Estimation methods for nonlinear state-space models in ecology

M.W. Pedersen [a,*], C.W. Berg [a,b], U.H. Thygesen [b], A. Nielsen [b], H. Madsen [a]

[a] Department for Informatics and Mathematical Modelling, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[b] National Institute of Aquatic Resources, Technical University of Denmark, 2920 Charlottenlund, Denmark

ARTICLE INFO

ABSTRACT

The use of nonlinear state-space models for analyzing ecological systems is increasing. A wide range of estimation methods for such models are available to ecologists, however it is not always clear, which is the appropriate method to choose. To this end, three approaches to estimation in the theta logistic model for population dynamics were benchmarked by Wang (2007). Similarly, we examine and compare the estimation performance of three alternative methods using simulated data. The first approach is to partition the state-space into a finite number of states and formulate the problem as a hidden Markov model (HMM). The second method uses the mixed effects modeling and fast numerical integration framework of the AD Model Builder (ADMB) open-source software. The third alternative is to use the popular Bayesian framework of BUGS. The study showed that state and parameter estimation performance for all three methods was largely identical, however with BUGS providing overall wider credible intervals for parameters than HMM and ADMB confidence intervals.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

State-space models (SSMs) have become the favored approach in modeling time varying ecological phenomena such as population dynamics (Wang, 2007; Gimenez et al., 2007), animal movement (Patterson et al., 2008) and animal behavior (Morales et al., 2004). SSMs come in a variety of classes depending on the problem type (linear or nonlinear) and the error structure of the data (Gaussian or non-Gaussian). In the linear and Gaussian case an exact solution to the SSM can be found using the Kalman filter (KF), which is the optimal estimator (Madsen, 2008). In case of minor departures from linearity, KF variants, such as the extended KF or unscented KF, can be employed. Both methods are reviewed and discussed by Wang (2007). In cases where the state-space equations are highly nonlinear, it is inappropriate to use any KF variant. For ecological problems Markov chain Monte Carlo (MCMC) is perhaps the most common approach to accommodate model nonlinearities owing to its flexibility and general applicability. In addition, free software for MCMC analysis is available, for example the widely used Win-BUGS (Gimenez et al., 2008). An example of non-WinBUGS MCMC population modeling is explained by Wang (2007).

We address three powerful methods for the analysis of nonlinear state-space models, two of which have only gained moderate attention previously within the field of ecology compared to the third. The idea of the first method we present is to discretize the continuous state-space and then reformulate the SSM as a hidden Markov model (HMM) (see Zucchini and MacDonald, 2009). A similar approach was described by Kitagawa (1987). The second method we consider is implemented in the open-source software AD Model Builder (ADMB-project, 2009a). In ADMB the SSM is formulated as a statistical model with mixed effects. A major advantage of ADMB is that it makes efficient use of available computer resources by so-called automatic differentiation. Thirdly, we apply OpenBUGS, which is the open-source version of WinBUGS (Spiegelhalter et al., 1996). BUGS is flexible and therefore widely used in modeling ecological systems (Gimenez et al., 2008).

To broaden the perspective of this study we apply the three methods to simulated data from the theta logistic population model, which is a nonlinear SSM. The same example was analyzed by Wang (2007). The performance of the three methods is summarized with respect to a range of aspects: complexity of implementation, computing time, estimation accuracy, limiting assumptions, and algorithmic design. Algorithmic design refers to the amount of subjective tuning required before actual estimation can begin. Because of reduced subjective influence, methods with fewer tuning parameters are often preferable. Finally, we discuss some differences between Bayesian (BUGS) and frequentist (HMM and ADMB) methods.

