. doi:10.1016/j.fishres.2009.07.006

Nielsen, A. & Berg, C. (2014). Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research*, *158*, 96–101. doi:10.1016/j.fishres.2014.01.014

Punt, A. E., Siddeek, M. S. M., Garber-Yonts, B., Dalton, M., Rugolo, L., Stram, D., … Zheng, J. (2012). Evaluating the impact of buffers to account for scientific uncertainty when setting TACs: application to red king crab in bristol bay, alaska. *ICES Journal of Marine Science*, *69*(4), 624–634. doi:10.1093/icesjms/fss047

Quinn, T. J., Deriso, R. B., & Neal, P. R. (1990). Migratory catch-age analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, *47*(12), 2315–2327. doi:10.1139/f90-258

Richards, R. A. & Jacobson, L. D. (2016). A simple predation pressure index for modeling changes in natural mortality: application to gulf of maine northern shrimp stock assessment. *Fisheries Research*, *179*, 224–236. doi:10.1016/j.fishres.2016.03.003

Senina, I., Sibert, J., & Lehodey, P. (2008). Parameter estimation for basin-scale ecosystem-linked population models of large pelagic predators: application to skipjack tuna. *Progress in Oceanography*, *78*(4), 319–335. doi:10.1016/j.pocean.2008.06.003

Tarnecki, J. H., Wallace, A. A., Simons, J. D., & Ainsworth, C. H. (2016). Progression of a gulf of mexico food web supporting atlantis ecosystem model development. *Fisheries Research*, *179*, 237–250. doi:10.1016/j.fishres.2016.02.023

Thygesen, U. H., Albertsen, C. M., Berg, C. W., Kristensen, K., & Nielsen, A. (2017). Validation of ecological state space models using the laplace approximation. *Environmental and Ecological Statistics*, *24*(2), 317–339. doi:10.1007/s10651-017-0372-4

Yodzis, P. (1998). Local trophodynamics and the interaction of marine mammals and fisheries in the benguela ecosystem. *Journal of Animal Ecology*, *67*(4), 635–658. doi:10.1046/j.1365-2656.1998.00224.x

II.A Unstructured covariance matrix

Let L be a lower triangular of parameters with diagonal 1, and normalized such that $\langle L_{i\cdot}, L_{i\cdot}\rangle = 1$ for all i. Then $\Sigma = LL^T$ is an arbitrarily scaled covariance matrix. Off-diagonal elements of the covariance matrix can be forced to be zero (e.g. https://github.com/admb-project/examples/tree/master/admb-tricks/parameterization/covariance-matrices; accessed: 06-03-2017) by noting that

$$\Sigma_{jk} = \langle L_{j\cdot}, L_{k\cdot}\rangle.$$

Hence, by defining $\tilde{L}$ such that $\tilde{L}_{j\cdot} = L_{j\cdot} - \langle L_{j\cdot}, L_{k\cdot}\rangle L_{k\cdot}$ and $\tilde{L}_{i\cdot} = L_{i\cdot}$ for $i \neq j$, then for the covariance matrix $\tilde{\Sigma} = \tilde{L}\tilde{L}^T$,

$$\begin{aligned}
\tilde{\Sigma}_{jk} &= \langle \tilde{L}_{j\cdot}, \tilde{L}_{k\cdot}\rangle \\
&= \langle L_{j\cdot} - \langle L_{j\cdot}, L_{k\cdot}\rangle L_{k\cdot}, L_{k\cdot}\rangle \\
&= \langle L_{j\cdot}, L_{k\cdot}\rangle - \langle L_{j\cdot}, L_{k\cdot}\rangle\langle L_{k\cdot}, L_{k\cdot}\rangle \\
&= \langle L_{j\cdot}, L_{k\cdot}\rangle - \langle L_{j\cdot}, L_{k\cdot}\rangle \\
&= 0
\end{aligned}$$

While this method allows us to impose constraints on the covariance matrix, it is, in practice, difficult to pre-determine which parameters will be redundant. Instead, the gradient and hessian of the resulting likelihood function can be examined to determine parameters that do not affect the likelihood value. These parameters must be fixed in optimization.

II.B Simulation study details

II.B.1 Stock structure

We simulated 100 years from the model and discarded the first half of the time series. Two stocks were simulated with six age groups each. The last age group was considered a plus group. Natural mortality was fixed at 0.2. The initial log-population size was set to 16 for the first age group and was 0.5 less for each subsequent age group.

II.B.2 Fishing mortality

log-fishing mortality for stock s, age a, and year y was simulated by

$$\log F_{s,a,y} = \cos\left(\frac{y}{10} + 10 \cdot 1_{\{2\}}(s)\right) - 1 + \epsilon_{s,a,y}$$

where

$$1_{\{2\}}(s) = \begin{cases} 1 & \text{if } s = 2 \\ 0 & \text{if } s \neq 2 \end{cases},$$

and $\epsilon_{s,a,y} \sim N(0, 0.01^2)$.

II.B.3 log-Population process

The log-population given fishing mortality was simulated from the models $\mathcal{M}_0$ to $\mathcal{M}_5$ above. The off-diagonal non-zero entries, $P_{i,j}$, of the precision, P, matrix for $\mathcal{M}_\alpha$ was constructed as $-(\alpha+1)/\left(\alpha-|a_i-a_j|+1\right)^2$, where a_i is the age corresponding to the ith index of the process. This construction ensures non-negligible partial correlations in the elements not included in the smaller models. Diagonal elements, $P_{i,i}$, were set to $\sum_{j\neq i} P_{i,j} + \exp(-\alpha)$ to ensure the precision matrix was positive-definite. The corresponding covariance matrix was scaled to have diagonal elements 0.1^2.

II.B.4 Observations

log-observations given fishing mortality and log-population were simulated from a multivariate normal with AR(1) covariance structure. The AR(1) parameter was set to 0. For commercial data, the variances for all ages were 1. A first quarter survey was simulated from a multivariate normal with AR(1) covariance structure with parameter 0 and variances $\exp(-4)$. The catchability was set to $\exp(-7)$.

II.B.5 Re-estimation

Re-estimation was performed assuming an AR(1) structure for the correlations in the log-F random walk, random walk recruitment, and univariate log-normal observations.

II.C Confidence interval for correlation parameters

Simulation based confidence intervals for correlation parameters were constructed through the profile likelihood. Splitting the parameter vector θ into correlation parameters θ_c and other parameters θ_o, we simulated proposal parameters $\theta_c^{(i)}$, $i = 1, \ldots, N$, from the hessian based Gaussian approximation of the distribution of the estimator. The proposal was accepted if the hypothesis $\theta_c = \theta_c^{(i)}$ could not be rejected by a likelihood ratio test with significance level α. Each accepted simulated correlation parameter vector was transformed to a partial correlation and a correlation matrix. For each matrix element, minimum and maximum values were used as $(1-\alpha)\cdot 100$ % confidence interval. We simulated $N = 1000$ proposal parameter vectors for each case study, and obtained acceptance rates of 0.80, 0.41, and 0.63 for Baltic Herring, Mix and Cod, respectively. For Baltic Mix and Cod, the variance of the proposal distribution was reduced to increase acceptance.

Paper III

Fast fitting of non-Gaussian state-space models to animal movement data via Template Model Builder

Albertsen, C. M.[1*], Whoriskey, K.[2], Yurkowski, D.[3], Nielsen, A.[1], and Flemming, J. M.[2]

[1] **National Institute of Aquatic Resources, Technical University of Denmark, Charlottenlund, Denmark**

[2] **Department of Mathematics and Statistics, Dalhousie University, Halifax, Canada**

[3] **Great Lakes Institute for Environmental Research, University of Windsor, Ontario, Canada**

[*] **Corresponding author. Email: cmoe@aqua.dtu.dk**

Post-print. Published in Ecology

Full citation:

Albertsen, C. M., Whoriskey, K., Yurkowski, D., Nielsen, A., and Flemming, J. M. (2015). Fast fitting of non-gaussian state-space models to animal movement data via template model builder. *Ecology*, *96*(10), 2598–2604. doi:10.1890/14-2101.1

Abstract

State-space models (SSM) are often used for analyzing complex ecological processes that are not observed directly, such as marine animal movement. When outliers are present in the measurements, special care is needed in the analysis to obtain reliable location and process estimates. Here we recommend using the Laplace approximation combined with automatic differentiation (as implemented in the novel R-package: Template Model Builder (`TMB`)) for the fast fitting of continuous time multivariate non-Gaussian SSMs. Through Argos satellite tracking data, we demonstrate that the use of continuous time t-distributed measurement errors for error-prone data is more robust to outliers and improves the location estimation compared to using discretized time t-distributed errors (implemented with a Gibbs sampler) or using continuous time Gaussian errors (as with the Kalman filter). Using `TMB`, we are able to estimate additional parameters compared to previous methods, all without requiring a substantial increase in computational time. The model implementation is made available through the R-package argosTrack.

Keywords:

Argos system, Template Model Builder, outliers, state-space models, animal movement, Laplace approximation

III.1 Introduction

Studying individual animal movement is fundamental to determining the biological processes that affect the dynamic structure and function of ecosystems (Nathan et al., 2008). Technological advances in satellite telemetry, such as Argos archival data loggers, have allowed researchers to track animal movements and behavior in difficult to study environments like marine systems. However, non-Gaussian measurement errors are inherent to this technology (Jonsen et al., 2005), with the magnitude of each of these errors associated with one of seven location classes. Thus our understanding of movement, which is already complicated by multiple spatiotemporal scales and interactions between habitat, life history, and physiology, is further obscured by errors in the observation process (Patterson et al., 2008).

State-space models (SSMs) estimate the parameters and true locations underlying stochastic animal movement processes (Jonsen et al., 2003; Sibert et al., 2003; Jonsen et al., 2005, 2012). A SSM can explicitly account for error inherent to the Argos system while simultaneously fitting a movement model to data, thereby separating the stochasticity in the movement from the stochasticity in the data collection (Jonsen et al., 2005; Johnson et al., 2008; McClintock et al., 2012). Extensive research and development in the past decade has produced widely applicable SSMs that can answer a variety of biological questions. SSMs are now capable of accounting for sources of movement variability related to environmental covariates (Jonsen et al., 2003; Bestley et al., 2012; McClintock et al., 2012), behavior (Jonsen et al., 2005; Flemming et al., 2010; Bestley et al., 2012; McClintock et al., 2012), bias or drift (Johnson et al., 2008; McClintock et al., 2012) and can also model multiple tracks simultaneously (Jonsen et al., 2006; Eckert et al., 2008; Bestley et al., 2012). The increasing popularity of SSMs is evident in the literature, and is no doubt partially driven by the availability of R-packages for fitting SSMs like `bsam` (Jonsen et al., 2005; Jonsen, 2015) and `crawl` (Johnson et al., 2008; Johnson, 2013).

Given the importance of animal movement and the flexibility of SSMs for estimating true locations

while simultaneously relating them to various environmental (and other) covariates of interest, these models and their implementations still warrant further development. In fact, despite widespread application in movement ecology, SSMs remain at times both difficult to understand (Tremblay et al., 2009) and to implement (Patterson et al., 2008). Current limitations arise from underlying difficulties of likelihood computation and maximization for non-Gaussian and non-linear models. Although Johnson et al. (2008) utilize the computationally efficient Kalman filter (applicable only to linear Gaussian SSM formulations) to evaluate their model likelihood, both Jonsen et al. (2005) and McClintock et al. (2012) rely on computationally expensive and comparatively slow Markov Chain Monte Carlo techniques to estimate locations and model parameters, while Pedersen et al. (2011) discretize the state-space and use a Hidden Markov Model. Features which add outlier robustness to models, specifically heavy-tailed error distributions, increase model complexity and computation time. For example, Johnson et al. (2008) are able to use the fast Kalman filter for evaluating their continuous-time SSM because they model the measurement errors with Gaussian distributions instead of the more robust t-distributions of Jonsen et al. (2005); while the former model takes only minutes to fit, the latter can take hours for a standard Argos track (an average of 21 observations per day for five months).

Template Model Builder (`TMB`; developed by Kasper Kristensen and freely available at www.tmb-project.org) is a recently developed R-package that is functionally similar to AD Model Builder (ADMB; Fournier et al., 2012): it uses a combination of reverse mode Automatic Differentiation and the Laplace approximation for high dimension integrals in order to perform parameter estimation. Such a combination allows for the efficient fitting of complex (nonlinear, non-Gaussian, and hierarchical) models to large multivariate datasets (Fournier et al., 2012).

Here we adapt the SSM formulated in Johnson et al. (2008) for fitting within the `TMB` framework. We use two versions: the original model of Johnson et al. (2008), and a modified model assuming t-distributed measurement errors as in Jonsen et al. (2005). Hereafter we interchangeably describe an SSM with t-distributed measurement errors as non-Gaussian or robust. We fit these models to simulated data as well as to real data collected from both a subadult and an adult ringed seal (*Pusa hispida*). Ecological differences in movement and behavior in relation to ontogeny occur in ringed seals (Harwood et al., 2015). Thus two age groups (i.e. adult and subadult) were chosen to illustrate model versatility with movement data reflecting different behavioral patterns where subadults often have periods of strong directional persistence while adults typically remain resident. Our aim is twofold. First, we demonstrate through simulation and the calculation of mean squared errors that within the `TMB` framework, SSMs with t-distributed errors can be fitted quickly, and second, we show the efficacy of using the Laplace approximation for fitting SSMs to real Argos telemetry data, specifically as implemented in `TMB`. We make all code readily available through the R-package `argosTrack` (www.github.com/calbertsen/argosTrack; also available in the online Supplement) to both facilitate the implementation of SSMs in `TMB` and to provide a "toolbox" that researchers can utilize to fit these SSMs to marine telemetry data.

III.2 Methods

III.2.1 Model Formulation

III.2.1 Velocity process:

We implement the continuous-time correlated random walk model described in Johnson et al. (2008). For each coordinate $c = 1, 2$ (1: latitude, 2: longitude) the instantaneous velocity of the animal $\nu_{c,i}$ at time i is described by the Ornstein-Uhlenbeck process:

$$\nu_{c,i} - \gamma_c = e^{-\beta_c \Delta_i}[\nu_{c,i-1} - \gamma_c] + \eta_{c,i}. \tag{III.1}$$

In this process, γ_c describes the long term mean velocity (can be interpreted as drift), while $\beta_c > 0$ determines how long a period with high or low velocities will last (i.e. the autocorrelation in the process). Further, Δ_i denotes the time interval between $\nu_{c,i-1}$ and $\nu_{c,i}$, and $\eta_{c,i}$ is a random error term which follows a Gaussian distribution with mean zero and variance $V(\eta_{c,i}) = \sigma_c^2(1-e^{-2\beta_c\Delta_i})/(2\beta_c)$, where σ_c^2 is a scale parameter controlling the variability in velocity. With respect to animal movement, γ_c can be interpreted as the directional preference. That is, if $\gamma_c = 0$, the animal has no directional preference, while a large γ_c corresponds to fast movement in a specific direction as determined by the sign of γ_c. Likewise β_c describes how much the animal will stray from the path. If β_c is large, the animal will stay close to the path.

III.2.2 Location process:

By integrating over the velocity process one obtains the location process:

$$\mu_{c,i} = \mu_{c,i-1} + \nu_{c,i-1}\left(1 - e^{-\beta_c\Delta_i}\right)/\beta_c + \zeta_{c,i}. \tag{III.2}$$

Here $\mu_{c,i}$ denotes the c coordinate of the location at time i and $\zeta_{c,i}$ follows a Gaussian distribution with zero mean and variance

$$V(\zeta_{c,i}) = \frac{\sigma_c^2}{\beta_c^2}\left(\Delta_i - 2\left(1 - e^{-\beta_c\Delta_i}\right)/\beta_c + \left(1 - e^{-2\beta_c\Delta_i}\right)/(2\beta_c)\right).$$

The joint distribution of $\eta_{c,i}$ and $\zeta_{c,i}$ is a bivariate Gaussian distribution with covariance

$$Cov(\eta_{c,i}, \zeta_{c,i}) = \frac{\sigma_c^2}{2\beta_c^2}\left(1 - 2e^{-\beta_c\Delta_i} + e^{-2\beta_c\Delta_i}\right).$$

These processes combine to form the unobserved states in our SSM.

III.2.3 Measurement equation:

As in Johnson et al. (2008), we consider the measurement equation

$$y_{c,i} = \mu_{c,i} + \epsilon_{c,i}. \tag{III.3}$$

where $y_{c,i}$ denotes the c coordinate of the observed location of an animal and $\epsilon_{c,i}$ is a measurement error term. We take the joint distribution of $\epsilon_{1,i}$ and $\epsilon_{2,i}$ to be modelled as a bivariate t-distribution with $d_{a(i)} > 3$ degrees of freedom (to ensure the distribution has both a mean, variance, and skewness), mean vector of zeros, and scale matrix H_i defined as $H_i = diag\left(\tau^2_{1,a(i)}, \tau^2_{2,a(i)}\right)$ where $a(i)$ is a factor determining the Argos location quality class of observation i, and $\tau^2_{1,a(i)}$ and $\tau^2_{2,a(i)}$ are scale parameters describing the variance of the measurement error for the corresponding quality class. Hence the measurement variance, through the scale and degrees of freedom parameters, for both coordinate directions is estimated for each location class within the model. This model formulation, denoted `TMB-t`, will be contrasted with the model denoted `TMB-g`, which instead takes the joint distribution of $\epsilon_{1,i}$ and $\epsilon_{2,i}$ to be bivariate Gaussian with mean vector of zeros and covariance H_i. The former can be considered more robust in that t-distributions are heavier tailed than Gaussian distributions which makes outlying observations (like those typical of Argos data) more likely under a t-distribution such that they have less impact on subsequent location and parameter estimation.

III.3 Estimation

In SSMs, maximum likelihood inference about parameters should be performed using the marginal distribution of the observations. This requires integration over the unobserved states in the joint distribution of the observed and unobserved (state) processes, which in our case correspond to the Argos observations and the true yet unobserved locations (and velocities) of the tagged animal, respectively. In the Kalman filter, the simplicity of the linear Gaussian SSM is exploited to reduce the integration to a recursive algorithm which is fast but inflexible. Inference with MCMC methods is performed by sampling from the posterior likelihood of the parameters and the unobserved states given the observed data, which makes MCMC methods flexible but slow. In `TMB` the integration is performed by using the Laplace approximation which also produces optimal state predictions and corresponding confidence intervals (Kass and Steffey, 1989). The deterministic approximation is fast compared to simulation methods while flexible compared to the Kalman filter. Similar to, for example, the Kalman smoother, the state predictions are smoothed with the Laplace approximation in the sense that they depend on the whole data set. For both consistency, simplicity, and easy comparability with the results of fitting similar models with the R-packages `crawl` based on the Kalman filter and `bsam` using MCMC with a Gibbs sampler, we impose on our model the restrictions $\beta_1 = \beta_2$ and $\gamma_1 = \gamma_2 = 0$. This implies that all simulated tracks experienced zero drift, and that autocorrelation was similar across coordinate axes. Note that the `crawl` and `TMB-g` implementations are not identical as the latter estimates additional variance parameters (associated with the Argos locations classes). The `bsam` model differs from the `TMB-t` implementation in three ways: the `bsam` model is formulated in discrete time, `bsam` models a turning angle instead of the velocity, and the `TMB-t` implementation estimates all parameters in the measurement equation. In `bsam` the ratio of the scale parameters and the degrees of freedom is fixed based on the results of Vincent et al. (2002), which reduced the number of estimated parameters in the measurement equation to one.

III.4 Simulation Study

We carried out an extensive simulation study to assess the gains in accuracy of the location estimates by fitting the Johnson et al. (2008) movement model with t-distributed measurement errors. We simulated data from our SSM as described by Equations III.1, III.2, and III.3. Specifically, the unobserved locations and velocities were simulated from Equations III.1 and III.2 and measurement errors were bootstrapped from the residuals of the `TMB-t` fit to mimic true heavy-tailed Argos measurement errors, without assuming a specific distribution (see Appendix A for details).

To measure the effect of using a SSM with t-distributed measurement errors we examined results of the following three `TMB` implementations: `TMB-g` - Gaussian measurement errors assumed; `TMB-gr` - Gaussian measurement errors assumed but with a swim speed filter applied (Freitas, 2012); and `TMB-t` - t-distributed measurement errors assumed. Further, the R-packages `crawl` (version 1.4-1; Johnson, 2013) and `bsam` (version 0.43.1; Jonsen, 2015) were applied to the simulated data. Since `bsam` only produces position estimates at regularized time-steps, these were interpolated by assuming linear movement between location estimates. This is the same way `bsam` handles observations at irregular time-steps (Jonsen et al., 2005). For each implementation the resulting estimated locations were compared to the true simulated locations by calculating the coordinate-wise root mean squared error (RMSE) along the track.

Using continuous time Gaussian measurement errors without pre-processing (the `crawl` and `TMB-g` implementations) resulted in an average RMSE of 0.012 in latitude and 0.031 in longitude (Figure III.1 and Appendix A), which corresponds to approximately 1400 meters in the North-South direction and 900 meters in the East-West direction. The RMSE was generally lowered by estimating all parameters in the model (`TMB-g` and `TMB-t`). When the data were pre-processed with a swim speed filter (`TMB-gr`) the RMSE decreased by an average of 22.3% (min: -2.1%; max: 46.0%) in latitude and 23.3% (min: -20.5%; max: 38.4%) in longitude. The negative decrease in RMSE indicates that in some cases the RMSE was higher for `TMB-gr` than for `TMB-g`. The lowest RMSE was achieved by assuming t-distributed measurement errors with the `TMB-t` implementation. Compared to `TMB-g`, `TMB-t` reduced the RMSE by 42.1% on average (min: 26.5%; max: 64.6%) in latitude and 43.6% (min: 34.7%; max: 54.3%) in longitude. `TMB-t` additionally reduced the RMSE compared to `bsam` by 10.4% (min: -9.7%; max: 24.7%) in latitude and 10.0% (min: 0.2%; max: 19.3%) in longitude.

III.5 Application

To further illustrate the utility of our approach we considered two previously unpublished movement tracks from a subadult and adult male ringed seal (*Pusa hispida*), each corresponding to a different life history stage. The datasets of the subadult and adult seals included 3,586 and 2,698 observations over a period of 168 and 219 days, respectively, of which 49.53% and 52.71% (respectively) were characterized by the two worst Argos quality classes.

We modelled the ringed seal data with the four full data approaches above: `crawl`, `bsam`, `TMB-g`, and `TMB-t` (see Appendix B for further details) to investigate the effect of using a continuous time model as opposed to a discrete time model, the effect of estimating all parameters in the model, and the effect of using t-distributed errors as opposed to Gaussian measurement errors. Hence pre-processing of data (`TMB-gr`) is not considered in this section. The location estimates from using `crawl` and `TMB-g` were similar (Figure III.2 and Appendix B), but `crawl` was 8x faster than `TMB-g` (Table III.1). Similar location estimates were also obtained from `bsam` and `TMB-t` (Figure III.2), but `TMB-t` was 60x faster than `bsam`,

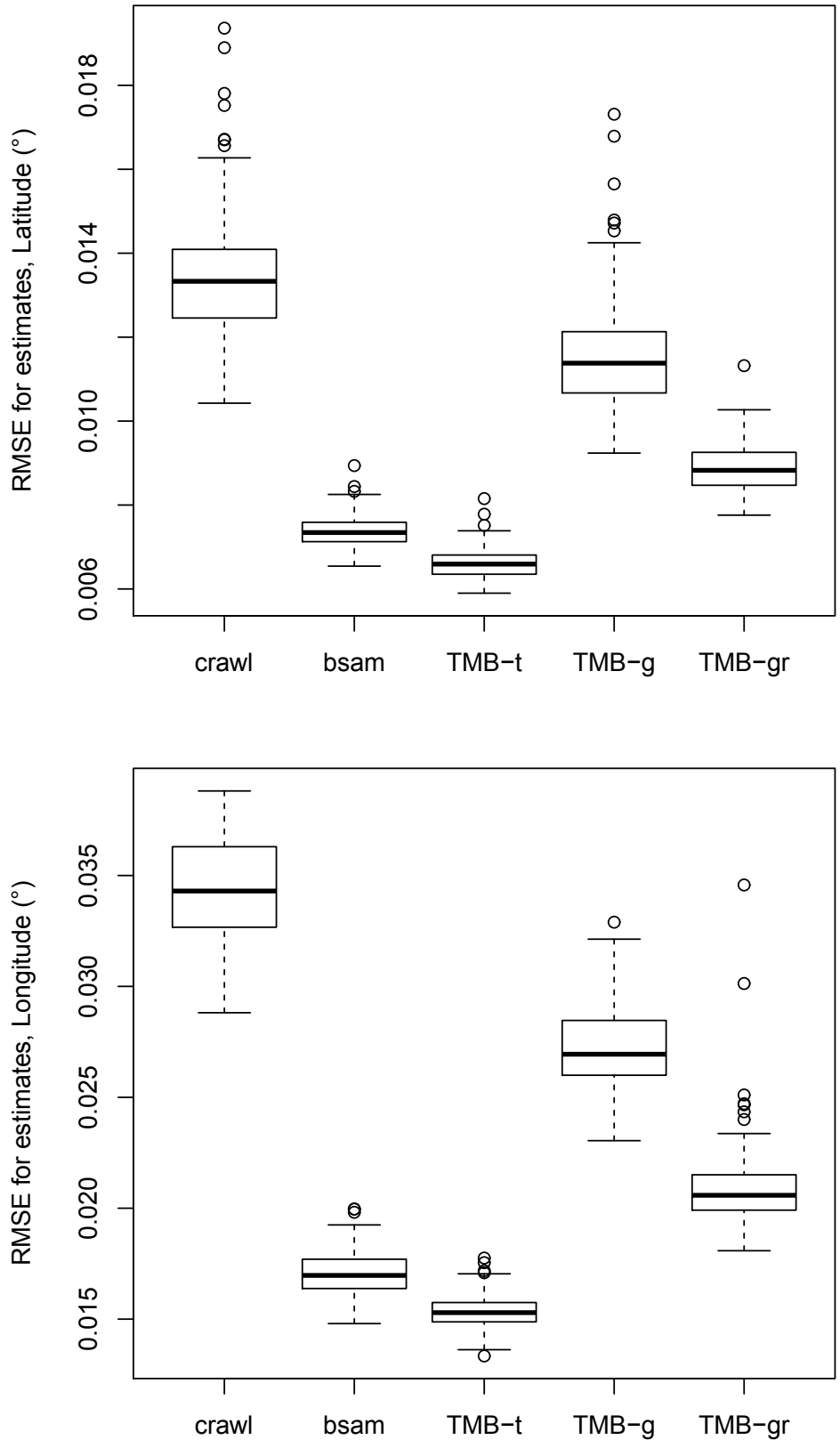


Figure III.1: Boxplot of the RMSE for `TMB-t`, `TMB-g`, and `TMB-gr` in latitude (A) and longitude (B). States were simulated parametrically and observations were bootstrapped using the residuals of the `TMB-t` model that was fitted to the subadult seal (so as to mimic true Argos errors)

Table III.1: Comparison of computation times in seconds across models applied to the subadult and adult seal tracks. All methods were implemented on a Dell Latitude E6410 laptop with 4GB RAM and an Intel Core i7 M640 processor at 2.8 Ghz. The operator system used was 64-bit Ubuntu 14.04 LTS.

		Computation time		Latent variables	
Method	Parameters	subadult	adult	subadult	adult
`crawl`	4	12.7	8.3	13680	10404
`bsam`	7	11109.8	11745.8	2698	3500
`TMB-g`	17	45.6	42.3	13680	10404
`TMB-t`	24	190.9	77.4	13680	10404

which reduced the estimation time from hours to three minutes.

Overall the robust implementations (`TMB-t` and `bsam`) and those with Gaussian measurement error assumptions gave similar location estimates, but with substantial differences near outliers. These differences are caused by the decreased sensitivity of the t-distribution to extreme observations compared to using the Gaussian distribution. This also lead to the low RMSE for `TMB-t` (Figure III.1). For both datasets, the `crawl` implementation was the fastest, while `TMB-g` took approximately 4 times as long. The additional complexity of using t-distributed measurement errors in `TMB-t` increased the execution time by 100% as compared to `TMB-g`, while `bsam` used a total of 3 hours.