2. Methods

A state-space model describes the dynamics of a latent state ($\mathbf{X}_t$) and how data ($\mathbf{Y}_t$) relate to this state. An important feature of SSMs is their ability to model random variations in the latent state and in

* Corresponding author. Tel.: +45 4525 3095; fax: +45 4588 2673.
E-mail addresses: mwp@imm.dtu.dk, wpsgodd@gmail.com (M.W. Pedersen), cbe@aqua.dtu.dk (C.W. Berg), uht@aqua.dtu.dk (U.H. Thygesen), an@aqua.dtu.dk (A. Nielsen), hm@imm.dtu.dk (H. Madsen).

data. For $t \in \{1, \ldots, N\}$ the general system and observation equations of the SSM are respectively $\mathbf{X}_t = g(t, \mathbf{X}_{t-1}, \mathbf{e}_t)$, and $\mathbf{Y}_t = h(t, \mathbf{X}_t, \mathbf{u}_t)$, where $\mathbf{e}_t \sim N(\mathbf{0}, \mathbf{Q}_t)$ is the system error and $\mathbf{u}_t \sim N(0, \mathbf{R}_t)$ is the observation error. Here, "$\sim N(\cdot)$" means Gaussian distributed. Because of the possible nonlinearity of g and h, advanced filtering and smoothing methods must be employed to gain meaningful estimates of $\mathbf{X}_t$. In this respect, the extended Kalman filter, the unscented Kalman filter, and Bayesian filtering e.g. using Markov chain Monte Carlo (MCMC) sampling or BUGS are common approaches. Alternative methods for nonlinear state estimation are hidden Markov models (HMMs, Zucchini and MacDonald, 2009) and mixed effects models using the software AD Model Builder (ADMB). ADMB is freely available and open-source (ADMB-project, 2009a).

2.1. Benchmarking of estimation methods

The log-transformed theta logistic population growth model (Wang, 2007) was used as benchmark example for assessing the estimation performance of HMM, ADMB and BUGS. The system and observation equations for this model are

$$X_t = X_{t-1} + r_0 \left(1 - \left(\frac{\exp(X_{t-1})}{K}\right)^\theta\right) + e_t, \tag{1}$$

$$Y_t = X_t + u_t, \tag{2}$$

where $e_t \sim N(0, Q)$ and $u_t \sim N(0, R)$.

Following Wang (2007), two different tests of the methods were carried out:

1. State estimation performance with known parameter values, i.e. the ability of the methods to estimate the population level x_t for all t. Obviously, this test is free of Bayesian prior assumptions on parameters.
2. Estimation of states and all five model parameters, $\lambda = (\log(\theta), \log(r_0), K, \log(Q), \log(R))$, simultaneously. This situation is common in practice if model parameters cannot be estimated from independent data. Notice that parameters that may yield estimates close to zero are log-transformed to avoid invalid parameter values.

Specifically for test 1, $T = 2000$ data replicates were simulated with $N = 200$, $K = 1000$, $Q = 0.01$, $R = 0.04$, and the initial state $x_0 = 3$ using 12 different sets of the θ and r_0 parameters (see Table 1). The performance of the methods was evaluated using an estimate of the state estimation error:

$$\text{RMSE} = \frac{1}{T} \sum_{i=1}^{T} \left(\frac{1}{N} \sum_{t=1}^{N} (\widehat{x}_{i,t} - x_t)^2\right)^{1/2}, \tag{3}$$

Table 1
Performance of state estimation as defined by Eq. (3) for HMM, ADMB, and BUGS.

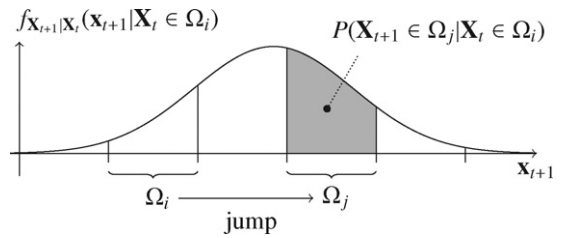| Sim. no.	r_0	θ	RMSE		
			HMM	ADMB	BUGS
1	0.1	0.5	0.100	0.100	0.100
2	0.5	0.5	0.099	0.099	0.100
3	0.75	0.5	0.097	0.097	0.097
4	1.0	0.5	0.095	0.095	0.095
5	0.1	1.0	0.100	0.100	0.100
6	0.5	1.0	0.095	0.095	0.095
7	0.75	1.0	0.091	0.092	0.092
8	1.0	1.0	0.090	0.090	0.090
9	0.1	1.5	0.100	0.100	0.100
10	0.5	1.5	0.092	0.092	0.092
11	0.75	1.5	0.091	0.091	0.091
12	1.0	1.5	0.096	0.096	0.096

Fig. 1. Probability of a jump (transition) from the state Ω_i to the state Ω_j in the time interval from t to $t + 1$ in a HMM. The shaded area corresponds to the integral in (4).

where $\widehat{x}_{i,t}$ is the state estimate for replicate i at time t, and x_t is the true state at time t.

Specifically for test 2, two datasets were simulated using two other sets of parameter values: $\lambda_1 = (\theta = 0.5, \ r_0 = 0.1, \ K = 900, Q = 0.01, \ R = 0.04)$ and $\lambda_2 = (\theta = 1.5, \ r_0 = 0.1, \ K = 900, Q = 0.01, R = 0.04)$ with the number of data points $N = 200$. Parameter estimates for these data using the three methods were found similarly to the study of Wang (2007). We further used these two parameter configurations to generate plots of the joint profile likelihood surfaces for r_0 and θ, which were transformed to confidence contours via a χ^2-distribution as in Polansky et al. (2009). The simulated data sets for λ_1 and λ_2 are available in the supplementary material to enable comparison of our results with future estimation methods. Additionally for test 2 we estimated all five model parameters along with 95% intervals using $T = 200$ of the data sets simulated for test 1. Inspired by Lambert et al. (2005), the purpose here was to evaluate the frequentist properties of the intervals provided by the three estimation methods.

2.2. Hidden Markov model with Matlab

The integrals involved in the prediction, filtering, and smoothing steps for nonlinear SSMs (see e.g. Eq. (2.2, 2.3 and 2.5) in Kitagawa, 1987) can, in general, not be solved analytically. However, by partitioning the continuous state-space uniformly into n parts the solution can be computed using hidden Markov models (HMMs) (Zucchini and MacDonald, 2009). See de Valpine and Hastings (2002) for an ecologically motivated study using a similar method. A state is denoted Ω_i, where $i \in \{1, 2, \ldots, n\}$. The probability distribution of the state given the observations $\mathcal{Y}_t$ available by time t is $P(\mathbf{X}_t \in \Omega_i | \mathcal{Y}_t) = p_t(i | \mathcal{Y}_t)$ which are collected in the row vector $\mathbf{p}_t(\mathcal{Y}_t) = \{p_t(i | \mathcal{Y}_t)\}$. The transition probability of jumping from Ω_i to Ω_j (see Fig. 1) is

$$p_t(i, j) = P(\mathbf{X}_{t+1} \in \Omega_j | \mathbf{X}_t \in \Omega_i) = \int_{\Omega_j} f_{\mathbf{x}_{t+1} | \mathbf{x}_t}(\mathbf{x}_{t+1} | \mathbf{X}_t \in \Omega_i) d\mathbf{x}_{t+1}. \tag{4}$$

For one-dimensional problems Ω_i are intervals on the line, in two dimensions Ω_i are areas, and analogously for higher dimensions. Note that the $n \times n$ probability transition matrix $\mathbf{P}_t = \{p_t(i, j)\}$ is not homogeneous, i.e. the transition probabilities may change as a function of time as indicated by (1). Now, the HMM prediction, filtering, and smoothing equations are respectively

$$\mathbf{p}_t(\mathcal{Y}_{t-1}) = \mathbf{p}_{t-1}(\mathcal{Y}_{t-1})\mathbf{P}_{t-1},$$
$$\mathbf{p}_t(\mathcal{Y}_t) = \psi_t^{-1}\mathbf{p}_t(\mathcal{Y}_{t-1}) \odot \mathbf{L}(y_t),$$

$$\mathbf{p}_t(\mathcal{Y}_N) = \mathbf{p}_t(\mathcal{Y}_t) \odot [\{\mathbf{p}_{t+1}(\mathcal{Y}_N) \oslash \mathbf{p}_{t+1}(\mathcal{Y}_t)\}\mathbf{P}_t^T]$$

where '$\odot$' and '$\oslash$' are elementwise matrix multiplication and division, respectively. The likelihood of the observations $\mathbf{L}(y_t)$ is a row vector with elements $p_t(y_t | i)$ and $\psi_t = \mathbf{p}_t(\mathcal{Y}_{t-1}) \cdot \mathbf{L}(y_t)^T$ is a normalization constant with '$\cdot$' denoting dot product. The estimate of the state given all N observations is simply the mean of the distribution $\mathbf{p}_t(\mathcal{Y}_N)$.