III.6 Discussion

In this paper we propose and demonstrate the use of `TMB` for fitting non-Gaussian SSMs to animal movement data, specifically that collected by the Argos satellite-based system. Our simulation results and application to two behaviorally distinct movement tracks make clear that the Laplace approximation as implemented in `TMB` is highly suitable for the parameter and state estimation required by complex SSMs. To illustrate how `TMB` can be used to fit non-Gaussian SSMs, we make available the R-package `argosTrack` (Supplement).

Using `TMB` we were able to improve location estimates compared to previous methods by estimating all model parameters from the data and by utilizing heavy-tailed t-distributions within the same model. In the simulation study, the location estimates were improved by estimating all measurement error parameters (`TMB-g` compared to `crawl`; Figure III.2 and Fig. A2), and further improvements were made by using t-distributed errors to account for outliers (`TMB-t` compared to `TMB-g`). Using the robust continuous time model on the full data even outperformed that of pre-processing the data with a standard swim speed filter (as proposed by Freitas et al. (2008)). The latter did indeed provide improvements over a full data Gaussian approach. Finally, `TMB-t` improved state estimation compared to `bsam`, which may be caused by estimating all parameters within the model and by using a continuous time formulation compared to a discrete time model, as the latter discards information from the observations by interpolating between state estimates at equidistant times.

When applied to Argos data collected on ringed seals, `TMB-t` did not substantialy decrease running time compared to `crawl` or `TMB-g`, but it did reduce computation time from the hours of `bsam` to only minutes. The fastest model we fitted during the application was `crawl`, which was a predictable result because the Kalman filter is an optimal estimation method for linear Gaussian SSMs (in that there is an analytical solution for these cases). Considering the simulation study together with the application,

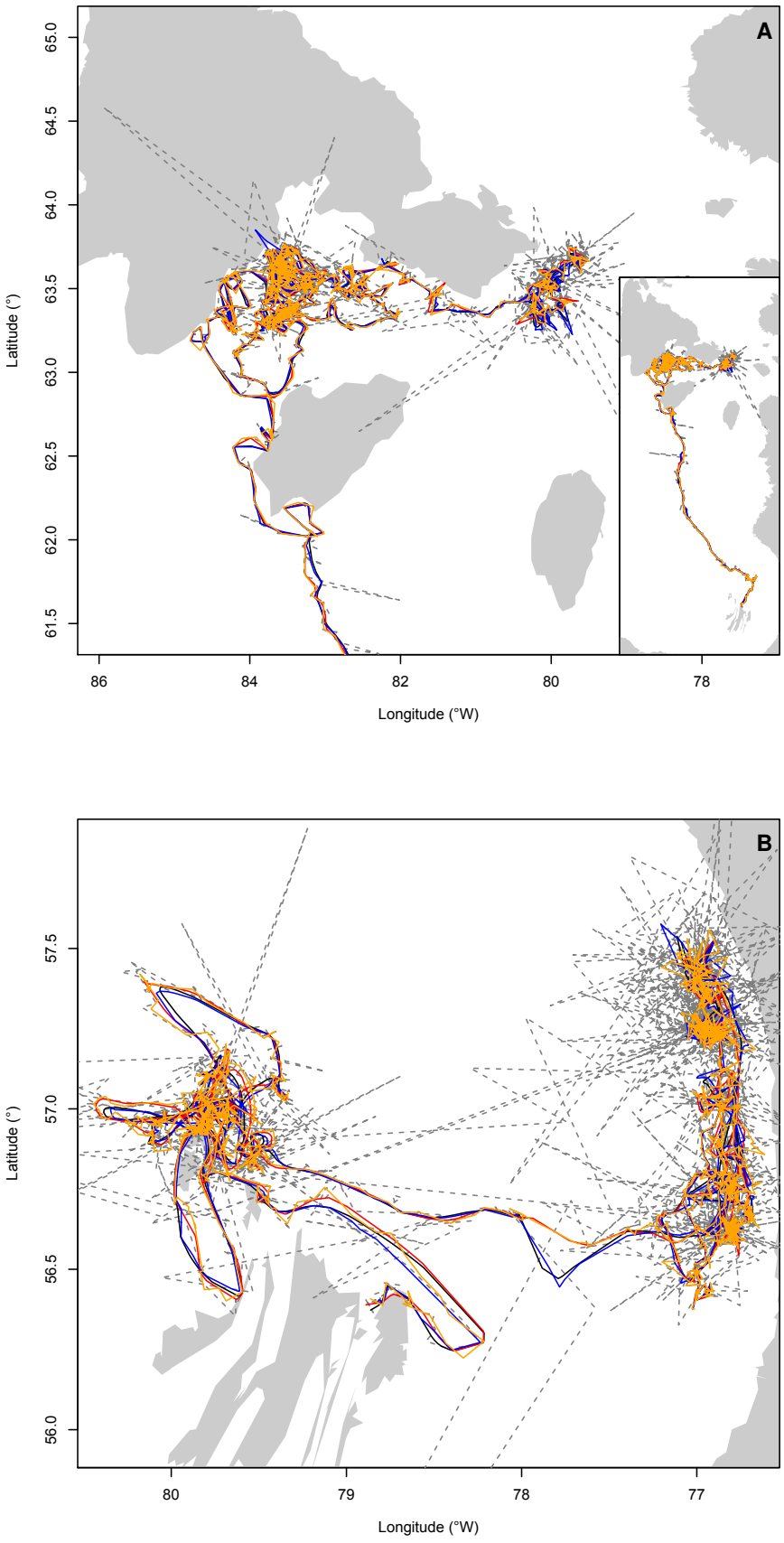


Figure III.2: Observed (grey, dashed) and estimated tracks from four models fit to Argos observations collected by a tagged subadult (A) and adult (B) ringed seal (*Pusa hispida*). The four models included `TMB-t` (red), `TMB-g` (black), `crawl` (blue), and `bsam` (orange).

the results are promising for using the Laplace approximation to estimate continuous time SSMs with t-distributed measurement errors for error-prone Argos tag data. The improved estimation achieved by using t-distributed errors agrees with previous results (Jonsen et al., 2005).

The power of TMB results from its flexibility allowing for the estimation of non-Gaussian models and additional parameters, all without substantially increasing computation time. For example, most SSM implementations to date have used existing estimates of Argos error distributions based on the 426 observations of Vincent et al. (2002; e.g. Johnson et al., 2008) or have used the Vincent et al. (2002) results to parametrize their error distributions (e.g. Jonsen et al., 2005). This is not necessary when fitting SSMs with TMB, where parameters describing the measurement error t-distributions are instead estimated within the SSM fitting procedure and using the entire animal track (which in our application included more than 2000 observations). Furthermore, TMB requires far fewer subjective choices (Bolker et al., 2013), which reduces estimation error introduced by imposing incorrect conditions on a SSM or its estimation procedure. For example, there are no priors, no pre-processing, no burn-in period, no time step, and no number of iterations for simulated annealing. Additionally one does not have to specify initial distributions for the velocity and location processes when using the Laplace approximation, unlike most filters and MCMC methods. Finally, the computational efficiency of the TMB framework allows us to more easily validate models through simulation or bootstrapping to, e.g., examine the validity of the standard errors of the estimated locations.

We have shown here that TMB is highly suitable for fitting SSMs, particularly non-Gaussian versions, to marine Argos telemetry data. However, we also note that the models we have fitted are readily extendable to data collected via Global Positioning System, which produces more accurate location estimates than those estimated by Argos, and additionally does not require location classes or associated error distributions. The results of this study are promising for improving the rapidity and robustness of future analyses of more complex models including covariates or multiple tracks for error-prone marine telemetry data.

Acknowledgements

This research has been supported by a National Sciences and Engineering Research Council Canada (NSERC) Discovery Grant to J. Mills Flemming, a Canadian Statistical Sciences Institute (CANSSI) Collaborative Team Project involving J. Mills Flemming and A. Nielsen, and an NSERC grant (NETGP 375118-08) to J. Mills Flemming via the Ocean Tracking Network Canada. K. Whoriskey has been supported by a NSERC Canada Graduate Scholarship and a Killam Predoctoral scholarship. D. Yurkowski has been supported by an Ontario Graduate Scholarship and the W. Garfield Weston Foundation. Handling was approved and conducted under a Fisheries and Oceans Animal Care Protocol (FWI-ACC-2010-006) and DFO Licence to Fish for Scientific Purposes (S-10/11-1006-NU).

Supplemental Material

Appendices A–C and the Supplement are available online: http://dx.doi.org/10.1890/14-2101.1.sm

Literature Cited

Bestley, S., Jonsen, I. D., Hindell, M. A., Guinet, C., & Charrassin, J.-B. (2012). Integrative modelling of animal movement: incorporating in situ habitat and behavioural information for a migratory marine predator. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1750), 2262–2262. doi:10.1098/rspb.2012.2262

Bolker, B. M., Gardner, B., Maunder, M., Berg, C. W., Brooks, M., Comita, L., … Zipkin, E. (2013). Strategies for fitting nonlinear ecological models in r, AD model builder, and BUGS. *Methods in Ecology and Evolution*, *4*(6), 501–512. doi:10.1111/2041-210x.12044

Eckert, S. A., Moore, J. E., Dunn, D. C., van Buiten, R. S., Eckert, K. L., & Halpin, P. N. (2008). Modeling loggerhead turtle movement in the mediterranean: importance of body size and oceanography. *Ecological Applications*, *18*(2), 290–308. doi:10.1890/06-2107.1

Flemming, J. M., Jonsen, I. D., Myers, R. A., & Field, C. A. (2010). Hierarchical state-space estimation of leatherback turtle navigation ability. *PLoS ONE*, *5*(12), e14245. doi:10.1371/journal.pone.0014245

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., … Sibert, J. (2012). AD model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, *27*(2), 233–249. doi:10.1080/10556788.2011.597854

Freitas, C., Lydersen, C., Fedak, M. A., & Kovacs, K. M. (2008). A simple new algorithm to filter marine mammal argos locations. *Marine Mammal Science*, *24*(2), 315–325. doi:10.1111/j.1748-7692.2007.00180.x

Freitas, C. (2012). *Argosfilter: argos locations filter*. R package version 0.63. Retrieved from https://CRAN.R-project.org/package=argosfilter

Harwood, L. A., Smith, T. G., Auld, J. C., Melling, H., & Yurkowski, D. J. (2015). Seasonal movements and diving of ringed seals, *Pusa hispida*, in the western canadian arctic, 1999–2001 and 2010–11. *ARCTIC*, *68*(2), 193. doi:10.14430/arctic4479

Johnson, D. S. (2013). *Crawl: fit continuous-time correlated random walk models to animal movement data*. R package version 1.4-1. Retrieved from https://CRAN.R-project.org/package=crawl

Johnson, D. S., London, J. M., Lea, M.-A., & Durban, J. W. (2008). Continuous–time correlated random walk model for animal telemetry data. *Ecology*, *89*(5), 1208–1215. doi:10.1890/07-1032.1

Jonsen, I. D. (2015). *Bsam: bayesian state-space models for animal movement*. with contributions from S. Luque, A. Winship and M.W. Pedersen. R package version 0.43.1.

Jonsen, I. D., Basson, M., Bestley, S., Bravington, M. V., Patterson, T. A., Pedersen, M. W., … Wotherspoon, S. J. (2012). State-space models for bio-loggers: a methodological road map. *Deep Sea Research Part II: Topical Studies in Oceanography*, *88-89*, 34–46. doi:10.1016/j.dsr2.2012.07.008

Jonsen, I. D., Flemming, J. M., & Myers, R. A. (2005). Robust state–space modeling of animal movement data. *Ecology*, *86*(11), 2874–2880. doi:10.1890/04-1852

Jonsen, I. D., Myers, R. A., & Flemming, J. M. (2003). Meta–analysis of animal movement using state–space models. *Ecology*, *84*(11), 3055–3063. doi:10.1890/02-0670

Jonsen, I. D., Myers, R. A., & James, M. C. (2006). Robust hierarchical state-space models reveal diel variation in travel rates of migrating leatherback turtles. *Journal of Animal Ecology*, *75*(5), 1046–1057. doi:10.1111/j.1365-2656.2006.01129.x

Kass, R. E. & Steffey, D. (1989). Approximate bayesian inference in conditionally independent hierarchical models (parametric empirical bayes models). *Journal of the American Statistical Association*, *84*(407), 717–726. doi:10.1080/01621459.1989.10478825

McClintock, B. T., King, R., Thomas, L., Matthiopoulos, J., McConnell, B. J., & Morales, J. M. (2012). A general discrete-time modeling framework for animal movement using multistate random walks. *Ecological Monographs*, *82*(3), 335–349. doi:10.1890/11-0326.1

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., & Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, *105*(49), 19052–19059. doi:10.1073/pnas.0800375105

Patterson, T., Thomas, L., Wilcox, C., Ovaskainen, O., & Matthiopoulos, J. (2008). State–space models of individual animal movement. *Trends in Ecology & Evolution*, *23*(2), 87–94. doi:10.1016/j.tree.2007.10.009

Pedersen, M. W., Patterson, T. A., Thygesen, U. H., & Madsen, H. (2011). Estimating animal behavior and residency from movement data. *Oikos*, *120*(9), 1281–1290. doi:10.1111/j.1600-0706.2011.19044.x

Sibert, J. R., Musyl, M. K., & Brill, R. W. (2003). Horizontal movements of bigeye tuna (thunnus obesus) near hawaii determined by kalman filter analysis of archival tagging data. *Fisheries Oceanography*, *12*(3), 141–151. doi:10.1046/j.1365-2419.2003.00228.x

Tremblay, Y., Robinson, P. W., & Costa, D. P. (2009). A parsimonious approach to modeling animal movement data. *PLoS ONE*, *4*(3), e4711. doi:10.1371/journal.pone.0004711

Vincent, C., Mcconnell, B. J., Ridoux, V., & Fedak, M. A. (2002). Assessment of argos location accuracy from satellite tags deployed on captive gray seals. *Marine Mammal Science*, *18*(1), 156–166. doi:10.1111/j.1748-7692.2002.tb01025.x

Paper IV

A Generalized First-Difference Correlated Random Walk for Animal Movement Data

Albertsen, C. M.*

National Institute of Aquatic Resources, Technical University of Denmark, Lyngby, Denmark

* **Corresponding author. Email: cmoe@aqua.dtu.dk**

Advance draft

Full citation:

Albertsen, C. M. (2017). A generalized first-difference correlated random walk for animal movement data. *Advance draft*. In Preparation

Abstract

Animal telemetry data are often analysed with discrete time movement models assuming rotation in the movement. These models are defined with equidistant distant time steps. However, telemetry data from marine animals are observed irregularly. To account for irregular data, a time-irregularised first-difference correlated random walk model with drift is introduced. The model generalizes the commonly used first-difference correlated random walk with regular time steps by allowing irregular time steps, including a drift term, and by allowing different autocorrelation in the two coordinates. The model is applied to data from a ringed seal collected through the Argos satellite system, and is compared to related movement models through simulations. Accounting for irregular data in the movement model results in accurate parameter estimates and reconstruction of movement paths. Measured by distance, the introduced model can provide more accurate movement paths than the regular time counterpart. Extracting accurate movement paths from uncertain telemetry data is important for evaluating space use patterns for marine animals, which in turn is crucial for management. Further, handling irregular data directly in the movement model allows efficient simultaneous analysis of several animals.

Keywords:

animal telemetry, movement ecology, correlated random walk, irregular data

IV.1 Introduction

Understanding animal movement behaviour, space use patterns and response to the environment relies on modelling animal telemetry data. So far, the collection of telemetry data has been hindered in marine environments. Satellite systems require animals to surface, storage tags require recapture, and acoustic networks require a dense array of receivers. Nonetheless, the scope and scales of movement studies are continuously increasing with rapid technological advances in tracking devices (Hussey et al., 2015).

In spite of technological advances leading to more accurate measurements and larger data sets, data from satellite tags have inherent measurement errors. Systems such as Fastloc GPS and Argos can only record data when the animal is above water, with an accuracy that depends on the surface time and diving behaviour (Costa et al., 2010). Typically, this results in data observed at irregular time steps with considerable uncertainty, which makes state-space models a valuable framework for analysing the data.

State-space models naturally separate the movement process from measurement errors. True animal locations are explicitly modelled separately from the observed measurements conditionally on the locations. Hence, the movement models developed can be combined with telemetry data from any source, for example, GPS, Argos, light based, or acoustic tags. However, inference in general state-space models can be challenging. Maximum likelihood estimation in state-space models requires integration of the joint distribution of locations and measurements over the locations to obtain the marginal distribution of the observations. Recent software such as AD Model Builder (Fournier et al., 2012) and Template Model Builder (`TMB`; Kristensen et al., 2016) utilizes the Laplace approximation to rapidly approximate the marginal distribution of observations in state-space models. While care is needed to ensure the approximation is appropriate, this framework is generally applicable. For instance, `TMB` has been used for both discrete time (Auger-Méthé et al., 2016, 2017) and continuous time (Albertsen et al., 2015) movement

models.

Although continuous time movement models can handle irregular data, discrete time models are often thought of as more intuitive (McClintock et al., 2014). Further, continuous time models can be difficult to construct or extend analytically, because they are often formulated through stochastic differential equations. Whereas continuous time models describe instantaneous change, discrete time models describe the change between two observations. In animal movement models, the change between observations are often modelled through a step length and an angle of movement (e.g., Morales et al., 2004; Jonsen et al., 2005; Gurarie et al., 2009; Tracey et al., 2010; McClintock et al., 2012; Michelot et al., 2016), which is considered a natural way to represent movement (Turchin, 1998). In discrete time models with regular time steps, irregular data can be handled by modifying the observational model to interpolate between states (Jonsen et al., 2005).

To facilitate the analysis of populations or meta-analysis of individuals, the model parameters for different individuals must be on the same scale. If the interpretation of parameter values depends on the chosen time steps between estimated locations (McClintock et al., 2014), the values must be corrected before comparison is possible. Likewise, estimating population parameters based on individuals either requires movement models that use the same time steps for all individuals (e.g. Jonsen, 2016), which can be sub-optimal, or movement models that handle irregular data directly.

The first-difference correlated random walk (DCRW; Jonsen et al., 2005) models the animal movement as a discrete time first order auto-regressive process on the difference between consecutive locations through a rotation matrix. A continuous time version of the DCRW without rotation is the continuous time correlated random walk (CTCRW; Johnson et al., 2008). The CTCRW models the velocity of an animal by an Ornstein-Uhlenbeck process: the continuous time counterpart of the first order auto-regressive process. Recently, an irregular time version of the DCRW was introduced without rotation including a time-varying parameter instead (Auger-Méthé et al., 2017).

In this paper, a generalization of the DCRW is introduced. The generalization of the DCRW is trifold: the model allows irregular time steps, a drift vector is included, and different auto-correlations in the two coordinates can be used. As a by-product, it is shown how parameters of the DCRW model depends on the selected time steps, and how to transform them to a common scale. The generalized DCRW is introduced as the discretization of a stochastic differential equation for the animal's velocity and location. Further, the close relation to a generalization of the CTCRW is shown. Through three simulation studies, the generalized DCRW is shown to perform well compared to the CTCRW and DCRW. Finally, The applicability and extendability of the model is illustrated through a real data set collected by the Argos system.

IV.2 Materials and Methods

IV.2.1 Movement model

The generalization of the DCRW movement model is constructed through a stochastic differential equation (SDE) for the velocity. From this SDE, the location process is the integrated velocity process. To obtain the time-irregular movement model, the SDE is solved analytically while the location process is a discrete time approximation.

IV.2.1 SDE

The bivariate SDE for the velocity, V_t, of an animal is

$$\begin{aligned} \mathrm{d}V_t &= -\begin{pmatrix} -\log\gamma_1 & \theta \\ -\theta & -\log\gamma_2 \end{pmatrix}(V_t-\mu)\mathrm{d}t + S\mathrm{d}B_t \\ &= -\Theta(V_t-\mu)\mathrm{d}t + S\mathrm{d}B_t \end{aligned} \tag{IV.1}$$

with initial condition $V_0 = v_0 \in \mathbb{R}^2$. In this model, γ_1 and γ_2 are autocorrelation parameters, μ is a vector of mean velocity parameters, and θ is the mean turning angle of the movement process. The matrix S is a lower triangular matrix with positive diagonal determining the covariance of the changes in velocity. Animals moving persistently in one direction will have high autocorrelation in the velocity and a mean turning angle close to zero, while animals foraging or in other ways displaying a tortuous movement will have a lower autocorrelation in the velocity and a mean turning angle different from zero (modulo 2π) (e.g. Jonsen et al., 2005; Whoriskey et al., 2017).

The solution to the SDE is a stochastic process with Gaussian increments (see e.g. Appendix IV.A). The mean of an increment is

$$E(V_t \mid V_s) = \exp(-\Theta(t-s))V_s + \left(I - e^{-\Theta(t-s)}\right)\mu, \quad s < t$$

where $\mu = (\mu_1, \mu_2)^T$, while the covariance is

$$\mathrm{vec}(Var\,(V_t \mid V_s)) = \mathrm{vec}(C) - e^{-\Theta(t-s)} \otimes e^{-\Theta(t-s)}\,\mathrm{vec}(C), \quad s < t$$

where $\mathrm{vec}(C) = (\Theta \oplus \Theta)^{-1}\,\mathrm{vec}(\Sigma)$, $\Sigma = SS^T$, $\oplus$ denotes the Kronecker sum, $A \oplus B = A \otimes I_B + I_A \otimes B$, and vec is an operator that stacks the columns of a matrix to a vector. Based on the velocity, the location process, X_t, follows the SDE

$$\mathrm{d}X_t = V_t\mathrm{d}t \tag{IV.2}$$

IV.2.2 Discretizing the SDE

Considering the N predetermined time points $\{t_i\}_{i\in\{1,\ldots,N\}}$, an Euler–Maruyama approximation to the location process is obtained by

$$X_{t_i} = X_{t_{i-1}} + V_{t_{i-1}}\Delta_i$$

where $\Delta_i = t_i - t_{i-1}$ is the length of the time step. Inserting the expression for $V_{t_{i-1}}$, the Generalized first-Differenced Correlated Random Walk (GDCRW) is obtained:

$$X_{t_i} = X_{t_{i-1}} + \Delta_i \exp(-\Theta\Delta_{i-1})(X_{t_{i-1}} - X_{t_{i-2}})/\Delta_{i-1} + \Delta_i\left(I - \exp(-\Theta\Delta_{i-1})\right)\mu + \Delta_i\epsilon_{t_i} \tag{IV.3}$$

The error terms, ϵ_{t_i}, follow a zero mean normal distribution with variance

$$Var(\epsilon_{t_i}) = C - \exp(-\Theta\Delta_i)C\exp(-\Theta^T\Delta_i)$$

where

$$\mathrm{vec}(C) = (\Theta \oplus \Theta)^{-1}\,\mathrm{vec}(\Sigma),$$

and $\Sigma = SS^T$ is a covariance matrix.

IV.2.3 Special cases

The GDCRW model is closely related to four other models: The CTCRW, the DCRW, the modified DCRW by Auger-Méthé et al. (2017), and a random walk.

IV.2.3.1 Relation to the CTCRW When the turning-angle parameter $\theta = 0$, and the increment covariance S is diagonal, the velocity model, V_t is identical to the velocity model of the CTCRW; however, the location models differ. While the location model of the CTCRW is an analytical solution to the two univariate integrated velocity models, the GDCRW is a discretization of the more general bivariate velocity model. Hence, when the length of the time steps approaches zero, the GDCRW approaches the CTCRW when the velocity models are identical. When $\theta \neq 0$ or S is not diagonal, the GDCRW approaches a generalization of the CTCRW when the length of the time steps approaches zero.

IV.2.3.2 Relation to the DCRW For regularly observed data, $\Delta_t = \Delta$, with $\mu = 0$ and $\gamma_1 = \gamma_2$,

$$\exp(-\Theta\Delta) = \begin{pmatrix} \cos(\theta\Delta) & -\sin(\theta\Delta) \\ \sin(\theta\Delta) & \cos(\theta\Delta) \end{pmatrix} \gamma^{\Delta}$$

Hence, the movement model simplifies to

$$X_i = X_{i-1} + \begin{pmatrix} \cos(\theta\Delta) & -\sin(\theta\Delta) \\ \sin(\theta\Delta) & \cos(\theta\Delta) \end{pmatrix} \gamma^{\Delta}(X_{i-1} - X_{i-2}) + \Delta\epsilon_i, \tag{IV.4}$$

which is the movement model of Jonsen et al. (2005) when $\Delta = 1$. Therefore, the GDCRW generalizes the DCRW in three ways: the restriction of regular time steps is relaxed, a drift vector is added, and different auto-correlations in the two coordinates are allowed. Further, equation (IV.4) allows scaling the DCRW parameters to a common time scale. That is, if parameters $\tilde{\gamma}$ and $\tilde{\theta}$ are obtained from a DCRW model fitted with time steps $\tilde{\Delta}$, scale independent parameters are obtained by $\gamma = \tilde{\gamma}^{1/\tilde{\Delta}}$ and $\theta = \tilde{\theta}/\tilde{\Delta}$. Being able to scale the DCRW parameters to a common scale allows comparison between model fits in meta or population studies even if different time scales are used.

IV.2.3.3 Relation to the modified DCRW Recently, Auger-Méthé et al. (2017) proposed a modified version of the DCRW for Fastloc-GPS data. The modified DCRW allowed irregularly observed data and included a time-varying autocorrelation parameter, γ_t. Ignoring the time-varying autocorrelation, the modified DCRW can be obtained as a special case of the GDCRW introduced here by letting $\gamma_1 = \gamma_2 = \gamma_t$, $\mu = 0$, $\theta = 0$, and by letting Σ be a diagonal matrix. Note, however, that the two models differ slightly in parameterisation. In the modified DCRW, γ_t is not scaled by the time differences and the variance of ϵ_{t_i} is parameterised as a diagonal matrix $diag(\sigma_1^2, \sigma_2^2)$. The GDCRW can easily be extended to have a time-varying autocorrelation parameter in the same way as Auger-Méthé et al. (2017).

IV.2.3.4 Relation to a random walk In the same manner as the three models above, the GDCRW can be reduced to a random walk model. From equation (IV.3) it follows that when the autocorrelation parameters γ_1 and γ_2 tend to zero, the model is reduced to a random walk. If $\mu \neq 0$, it will be a random walk with drift.

IV.2.2 Measurement equation

Besides the movement process describing the latent states, $\{X_{t_i}\}_{i\in\{1,\ldots,N\}}$, a measurement equation is needed to form a state-space model. For the DCRW, measurement errors are often included by linearly interpolating between the latent states. For an observation Y_{s_j} at time s_j,

$$Y_{s_j} = (1-q)X_{t_i} + qX_{t_{i+1}} + \eta_j$$

such that $t_i \leq s_j < t_{i+1}$, $q = \frac{s_j - t_i}{\Delta_{i+1}}$, and η_i is a zero mean bivariate random variable. However, when the movement model allows irregular time steps, it is more natural to ensure that the latent states align with the observations, such that the measurement equation simplifies to

$$Y_{s_j} = X_{s_j} + \eta_j$$

Then the observation at time s_j does not depend on a latent state in the future. Note that the state-space model can include additional latent states between observations. These latent states only contributes to the likelihood function through the movement process. Below, η_j will either be modelled as a bivariate normal distribution (Johnson et al., 2008) or a bivariate t-distribution (Albertsen et al., 2015).

IV.2.3 Simulation study: Comparing the GDCRW and the CTCRW

Since the location process is constructed as an Euler–Maruyama discrete time approximation to an underlying continuous time model, increasing the number of latent states between observations should improve the approximation. This was investigated by simulating from the underlying continuous time model in the special case that reduces to the CTCRW. On each simulated data set, the CTCRW was compared to the GDCRW with latent states at the observations, and the GDCRW$_i$ with additional latent states every $i = 8, 4, 2$, and 1 time unit, respectively.