Using the above scheme we can estimate the unknown parameters (λ) of the SSM by maximizing the likelihood function

$$\mathcal{L}(\lambda|y_N) = f_{y_N}(y_N|\lambda) = [\mathbf{L}(y_1) \cdot \mathbf{1}]\prod_{t=2}^{N}\psi_t, \tag{5}$$

as in Kitagawa (1987), where **1** is a column vector of ones. The maximum likelihood (ML) estimate of the model parameters $\widehat{\lambda}$ is found by optimizing (5) as a function of λ. The covariance matrix of $\widehat{\lambda}$ is approximated by the inverse Hessian of the likelihood function at the optimum $\widehat{\lambda}$. This approximation is appropriate because the ML estimate is asymptotically Gaussian under certain regularity conditions (Cappé et al., 2005). Thus, confidence intervals can be constructed using the approximated covariance matrix. Under parameter transformations it is important to construct the confidence intervals in the transformed parameters and then reverse transform the computed confidence limits.

When analyzing the theta logistic model we set $n = 251$. The bounds of the discrete state-space are chosen such, that the probability of the true state falling outside the grid is negligible. That is, we use the observation model (2) to determine bounds that envelope the true latent state with a probability close to 1. This approach is similar to the one used in de Valpine and Hastings (2002). Details on grid specification can be found in the supplementary material containing model code.

The HMM code provided in the supplementary material was written in Matlab, but the method is not language specific. Matlab was chosen because it is widely used and has a syntax which is relatively easy to understand even for non-Matlab users.

2.3. Mixed effects model with AD Model Builder

Hierarchical mixed effects models are an alternative framework for analyzing nonlinear SSMs. The states are the random effects of the model and are collectively referred to as $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$. Here, as in Madsen and Thyregod (2010), we specify a model for the data conditional on the unobserved random effects, $f_{y_N|\mathcal{X}}(y_N|\mathcal{X}, \lambda_a)$ which corresponds to (2). We also specify a model for the random effects, $f_\mathcal{X}(\mathcal{X}|\lambda_b)$ which corresponds to (1). The joint density of random effects and observations conditional on the parameters is

$$f_{\mathcal{X}, y_N}(\mathcal{X}, y_N|\lambda) = f_\mathcal{X}(\mathcal{X}|\lambda_b)f_{y_N|\mathcal{X}}(y_N|\mathcal{X}, \lambda_a).$$

To obtain the marginal likelihood for estimating $\lambda = \{\lambda_a, \lambda_b\}$ we integrate over the unobserved random effects

$$\mathcal{L}(\lambda|y_N) = f_{y_N}(y_N|\lambda) = \int_{\mathbb{R}^N} f_{\mathcal{X}, y_N}(\mathcal{X}, y_N|\lambda)d\mathcal{X}. \tag{6}$$

The N-dimensional integral in (6) is generally challenging to solve, and for nonlinear mixed models we must resort to numerical approximation methods for estimating the model parameters. An efficient and widely used method for this is the Laplace approximation (Wolfinger and Xihong, 1997), which replaces the integrand with a second order Taylor expansion around the optimum of the log-likelihood function. This allows for elimination of the integral, because the second-order Taylor expansion can be formulated as a known constant multiplied by a multivariate Gaussian density, which integrates to unity. For nonlinear models the distribution of the random effect may not be Gaussian. Then the Laplace approximation is not exact. In particular for multi modal distributions one should use the Laplace approximation with caution. Still, when analyzing nonlinear models with moderately skewed unimodal distributions good results can be obtained with the Laplace approximation (Vonesh, 1996; Mortensen, 2009). In any case it is important to investigate if the approximation is critically violated

e.g. by Monte Carlo sampling from the random effects distribution.