To compare the GDCRW with the CTCRW, 200 data sets were simulated from equations (IV.1) and (IV.2) with $\theta = 0$ and diagonal S; that is, with the CTCRW as the true model. To obtain locations, 250 time steps were simulated from a mixture of an exponential and a normal distribution (see Appendix IV.B) to ensure both short and long time steps between observations. Between the simulated time points, the velocity and location processes were simulated using the Euler–Maruyama method with 200 steps. For each of the 250 simulated locations, observations were simulated from a bivariate normal with covariance $0.1^2 I$. The processes were simulated with $\gamma_1 = \gamma_2 = 0.9$, $\mu = 0$, and $S_{11} = S_{22} = \exp(-2) \approx 0.135$. The models were fitted to the data by maximum likelihood through the Laplace approximation using the `R` package `TMB` (Kristensen et al., 2016). For simplicity, the model was fitted with $\gamma_1 = \gamma_2$, $\theta = 0$, $\mu = 0$, and $S_{12} = S_{21} = 0$.

IV.2.4 Simulation study: Comparing the GDCRW and the DCRW

Above it was established that the DCRW was a special case of the GDCRW. One of the ways the DCRW was generalized was that the GDCRW allowed latent states at arbitrary time points. In contrast, the DCRW only allowed equidistant time steps in discretising the underlying SDE. In this simulation study the effect of the choice of discretization on parameter estimates was investigated.

Data sets were simulated from two behavioural scenarios using the same procedure as for the previous simulation study. For each scenario, 200 data sets were simulated from the continuous time model determined by equations (IV.1) and (IV.2). The first scenario emulated tortuous movement simulated with $\phi = \frac{\pi}{3} \approx 1.05$, $\mu = 0$, and $\gamma_1 = \gamma_2 = 0.6$. The movement was simulated with diagonal S such that $S_{11} = S_{22} = \exp(-2)$. Measurements were simulated from a normal distribution with covariance $\Sigma_o = 0.1^2 I_2$. The second scenario mimicked a more persistent movement. Movement was simulated with $\phi = 0$ and $\gamma = 0.7$. Similar to the first scenario, the movement was simulated with $S_{11} = S_{22} = \exp(-2)$ and the observations with $\Sigma_o = 0.1^2 I_2$.

For each data set, a total of 10 models were fitted consisting of the GDCRW and the DCRW with $N = 250, 500, 750, 1000$, and 1250 latent locations. In the DCRW, the latent locations were included

equidistantly. In the GDCRW, 250 latent locations were aligned with the observations. Remaining latent locations were included recursively in the middle of the longest time step. The models were fitted to the data by maximum likelihood through the Laplace approximation. For simplicity, the DCRW models were fitted with $S_{11} = S_{22}$, and $S_{12} = S_{21} = 0$. The GDCRW models were fitted with $\gamma_1 = \gamma_2$, $\mu = 0$, $S_{11} = S_{22}$, and $S_{12} = S_{21} = 0$ to reduce to the same parameters as the DCRW models.

IV.2.5 Simulation study: Effect of measurement error on choice of time steps

In the final simulation study, the GDCRW and DCRW is compared once more. While the previous comparison focused on parameter estimates, this simulation study compares the ability to accurately obtain location estimates for different ratios of process variability and measurement variability. Again, 200 data sets with 250 observations were simulated from two behavioural scenarios using the same procedure as for the two previous simulation studies. The same movement parameters as in the previous simulation study was used, and the GDCRW and DCRW with 250 latent locations were fitted in the same manner as before; however, the observations were simulated with covariances $\exp(-6)^2 I, \exp(-5.5)^2 I, \ldots, \exp(1.5)^2 I, \exp(2)^2 I$, respectively. For each location for each track the relative distance to the true locations were calculated for the GDCRW compared to the DCRW. The DCRW locations were interpolated linearly to get values at the time of observations.

IV.2.6 Case study: Ringed seal

To illustrate its practical applicability, the GDCRW model was fitted to the subadult ringed seal *Pusa hispida* data of Albertsen et al. (2015) in this case study. The data set was chosen to illustrate the movement model when substantial measurement errors are present. The ringed seal data consists of $N = 3583$ locations collected by the Argos satellite system, which is known to have substantial non-Gaussian measurement errors. Three observations with location class Z were excluded. In the first period of the track, the seal was moving north-west, followed by two periods with restricted space use close to land. Consequently, Thygesen et al. (2017) found indications that a non-constant drift parameter was needed. To account for this behaviour, the mean velocity parameters of the GDCRW were set to be location dependent:

$$\begin{aligned} X_{t_i} =& X_{t_{i-1}} + \Delta_{t_i} \exp\left(- \begin{pmatrix} -\log\gamma_1 & \theta \\ -\theta & -\log\gamma_2 \end{pmatrix} \Delta_{t_{i-1}} \right) (X_{t_{i-1}} - X_{t_{i-2}})/\Delta_{t_{i-1}} \\ &+ \Delta_{t_i} \left(I - \exp(-\Theta \Delta_{t_{i-1}}) \right) \mu(X_{t_{i-1}}) + \Delta_{t_i} \epsilon_{t_i} \end{aligned}$$

The bivariate mean velocity field $\mu(X_t) = (\mu_1(X_t), \mu_2(X_t))^T$ was modelled through a 7×7 grid of random effects. Each grid point was related to two random effects: one for the mean latitudinal velocity, μ_1, and one for the mean longitudinal velocity, μ_2. Between the grid points, the random effect values were interpolated by a search tree approximation to local polynomial regression (Albertsen, 2017). In this case study, the number of random effects to construct the grids were selected as a trade-off between computational cost and flexibility. Locations of the movement process were estimated at every observation and every three hours from the first observation, giving a total of 4764 location estimates each with two latent states. Following Albertsen et al. (2015), the measurements were modelled by a bivariate t-distribution with scale matrices and degrees of freedom depending on the location class of the observations. The model was fitted to the data by maximum likelihood through the Laplace approximation using `TMB`.

Having a location dependent drift in the model, the animal movement may be related to available resources or hydrographic variables. The model cannot be estimated using the DCRW, since the DCRW does not include a drift parameter. Furthermore, the model cannot easily be estimated using CTCRW without assuming constant movement between states, in which case the CTCRW is a special case of the GDCRW (see subsection IV.2.3).

IV.3 Results

IV.3.1 Simulation study: Comparing the GDCRW and the CTCRW

In the first simulation study, the GDCRW was compared to the CTCRW, when the CTCRW was the true data generating model. The CTCRW models both the velocity and location giving a total of 1000 latent variables: 4 for each location. In contrast, the GDCRW only has 500 latent variables: 2 for each location. The mean of the time step distribution used in the simulation study was approximately 1.99, giving an average length of a simulated trajectory of 497.5 time units. Therefore, the GDCRW_8 would include 63 additional location estimates on average, corresponding to a total of 313 latent variables. Likewise, the GDCRW_4, GDCRW_2, and GDCRW_1 would on average include 125, 249, and 498 additional locations, respectively, corresponding to 750, 998, and 1496 latent variables.

Regardless of the different number of latent variables, all six estimation models provided γ estimates close to the true value (Figure IV.1). For the true continuous time model, the CTCRW, the average of the 200 estimates was 0.894, whereas it was 0.897 for the discrete time approximation, the GDCRW, with latent states only at the time of the observations. When additional latent states were included every 8, 4, 2, and 1 time unit, the average of the estimated γs were 0.901, 0.903, 0.9, and 0.897, respectively. Although all models gave estimates close to the true value, the standard deviation of the estimates was higher for the GDCRW than for the five other models. For the GDCRW the standard deviation of the 200 estimates was 0.028, whereas it was 0.019 for the CTCRW, and 0.018, 0.017, 0.017, and 0.018 for the GDCRW_8, GDCRW_4, GDCRW_2, and GDCRW_1, respectively.

Besides the ability to re-obtain the true parameters, the models were compared on the distance between the estimated locations and the true simulated locations Figure IV.2. To compare the models, this distance was calculated for each location for each track for the GDCRW, GDCRW_8, GDCRW_4, GDCRW_2, and GDCRW_1, and divided by the distance obtained for the CTCRW. Finally, the median relative distance was obtained. Whereas the models performed similarly in estimating γ, the CTCRW performed better than the other models in obtaining precise location estimates; however, the GDCRW models improved when additional latent states were included, as expected. For the GDCRW, 16% of the simulations had a relative distance compared to the CTCRW that was less than one; that is, where at least half of the estimated locations were closer to the true location than for the estimated locations from the CTCRW. For the GDCRW_8, 24% of the simulations had a relative distance less than one, while it was 31, 36, and 47%, respectively, for the GDCRW_4, GDCRW_2, and GDCRW_1. Further, the range of median relative distances narrowed when the number of latent states increased. When no additional latent states were included, the length of the range was 0.188 (0.958;1.146). Based on these 200 simulations, the range gradually narrowed for the GDCRW_8 (0.143; min: 0.958; max: 1.146), GDCRW_4 (0.116; min: 0.953; max: 1.069), GDCRW_1 (0.099; min: 0.962; max: 1.061), and GDCRW_1 (0.072; min: 0.964; max: 1.036).

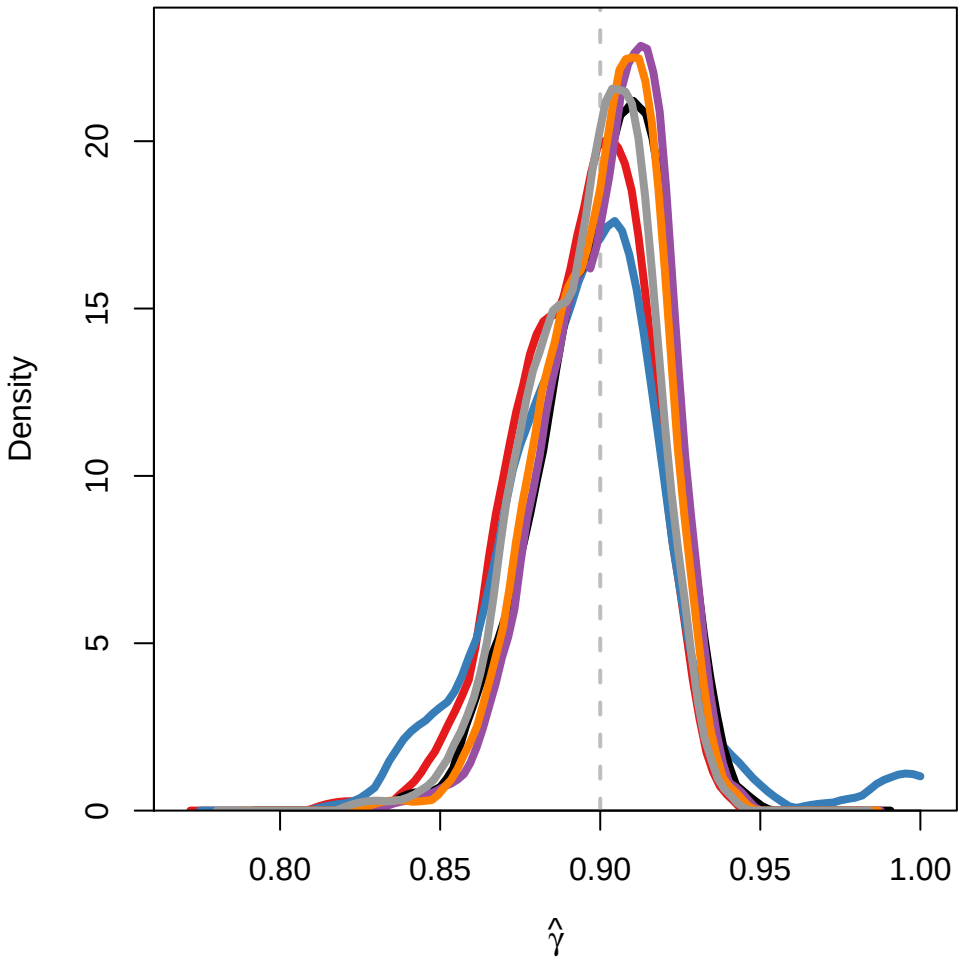


Figure IV.1: Density of γ estimates in the simulation study for the models: CTCRW (red line), GDCRW (blue line), GDCRW_8 (black line), GDCRW_4 (purple line), GDCRW_2 (orange line), and GDCRW_1 (grey line). Grey dashed line indicates the true parameter value.

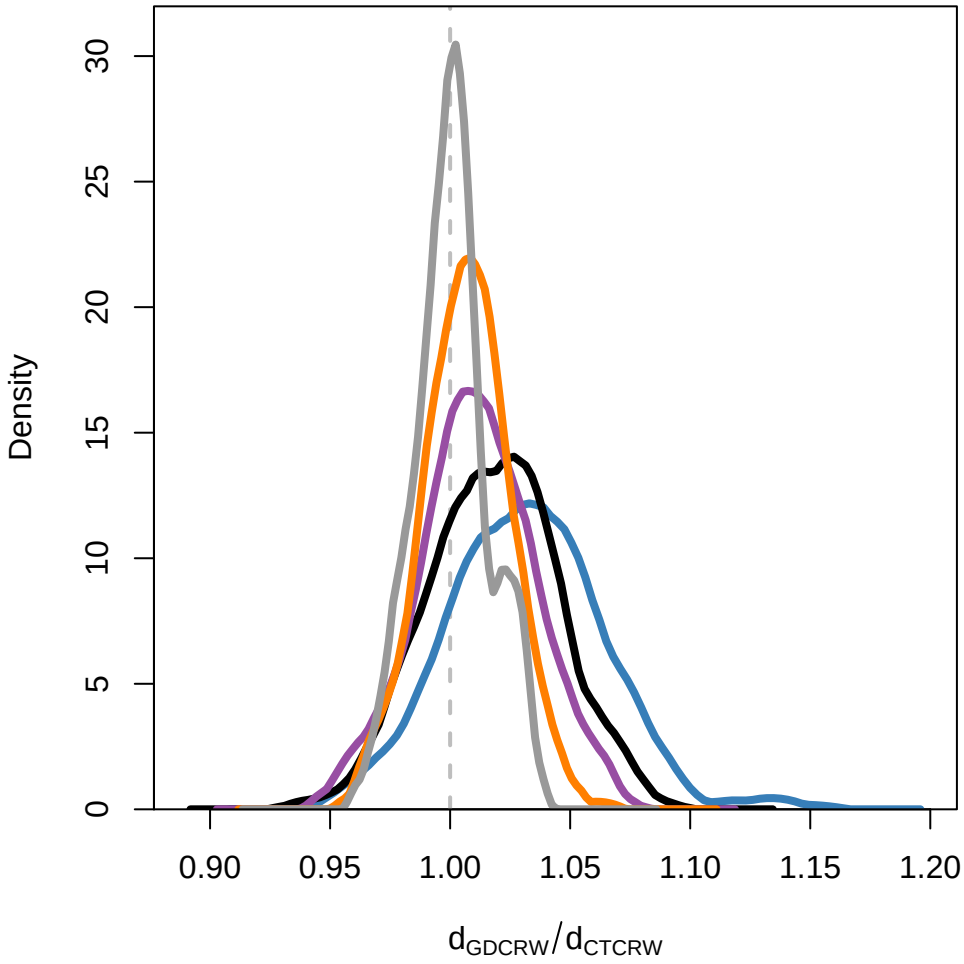


Figure IV.2: Density of track-wise median distance between estimated locations and true simulated locations in the simulation study for the models: GDCRW (blue line), GDCRW_8 (black line), GDCRW_4 (purple line), GDCRW_2 (orange line), and GDCRW_1 (grey line) relative to the CTCRW. A value of one (grey dashed line) indicates the median distance is the same as for the CTCRW.

IV.3.2 Simulation study: Comparing the GDCRW and the DCRW

In the second simulation study, the parameter estimates from the GDCRW and DCRW were compared in two movement scenarios. In the tortuous movement scenario, the GDCRW provided γ estimates close to the true value for all five number of latent locations with a slight tendency to overestimate the value (Figure IV.3). The average γ estimates (standard deviation) were 0.622 (0.059), 0.643 (0.061), 0.625 (0.062), 0.614 (0.066), and 0.611 (0.067) for the GDCRW with $N = 250, 500, 750, 1000$, and 1250 latent locations, respectively. A difference was, however, seen in for the θ estimates. Their accuracy increased with the number of latent locations. The average estimates were 0.729 (0.139), 0.956 (0.124), 0.999 (0.124), 1.028 (0.129), and 1.033 (0.134), respectively. Unlike for the GDCRW, the estimates obtained directly from the DCRW were not close to the true values. For these models, the average γ estimates were 0.56 (0.067), 0.654 (0.061), 0.734 (0.053), 0.787 (0.046), and 0.823 (0.041) with $N = 250, 500, 750, 1000$, and 1250 latent locations, respectively, while the average θ estimates were 1.841 (0.237), 1.059 (0.178), 0.705 (0.124), 0.528 (0.094), and 0.427 (0.113). Correcting the estimates obtained from the DCRW by equation (IV.4), the accuracy of the estimates increased with the number of latent locations. Then, the average γ estimates were 0.746 (0.046), 0.653 (0.059), 0.63 (0.064), 0.621 (0.066), and 0.616 (0.069), respectively, while the average θ estimates were 0.921 (0.114), 1.057 (0.133), 1.055 (0.136), 1.054 (0.138), and 1.066 (0.241).

In the persistent movement scenario, the GDCRW and the DCRW corrected by equation (IV.4) provided estimates close to the true values for both parameters for all five number of latent locations (Figure IV.4). For the GDCRW, the average γ estimates were 0.897 (0.028), 0.897 (0.028), 0.899 (0.019), 0.897 (0.019), and 0.896 (0.019), respectively, and the average θ estimates were 0.002 (0.02), 0.002 (0.02), 0.002 (0.019), 0.002 (0.019), and 0.002 (0.019). Likewise, the average DCRW estimates corrected by equation (IV.4) were 0.9 (0.019), 0.898 (0.019), 0.898 (0.019), 0.897 (0.019), and 0.897 (0.019), respectively, for γ, and 0.002 (0.019), 0.002 (0.019), 0.002 (0.019), 0.002 (0.019), and 0.002 (0.019), respectively, for θ. As for the tortuous movement scenario, the estimates obtained directly from the DCRW were not close to the true values. For the DCRW models, the average γ estimates were 0.812 (0.037), 0.898 (0.022), 0.931 (0.015), 0.948 (0.011), and 0.958 (0.009) with $N = 250, 500, 750, 1000$, and 1250 latent locations, respectively, while the average θ estimates were 0.006 (0.08), 0.003 (0.038), 0.002 (0.019), 0.001 (0.012), 0.001 (0.009), and 0.001 (0.007).

IV.3.3 Simulation study: Effect of measurement error on choice of time steps

In the third simulation study, the GDCRW and DCRW were compared again. This simulation study compared location estimates for different levels of measurement error. When the measurement standard deviation was low, the GDCRW performed notably better than the DCRW (Figure IV.5). In both scenarios, all 200 simulations had a median relative distance less than 1; that is, at least half of the locations were better estimated by the GDCRW than the DCRW. As the measurement standard deviation increases relative to the movement standard deviation, the median relative distance tends to one, suggesting that the GDCRW and the DCRW estimates similarly. When the measurement and movement standard deviation were equal, the DCRW performed slightly better than the GDCRW.

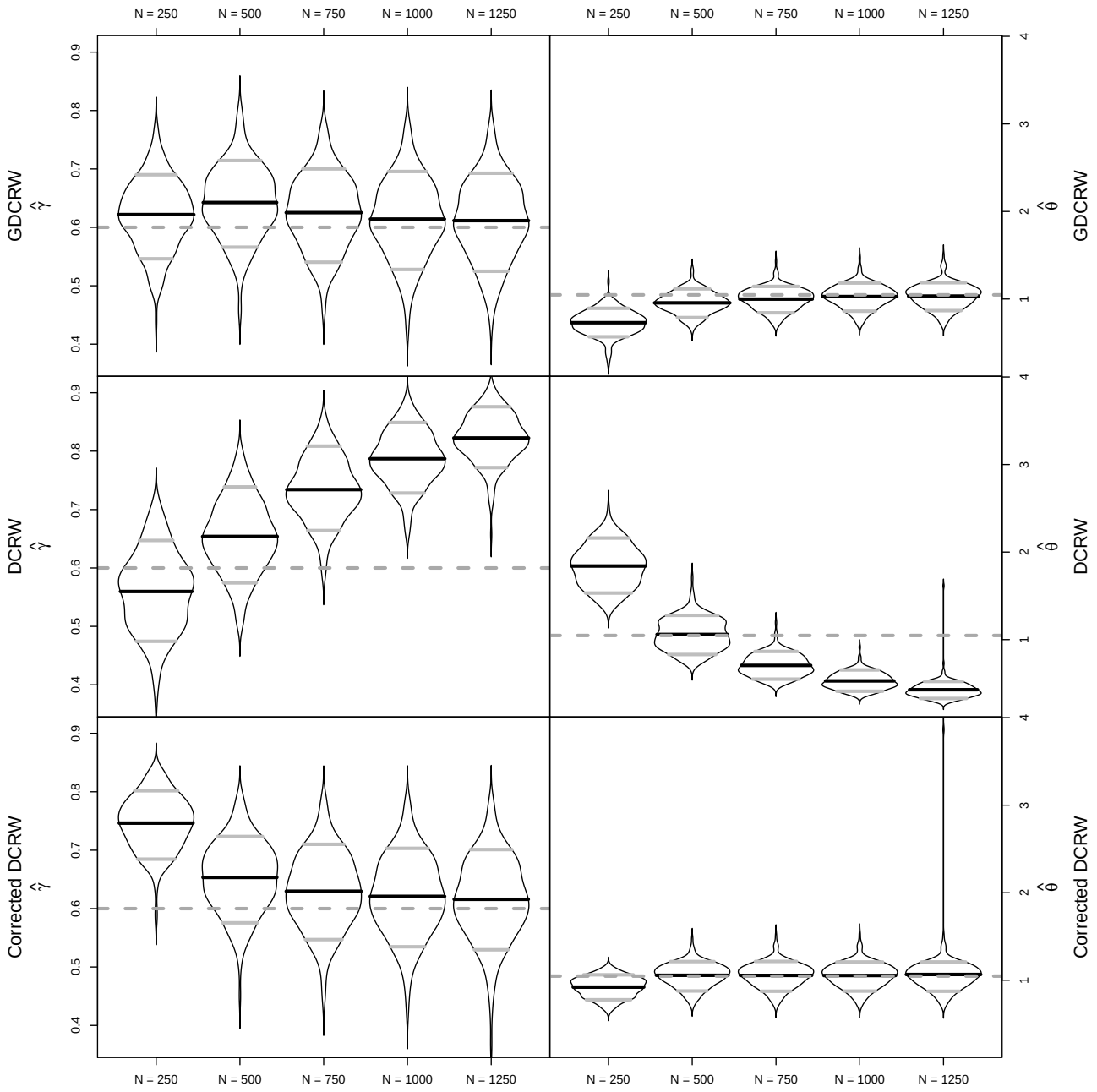


Figure IV.3: Violin plots of γ and θ estimates in the tortuous movement simulations for the GDCRW, DCRW, and the DCRW corrected by equation (IV.4) with $N = 250, 500, 750, 1000, 1250$ latent states and 250 observations. Grey dashed lines indicate true parameter values. In the violin plot, black lines indicate the mean of estimates and grey lines indicate 10 and 90 % quantiles of estimates.

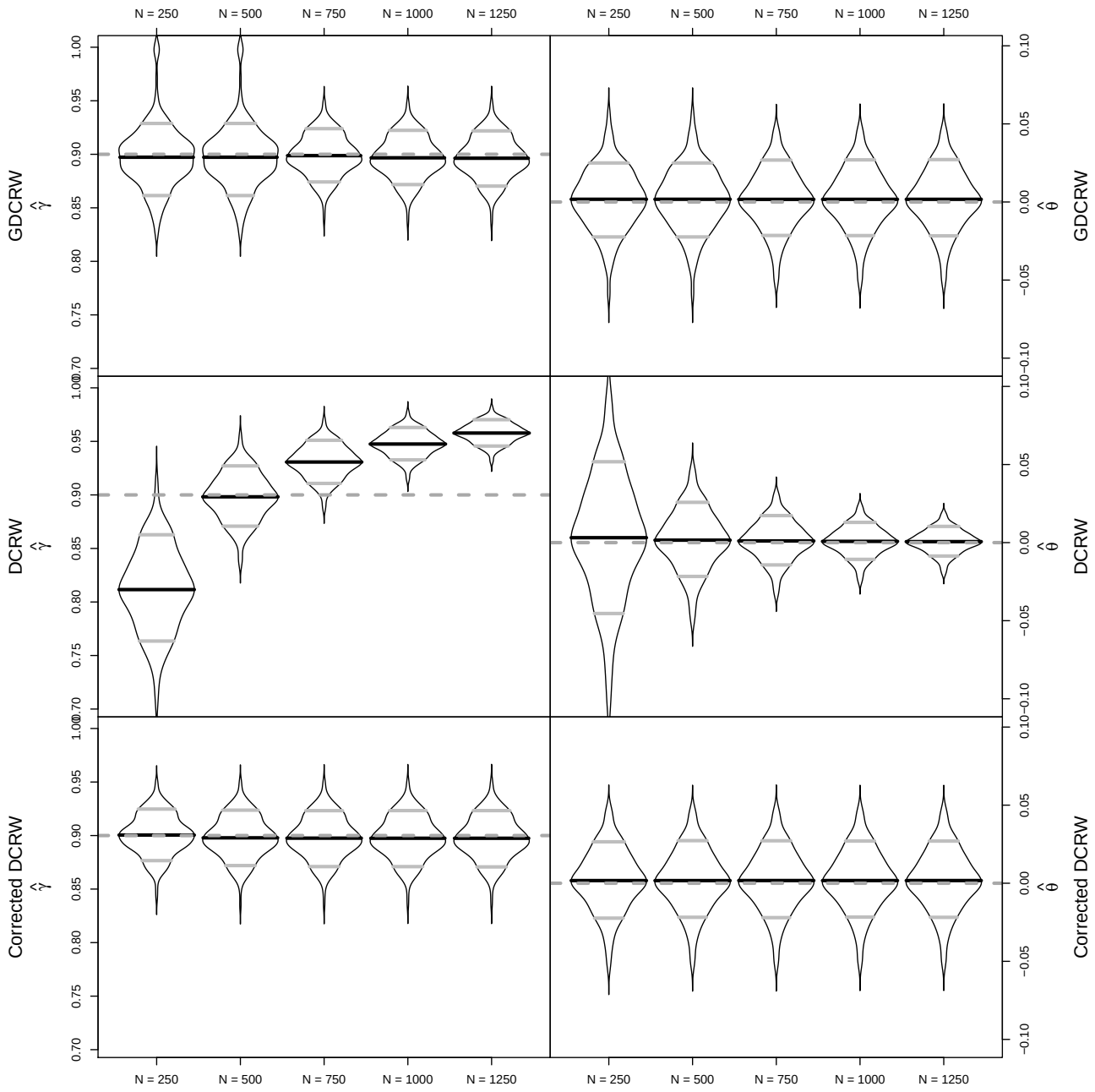


Figure IV.4: Violin plots of γ and θ estimates in the persistent movement simulations for the GDCRW, DCRW, and the DCRW corrected by equation (IV.4) with $N = 250, 500, 750, 1000, 1250$ latent states and 250 observations. Grey dashed lines indicate true parameter values. In the violin plot, black lines indicate the mean of estimates and grey lines indicate 10 and 90 % quantiles of estimates.

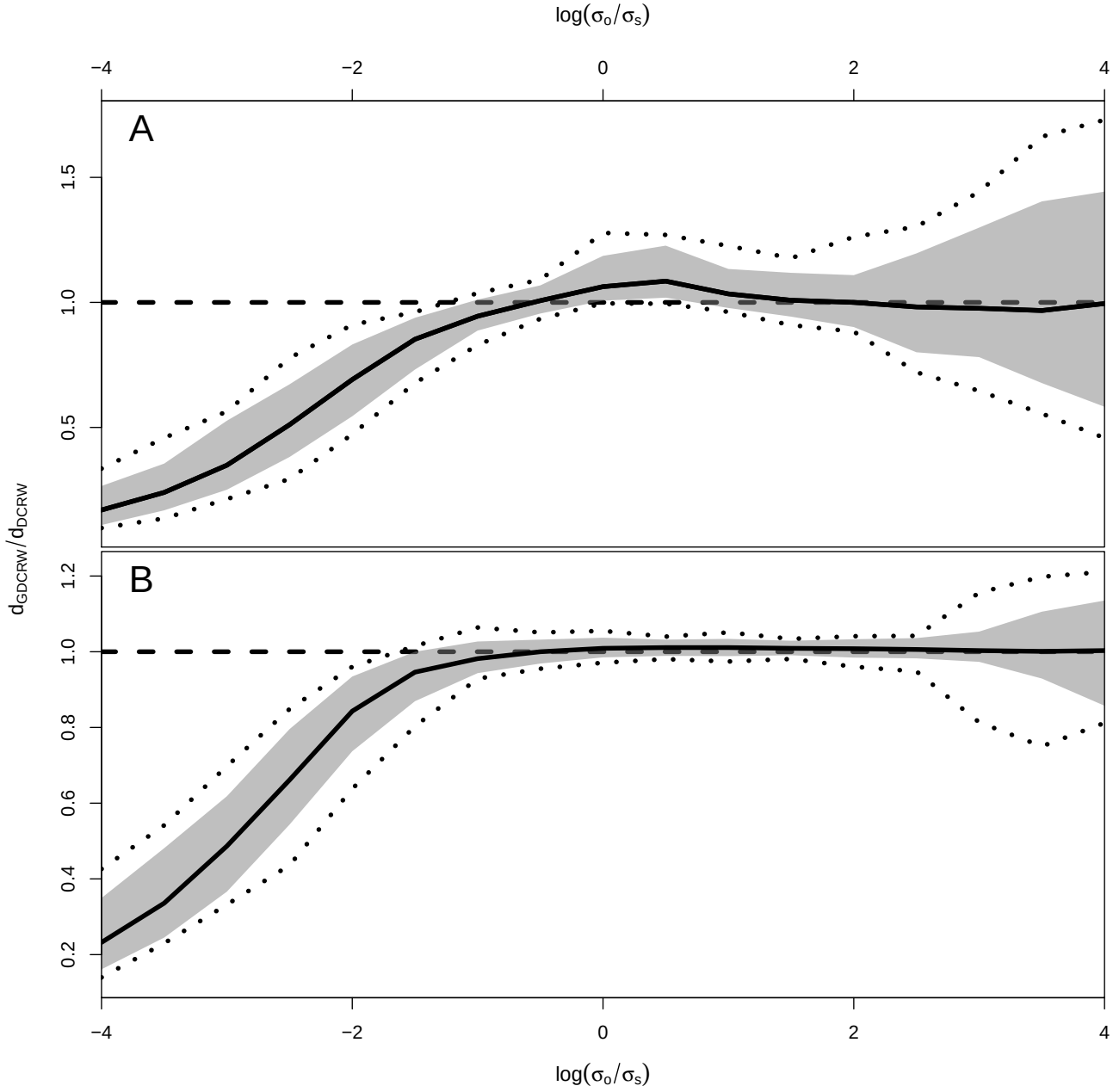


Figure IV.5: Minimum (lower dotted black line), maximum (upper dotted black line), median (full black line), and 95% range (grey area) of track-wise median distances between estimated locations and true simulated locations for the GDCRW relative to the DCRW in the tortuous (panel A) and persistent (panel B) simulated movement scenarios as a function of the ratio between observational standard deviation (σ_o) and movement standard deviation (σ_s). Values below 1 (black dashed line) indicates that at least half of the estimated location in the GDCRW were closer to the true value than for the DCRW.

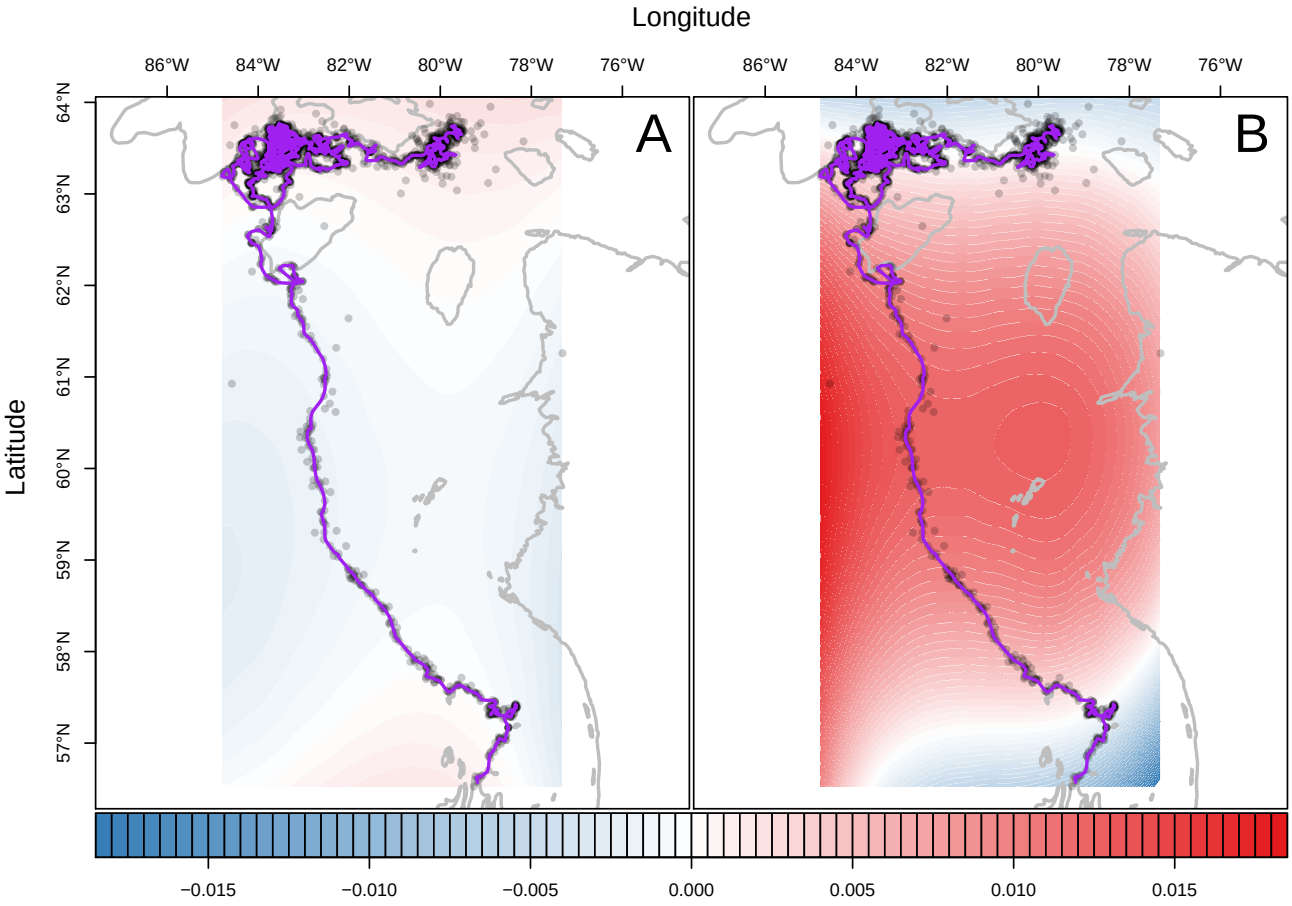


Figure IV.6: Mean longitudinal (A) and latitudinal (B) velocity field of the subadult ringed seal track with land (outlined by grey lines), Argos satellite observations (dark grey dots), and fitted most probable track given all observations (purple line). In the longitudinal field, red values indicate an eastward attraction, while blue values indicate a westward attraction. In the latitudinal field, red values indicate a northward attraction, while blue values indicate a southward attraction.

Table IV.1: Estimated movement parameters in the ringed seal case study.

Parameter	Estimate	Standard Error
γ_{lat}	0.608	0.027
γ_{lon}	0.601	0.028
θ	0.027	0.019
$\Sigma_{lat,lon}/\sqrt{\Sigma_{lat,lat}\Sigma_{lon,lon}}$	-0.018	0.052
$\log \Sigma_{lat,lat}$	-9.189	0.114
$\log \Sigma_{lon,lon}$	-7.486	0.114

IV.3.4 Case study: Ringed seal

In the case study, the GDCRW was fitted to a real data set collected by the Argos system. The GDCRW was modified to include location dependent drift and turning angle. The fitted model resulted in small estimated longitudinal mean velocity field values (Figure IV.6). On the contrary, the latitudinal mean velocity field was estimated to have a southward drift when the seal was south of the 57.5°N parallel. Despite the southward drift, the seal then entered a large area of northward drift between 57.5°N and 63.5°N. North of the 63.5°N parallel, the latitudinal mean velocity field was estimated to have a southward attraction, keeping the seal close to the 63.5°N parallel. Further the model estimated an overall mean turning angle, θ. This parameter was estimated to be close to zero (Table IV.1), whereas both the latitudinal and longitudinal autocorrelation parameters were estimated to be close to 0.6.

IV.4 Discussion

The GDCRW introduced here generalizes the DCRW movement model of Jonsen et al. (2005) in three ways. Firstly, whereas the DCRW handles irregular observations by linear interpolation in the observational model, the GDCRW allows modelling irregular observations directly in the movement process. Through this construction, time-scale corrections of the DCRW parameters were found by considering a time regular GDCRW. Secondly, the GDCRW generalizes the DCRW by allowing different auto-correlation parameters in the two coordinates. Hence, when the rotation parameter is zero, the GDCRW resembles the integrated velocity model of the CTCRW of Johnson et al. (2008). Like the CTCRW and DCRW, the GDCRW includes the random walk as a limiting case. When the auto-correlation parameter approaches zero, the movement model approaches that of a random walk, which has been used in several studies, in particular when the data includes only daily observations (e.g. Nielsen et al., 2006). Thirdly, the GDCRW extends the DCRW by including a drift term.

Including a drift term in the model is useful for animals moving persistently between areas, such as the ringed seal analysed in the case study. In the case study, the drift term was modelled by a location dependent field. Having location or time dependent parameters in the SDE greatly complicates the calculations to obtain an analytical solution and in turn a fully continuous time model. Therefore, location dependent mean velocities could not easily be implemented in the CTCRW which has previously been used to analyse the track (Albertsen et al., 2015; Thygesen et al., 2017). Nonetheless, using the discrete time model, or a discrete time approximation, it is straightforward to extend the movement model with a location dependent drift. Using the same methods, spatial covariates such as sea surface temperature or depth could be included.

While the mean turning angle was estimated to be close to zero using the entire track, this may not be the case if, for instance, a time varying parameter or behavioural switching was included. Changes in behavioural states, $S_{t_i} \in \{1, 2, \ldots, n\}$, can be modelled by a Markov Chain such that

$$P(S_{t_i} = k \mid S_{t_{i-1}} = j) = (\exp(Q\Delta_{t_i}))_{jk}$$

where Q is an $n \times n$ matrix such that $Q_{jk} > 0$ for $j \neq k$ and $Q_{jj} = -\sum_{\{k:j\neq k\}} Q_{jk}$. The movement model is modified to have different parameters depending on the current behavioural state

$$X_{t_i} = X_{t_{i-1}} + \Delta_{t_i} \exp(-\Theta_{S_{t_i}}\Delta_{t_i})(X_{t_{i-1}} - X_{t_{i-2}})/\Delta_{t_{i-1}} + \Delta_{t_i}\left(I - \exp(-\Theta_{S_{t_i}}\Delta_{t_i})\right)\mu_{S_{t_i}} + \Delta_{t_i}\epsilon_{t_i}.$$

This model generalizes the behavioural switching movement models of Jonsen et al. (2005) and Whoriskey et al. (2017). It assumes that switching only occurs at the predetermined time points; however, as it has been shown in both the case study and simulation studies, latent locations can be included either at regular time points, or at the time of observations and any time between them. While the extension to include behavioural switching is simple, maximum likelihood estimation is greatly complicated when measurement error is present.

The GDCRW can be constructed as a discretization of a generalization of the CTCRW. The first simulation study comparing the GDCRW and CTCRW suggests that using the analytical continuous time solution, when it can be found, generally provides better results than a discrete time approximation. Nonetheless, the discrete time approximation provided accurate parameter estimates, even when no additional latent states were included to improve the approximation. While the GDCRW was outperformed by the CTCRW in estimating true locations when the CTCRW was the true model, the first simulation study illustrated that the performance of the GDCRW could be improved to closely resemble the CTCRW by including additional latent states between observations.

The second simulation study compared the DCRW and the GDCRW in their ability to re-estimate the movement parameters of the underlying continuous time model. In both the tortuous and persistent movement scenarios, the GDCRW performed well; however, in the tortuous movement scenario, the performance was notably improved by adding additional latent locations between observations. Further, it became evident that the parameter estimates of the DCRW are highly dependent on the selected time steps between latent locations. Nevertheless, the estimates could be corrected by equation (IV.4): the GDCRW with regular time steps. This correction provided estimates close to the true values in all cases except in the tortuous movement case with 250 and 500 latent locations. Clearly, the ability to have time scale independent movement parameters is a key feature of the GDCRW, regardless of whether it is used with regular or irregular time steps. Time scale independent movement parameters allow analysis of several animals without using the same time steps, and further allow comparing results from previous studies using the DCRW even if different time steps are used.

Besides accuracy of the estimates, the DCRW and GDCRW were compared on their ability to reconstruct the movement tracks. The third simulation study showed that modelling time-irregular data through a time-irregular movement could increase the accuracy of predicted locations. This is consistent with previous findings comparing the DCRW and CTCRW, which showed that the continuous time model performed better if both movement models were combined with t-distributed measurement errors (Albertsen et al., 2015). Overall, using the GDCRW with irregular time steps performed well in both scenarios compared to the DCRW. For small measurement standard deviations, the irregular time steps outperformed the regular time steps with interpolation. For medium measurement standard deviations, the two approaches performed equally in the persistent movement scenario, whereas the DCRW slightly outperformed the GDCRW in the tortuous movement case. For large measurement standard deviations, the two approaches performed equally in both scenarios. Combined with the previous simulation study, these results suggest that having latent locations at the time of observations generally performs well compared to the DCRW with equidistant time steps in the movement; keeping in mind that both improves as

approximations to the same underlying continuous time movement model as the number of latent locations increases.

Directly modelling irregular data eases interpretation and comparison of movement parameters because they are scaled by the time steps. From equation (IV.4), parameters of the DCRW and GDCRW can be transformed to any time scale. However, the simulation study shows that fitting the DCRW with a poorly chosen time step introduces bias. This poses a problem in population and meta studies. With a discrete time model, all animals must be fitted with the same time steps to compare movement parameters. However, if the same time step is not optimal for all animals, bias can be introduced. Using an irregular or continuous time model scales the parameters to a common time scale, even when the animals are observed, or behave, at different time scales.

References

Albertsen, C. M. (2017). *Covafillr: local polynomial regression of state dependent covariates in state-space models*. R package version 0.4.1. https://CRAN.R-project.org/package=covafillr. Retrieved from https://CRAN.R-project.org/package=covafillr

Albertsen, C. M., Whoriskey, K., Yurkowski, D., Nielsen, A., & Flemming, J. (2015). Fast fitting of non-gaussian state-space models to animal movement data via template model builder. *Ecology*, *96*(10), 2598–2604. doi:10.1890/14-2101.1

Auger-Méthé, M., Albertsen, C. M., Jonsen, I. D., Derocher, A. E., Lidgard, D. C., Studholme, K. R., … Flemming, J. M. (2017). Spatiotemporal modelling of marine movement data using template model builder. *Marine Ecology Progress Series*, *565*, 237–249. 0171-8630. doi:10.3354/meps12019

Auger-Méthé, M., Field, C., Albertsen, C. M., Derocher, A., Lewis, M. A., Jonsen, I. D., & Flemming, J. M. (2016). State-space models' dirty little secrets: even simple linear gaussian models can have parameter and state estimation problems. *Scientific Reports*, *6*, 26677. doi:10.1038/srep26677

Costa, D. P., Robinson, P. W., Arnould, J. P. Y., Harrison, A.-L., Simmons, S. E., Hassrick, J. L., … Crocker, D. E. (2010). Accuracy of ARGOS locations of pinnipeds at-sea estimated using fastloc GPS. *PLoS ONE*, *5*(1), e8677. doi:10.1371/journal.pone.0008677

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., … Sibert, J. (2012). Ad model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, *27*(2), 233–249. doi:10.1080/10556788.2011.597854

Gurarie, E., Andrews, R. D., & Laidre, K. L. (2009). A novel method for identifying behavioural changes in animal movement data. *Ecology Letters*, *12*(5), 395–408. doi:10.1111/j.1461-0248.2009.01293.x

Hussey, N. E., Kessel, S. T., Aarestrup, K., Cooke, S. J., Cowley, P. D., Fisk, A. T., … Whoriskey, F. G. (2015). Aquatic animal telemetry: a panoramic window into the underwater world. *Science*, *348*(6240). doi:10.1126/science.1255642

Johnson, D. S., London, J. M., Lea, M.-A., & Durban, J. W. (2008). Continuous-time correlated random walk model for animal telemetry data. *Ecology*, *89*(5), 1208–1215. doi:10.1890/07-1032.1

Jonsen, I. D. (2016). Joint estimation over multiple individuals improves behavioural state inference from animal movement data. *Scientific Reports*, *6*, 20625. doi:10.1038/srep20625

Jonsen, I. D., Flemming, J. M., & Myers, R. A. (2005). Robust state–space modeling of animal movement data. *Ecology*, *86*(11), 2874–2880. doi:10.1890/04-1852

Kristensen, K., Nielsen, A., Berg, C., Skaug, H., & Bell, B. (2016). Tmb: automatic differentiation and laplace approximation. *Journal of Statistical Software*, *70*(1), 1–21. doi:10.18637/jss.v070.i05

McClintock, B. T., Johnson, D. S., Hooten, M. B., Ver Hoef, J. M., & Morales, J. M. (2014). When to be discrete: the importance of time formulation in understanding animal movement. *Movement Ecology*, *2*(1), 21. doi:10.1186/s40462-014-0021-6

McClintock, B. T., King, R., Thomas, L., Matthiopoulos, J., McConnell, B. J., & Morales, J. M. (2012). A general discrete-time modeling framework for animal movement using multistate random walks. *Ecological Monographs*, *82*(3), 335–349.

Michelot, T., Langrock, R., & Patterson, T. A. (2016). moveHMM: an r package for the statistical modelling of animal movement data using hidden markov models. *Methods in Ecology and Evolution*, *7*(11), 1308–1315. doi:10.1111/2041-210x.12578

Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., & Fryxell, J. M. (2004). Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology*, *85*(9), 2436–2445. doi:10.1890/03-0269

Nielsen, A., Bigelow, K. A., Musyl, M. K., & Sibert, J. R. (2006). Improving light-based geolocation by including sea surface temperature. *Fisheries Oceanography*, *15*(4), 314–325. doi:10.1111/j.1365-2419.2005.00401.x

Thygesen, U. H., Albertsen, C. M., Berg, C. W., Kristensen, K., & Nielsen, A. (2017). Validation of ecological state space models using the laplace approximation. *Environmental and Ecological Statistics*, *24*(2), 317–339. doi:10.1007/s10651-017-0372-4

Tracey, J. A., Zhu, J., & Crooks, K. R. (2010). Modeling and inference of animal movement using artificial neural networks. *Environmental and Ecological Statistics*, *18*(3), 393–410. doi:10.1007/s10651-010-0138-8

Turchin, P. (1998). *Quantitative analysis of movement: measuring and modeling population redistribution in animals and plants*. Sunderland, MA, U.S.A: Sinauer.

Whoriskey, K., Auger-Méthé, M., Albertsen, C. M., Whoriskey, F. G., Binder, T. R., Krueger, C. C., & Flemming, J. M. (2017). A hidden markov movement model for rapidly identifying behavioral states from animal tracks. *Ecology and Evolution*, *7*(7), 2112–2121. doi:10.1002/ece3.2795

IV.A Solving the SDE

To solve the SDE,

$$\mathrm{d}V_t = -\Theta(V_t - \mu)\mathrm{d}t + SdB_t,$$

consider the process

$$Y_t = e^{\Theta t}(V_t - \mu)$$

Itô's formula yields

$$\begin{aligned} dY_t &= \Theta e^{\Theta t} V_t d_t + e^{\Theta t} dV_t \\ &= \Theta e^{\Theta t} V_t d_t + e^{\Theta t}\left(-\Theta(V_t - \mu)d_t + SdB_t\right) \\ &= e^{\Theta t} SdB_t \end{aligned}$$

Hence,

$$e^{\Theta t}(V_t - \mu) = (V_0 - \mu) + \int_0^t e^{\Theta u} SdB_u$$

which implies

$$\begin{aligned} V_t - \mu &= e^{-\Theta t}(V_0 - \mu) + e^{-\Theta t}\int_0^t e^{\Theta s} SdB_s \\ &= e^{-\Theta t} V_0 + \left(I - e^{-\Theta t}\right)\mu + \int_0^t e^{-\Theta(t-s)} SdB_s \end{aligned}$$

Since this is a sum of a deterministic term and an integral of a deterministic function with respect to a Wiener process with Gaussian increments, the distribution is Gaussian. The mean of the increment is

$$\begin{aligned} E(V_t \mid V_0) &= E\left(e^{-\Theta t}V_0 + \left(I - e^{-\Theta t}\right)\mu \mid V_0\right) + E\left(\int_0^t e^{-\Theta(t-s)}SdB_s \mid V_0\right) \\ &= e^{-\Theta t}V_0 + \left(I - e^{-\Theta t}\right)\mu \end{aligned}$$

Using Itô isometry, the variance is

$$\begin{aligned} Var\left(V_t \mid V_0\right) &= Var\left(e^{-\Theta t}V_0 + \left(I - e^{-\Theta t}\right)\mu + \int_0^t e^{-\Theta(t-s)}SdB_s \mid V_0\right) \\ &= Var\left(\int_0^t e^{-\Theta(t-s)}SdB_s \mid V_0\right) \\ &= Var\left(\int_0^t e^{-\Theta(t-s)}SdB_s\right) \\ &= E\left(\left(\int_0^t e^{-\Theta(t-s)}SdB_s\right)^2\right) - E\left(\int_0^t e^{-\Theta(t-s)}SdB_s\right)^2 \\ &= E\left(\left(\int_0^t e^{-\Theta(t-s)}SdB_s\right)^2\right) \\ &= \int_0^t e^{-\Theta(t-s)}SS^T\left(e^{-\Theta(t-s)}\right)^T ds \\ &= \int_0^t e^{-\Theta(t-s)}\Sigma\left(e^{-\Theta(t-s)}\right)^T ds \end{aligned}$$

Now (Gentle, 2007),

$$\begin{aligned} \text{vec}(Var\left(V_t \mid V_0\right)) &= \int_0^t e^{-\Theta(t-s)} \otimes e^{-\Theta(t-s)}\,\text{vec}(\Sigma)ds \\ &= \int_0^t e^{-\Theta\oplus\Theta(t-s)}ds\,\text{vec}(\Sigma) \\ &= (\Theta\oplus\Theta)^{-1}\left(I - e^{-\Theta\oplus\Theta t}\right)\text{vec}(\Sigma) \end{aligned}$$

where $\oplus$ denotes the Kronecker sum, $A \oplus B = A \otimes I_B + I_A \otimes B$.

Defining the matrix C such that $\text{vec}(C) = (\Theta\oplus\Theta)^{-1}\,\text{vec}(\Sigma)$,

$$\begin{aligned} \text{vec}(Var\left(V_t \mid V_0\right)) &= (\Theta\oplus\Theta)^{-1}\,\text{vec}(\Sigma) - (\Theta\oplus\Theta)^{-1}e^{-\Theta\oplus\Theta t}\,\text{vec}(\Sigma) \\ &= (\Theta\oplus\Theta)^{-1}\,\text{vec}(\Sigma) - e^{-\Theta\oplus\Theta t}(\Theta\oplus\Theta)^{-1}\,\text{vec}(\Sigma) \\ &= \text{vec}(C) - e^{-\Theta\oplus\Theta t}\,\text{vec}(C) \\ &= \text{vec}(C) - e^{-\Theta t} \otimes e^{-\Theta t}\,\text{vec}(C) \end{aligned}$$

Hence,

$$Var(V_t \mid V_0) = C - e^{-\Theta t}Ce^{-\Theta^T t}$$

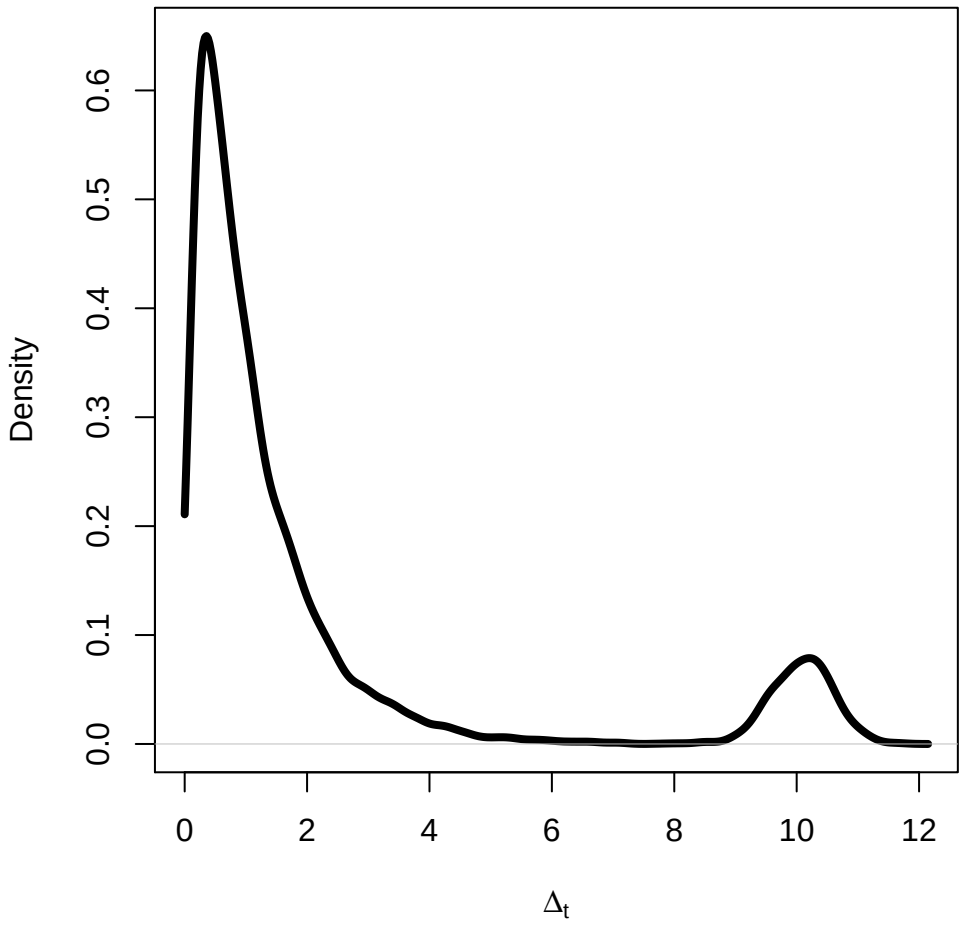


Figure IV.7: Probability density function used to simulate time steps.

References

Gentle, J. (2007). *Matrix algebra: theory, computations, and applications in statistics*. Springer Texts in Statistics. Springer.

IV.B Simulation method details

To simulate from the process,

$$\mathrm{d}V_t = -\Theta(V_t - \mu)\mathrm{d}t + SdB_t \mathrm{d}X_t \quad = V_t \mathrm{d}t$$

time steps,$\Delta_{t_i} = t_i - t_{i-1}$, between observations where simulated from a mixture of an exponential distribution and a normal distribution. To simulate from the mixture, a uniform random variable, $U \sim unif(0,1)$, an exponentially distributed random variable, $\delta_0 \sim exp(1)$, and a Gaussian random variable, $\delta_1 \sim N(10, 0.5^2)$, were simulated. Then the time step was

$$\Delta_{t_i} = \begin{cases} 0.1 + \delta_0 & U < 0.9 \\ 0.1 + \max(\delta_1, 0.01) & U \geq 0.9 \end{cases}$$

The resulting density function is seen in Figure IV.7.