Even with the Laplace approximation maximization of the marginal log-likelihood with respect to λ is challenging. A computationally efficient method is to combine the Laplace approximation with so-called automatic differentiation (AD Skaug and Fournier, 2006). AD is a technique for finding the gradient of a function h (in our case the log-likelihood), provided that h can be expressed in computed code. Evaluating h using AD gives the function value along with the gradient of h at the point of evaluation. The gradient is computed using the chain rule of calculus on every operation in the code that contributes to the value of h. For efficient maximization of the Laplace approximation of the marginal log-likelihood with respect to λ, up to third order partial derivatives must be found. Skaug and Fournier (2006) show how this can be accomplished by repeated use of AD.

The above procedure is implemented in AD Model Builder (ADMB), which we use to analyze the theta logistic model. ADMB is an open-source software package and programming language based on C++. It includes a function minimizer for ML parameter estimation and a random effects module, which utilizes the Laplace approximation for integration of random effects. Standard deviations for constructing confidence intervals are calculated using the delta method (Oehlert, 1992) and automatically reported on all estimated quantities. The covariance matrix for all states in an SSM is a banded matrix (Skaug and Fournier, 2006). ADMB can exploit this property by using the SEPARABLE FUNCTION construct (ADMB-project, 2009b) to gain significant speed improvements. Other than this useful property ADMB has no tuning parameters as such.

2.4. Monte Carlo estimation with BUGS

Finally, we analyze the theta logistic model using the Bayesian modeling language BUGS, which is an MCMC estimation method (Spiegelhalter et al., 1996). BUGS is a popular tool in ecological modeling (e.g. Gimenez et al., 2007; Jonsen et al., 2005; Schofield et al., 2009). BUGS is best known in the WinBUGS form which has a graphical user interface. Here, however, we use the open-source alternative OpenBUGS, yet the BUGS code provided in the supplementary material is compatible with WinBUGS.

A Bayesian analysis requires that prior distributions are specified for the model parameters. The type of prior distributions and parameter values related to these distributions should reflect the *a priori* knowledge that is available about the model parameters. BUGS then uses Gibbs sampling (Casella and George, 1992) to explore the posterior distribution of the parameter and state-space by incorporating the information specified by the priors, the state-space model, and the observed data. The Gibbs algorithm exploits that sampling the posterior is sometimes simpler via its conditional distributions rather than directly from the joint distribution. This is the case for state-space models where direct sampling of the posterior for states and parameters is difficult. Instead, sampling model parameters from priors and then sampling X_t conditional on model parameters and remaining states $(X_1, \ldots, X_{t-1}, X_{t+1}, \ldots, X_N)$ for all t is simple using (1). The ampling algorithm applied by BUGS in specific cases depends on the form and type of the conditional distribution, and also on the composition of priors on model parameters (see Spiegelhalter et al., 1996, 2003, for details).

We consider the common practical situation where *a priori* knowledge is unavailable and estimation therefore relies entirely on information in data. How to specify vague (or uninformative) priors is a topic of on-going research (Gelman, 2006; Lambert et al., 2005), which is outside the scope of this study. One suggested vague prior is a uniform distribution with wide support (Spiegelhalter et al., 1996). So, we choose a uniform prior for K, and uniform priors for $\log\theta$ and $\log r_0$ that were much wider than the natural biological bounds for the parameter values. By log-transforming θ and r_0

Table 2
Computing times for HMM, ADMB, BUGS1 (inverse-Gamma prior on variances), and BUGS2 (uniform prior on log-standard deviations). All times are for a single dataset run on the same computer.

	HMM	ADMB	BUGS1	BUGS2
State est.	6.12 s	0.49 s	58 s	58 s
Par. est.	225 s	2.5 s	118 s	614 s

biological meaningful (i.e. positive) parameter values are ensured. The state-space formulation implies that the variance parameters Q and R are non-zero and therefore also require prior distributions. It is common to assign vague inverse-gamma distributed priors to variance parameters (Spiegelhalter et al., 2003; Lambert et al., 2005). Gelman (2006), however, recommends using a uniform prior on the log-transformed standard deviation. Therefore, to asses the sensitivity of the estimation results to the choice of prior we perform BUGS estimation in two separate cases: BUGS1 using an inverse-gamma distribution for Q and R, and BUGS2 using a uniform distribution on the log-transformed standard deviation, i.e. $0.5\log(Q)$ and $0.5\log(R)$.