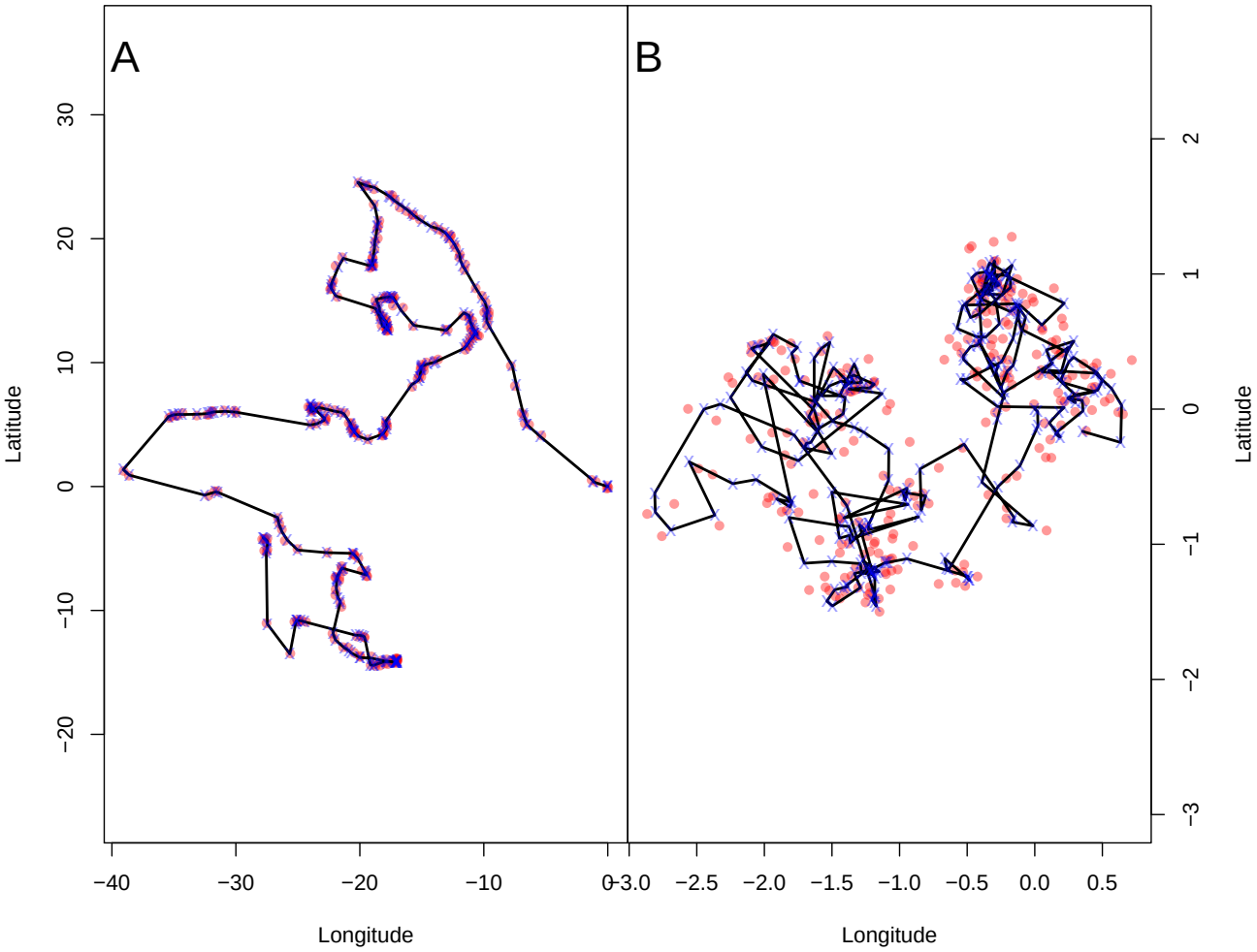


Figure IV.8: Simulated trajectories (blue crosses connected by black lines) and observations (red points) from the SDE model in the two scenarios: persistent (A) and tortuous (B) movement.

Between two time points t_i and t_{i+1}, the processes were simulated using the Euler–Maruyama approximation,

$$V_{s(j+1)} = V_{s(j)} - \Theta(V_n - \mu)\Delta_{s(j+1)} + S\eta_{s(j+1)}X_{s(j+1)} \quad = X_{s(j)} + V_{s(j+1)}\Delta_{s(j+1)}$$

with $s(j) = t_i + (n)/199 \cdot \Delta_{t_{i+1}}, j = 0, 1, \ldots 199$. Simulated examples can be seen in Figure IV.8

Paper V

An approximate filter and smoother for general regime-switching state-space models

Albertsen, C. M.*

National Institute of Aquatic Resources, Technical University of Denmark, Lyngby, Denmark

* **Corresponding author. Email: cmoe@aqua.dtu.dk**

Advance draft

Full citation:

Albertsen, C. M. (2017). An approximate filter and smoother for general regime-switching state-space models. *Advance draft*. In Preparation

Abstract

State-space models with Markov-switching parameters are applicable in many areas including modelling of animal movement, population dynamics, and stochastic volatility. A general filter and smoother is introduced for quickly retrieving the latent regimes and states based on an observed time series. The filter and smoother recursively calculates the likelihood of data and approximates the distribution of latent states given observations through recursive use of the Laplace approximation. The proposed method is illustrated in three examples: a linear Gaussian mode, a stochastic volatility model with t-distributed measurements, and a non-linear population model with Poisson measurements. Further, the supplemental material evaluates the methods through a simulation study. For each simulation, the filter and smoother is compared to a particle filter, and maximum likelihood estimates are found. The proposed method provides accurate maximum likelihood estimates, and state and regime estimates comparable to those of the computationally demanding particle filter.

Keywords:

Laplace approximation, non-linear, non-Gaussian, maximum likelihood inference, stochastic volatility, population model

V.1 Introduction

State-space models provide a valuable framework for analysing time series data from a system, where the process at interest can not be measured directly or exactly. In general, obtaining the likelihood function analytically is intractable. Simulation based methods such as particle filters or Markov Chain Monte Carlo (MCMC) can be used, but is often time-consuming and computationally expensive. Lately, the Laplace approximation has gained popularity through software such as AD Model Builder (Fournier et al., 2012), INLA (Rue et al., 2009), and Template Model Builder (Kristensen et al., 2016). Like many versions of MCMC, all three packages ignore the temporal structure of the state-space model to be applicable in more general settings. In a state-space setting, the Laplace Gaussian Filter (Koyama et al., 2010) utilizes the Laplace approximation to approximate the filtering distribution, while the predictive distribution and smoothing distribution is calculated.

In many systems, it is natural to consider several regimes for the process, where the process exhibits different behaviours, modelled by a finite-state Markov chain. In econometrics, models for economic activity can consider periods of recession and expansion; in behavioural ecology, animal movement models can consider periods of foraging, travelling or nesting; and in stochastic volatility models, the volatility of stock returns can have periods of low or high volatility. When the current regime is not known, calculating the likelihood function is complicated even further, as we must sum over all possible regime histories. In this situation, the Laplace approximation can not be used on the joint likelihood of observations and states. As an alternative to simulation based methods, the Kim filter (Kim, 1994; Kim and Nelson, 1999) combines the Kalman filter (Kalman, 1960) and the Hamilton filter (Hamilton, 1989) for inference in Markov switching linear Gaussian state-space models.

In the following, a general filter and smoother is introduced for approximate maximum likelihood inference in state-space models with regime switching (Figure V.1) based on successive Laplace approxi-

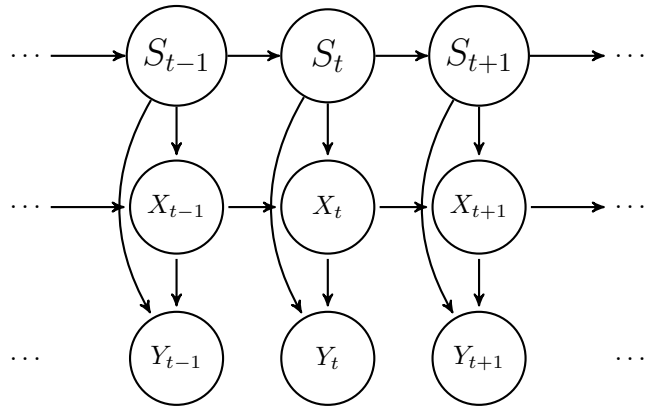


Figure V.1: Dependency graph of the regime switching state-space model. $S_t \in \{1, \ldots, n\}, X_t \in \mathbb{R}^d, Y_t \in \mathcal{Y}$

mations and the Kim filter. The filter and smoother only require specifying the regime-switching probabilities, the state transition densities conditional on the regime, and the measurement distribution conditional on the regime and the states. The remaining calculations can be fully automated. The procedures are illustrated by three examples.

V.2 The state-space model

Consider the general state-space model with Markov switching (Figure V.1) defined by the measurement and transition equations:

$$\begin{aligned} X_t &= f(X_{t-1}, \theta_{S_t}, \epsilon_t) \\ Y_t &= g(X_t, \theta_{S_t}, \nu_t). \end{aligned}$$

Here, f and g are functions of the state process, θ_{S_t} is a parameter vector, and ϵ_t and ν_t are independent random variables. The parameter vector used depends on the current state of an n-state, first-order Markov chain, S_t. The probability of a transition from state j to state k is referred to by $P(S_t = k \mid S_{t-1} = j)$. The state process, X_t, is a real valued vector, while the observations, Y_t, can take values in any space $\mathcal{Y}$.

In the following sections, the aim for this model is twofold: to obtain estimates of the state process given the observations at hand, and to calculate the marginal likelihood function of the observations. Below, p will be used for densities; the distribution referred to will be evident from the context. Note that for observations, the densities can be with respect to either the counting measure or the Lebesgue measure, while it will be with respect to the Lebesgue measure for the states. For simplicity of presentation, we will suppress any dependence on parameters and, further, adopt the short hand notation $X_{i:j} = \{X_i, X_{i+1}, \ldots, X_{j-1}, X_j\}$.

In time series models, it is natural to factorize the likelihood in terms of one-step predictions, $p(Y_t \mid Y_{0:t-1})$. That is,

$$\begin{aligned} L(\theta; Y_{0:T}) &= p(Y_{0:T}) \\ &= p(Y_0) \prod_{t=1}^{T} p(Y_t \mid Y_{0:t-1}). \end{aligned} \tag{V.1}$$

In models with a dependency structure as Figure V.1, $p(Y_t \mid Y_{0:t-1})$ can be calculated by

$$\begin{aligned} p(Y_t \mid Y_{0:t-1}) = \sum_{j=1}^{n}\sum_{k=1}^{n} & p(Y_t \mid Y_{0:t-1}, S_t = j, S_{t-1} = k) \\ & \times P(S_t = j \mid S_{t-1} = k) \\ & \times P(S_{t-1} = k \mid Y_{0:t-1}), \end{aligned} \tag{V.2}$$

where

$$\begin{aligned} p(Y_t \mid Y_{0:t-1}, S_t = j, S_{t-1} = k) = \int_{\mathbb{R}^{2d}} & p(Y_t \mid X_t, S_t = j) \\ & \times p(X_t \mid X_{t-1}, S_t = j, S_{t-1} = k) \\ & \times p(X_{t-1} \mid Y_{0:t-1}, S_{t-1} = k) dX_{t-1:t}, \end{aligned} \tag{V.3}$$

and similarly for $p(Y_0)$.

V.3 Filtering

Equations (V.1)-(V.3) provide the basis for recursively calculating the likelihood function of the regime-switching state-space model. Letting

$$\begin{aligned} g_{Y_t,S_t,S_{t-1}}(X_{t-1}, X_t) = & \log p(Y_t \mid X_t, S_t = j) \\ & + \log p(X_t \mid X_{t-1}, S_t = j, S_{t-1} = k) \\ & \log p(X_{t-1} \mid Y_{0:(t-1)}, S_{t-1} = k), \end{aligned}$$

the density $p(Y_t \mid Y_{0:(t-1)}, S_t = j, S_{t-1} = k)$ can be approximated by the Laplace approximation (See e.g. Skaug and Fournier, 2006). Let $(\hat{X}^{kj}_{t-1|0:t}, \hat{X}^{kj}_{t|0:t})$ be the mode of $g_{Y_t,S_t,S_{t-1}}$, and $H_{Y_t,S_t,S_{t-1}}$ be the hessian of $g_{Y_t,S_t,S_{t-1}}$ evaluated at the mode. Then,

$$\log\left(p(Y_t \mid Y_{0:(t-1)}, S_t = j, S_{t-1} = k)\right) \approx g_{Y_t,S_t,S_{t-1}}(\hat{X}^{kj}_{t-1|0:t}, \hat{X}^{kj}_{t|0:t}) - \frac{1}{2}\log\det H_{Y_t,S_t,S_{t-1}} + d\log(2\pi).$$

Ignoring uncertainty in the parameters, the filtering distribution of the state at time t, for given regimes at time $t-1$ and t, can be approximated by

$$X_t \mid Y_{0:t}, S_t = j, S_{t-1} = k \overset{approx}{\sim} \mathcal{N}\left(\hat{X}^{kj}_{t|0:t}, \hat{V}^{kj}_{t|0:t}\right),$$

where

$$\hat{V}^{kj}_{t|0:t} = \{H_{Y_t,S_t,S_{t-1}}\}^{-1}_{(d+1):2d,(d+1):2d};$$

that is, the marginals of the inverse hessian of $g_{Y_t,S_t,S_{t-1}}$ corresponding to $\hat{X}^{kj}_{t|0:t}$.

Calculating

$$\begin{aligned} P(S_t = j, S_{t-1} = k \mid Y_{0:t}) = & p(Y_t \mid Y_{0:t-1}, S_t = j, S_{t-1} = k) \\ & \times \frac{P(S_t = j \mid S_{t-1} = k) P(S_{t-1} = k \mid Y_{0:t-1})}{p(Y_t \mid Y_{0:t-1})} \end{aligned} \tag{V.4}$$

with

$$\begin{aligned} p(Y_t \mid Y_{0:t-1}) = \sum_{j=1}^{n}\sum_{k=1}^{n} & p(Y_t \mid Y_{0:t-1}, S_t = j, S_{t-1} = k) \\ & \times P(S_t = j \mid S_{t-1} = k) \\ & \times P(S_{t-1} = k \mid Y_{0:t-1}) \end{aligned} \tag{V.5}$$

and

$$P(S_t = j \mid Y_{0:t}) = \sum_{k=1}^{n} P(S_t = j, S_{t-1} = k \mid Y_{0:t}), \tag{V.6}$$

the state estimates can be collapsed from n^2 values, depending on the last two regimes, to n values, depending only on the current regime. Since $\hat{X}^{kj}_{t|0:t}$ represents the mean, $E\left(X_t \mid Y_{0:t}, S_t = j, S_{t-1} = k\right)$, of the approximate distribution of $X_t \mid Y_{0:t}, S_t = j, S_{t-1} = k$, the tower property,

$$E\left(X_t \mid Y_{0:t}, S_t = j\right) = E\left(E\left(X_t \mid Y_{0:t}, S_t = j, S_{t-1} = k\right) \mid Y_{0:t}, S_t = j\right),$$

can be used to calculate

$$\hat{X}^{j}_{t|0:t} = \sum_{k=1}^{n} \frac{\hat{X}^{kj}_{t|0:t} P(S_t = j, S_{t-1} = k \mid Y_{0:t})}{P(S_t = j \mid Y_{0:t})}$$

and

$$\hat{V}^{j}_{t|0:t} = \sum_{k=1}^{n} \frac{P(S_t = j, S_{t-1} = k \mid Y_{0:t}) \left(\hat{V}^{jk}_{t|0:t} + (\hat{X}^{j}_{t|0:t} - \hat{X}^{jk}_{t|0:t})(\hat{X}^{j}_{t|0:t} - \hat{X}^{jk}_{t|0:t})^T\right)}{P(S_t = j \mid Y_{0:t})},$$

thus obtaining an approximation of the regime-wise filtering distributions

$$X_t \mid Y_{0:t}, S_t = j \overset{approx}{\sim} \mathcal{N}\left(\hat{X}^{j}_{t|0:t}, \hat{V}^{j}_{t|0:t}\right).$$

The tower property can be used once more to get an approximate regime-independent filtering distribution.

V.4 Smoothing

While the filter approximates the distributions $X_t \mid Y_{0:t}, S_t = j$, the aim of the smoother is to approximate the distribution of the states X_t given all available information, $Y_{0:T}$:

$$X_t \mid Y_{0:T}, S_t = j.$$

The density of $X_t \mid Y_{0:T}, S_t = j, S_{t+1} = k$ can be calculated recursively, going backwards in time, by the smoothing formula (e.g., Kitagawa, 1987):

$$\begin{aligned} p(X_t \mid Y_{0:T}, S_t = j, S_{t+1} = k) = \int_{\mathbb{R}^d} & p(X_t \mid Y_{0:t}, S_t = j) p(X_{t+1} \mid Y_{0:T}, S_{t+1} = k) \\ & \times \frac{p(X_{t+1} \mid X_t, S_{t+1} = k, S_t = j)}{p(X_{t+1} \mid Y_{0:t}, S_{t+1} = k, S_t = j)} dX_{t+1} \end{aligned} \tag{V.7}$$

where

$$\begin{aligned} p(X_{t+1} \mid Y_{0:t}, S_{t+1} = k, S_t = j) = \int_{\mathbb{R}^d} & p(X_{t+1} \mid X_t, S_{t+1} = k, S_t = j) \\ & \times p(X_t \mid Y_{0:t}, S_t = j) dX_t. \end{aligned} \tag{V.8}$$

However, using equations (V.7) and (V.8) to obtain the mode and hessian at the mode of $p(X_t \mid Y_{0:T}, S_t = j, S_{t+1} = k)$ requires two applications of the Laplace approximation for each iteration

in an optimization procedure for finding the mode. Instead, by calculating the integral

$$\begin{aligned}\int_{\mathbb{R}^d} p(X_t \mid Y_{0:T}, S_{t+1}=k, S_t=j)dX_t &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} p(X_t \mid Y_{0:t}, S_t=j)p(X_{t+1} \mid Y_{0:T}, S_{t+1}=k) \\ &\quad \times \frac{p(X_{t+1} \mid X_t, S_{t+1}=k, S_t=j)}{p(X_{t+1} \mid Y_{0:t}, S_{t+1}=k, S_t=j)} dX_{t+1}dX_t \\ &= \int_{\mathbb{R}^{2d}} p(X_t \mid Y_{0:t}, S_t=j)p(X_{t+1} \mid Y_{0:T}, S_{t+1}=k) \\ &\quad \times \frac{p(X_{t+1} \mid X_t, S_{t+1}=k, S_t=j)}{p(X_{t+1} \mid Y_{0:t}, S_{t+1}=k, S_t=j)} dX_{t:t+1}\end{aligned} \tag{V.9}$$

through the Laplace approximation, we obtain the mode, $\bar{X}^{(jk)}_{t|0:T}$, and the inverse hessian at the mode, $\bar{V}^{(jk)}_{t|0:T}$, with only two applications of the Laplace approximation. From the mode and the inverse hessian, the Gaussian approximation,

$$X_t \mid Y_{0:T}, S_{t+1}=k, S_t=j \overset{approx}{\sim} \mathcal{N}\left(\bar{X}^{(jk)}_{t|0:T}, \bar{V}^{(jk)}_{t|0:T}\right),$$

is constructed. Like the filter, the Laplace approximation works on the joint distribution,

$$X_t, X_{t+1} \mid Y_{0:T}, S_{t+1}=k, S_t=j,$$

from which the marginals corresponding to

$$X_t \mid Y_{0:T}, S_{t+1}=k, S_t=j$$

are extracted.

Approximating the smoothing distribution iteratively through equations (V.8) and (V.9) only requires three quantities: $p(X_t \mid Y_{0:t}, S_t=j)$, which is known from the filter; $p(X_{t+1} \mid Y_{0:T}, S_{t+1}=k)$; which is known from the previous step of the smoother; and $p(X_{t+1} \mid X_t, S_{t+1}=k, S_t=j)$, which is known from the model. After each iteration the states must be collapsed from n^2 to n values, as in the filter, to obtain an approximation of

$$X_t \mid Y_{0:T}, S_t=j.$$

To collapse the states, the smoothed regime probabilities are calculated as

$$P(S_t=j \mid Y_{0:T}) = \sum_k P(S_t=j, S_{t+1}=k \mid Y_{0:T}),$$

where

$$\begin{aligned}P(S_t=j, S_{t+1}=k \mid Y_{0:T}) &= P(S_{t+1}=k \mid Y_{0:T})P(S_t=j \mid S_{t+1}=k, Y_{0:T}) \\ &\approx P(S_{t+1}=k \mid Y_{0:T})P(S_t=j \mid S_{t+1}=k, Y_{0:t}) \\ &= \frac{P(S_{t+1}=k \mid Y_{0:T})P(S_t=j \mid Y_{0:t})P(S_{t+1}=k \mid S_t=j)}{P(S_{t+1}=k \mid Y_{0:t})}\end{aligned}$$

From these, the states and their variances are collapsed as in the filter,

$$\bar{X}^{(j)}_{t|0:T} = \frac{\sum_k P(S_t=j, S_{t+1}=k \mid Y_{0:T})\bar{X}^{(jk)}_{t|0:T}}{P(S_t=j \mid Y_{0:T})}$$

and

$$\bar{V}^{(j)}_{t|0:T} = \sum_{k=1}^{n} \frac{P(S_t = j, S_{t+1} = k \mid Y_{0:T}) \left(\bar{V}^{(jk)}_{t|0:T} + (\bar{X}^{(j)}_{t|0:T} - \bar{X}^{(jk)}_{t|0:T})(\bar{X}^{(j)}_{t|0:T} - \bar{X}^{(jk)}_{t|0:T})^T \right)}{P(S_t = j \mid Y_{0:T})},$$

to obtain the approximate regime-wise smoothing distribution

$$X_t \mid Y_{0:T}, S_t = j \overset{approx}{\sim} N(\bar{X}^{(j)}_{t|0:T}, \bar{V}^{(j)}_{t|0:T}).$$

By collapsing the states once more, an approximate regime-independent smoothing distribution can be obtained.

V.5 Automated maximum likelihood inference

Given the regime transition probabilities $P(S_t = j \mid S_{t-1} = k)$, the regime-wise state transition densities $p(X_t \mid X_{t-1}, S_t = j)$, and the regime-wise measurement densities $p(Y_t \mid X_t, S_t = j)$, the filter and smoother can be completely automated using algorithmic differentiation (See, e.g., Skaug and Fournier, 2006; or Neidinger, 2010, for a recent introductory review). Algorithmic (or automatic) differentiation (AD) calculates derivatives of a computer implemented function using the chain rule by going through the computational graph. For a given parameter vector, AD can be used to obtain the gradient and hessian of g, and Newton's method can be used to retrieve the mode and subsequently the Laplace approximated integral value.

Implementing the entire filter to calculate the negative log-likelihood by equation (V.5),

$$\ell(\theta) = -\sum_{t=0}^{T} \log \left(p_\theta \left(Y_t \mid Y_{0:t} \right) \right),$$

AD can be used on the likelihood implementation to calculate the score and observed information for a given parameter vector. From the approximate negative log-likelihood, score, and observed hessian, the maximum likelihood estimates can be found by a standard optimization procedure. The smoother, which is computationally expensive compared to the filter, only needs to be evaluated once at the parameter estimates.

V.6 Examples

To illustrate the proposed procedure, this section considers three simulated examples of regime switching state-space models: a linear Gaussian model, a stochastic volatility model with t-distributed measurements, and a non-linear population model with Poisson distributed measurements. All three examples were simulated with two regimes, $n = 2$, and 1,000 time steps, $T = 999$.

The models were implemented in TMB using CppAD (Bell, 2005) for algorithmic differentiation. Since TMB only builds the computational graph once, the number of iterations in the Newton optimizer in the Laplace approximations was fixed. For the first time step, 100 iterations were used, while 10 iterations were used for subsequent time steps. The optimization for the Laplace approximation was started at 0 for the first time step, while the remaining time steps used the previous state estimate as the starting point. Note, however, that this is not a requirement for the proposed procedure, but an implementation detail for the specific tools used here.

V.6.1 Linear Gaussian state-space model

The first example is a linear Gaussian state-space model, where the parameters depend on the current regime:

$$\begin{aligned} P(S_t = j \mid S_{t-1} = k) &= p_{kj}, \quad t > 0, \\ X_t \mid X_{t-1}, S_t = j &\sim N(\mu_j + \alpha_j(X_{t-1} - \mu_j), \sigma_j^2), \quad t > 0, \text{ and} \\ Y_t \mid X_t, S_t = j &\sim N(X_t, \tau_j^2), \quad t \geq 0. \end{aligned}$$

That is, the state dynamics, X_t, follow an AR(1) process with regime dependent mean and autocorrelation, while observations, Y_t, given the current state follow a normal distribution with regime dependent variance.

In the simulation, X_0 was fixed at zero, while the filter was initialized with $X_0 \mid S_0 = j \sim N(0, 10^2)$. Likewise, S_0 was fixed to the first regime in the simulations, while the filter was initialized with the stationary distribution of the Markov chain given by the regime transition probabilities. The model was simulated with regime transition probabilities such that $p_{11} = p_{22} = 0.95$, while the regime-wise process means were $\mu_1 = -\frac{1}{2}$ and $\mu_2 = -\frac{1}{2}$. Both the autocorrelations and variance parameters were equal amongst the regimes: $\alpha_1 = \alpha_2 = 0.8$, and $\sigma_1 = \sigma_2 = \tau_1 = \tau_2 = 0.1$.

The simulated data and smoothed state and regime estimates are shown in Figure V.2. The filter correctly identified 91% of the regimes. Of the remaining time steps, 4% were wrongly identified as regime one, while 5% were wrongly identified as regime two. Using the smoother, the regime decoding improved. The smoother correctly identified 96% of the regimes, while 2% were wrongly identified as regime one, and 2% were wrongly identified as regime two.

Over the entire time series, both the filter and smoother provided accurate state estimates. On average, the difference between the filtered states and the true states, $\hat{X}_{t|0:t} - X_t$, was 0.0034. For the smoother, the average difference, $\bar{X}_{t|0:T} - X_t$, was 0.0001. The average distances correspond to 0.19 and 0.01% of the range of true state values, respectively. Likewise, the uncertainty assessments were accurate for both procedures. The true state value was contained in 95.7% of the point-wise confidence intervals provided by the filter, while the corresponding coverage for the smoother was 96.2%.

V.6.2 Stochastic volatility model

The second example is a stochastic volatility model. Again, the state dynamics, X_t, follow an AR(1) process with regime dependent parameters; however, while the state was used to model the mean of the observations in the previous example, it is now used to model the scale. Scaled by the exponential of the states and a scale parameter, the observations follow a Student's t-distribution with regime dependent degrees of freedom:

$$\begin{aligned} P(S_t = j \mid S_{t-1} = k) &= p_{kj}, \quad t > 0, \\ X_t \mid X_{t-1}, S_t = j &\sim N(\mu_j + \alpha_j(X_{t-1} - \mu_j), \sigma_j^2), \quad t > 0, \text{ and} \\ Y_t \tau_j^{-1} \exp(-0.5 \cdot X_t) \mid X_t, S_t = j &\sim t(\nu_j), \quad t \geq 0. \end{aligned}$$

Since the scale parameter τ_j and the state variance σ_j^2 can not both be estimated, the scale parameter was fixed at $\tau_1 = \tau_2 = \exp(-5)$. As in the previous example, the state autocorrelations and variances were the same in the two regimes; this time with $\alpha_1 = \alpha_2 = 0.9$ and $\sigma_1 = \sigma_2 = 0.2$. The regime-wise means were $\mu_1 = 0$ and $\mu_2 = 1$, while the degrees of freedom were $\nu_1 = 16$ and $\nu_2 = 4$, respectively.

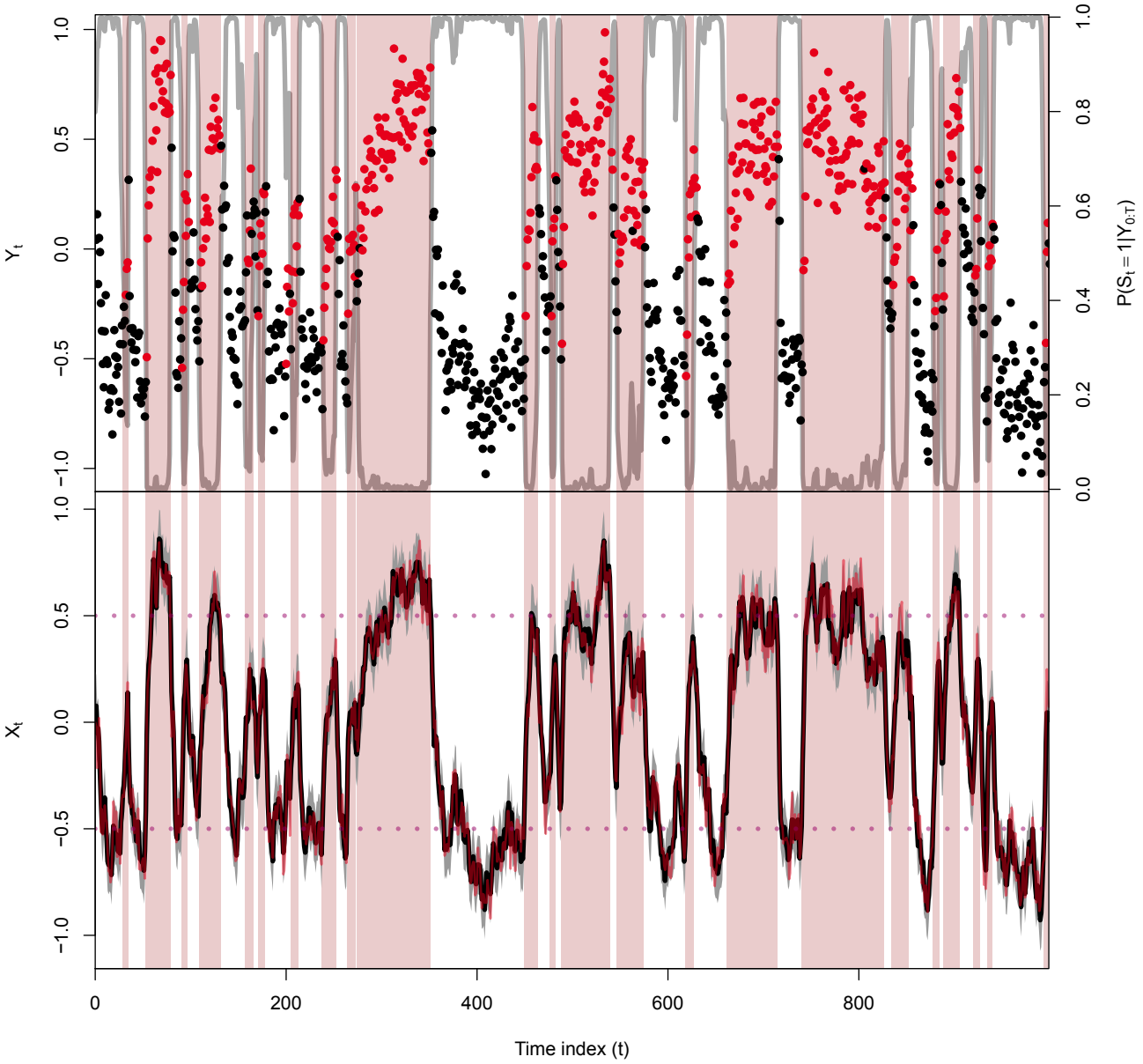


Figure V.2: Simulated data with smoothed state and regime estimates in the linear Gaussian example. Upper panel: Simulated observations (points), true regimes (black: regime 1; red: regime 2), and smoothed regime probability $P(S_t = 1 \mid Y_{0:T})$ (grey line, right axis). Lower panel: Simulated true state trajectory (red line), smoothed state estimates (black line) and their confidence interval (grey region). Dashed purple lines indicate μ_1 and μ_2. In both panels, the background is white in periods with $P(S_t = 1 \mid Y_{0:T}) \geq 0.5$ and red in periods with $P(S_t = 1 \mid Y_{0:T}) < 0.5$.

Hence, the first regime represents periods with nearly Gaussian observations with a small variance, while the second regime represents periods of high volatility with heavy tailed observations and high variance.

Figure V.3 depicts the simulated data, smoothed state estimates and regime estimates. In this example, the regimes were more difficult to decode. The simulated time series had several short regime periods which were not correctly decoded. Consequently, the filter correctly identified 72% of the regimes, 14% were wrongly identified as regime one, and 14% were wrongly identified as regime two. Again, the regime decoding success was higher for the smoother, which correctly identified 80% of the regimes. Of the remaining time points, 9 and 11% were wrongly identified as regime one and two, respectively.

On average, the filter and smoother once more provided state estimates close to the true values. The average difference between the estimated states and the true states in the simulation was -0.0381 for the filter and -0.0545 for the smoother. Compared to the range of the true state values, the average distances correspond to 1.24 and 1.77%, respectively. Furthermore, the confidence intervals provided accurate assessments of the uncertainty in state estimates for both the filter and smoother. For the filter, the true state value was contained in 96.4% of the point-wise confidence intervals, while the true value was contained by the confidence interval provided by the smoother for 96.7% of the time steps.

V.6.3 Non-linear population model

The final example is a non-linear population model with Poisson measurements. The population is exposed to changes in the environment, which result in regime shifts in the population dynamics (Munch and Kottas, 2009). The model is a discretised special case of the log-population model considered by Pedersen and Berg (2017):

$$
\begin{aligned}
P(S_t = j \mid S_{t-1} = k) &= p_{kj} \\
X_t \mid X_{t-1}, S_t = j &\sim N(X_{t-1} + \frac{R_j}{K_j} - \frac{R_j}{K_j}\frac{\exp(X_{t-1})}{K_j} - \frac{1}{2}\sigma_j^2, \sigma_j^2) \\
Y_t \mid X_t, S_t = j &\sim Pois(\tau_j \cdot \exp(X_t))
\end{aligned}
$$

Here, K_j is the regime-wise equilibrium, or carrying capacity, in the absence of process noise, while $\frac{R_j}{K_j}$ is the population growth rate. That is, the process will alternate between tending to each of the carrying capacities, K_j, with a rate determined by $\frac{R_j}{K_j}$; emulating a population subject to, for instance, epizootics that causes the population to decrease rapidly for a period before recovering. The $-\frac{1}{2}\sigma_j^2$ term in the transition density mean results from the transformation from a population model with multiplicative noise to a log-population model with additive noise (Pedersen and Berg, 2017).

In the simulation, the regime transition probabilities were given by $p_{1,1} = p_{2,2} = 0.98$. The population carrying capacities were $K_1 = \exp(-1)$ and $K_2 = \exp(1)$, while the population growth rates were $\frac{R_1}{K_1} = \frac{R_2}{K_2} = 0.04$. Finally, the process noise parameters were $\sigma_1 = \sigma_2 = 0.1$. Like in the previous example, τ_j is not identifiable, but determines the transformation from process scale to observation scale. Here, τ_1 and τ_2 were fixed at 100.

Data simulated from the model along with smoothed state and regime estimates are shown in Figure V.4. In this example, the regime decoding success was higher than in the volatility example. The filter decoded 84% of the regimes correctly, while it was 97% for the smoother. In the filter, 12 and 4% of the states were erroneously identified as regime one and two, respectively. Likewise, 2 and 1% were wrongly identified as regime one and two, respectively, in the smoother. There is, in particular, one period where the regime is wrongly decoded. Near the beginning of the time series, the observed population decreases

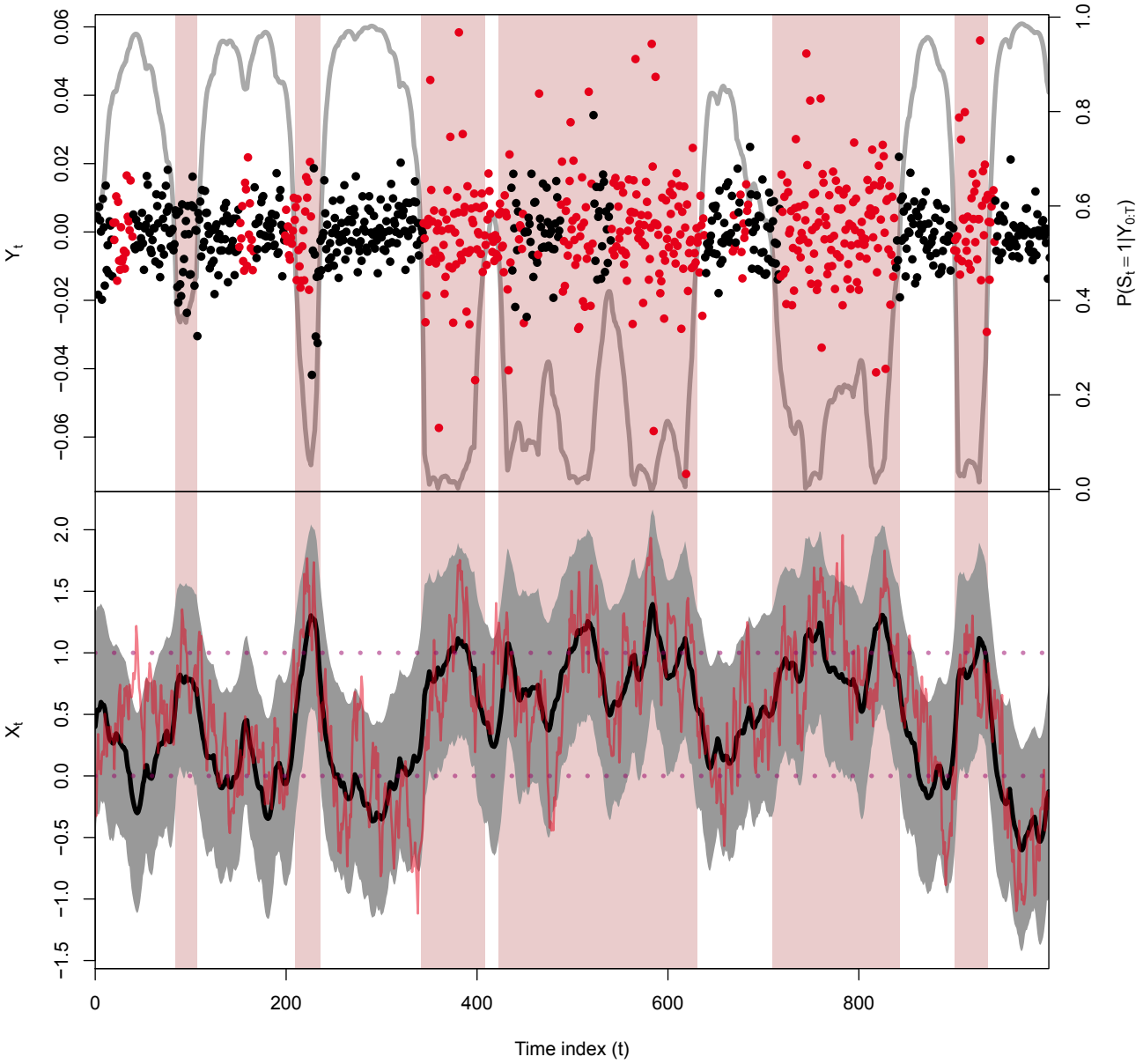


Figure V.3: Simulated data with smoothed state and regime estimates in the stochastic volatility example. Upper panel: Simulated observations (points), true regimes (black: regime 1; red: regime 2), and smoothed regime probability $P(S_t = 1 \mid Y_{0:T})$ (grey line, right axis). Lower panel: Simulated true state trajectory (red line), smoothed state estimates (black line) and their confidence interval (grey region). Dashed purple lines indicate μ_1 and μ_2. In both panels, the background is white in periods with $P(S_t = 1 \mid Y_{0:T}) \geq 0.5$ and red in periods with $P(S_t = 1 \mid Y_{0:T}) < 0.5$.

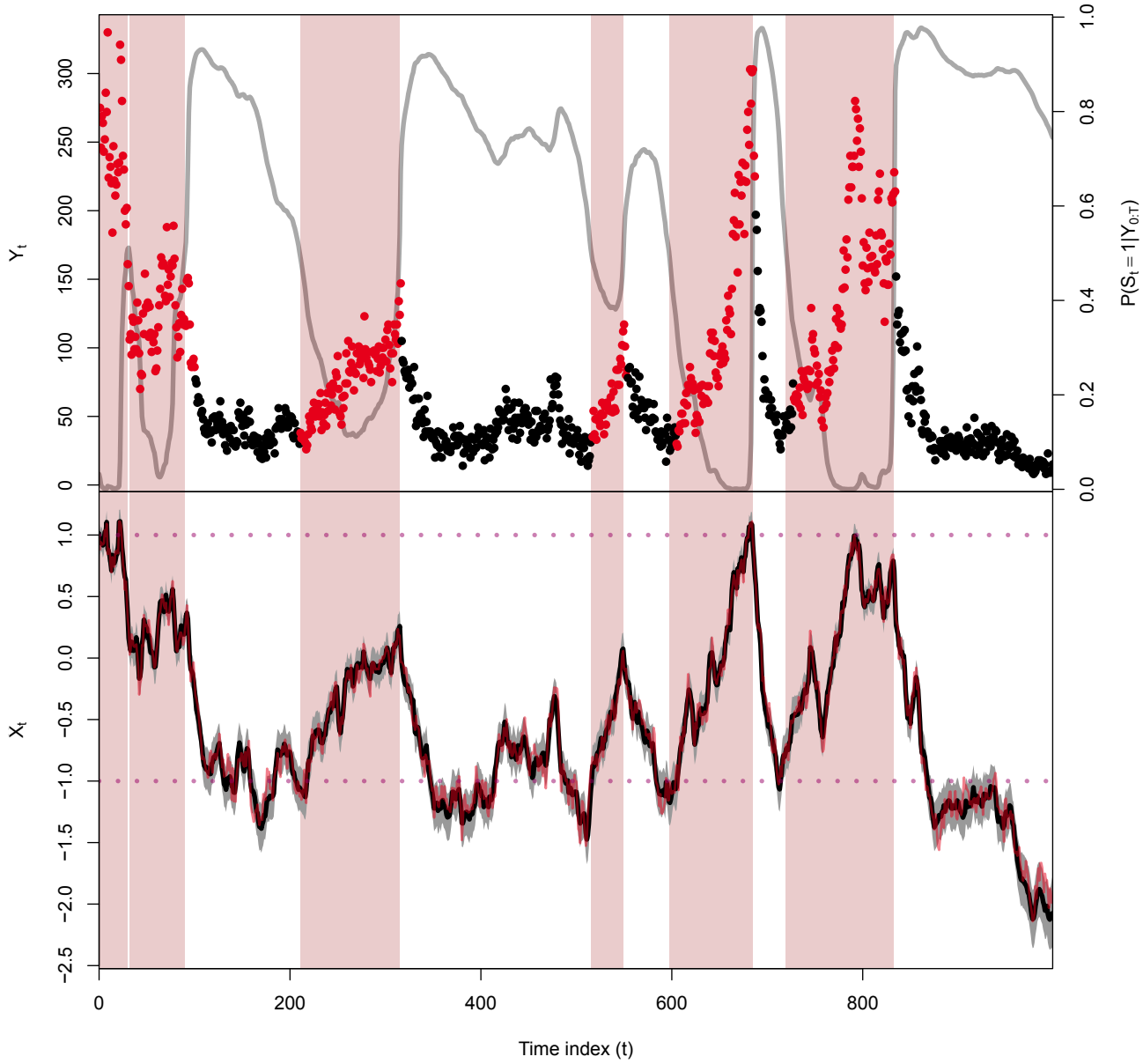


Figure V.4: Simulated data with smoothed state and regime estimates in the non-linear population example. Upper panel: Simulated observations (points), true regimes (black: regime 1; red: regime 2), and smoothed regime probability $P(S_t = 1 \mid Y_{0:T})$ (grey line, right axis). Lower panel: Simulated true state trajectory (red line), smoothed state estimates (black line) and their confidence interval (grey region). Dashed purple lines indicate $\log(K_1)$ and $\log(K_2)$. In both panels, the background is white in periods with $P(S_t = 1 \mid Y_{0:T}) \geq 0.5$ and red in periods with $P(S_t = 1 \mid Y_{0:T}) < 0.5$.

although the true process is still in the high population regime. This period is decoded as regime one; that is, as a population decline towards the lower carrying capacity.

Once more, the filter and smoother provided accurate state estimates. On average, the difference between the filtered states and the true states was -0.0117. Likewise, the average difference for the smoother was -0.0106. These average distances correspond to 0.36 and 0.33% of the range of true state values, respectively. Again, the state uncertainty assessments were accurate for both the filter and smoother. The true location was contained in 95.2 and 94.8% of the point-wise confidence intervals provided by the filter and smoother, respectively.

V.7 Conclusions

The approximate filter and smoother extends the Kim filter (Kim, 1994) to non-linear non-Gaussian regime-switching state-space models. Through three simulated examples, it was shown that the procedures provide accurate regime and state estimates. Further, simulation studies (Appendix V.A) show that the filter could be used to attain maximum likelihood estimates close to the true value, although it can be difficult in some cases. The simulation studies also compared the accuracy of the filter and smoother to a particle filter. In all three examples, the approximate filter provided better regime and state estimates than a particle filter with 100 particles, while it was comparable to particle filters with 1,000 and 5,000 particles. As expected, the smoother provided better regime and state estimates than all four filters.

An advantage of filters over methods ignoring the temporal structure is the ability to calculate residuals during the recursion. In the approximate filter, quantile residuals may be calculated recursively through $P(Y_t \leq y_t \mid Y_{0:t-1})$ (Früiiwirth-Schnatter, 1996). Thygesen et al. (2017) outlines the calculation via the Laplace approximation for general mixed-effects models. A similar approach may be used for the regime-wise calculations, $P(Y_t \leq y_t \mid Y_{0:t-1}, S_t = j, S_{t-1} = k)$, recognizing the simplifying structure of the time series, which can then be summed as in Equation (V.5).

The quality of the approximate inference will depend on the quality of each Laplace approximation. While the Laplace approximation is often a poor point-wise approximation, it is an accurate integral approximation in many cases, in particular if the model is suitably parameterised. In cases where the accuracy is questionable, the proposed filter and smoother can be altered to use more accurate integration methods such as Gaussian quadrature. While using, for instance, Gaussian quadrature to obtain the marginal likelihood of observations is infeasible on the entire time series, the proposed filter reduces the integral dimension. That is, instead of one high-dimensional integral, the proposed filter calculates several low-dimensional integrals; therefore, making numerical integration possible. Further, the filter and smoother can be used to initialize more computationally demanding methods, where the accuracy can be increased, such as particle filters or MCMC.

References

Bell, B. M. (2005). *CppAD: a package for C++ algorithmic differentiation*. Retrieved from http://www.coin-or.org/CppAD

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., … Sibert, J. (2012). AD model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, *27*(2), 233–249. doi:10.1080/10556788.2011.597854

Früiiwirth-Schnatter, S. (1996). Recursive residuals and model diagnostics for normal and non-normal state space models. *Environmental and Ecological Statistics*, *3*(4), 291–309. doi:10 . 1007 / BF00539368

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, *57*(2), 357. doi:10.2307/1912559

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, *82*(1), 35. doi:10.1115/1.3662552

Kim, C.-J. (1994). Dynamic linear models with markov-switching. *Journal of Econometrics*, *60*(1-2), 1–22. doi:10.1016/0304-4076(94)90036-1

Kim, C.-J. & Nelson, C. R. (1999). *State-space models with regime switching: classical and gibbs-sampling approaches with applications*. MIT Press.

Kitagawa, G. (1987). Non-gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association*, *82*(400), 1032. doi:10.2307/2289375

Koyama, S., Pérez-Bolde, L. C., Shalizi, C. R., & Kass, R. E. (2010). Approximate methods for state-space models. *Journal of the American Statistical Association*, *105*(489), 170–180. doi:10.1198/jasa.2009.tm08326

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: automatic differentiation and laplace approximation. *Journal of Statistical Software*, *70*(5). doi:10.18637/jss.v070.i05

Munch, S. B. & Kottas, A. (2009). A bayesian modeling approach for determining productivity regimes and their characteristics. *Ecological Applications*, *19*(2), 527–537. doi:10.1890/07-2116.1

Neidinger, R. D. (2010). Introduction to automatic differentiation and MATLAB object-oriented programming. *SIAM Review*, *52*(3), 545–563. doi:10.1137/080743627

Pedersen, M. W. & Berg, C. W. (2017). A stochastic surplus production model in continuous time. *Fish and Fisheries*, *18*(2), 226–243. doi:10.1111/faf.12174

Rue, H., Martino, S., & Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319–392. doi:10.1111/j.1467-9868.2008.00700.x

Skaug, H. J. & Fournier, D. A. (2006). Automatic approximation of the marginal likelihood in non-gaussian hierarchical models. *Computational Statistics & Data Analysis*, *51*(2), 699–709. doi:10.1016/j.csda.2006.03.005

Thygesen, U. H., Albertsen, C. M., Berg, C. W., Kristensen, K., & Nielsen, A. (2017). Validation of ecological state space models using the laplace approximation. *Environmental and Ecological Statistics*, *24*(2), 317–339. doi:10.1007/s10651-017-0372-4

V.A Simulation studies

To evaluate the proposed filter and smoother, a simulation study was conducted. Data were simulated for each example with 5,000 time points and $n = 1, 2, 3$ regimes. For each data set, the regime decoding success and root mean squared error between estimated states and true states were calculated for the filter, smoother, and bootstrap particle filters with 100, 1000, and 5000 particles, respectively. Further, the filter was used to obtain maximum likelihood estimates.

In the linear Gaussian example, data were simulated with $\mu_1 = \frac{-1}{4}$ for $n = 1$; $\mu_1 = \frac{-1}{2}$, $\mu_2 = \frac{1}{2}$ for $n = 2$; and $\mu_3 = \frac{-3}{4}$, $\mu_2 = 0$, and $\mu_3 = \frac{3}{4}$ for $n = 3$. The probabilities of staying in a regime were $p_{11} = p_{22} = p_{33} = 0.98$, while the remaining transition probabilities were $\frac{0.02}{n-1}$. All other parameters were defined as in the example.

In the stochastic volatility example, data were simulated with $\mu_1 = 0$ for $n = 1$; $\mu_1 = 0$, $\mu_2 = 1$ for $n = 2$; and $\mu_1 = 0$, $\mu_2 = \frac{3}{4}$, and $\mu_3 = \frac{3}{2}$ for $n = 3$. The probabilities of staying in a regime were $p_{11} = p_{22} = p_{33} = 0.98$, while the remaining transition probabilities were $\frac{0.02}{n-1}$. The degrees of freedom were $\nu_1 = 16$ for $n = 1$; $\nu_1 = 16$ and $\nu_2 = 4$ for $n = 2$; $\nu_1 = 16$, $\nu_2 = 10$, and $\nu_3 = 4$ for $n = 3$. All other parameters were defined as in the example.

In the non-linear population example, data were simulated with $K_1 = \frac{-1}{2}$ for $n = 1$; $K_1 = \frac{-1}{2}$, $K_2 = 1$ for $n = 2$; and $K_1 = \frac{-1}{2}$, $K_2 = \frac{1}{4}$, and $K_3 = 1$ for $n = 3$. The probabilities of staying in a regime were $p_{11} = p_{22} = p_{33} = 0.99$, while the remaining transition probabilities were $\frac{0.01}{n-1}$. All other parameters were defined as in the example.

The results are shown below. Comparisons with particle filters are summarized in bean plots (Kampstra, 2008), while the estimates are summarized in tables. Parameter estimates are compared to the true values on the scale at which they were estimated; not the scale specified above.

References

Kampstra, P. (2008). Beanplot: a boxplot alternative for visual comparison of distributions. *Journal of Statistical Software, Code Snippets*, *28*(1), 1–9. doi:10.18637/jss.v028.c01

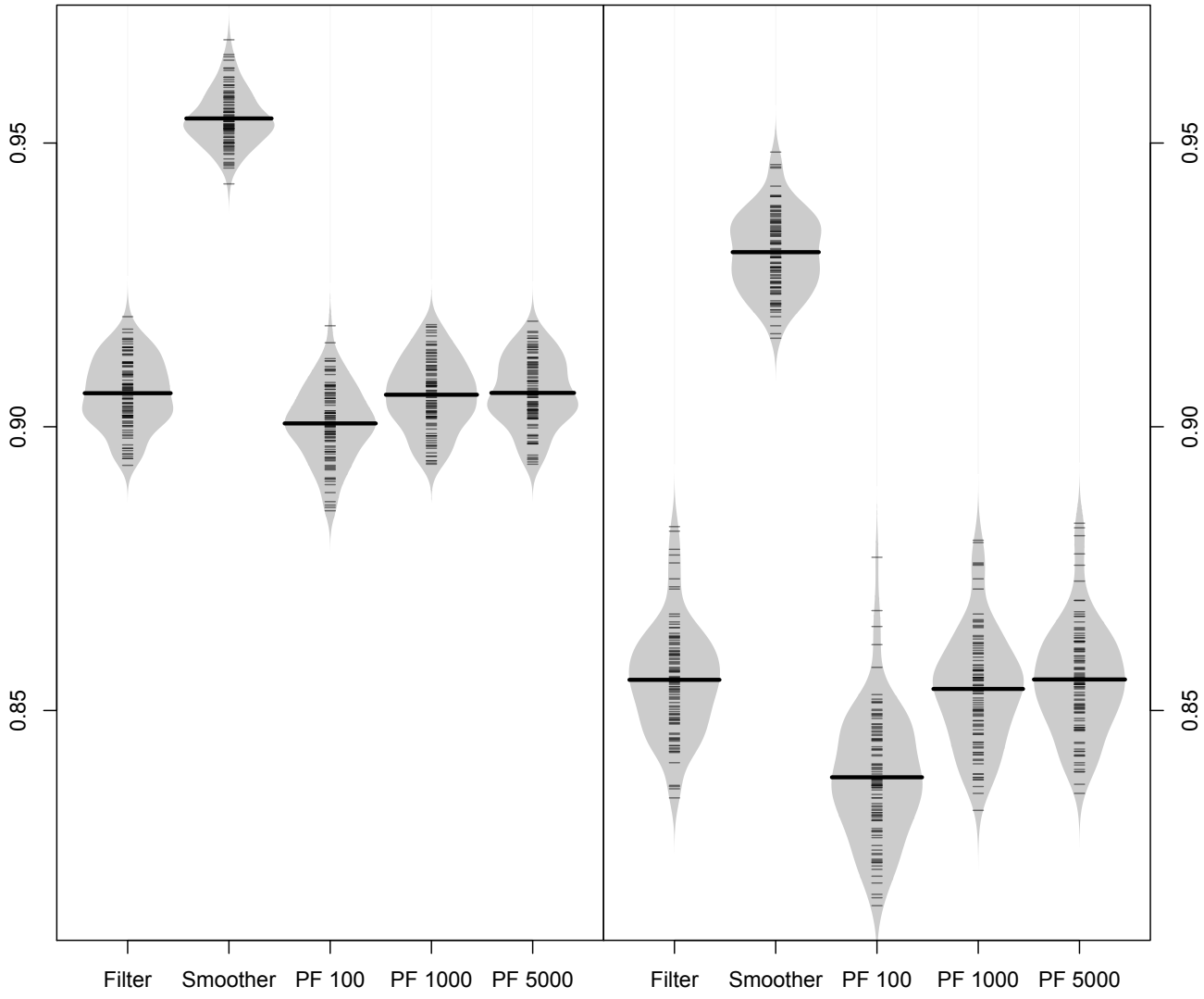


Figure V.5: Regime decoding success from the approximate filter, approximate smoother, a particle filter with 100 particles (PF 100), a particle filter with 1,000 particles (PF 1000), and a particle filter with 5,000 particles (PF 5000) in the linear simulation study with two regimes (left panel) and three regimes (right panel). In each bean, the grey area shows a kernel density estimate based on each value from the simulation study, which are shown by short lines. The thick line spanning the grey area indicates the mean value.