Estimation using BUGS involves a number of tuning parameters: the initial values for the sampling scheme can be found in the supplementary material online along with the specifics of the priors. The total number of generated samples was 100,000 with 50,000 used for burn-in. The appropriate number of samples was found iteratively by repeated application of Geweke Z score test for convergence (Geweke, 1992). The BUGS thinning rate was 50 (for reducing sample autocorrelation, which was apparent for θ and r_0 at lower thinning rates). With these values of the tuning parameters we get an effective sample size of 1000. For summarizing the estimation results the maximum *a posteriori* (MAP) parameter estimates along with 95% credible intervals are reported (where the lower bound equals the 2.5% quantile and the upper bound equals the 97.5% quantile of the posterior distribution).

3. Results

State estimation results for the three methods using known parameter values were practically identical (Table 1). ADMB was an order of magnitude faster than HMM, which, in turn, was an order of magnitude faster than BUGS (Table 2). State estimation using estimated parameter values also gave practically identical results for all three methods (Fig. 2). Regarding ML parameter estimation and confidence intervals (CIs) for λ_1 and λ_2, HMM and ADMB performed almost identically (Table 3). Likewise, MAP estimates and credible intervals provided by BUGS1 and BUGS2 were overall similar in the λ_2 case. In the λ_1 case, however, BUGS1 MAP estimates of θ and r_0 were markedly lower and higher respectively than the estimates provided by HMM, ADMB, and BUGS2. Perhaps most surprisingly was the upper limit of the credible interval for K seemingly quite sensitive to the choice of prior employed by BUGS, and in both cases considerably higher than the HMM and ADMB CI upper limits. Some notable differences between CIs and credible intervals were present for θ, K, and r_0 in the λ_1 case (Table 3), with BUGS generally being more conservative and providing wider intervals (in the log domain). Inspection of the joint profile likelihood surfaces for θ and r_0 revealed that contour lines closely approximated elliptical shapes for λ_2 (Fig. 4, panel B), thus indicating that the quadratic approximation used by HMM and ADMB was appropriate. For λ_1, on the other hand, the quadratic approximation was only appropriate until the 65% confidence limit where the contour shape started to diverge from the elliptical shape (Fig. 4, panel A). If comparing the limits of the intervals provided by all three methods for the λ_1 case (Table 3) with the extents of the likelihood surface (Fig. 4, panel A), it is clear that neither credible intervals nor CIs captured the actual range of plausible parameter values.

Visualizing the empirical distributions of the $T = 200$ parameter estimates (Fig. 3) showed largely identical results for all three methods. For all parameters the average 95% CIs provided by HMM and ADMB closely approximated the 2.5% and 97.5% quantiles of the corresponding empirical distribution. Similar results were observed for BUGS1 and BUGS2 for parameters R and Q. Regarding the three remaining parameters θ, K, and r_0, on the other hand, the average credible intervals were markedly wider than the corresponding quantiles of their empirical distribution, and therefore also wider than their CI counterparts. The difference in results between the two vague priors (BUGS1 and BUGS2) was minimal except for the credible intervals for K where BUGS2 gave wider intervals than BUGS1. Since both priors have been regarded in the literature as vague their influence on the resulting intervals is surprising. Computing times for parameter estimation showed that ADMB, again, was significantly faster than HMM and BUGS (Table 2). Interestingly, BUGS1 was considerably (six times) faster than BUGS2. This results can most likely be ascribed to BUGS using different sampling algorithms in the two cases.