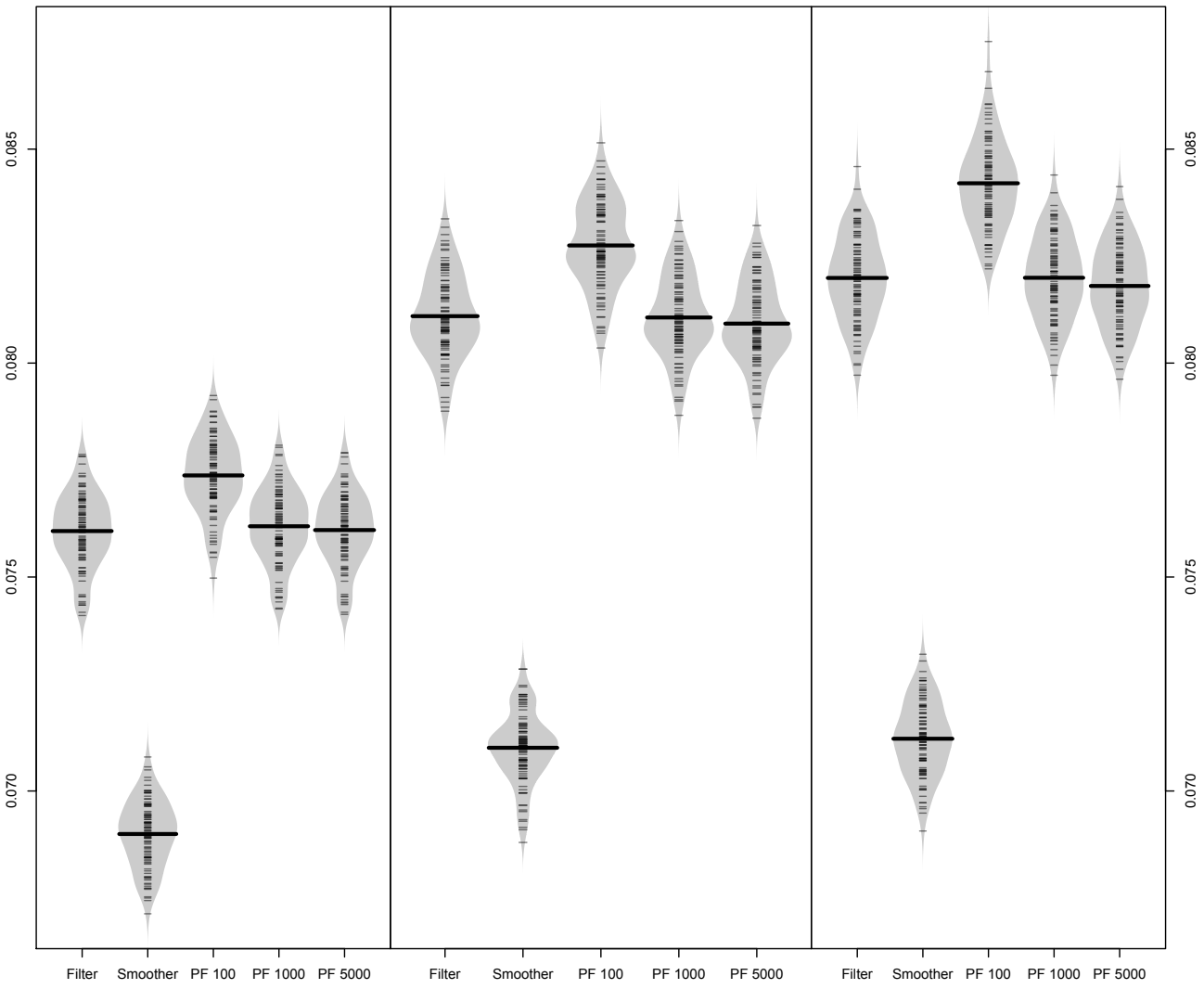


Figure V.6: Beanplot of the root mean squared error of state predictions from the approximate filter, approximate smoother, a particle filter with 100 particles (PF 100), a particle filter with 1,000 particles (PF 1000), and a particle filter with 5,000 particles (PF 5000) in the linear simulation study with one regime (left panel), two regimes (middle panel), and three regimes (right panel). In each bean, the grey area shows a kernel density estimate based on each value from the simulation study, which are shown by short lines. The thick line spanning the grey area indicates the mean value.

Table V.1: Summary of parameter estimates from the linear Gaussian simulation study with one regime

Parameter	True	Median	Mean	Min.	1st Qu.	3rd Qu.	Max
$\log \tau$	-2.303	-2.300	-2.308	-2.483	-2.355	-2.266	-2.166
$\log \sigma$	-2.303	-2.319	-2.311	-2.506	-2.355	-2.272	-2.120
μ_1	-0.250	-0.251	-0.250	-0.293	-0.263	-0.237	-0.214
$\text{logit}\, \alpha$	1.386	1.389	1.389	0.908	1.251	1.508	1.808

Table V.2: Summary of parameter estimates from the linear Gaussian simulation study with two regimes

Parameter	True	Median	Mean	Min.	1st Qu.	3rd Qu.	Max
$\log \tau$	-2.303	-2.317	-2.321	-2.550	-2.359	-2.277	-2.155
$\log \sigma$	-2.303	-2.284	-2.289	-2.457	-2.346	-2.240	-2.049
$\tilde{p}_1$	2.944	2.932	2.945	2.324	2.782	3.094	3.475
$\tilde{p}_2$	-2.944	-2.895	-2.917	-3.811	-3.054	-2.758	-2.298
μ_1	-0.500	-0.500	-0.504	-0.573	-0.524	-0.481	-0.427
$\log(\mu_2 - \mu_1)$	0.000	-0.006	-0.003	-0.120	-0.030	0.028	0.128
$\text{logit}\, \alpha$	1.386	1.289	1.294	0.942	1.214	1.376	1.581

Table V.3: Summary of parameter estimates from the linear Gaussian simulation study with three regimes

Parameter	True	Median	Mean	Min.	1st Qu.	3rd Qu.	Max
$\log \tau$	-2.303	-2.326	-2.332	-2.656	-2.364	-2.287	-2.205
$\log \sigma$	-2.303	-2.296	-2.284	-2.480	-2.361	-2.229	-2.021
$\tilde{p}_1$	3.638	3.590	3.343	-17.278	3.074	3.910	20.212
$\tilde{p}_2$	0.000	-0.212	-3.491	-23.079	-1.409	0.384	19.538
$\tilde{p}_3$	-3.638	-3.405	-3.608	-21.863	-3.726	-3.101	-1.355
$\tilde{p}_4$	0.000	-0.100	-0.456	-18.975	-0.649	0.723	18.980
$\tilde{p}_5$	3.638	3.550	4.111	2.598	3.295	3.975	22.945
$\tilde{p}_6$	-3.638	-3.822	-6.095	-25.639	-4.439	-3.350	-2.204
μ_1	-0.750	-0.756	-0.897	-3.068	-0.787	-0.731	-0.646
$\log(\mu_2 - \mu_1)$	-0.288	-0.306	-1.317	-22.493	-0.367	-0.243	0.981
$\log(\mu_3 - \mu_1)$	-0.288	-0.254	-0.147	-0.472	-0.306	-0.157	0.484
$\text{logit}\, \alpha$	1.386	1.323	1.448	1.035	1.250	1.506	2.378

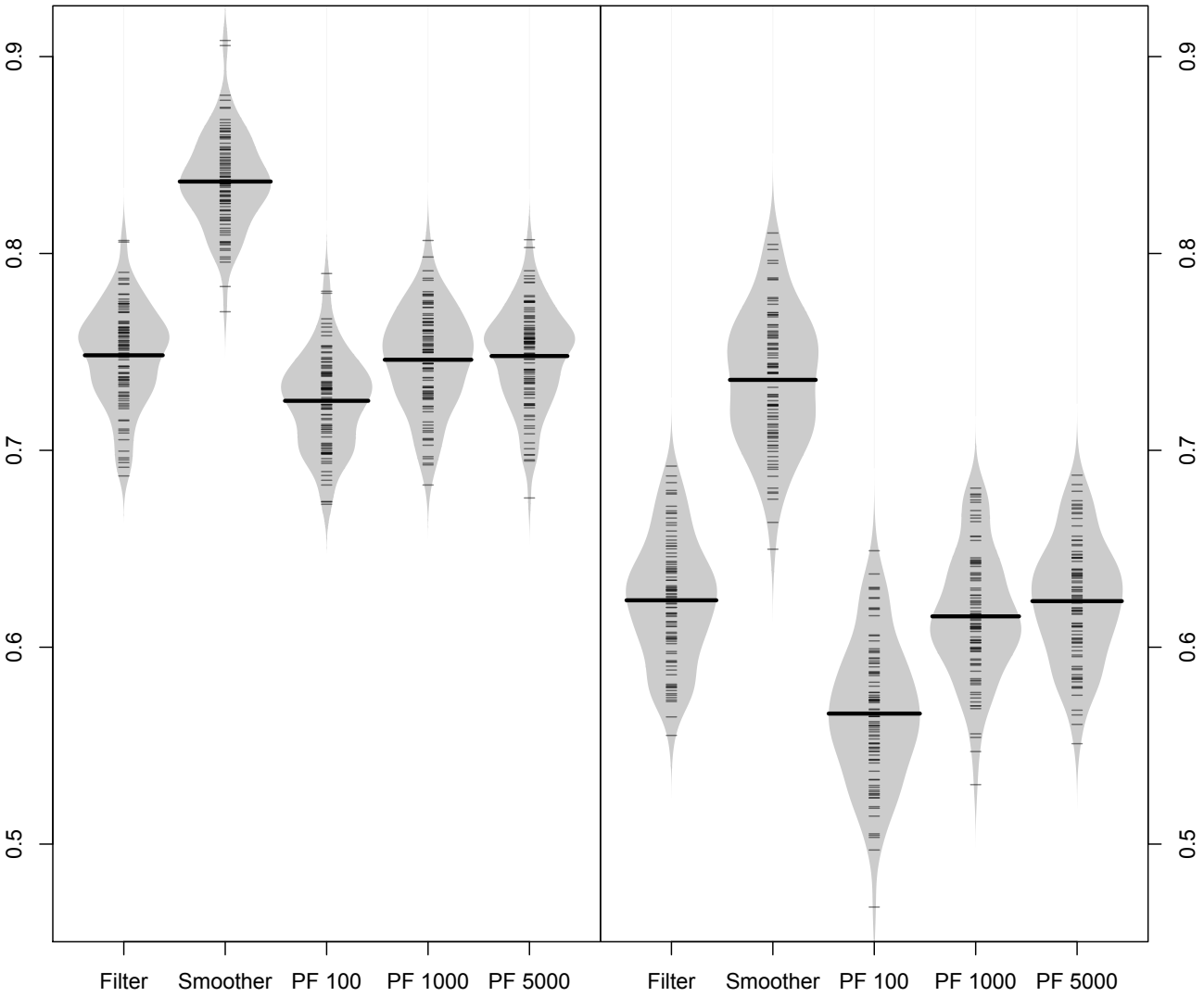


Figure V.7: Regime decoding success from the approximate filter, approximate smoother, a particle filter with 100 particles (PF 100), a particle filter with 1,000 particles (PF 1000), and a particle filter with 5,000 particles (PF 5000) in the stochastic volatility model simulation study with two regimes (left panel) and three regimes (right panel). In each bean, the grey area shows a kernel density estimate based on each value from the simulation study, which are shown by short lines. The thick line spanning the grey area indicates the mean value.

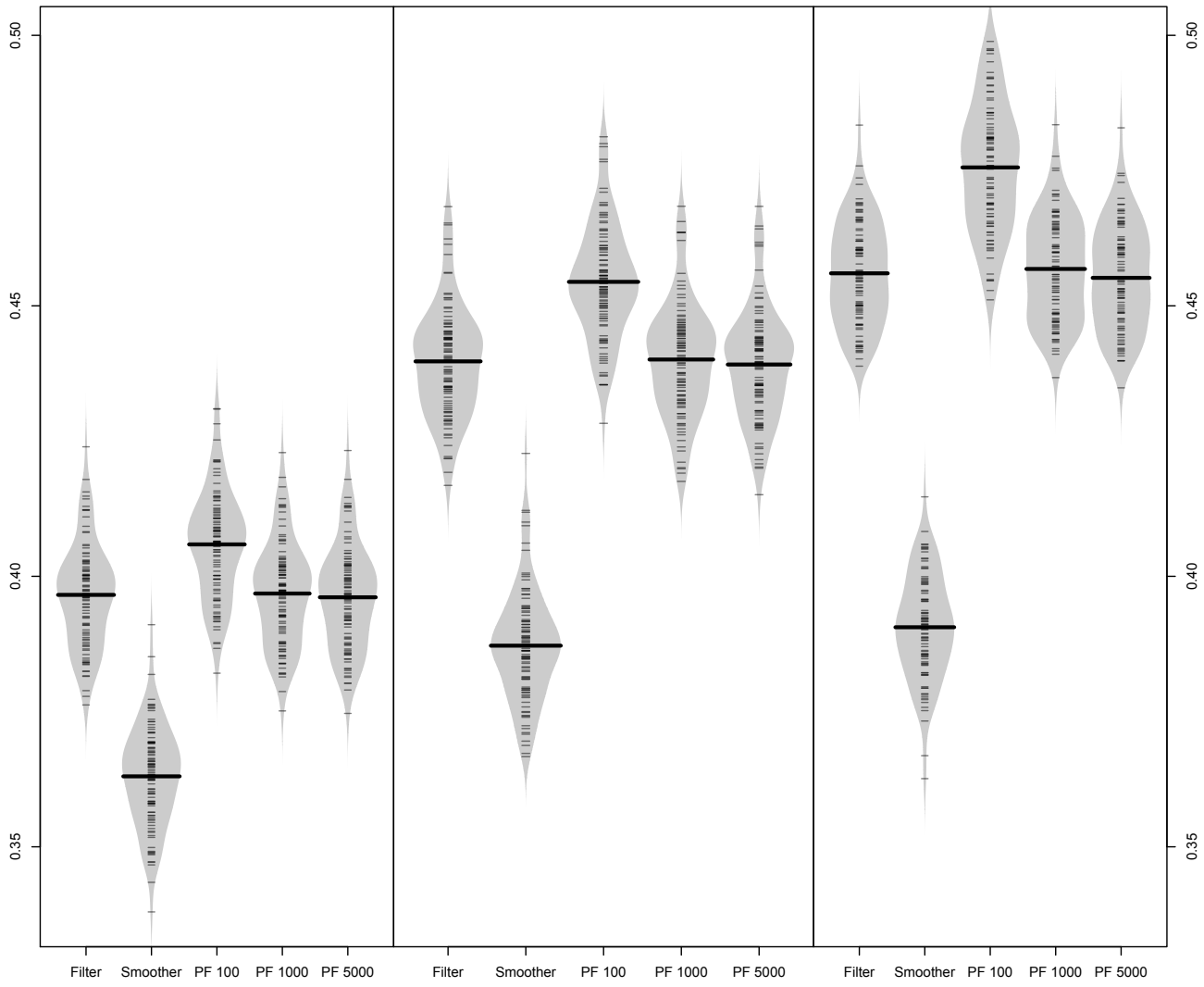


Figure V.8: Beanplot of the root mean squared error of state predictions from the approximate filter, approximate smoother, a particle filter with 100 particles (PF 100), a particle filter with 1,000 particles (PF 1000), and a particle filter with 5,000 particles (PF 5000) in the stochastic volatility simulation study with one regime (left panel), two regimes (middle panel), and three regimes (right panel). In each bean, the grey area shows a kernel density estimate based on each value from the simulation study, which are shown by short lines. The thick line spanning the grey area indicates the mean value.

Table V.4: Summary of parameter estimates from the stochastic volatility simulation study with one regime

Parameter	True	Median	Mean	Min.	1st Qu.	3rd Qu.	Max
$\log \sigma$	-1.609	-0.850	-0.938	-2.205	-1.102	-0.601	-0.280
μ_1	0.000	0.093	0.089	-0.086	0.038	0.142	0.291
$\text{logit}\, \alpha$	2.197	1.105	0.952	-12.372	0.357	1.710	3.534
$\log \nu$	2.773	16.049	15.360	2.133	15.567	16.759	32.478

Table V.5: Summary of parameter estimates from the stochastic volatility simulation study with two regimes

Parameter	True	Median	Mean	Min.	1st Qu.	3rd Qu.	Max
$\log \sigma$	-1.609	-0.554	-0.950	-9.125	-1.379	-0.253	-0.007
$\tilde{p}_1$	3.892	4.115	4.309	2.404	3.664	4.765	9.372
$\tilde{p}_2$	-3.892	-3.965	-4.017	-6.914	-4.406	-3.560	-2.648
μ_1	0.000	0.090	0.099	-0.462	-0.041	0.199	0.898
$\log(\mu_2 - \mu_1)$	0.000	0.108	-0.141	-13.579	-0.101	0.289	0.520
$\text{logit}\, \alpha$	2.197	0.486	-0.459	-15.169	-1.218	1.873	3.589
$\log \nu_1$	2.773	16.307	15.200	1.436	15.470	16.759	75.285
$\log \nu_2$	1.386	7.827	9.056	0.782	1.438	15.883	39.102

Table V.6: Summary of parameter estimates from the stochastic volatility simulation study with three regimes

Parameter	True	Median	Mean	Min.	1st Qu.	3rd Qu.	Max
$\log \sigma$	-1.609	-0.609	-1.082	-9.399	-1.592	-0.335	-0.106
$\tilde{p}_1$	4.585	5.329	9.438	-1.928	4.450	16.418	21.047
$\tilde{p}_2$	0.000	-1.262	-3.748	-20.157	-13.927	0.980	16.102
$\tilde{p}_3$	-4.585	-4.340	-8.661	-21.367	-16.533	-3.517	-1.544
$\tilde{p}_4$	0.000	-0.094	0.241	-19.308	-12.437	12.125	16.226
$\tilde{p}_5$	4.585	4.334	5.903	-10.870	3.509	5.419	20.096
$\tilde{p}_6$	-4.585	-3.966	-6.603	-25.785	-13.863	-2.962	12.970
μ_1	0.000	0.079	0.031	-2.638	-0.142	0.277	0.603
$\log(\mu_2 - \mu_1)$	-0.288	0.012	-0.174	-13.671	-0.172	0.169	1.238
$\log(\mu_3 - \mu_1)$	-0.288	-0.263	-2.474	-16.210	-1.049	0.188	1.760
$\text{logit}\, \alpha$	2.197	-0.067	-1.570	-15.879	-2.108	1.903	4.189
$\log \nu_1$	2.773	16.305	14.455	0.510	15.211	16.792	29.533
$\log \nu_2$	2.303	15.978	10.962	0.720	1.902	16.511	19.138
$\log \nu_3$	1.386	9.646	8.616	-0.193	1.405	15.878	19.954

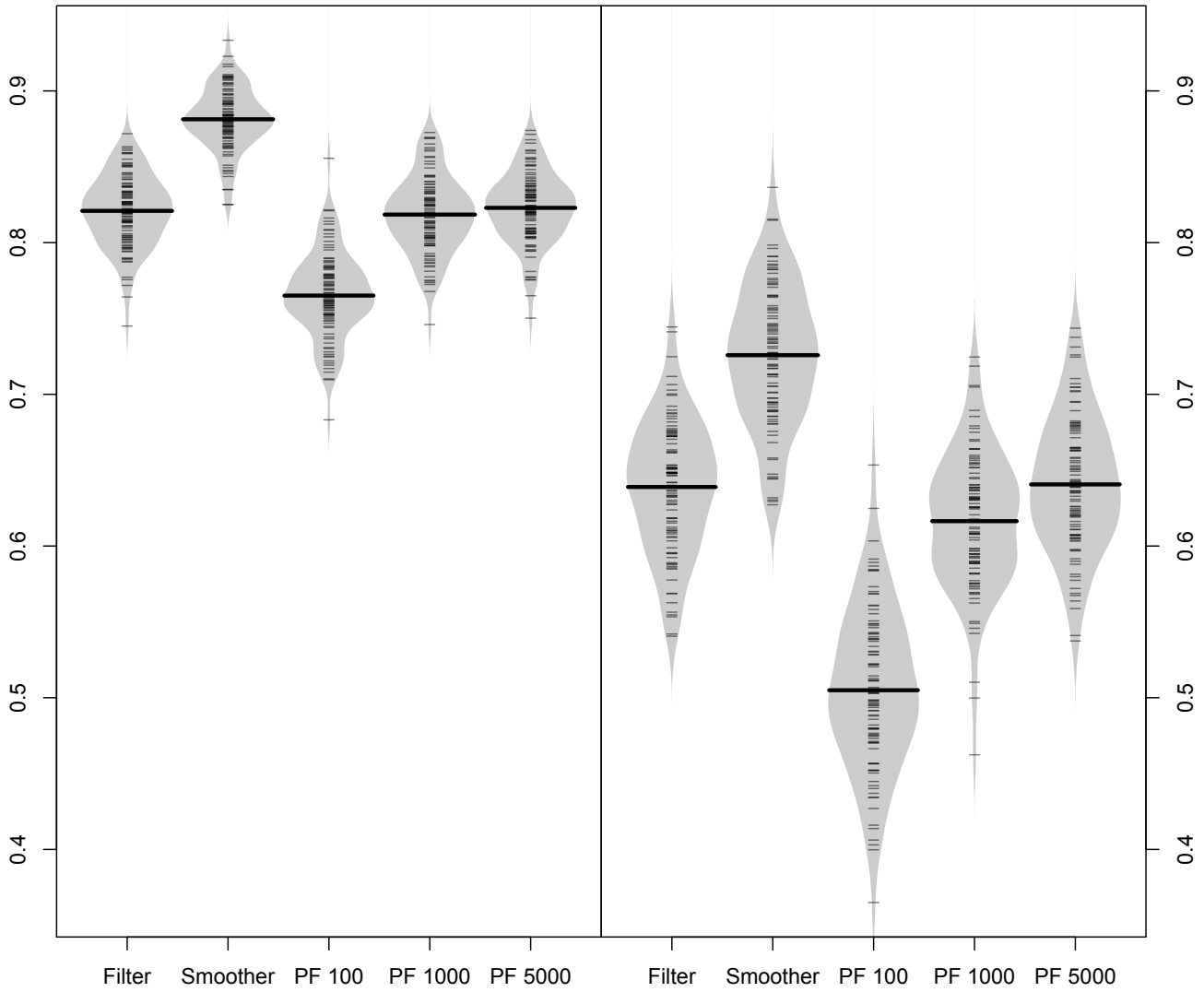


Figure V.9: Regime decoding success from the approximate filter, approximate smoother, a particle filter with 100 particles (PF 100), a particle filter with 1,000 particles (PF 1000), and a particle filter with 5,000 particles (PF 5000) in the non-linear population model simulation study with two regimes (left panel) and three regimes (right panel). In each bean, the grey area shows a kernel density estimate based on each value from the simulation study, which are shown by short lines. The thick line spanning the grey area indicates the mean value.

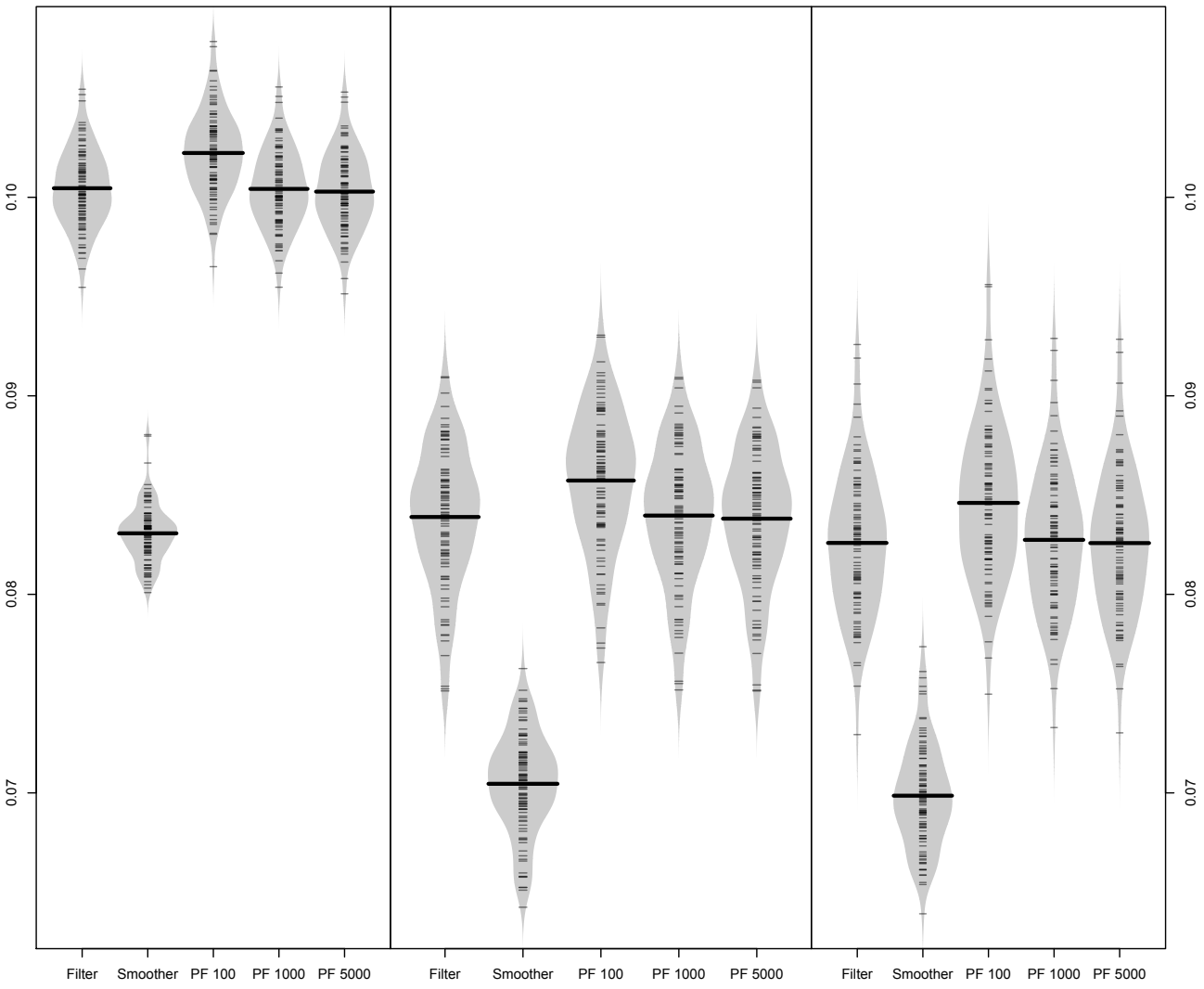


Figure V.10: Beanplot of the root mean squared error of state predictions from the approximate filter, approximate smoother, a particle filter with 100 particles (PF 100), a particle filter with 1,000 particles (PF 1000), and a particle filter with 5,000 particles (PF 5000) in the non-linear population model simulation study with one regime (left panel), two regimes (middle panel), and three regimes (right panel). In each bean, the grey area shows a kernel density estimate based on each value from the simulation study, which are shown by short lines. The thick line spanning the grey area indicates the mean value.

Table V.7: Summary of parameter estimates from the non-linear population simulation study with one regime

Parameter	True	Median	Mean	Min.	1st Qu.	3rd Qu.	Max
$\log \sigma$	-2.303	-2.297	-2.300	-2.469	-2.333	-2.263	-2.182
$\log K_1$	-0.500	-0.560	-0.562	-0.803	-0.629	-0.477	-0.351
$\log R_1$	-4.400	-3.739	-3.826	-13.666	-3.910	-3.575	-3.067

Table V.8: Summary of parameter estimates from the non-linear population simulation study with two regime

Parameter	True	Median	Mean	Min.	1st Qu.	3rd Qu.	Max
$\log \sigma$	-2.303	-2.271	-2.231	-2.428	-2.305	-2.246	0.275
$\tilde{p}_1$	3.892	5.577	6.288	-6.780	2.650	9.874	17.153
$\tilde{p}_2$	-3.892	-1.891	3.288	-14.811	-3.862	4.558	120.884
$\log K_1$	-0.500	0.405	0.722	-1.431	0.104	0.687	15.555
$\log(K_2 - K_1)$	0.405	-5.331	-7.746	-53.539	-11.761	-2.783	7.499
$\log R_1$	-4.400	-4.199	-5.609	-52.433	-5.382	-3.670	-0.531
$\log R_2$	-2.856	-1.424	-2.059	-48.814	-3.581	-0.364	17.508

Table V.9: Summary of parameter estimates from the non-linear population simulation study with three regime

Parameter	True	Median	Mean	Min.	1st Qu.	3rd Qu.	Max
$\log \sigma$	-2.303	-2.268	-1.971	-2.437	-2.324	-1.683	-0.045
$\tilde{p}_1$	4.585	3.203	6.124	-16.846	2.738	7.330	34.317
$\tilde{p}_2$	0.000	-3.030	-1.693	-61.093	-3.050	-2.593	81.194
$\tilde{p}_3$	-4.585	-3.385	0.617	-97.207	-3.449	0.567	152.301
$\tilde{p}_4$	0.000	-2.987	-4.344	-77.572	-4.420	-2.978	27.269
$\tilde{p}_5$	4.585	3.223	2.605	-106.332	-0.527	3.549	86.073
$\tilde{p}_6$	-4.585	-2.869	-5.657	-196.357	-5.504	-2.821	50.489
$\log K_1$	-0.500	0.499	0.598	-3.269	0.190	0.990	2.974
$\log(\log K_2 - \log K_1)$	-0.288	-1.819	-5.163	-347.639	-2.512	-0.108	7.268
$\log(\log K_3 - \log K_1)$	-0.288	-1.022	-0.988	-25.943	-1.644	0.305	9.824
$\log R_1$	-4.400	-3.283	-3.448	-27.301	-3.730	-1.654	-0.584
$\log R_2$	-3.636	-2.000	-3.277	-66.893	-3.824	1.138	2.915
$\log R_3$	-2.856	0.289	-0.972	-105.047	-2.166	2.681	9.275

Software I

argosTrack: Fit Movement Models to Argos Data for Marine Animals

The `argosTrack` R package[1] was originally developed for the analysis in Paper III (Albertsen et al., 2015); however, several movement models have been included since the beginning. The package uses reference classes to allow a modular approach to modelling the data.

The package can be loaded as any other R package, and data can be imported. Below, the example above uses a movement track included in the package.

```
library(argosTrack)
data(subadult_ringed_seal)

d <- subadult_ringed_seal
```

In the `argosTrack` package, data is contained in the `Observation` class.

```
obs <- Observation(lon = d$lon,
                   lat = d$lat,
                   dates = as.POSIXct(d$date),
                   locationclass = d$lc)

obs
```

```
## Observations:
## -------------
##
## Number of observations: 3586
## First date: 2010-10-20
## Last date: 2011-04-06
## Location classes:
## qual
##  GPS    3    2    1    0    A    B    Z
##    0   92  209  392  263  854 1773    3
```

[1] available from https://github.com/calbertsen/argosTrack

Likewise, a measurement model is specified by the `Measurement` class. A measurement model can be specified as a normal distribution or a t-distribution. For simplicity, the normal distribution is used for this example.

```
meas <- Measurement("n")

meas
```

```
##
##
## Measurement model
## -----------------
##
## Measurement distribution: n
## Use nautical observations: FALSE
## Variance parameters:
##           GPS        3        2         1         0        A        B
## Latitude    1 3.674437 108.4620 2214.5400 2113.2874 7211.267 69064.77
## Longitude   1 5.187231  60.7277  372.5607  442.5706 1620.030 61685.24
##                 Z
## Latitude  8886111
## Longitude 8886111
```

The final component in the model is the movement model. Currently, the CTCRW, DCRW, random walk (RW), no movement (DIRAC), step-length/bearing model with half-normal steps, and the step-length/bearing model with Weibull steps are implemented and exported. As experimental features, a continuous time step-length/bearing model, the GDCRW, a bivariate Ornstein-Uhlenbeck on locations, and a bivariate Ornstein-Uhlenbeck on velocities are included but not exported. For simplicity, the random walk is used in this example. The latent states will be at the same time as the observations.

```
mov <- RW(unique(obs$dates))

mov
```

```
##
##
## Movement model:
## -------------
##
## Model: Random Walk (RW)
## Movement parameters:
## numeric(0)
## Movement variance parameters:
## log(sigma_lat) log(sigma_lon)
##              0              0
## Number of latent variables: 6840
## Using nautical states: FALSE
## Using time unit: hours
```

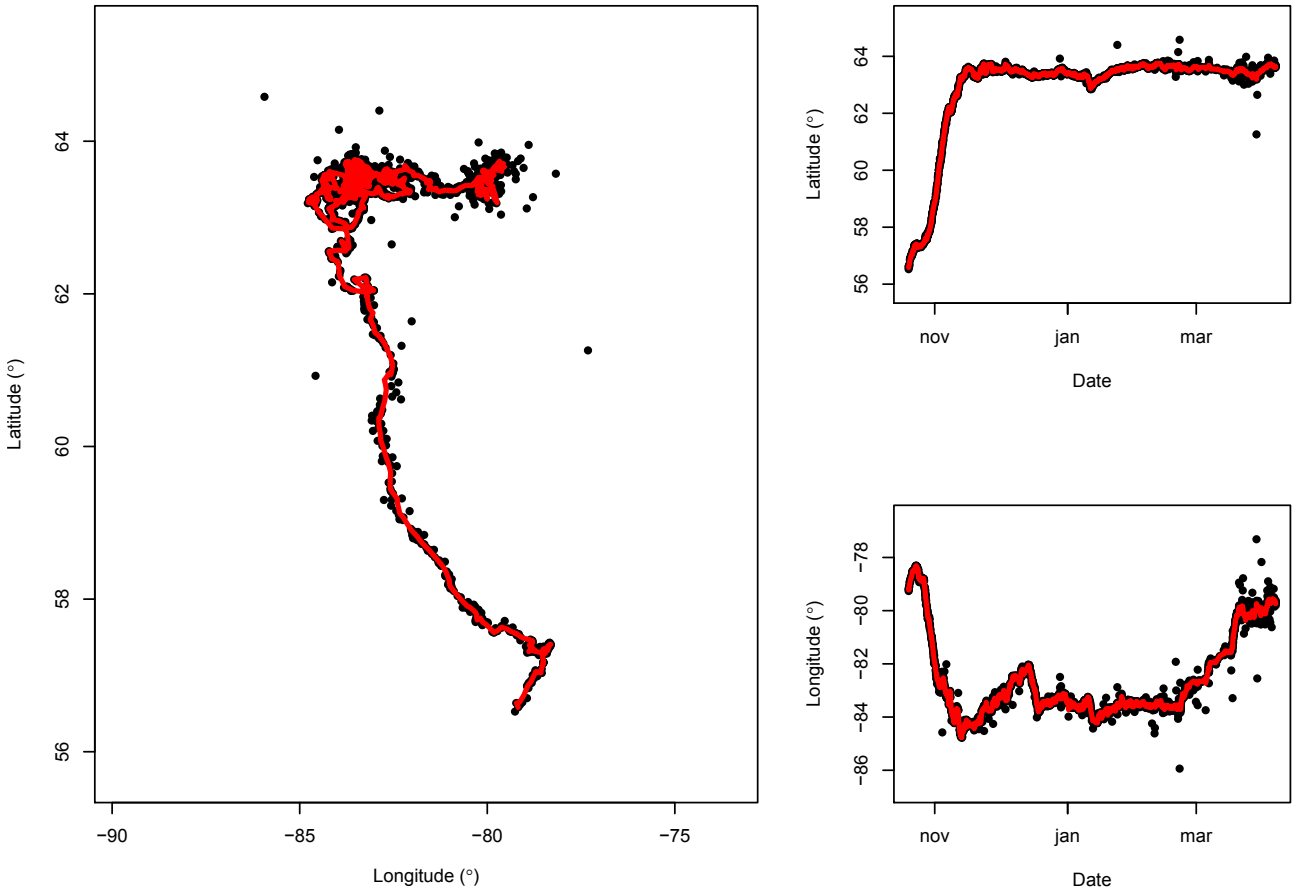


Figure SI.1: Estimated track

The three components are collected by the `Animal` class.

```
anim <- Animal(observation = obs,
               measurement = meas,
               movement = mov,
               name = "Random walk example")
```

Now, the track can be fitted through TMB (Kristensen et al., 2016) by the `fitTrack` function.

```
fitTrack(anim)
```

Afterwards, the result can be plottet.

```
plot(anim)
```

Further, trajectory information can be extracted by the `getTrack` function,

```
head(getTrack(anim))
```

```
##                 dates                  id obs.lat obs.lon obs.lc  est.lat
## 1 2010-10-20 01:31:57 Random walk example  56.526 -79.261      B 56.58090
## 2 2010-10-20 02:13:06 Random walk example  56.583 -79.174      1 56.58361
## 3 2010-10-20 03:12:42 Random walk example  56.630 -79.178      A 56.61716
## 4 2010-10-20 03:54:43 Random walk example  56.638 -79.228      1 56.63749
```

```
## 5 2010-10-20 06:03:39 Random walk example  56.616 -79.126      A 56.64588
## 6 2010-10-20 07:23:00 Random walk example  56.646 -79.094      B 56.66567
##     est.lon       sd.lat       sd.lon
## 1 -79.17863   0.03004365   0.03413221
## 2 -79.17597   0.10863630 595.29038910
## 3 -79.20020 940.95550394   0.16959218
## 4 -79.22331   0.11822253   0.08135140
## 5 -79.14986   0.07037241   0.03503249
## 6 -79.11689   0.03848183   0.41115953
```

while new trajectories can be simulated by `simTrack`.

References

Albertsen, C. M., Whoriskey, K., Yurkowski, D., Nielsen, A., & Flemming, J. M. (2015). Fast fitting of non-gaussian state-space models to animal movement data via template model builder. *Ecology*, *96*(10), 2598–2604. doi:10.1890/14-2101.1

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: automatic differentiation and laplace approximation. *Journal of Statistical Software*, *70*(5). doi:10.18637/jss.v070.i05

Software II

covafillr: Local Polynomial Regression of State Dependent Covariates in State-Space Models

The `covafillr` R package[1] is intended for including state dependent covariates in state-space models by interpolating with local polynomial regression. The package can be used with `Rcpp`, `TMB`, `JAGS`, and `ggplot2`; however, only the R functions will be demonstrated here.

Given coordinates, `x`, and corresponding observations, `y`,

```
set.seed(8723)

n <- 500
m <- function(x) x * sin(2 * pi * x)
sigma <- function(x) (2 + cos(2 * pi * x)) / 10

x <- runif(n,0,1)
y <- m(x) + sigma(x) * rnorm(n)
```

the `covafill` reference class can be used to set up the local polynomial regression. In this case, the polynomial degree is 5, while the bandwidth is suggested by default.

```
library(covafillr)
cf <- covafill(coord = x,
               obs = y,
               p = 5L)
cf
```

Besides the local polynomial regression, a search three approximation is also implemented, `covatree`. It requires an additional parameter determining when the tree splitting is stopped, and it can only be used for coordinates with a dimension no more than three. Otherwise, the `covatree` reference class can be used in the same way as `covafill`. Estimates of the mean function can be obtained by the `predict` member function.

[1] available from https://cran.r-project.org/package=covafillr or https://github.com/calbertsen/covafillr

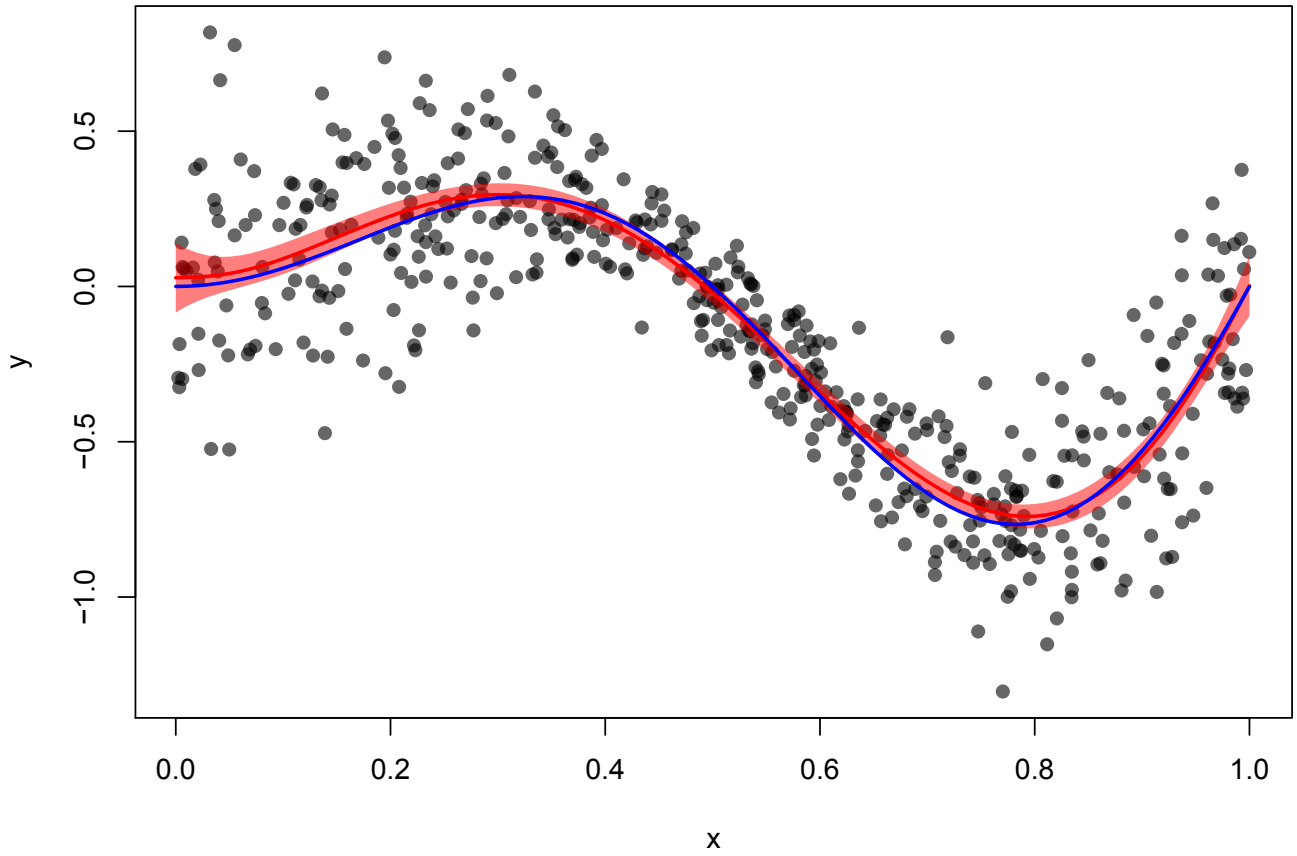


Figure SII.1: Simulated data (points) with the true mean function (blue line) and the local polynomial estimate (red line) with 95% confidence interval (red area).

```
x0 <- seq(0,1,0.01)
pred <- cf$predict(x0,se.fit=TRUE)
```

The local polynomial fit can be added directly to a `ggplot`. This is done with `stat_covafill`.

```
library(ggplot2)
ggplot(aes(x=x,y=y),data=data.frame(x=x,y=y)) +
    geom_point() +
    stat_covafill()
```

A special case of local polynomial regression is kernel density estimation. The `covafillr` package includes the `kde` function, which can be used for kernel density estimation in any dimension.

Given a matrix of coordinates,

```
Sigma <- matrix(c(1,0.9,0.9^2,
                  0.9,1,0.9,
                  0.9^2,0.9,1),3,3,byrow=TRUE)
CSigma <- t(chol(Sigma))
sim <- function() as.vector(CSigma %*% rnorm(3))
distfun <- function(x){U <- pnorm(x);
    y <- c(qgamma(U[1],shape=5,scale=0.5),
           qnorm(U[2],0,2),
           qbeta(U[3],0.5,0.5))
    return(y)
```

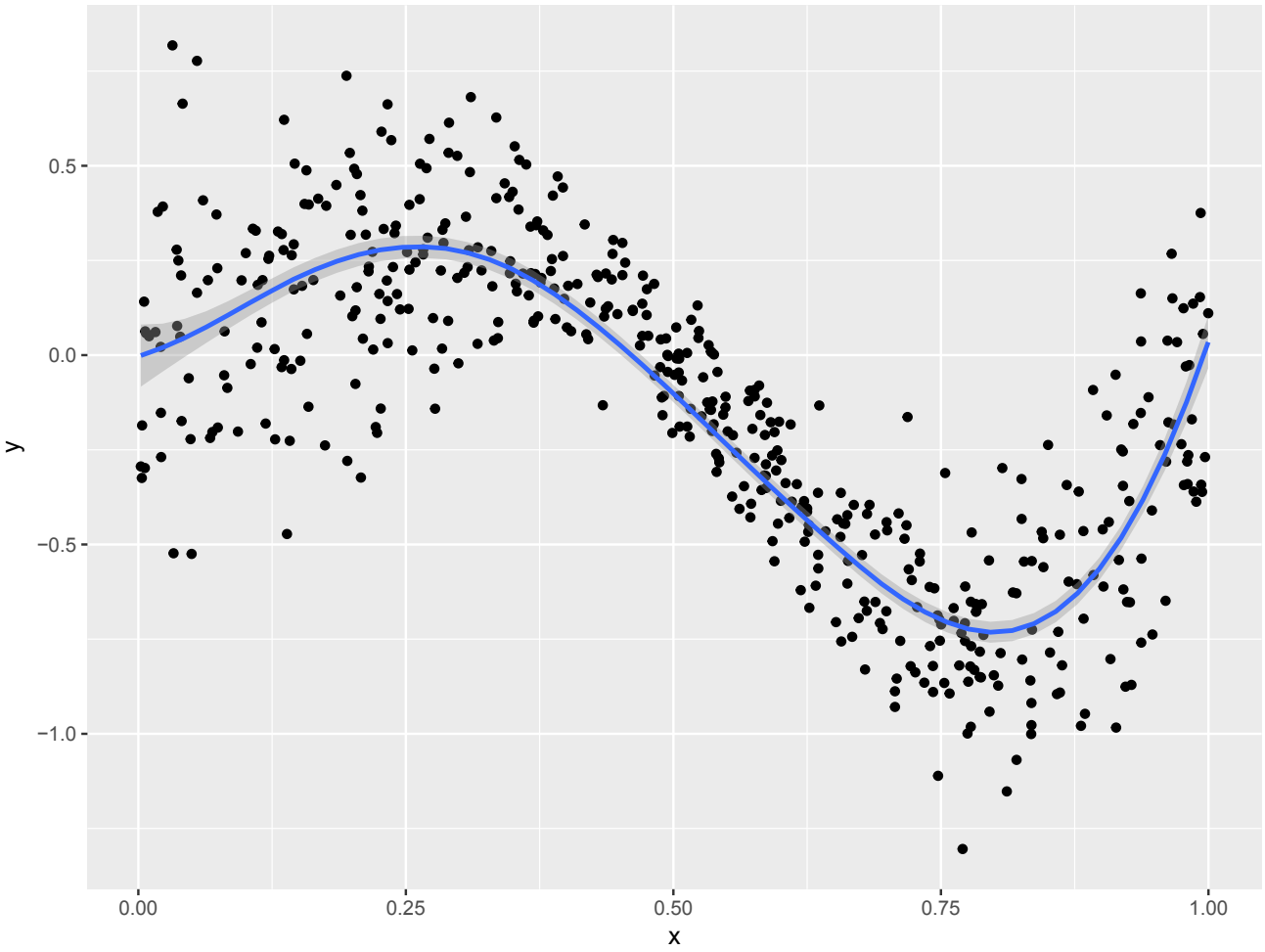


Figure SII.2: covafillr used with ggplot2.

```
}
X0 <- t(replicate(1000,sim()))
X <- t(apply(X0,1,distfun))
```

the kde function is called to estimate the density function on a regular grid.

```
z <- kde(X,
         npred = 100,
         from=apply(X,2,min)-0.5,
         to=apply(X,2,max)+0.5)
```

For three-dimensional data, the result can be plotted with, for instance, the plot3D package.

```
library(plot3D)
addTrans <- Vectorize(function(col,alpha){
    nc <- col2rgb(col,TRUE)/255
    ncl <- as.list(nc)
    names(ncl) <- rownames(nc)
    ncl$alpha <- alpha
    do.call("rgb",ncl)
},"col")
Y <- as.matrix(z$coord)
cv <- z$density
cv2 <- cv
cv2[cv2 < 0.01] <- NA
slice3D(unique(Y[,1]),
        unique(Y[,2]),
        unique(Y[,3]),
        array(cv2,dim=c(100,100,100)),
        xs=NA,zs=NA,ys=seq(-6,6,0.75),
        phi=45,theta=65,NAcol=rgb(0,0,0,0),
        ticktype= "detailed")
```

Examples on how to use the package from C++ or JAGS is available from the GitHub page.

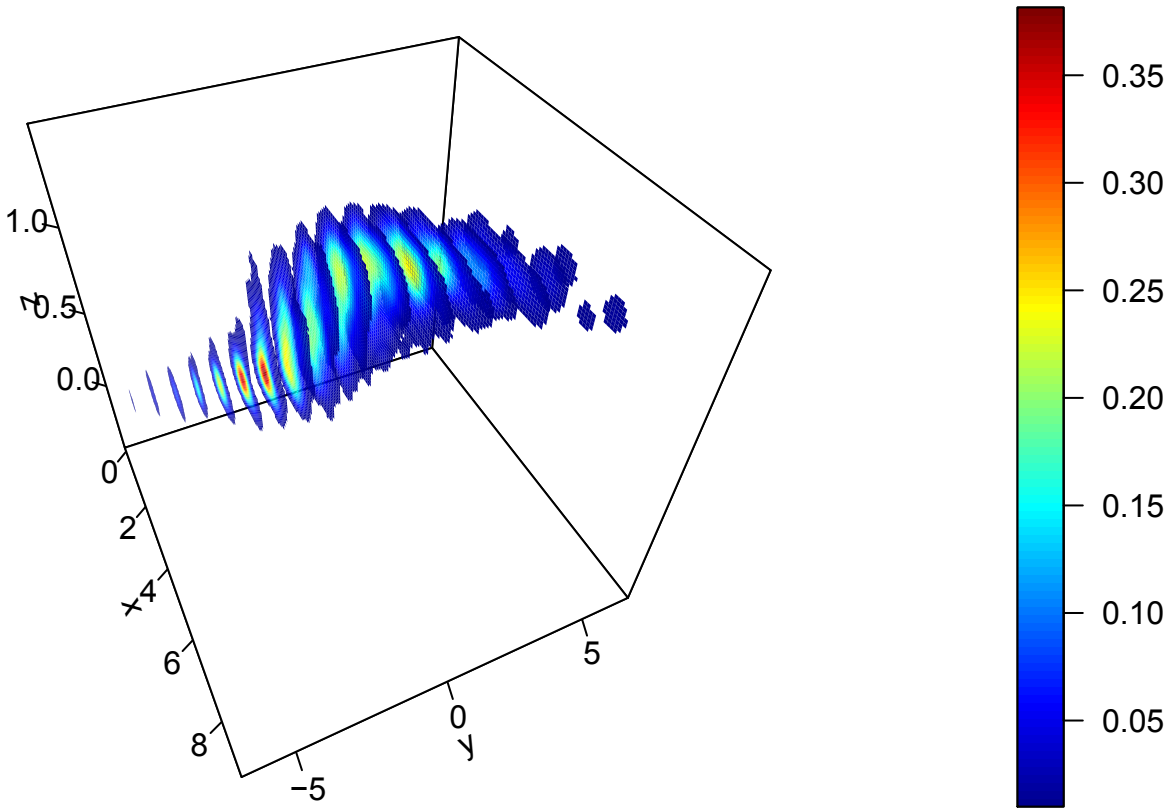


Figure SII.3: 3D kernel density estimate.

Software III

GrAD: Graph based reverse mode automatic differentiation

GrAD is a lightweight header-only C++ library for automatic differentiation. The library represents the computational graph as a tree, which keeps the structure of the objective function. Most C++ libraries build a sequential stack or tape which only have references to the input parameters. In contrast, GrAD can access both parents and children in the tree representation of the computational graph.

Below is a simple example of how to use the library. The ADparlist class registers active parameters using the Independent member function. The AD class indicates that the variable should be differentiated.

```
#include <GrAD/GrAD>
#include <iostream>
#include <iomanip>

using namespace GrAD;

template<class T>
T fn(T x,T y){
    return exp(-pow(x,T(2.0))/y) + y;
}

int main(){

    ADparlist<double>* grd = new ADparlist<double>();

    AD<double> x(3.0);
    AD<double> y(2.0);

    grd->Independent(x);
    grd->Independent(y);

    AD<double> z = fn(x,y);
    double fn0 = z.fn();
    std::cout << std::setprecision(12) <<
      "Function value: " <<
```

```
      fn0 <<
      std::endl;
    vector<double> gr0 = z.gr();
    std::cout << std::setprecision(12) <<
      "Gradient: " <<
      gr0[0] << ", " << gr0[1] <<
      std::endl;

    return 0;
}
```

After compilation, the object can be called to output the gradient:

```
./synopsis_files/grad/grad.so
```

```
## Function value: 2.01110899654
## Gradient: -0.0333269896147, 1.02499524221
```

The library provides a simple interface for automatic differentiation and can be used for derivatives of any order by nesting the classes. That is, second order derivatives can obtained by using

```
AD<AD<double> > x;
```

For small examples, the performance is comparable to more developed libraries; however, for complex calculations `GrAD` is still immature. Implementation details can be seen in the source code.

Software IV

multiStockassessment: Connecting single stock assessment models through correlated survival

The `multiStockassessment` R package[1] implements the model described in Paper II (Albertsen et al., 2017). The package extends the `stockassessment` package[2], to allow fitting several correlated stocks.

To use the package, the stocks must be fitted individually by the `stockassessment` package[3]. The following code simulates and fits two independent stocks with the `stockassessment` package.

```
library(stockassessment)

data(nscodData)
data(nscodConf)
data(nscodParameters)
fit <- sam.fit(nscodData, nscodConf, nscodParameters,
               sim.condRE = FALSE)

set.seed(238)
datSim <- simulate(fit,nsim=2)
fitSim <- do.call("c",
                  lapply(datSim,
                         function(x)sam.fit(x,
                           defcon(x),
                           defpar(x,defcon(x)))))
```

The combined AIC of the two assessments are

```
do.call("sum",lapply(fitSim,AIC))
```

```
## [1] 1253.106
```

[1] available from https://github.com/calbertsen/multi_SAM

[2] available from https://github.com/fishfollower/SAM

[3] This was not done in the manuscript introducing the model.

In the `multiStockassessment` package, a correlation structure can be constructed with the `suggestCorStructure` function. With `nAgeClose = 0`, independent assessments are fitted, with `nAgeClose = 1` the $\mathcal{M}_1$ model of Paper II (Albertsen et al., 2017) is fitted, and so on. Additional options are available.

```
library(multiStockassessment)
cs <- suggestCorStructure(fitSim,nAgeClose=0)
m0 <- multisam.fit(fitSim,cs)
```

The AIC of the combined (independent) assessments are identical to the results from the `stockassessment` package

```
AIC(m0)
```

```
## [1] 1253.106
```

A model with correlations can be fitted by

```
cs <- suggestCorStructure(fitSim,nAgeClose=4)
m4 <- multisam.fit(fitSim,cs)
```

with the resulting AIC,

```
AIC(m4)
```

```
## [1] 1379.561
```

References

Albertsen, C. M., Nielsen, A., & Thygesen, U. H. (2017). Connecting single-stock assessment models through correlated survival. *ICES Journal of Marine Science*. In Press. doi:10.1093/icesjms/fsx114

DTU Aqua – National Institute of Aquatic Resources – is an institute at the Technical University of Denmark. DTU Aqua's mission is to conduct research, provide advice, educate at university level and contribute to innovation in sustainable exploitation and management of aquatic resources. We investigate the biology and population ecology of aquatic organism, aquatic physics and chemical processes, ecosystem structure and dynamics, taking account of all relevant natural and anthropogenic drivers.

DTU Aqua
National Institute of Aquatic Resources
Technical University of Denmark

Kemitorvet
Building 202
2800 Kgs. Lyngby
Denmark
Tlf. +45 35 88 33 00
www.aqua.dtu.dk