4. Discussion

Dynamical processes are prevalent in ecology. State-space models are commonly used in the analysis of such nonlinear processes because they join separate models of the ecological system and the observation process. This paper assessed the performance of three methods for estimation in nonlinear state-space models: an approach using hidden Markov models (HMM), the open-source AD Model Builder framework (ADMB), and the BUGS language. HMM and ADMB are frequentist (non-Bayesian) methods, while BUGS is Bayesian. To facilitate a transparent comparison among available estimation methods we considered the theta logistic population model, which Wang (2007) analyzed with three other methods (extended Kalman filter, the unscented Kalman filter and a Metropolis–Hastings approach). To increase accessibility, the computer code for our three methods can be found in the online supplementary material.

The state estimation root mean square errors (RMSEs) of HMM, ADMB, and BUGS (Table 1) were lower than those for the three methods presented by Wang (2007), his Table 1. The 95% intervals for the parameter estimates of θ provided by our three methods all included the true values (Table 3). Note that they also included $\theta = 1$, which means that the models could not distinguish between
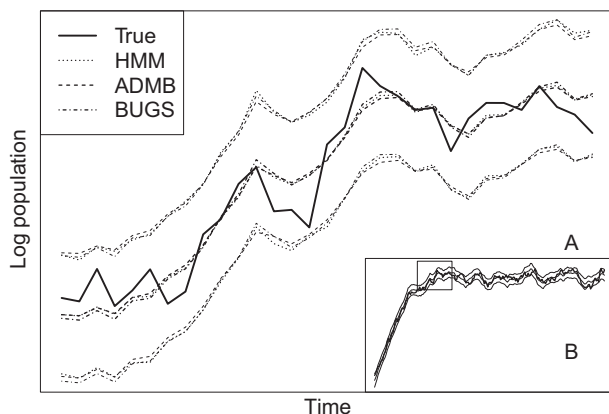

Fig. 2. State estimation of the theta logistic model with 95% intervals using the estimated parameter values in Table 3. True states were generated using $\lambda_2(\theta = 1.5)$. Panel A is a zoom of a part of the full time series indicated by the small box in panel B. Clearly in this case, HMM, ADMB, and BUGS gave close to identical state estimation results.

Table 3
Parameter values estimated by HMM, ADMB, BUGS1 (inverse-Gamma prior on variances), and BUGS2 (uniform prior on log-standard deviations) with related 95% intervals. Data were simulated with the listed true parameter values: $\lambda_1 = (\theta = 0.5, r_0 = 0.1, K = 900, Q = 0.01, R = 0.04)$ and $\lambda_2 = (\theta = 1.5, r_0 = 0.1, K = 900, Q = 0.01, R = 0.04)$ of the theta logistic model.

	HMM		ADMB		BUGS1		BUGS2	
	ML est.	95% conf. intv.	ML est.	95% conf. intv.	MAP est.	95% cred. intv.	MAP est.	95% cred. intv.
λ_1								
θ	0.588	0.134–2.588	0.583	0.129–2.640	0.374	0.0210–1.446	0.538	0.020–1.496
K	829.3	643.3–1015	829.5	639.2–1020	860.0	629.2–1900	834.0	638.3–4957
r_0	0.116	0.046–0.298	0.117	0.045–0.305	0.135	0.053–1.667	0.118	0.045–1.666
R	0.041	0.032–0.053	0.041	0.032–0.053	0.042	0.031–0.054	0.041	0.031–0.052
Q	0.0092	0.0052–0.016	0.0092	0.0051–0.017	0.011	0.0055–0.017	0.0099	0.0060–0.018
λ_2								
θ	1.098	0.412–2.926	1.079	0.402–2.902	1.006	0.043–2.551	1.037	0.043–2.869
K	886.9	792.7–981.0	887.0	790.5–983.5	891.3	769.3–1121	910.0	774.9–1097
r_0	0.128	0.082–0.201	0.129	0.081–0.203	0.127	0.078–1.136	0.134	0.074–1.032
R	0.043	0.032–0.056	0.043	0.032–0.056	0.043	0.031–0.056	0.044	0.032–0.056
Q	0.0082	0.0038–0.018	0.0081	0.0045–0.015	0.0094	0.0041–0.018	0.0086	0.0043–0.019

a concave and convex relation between population size and growth rate. This is in contrast with the credible intervals in Wang (2007), his Table 2, that excluded $\theta = 1$, however three out of six of his credible intervals also excluded the true parameter value, which is of some concern.

Recent studies have indicated that θ and r_0 of the theta logistic model (1) can be difficult to identify for certain data sets (Polansky et al., 2009). This is the case because given $\theta < 1$ similar model dynamics can be generated for different values of θ (Clark et al., 2010). Supporting this, a joint profile likelihood surface for $\log\theta$ and $\log r_0$ showed that combinations of different values for the two

parameters may fit data equally well, i.e. result in practically identical model likelihoods (Fig. 4, panel A, data generated with $\theta = 0.5$). Still parameters estimated by HMM and AMDB were reasonably accurate (Table 3, case λ_1), however the confidence intervals (CIs) were too narrow when compared to the contours of the confidence regions in Fig. 4, panel A. This result underlines the importance of validating the quadratic approximation to the log-likelihood function employed by HMM and ADMB before using it to construct CIs. The credible intervals from BUGS were wider and therefore more realistic than the CIs provided by HMM and ADMB, yet the interval bounds were narrower than the range of plausible models indi-
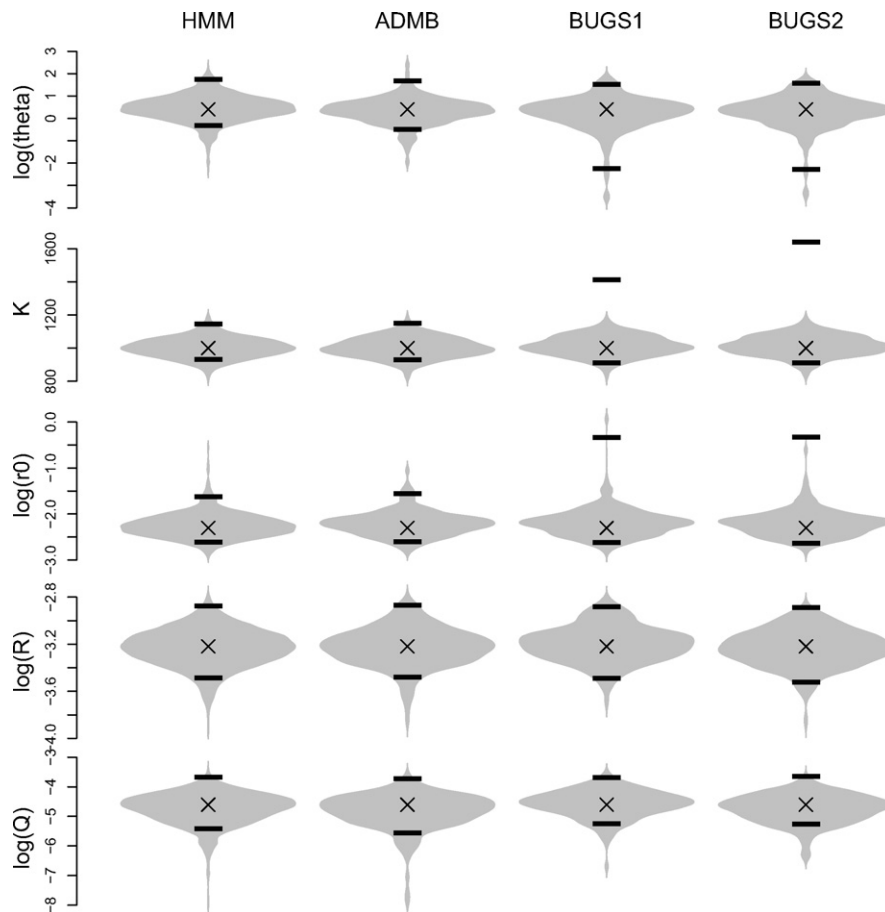

Fig. 3. Violin plots showing the empirical distribution of $T = 200$ parameter estimates. Data used for estimation were simulated with the parameter configuration $\lambda = (\theta = 1.5, r_0 = 0.1, K = 1000, Q = 0.01, R = 0.04)$. Crosses indicate the true parameter values, λ. Horizontal lines indicate the average limits of the 200 individual 95% intervals.
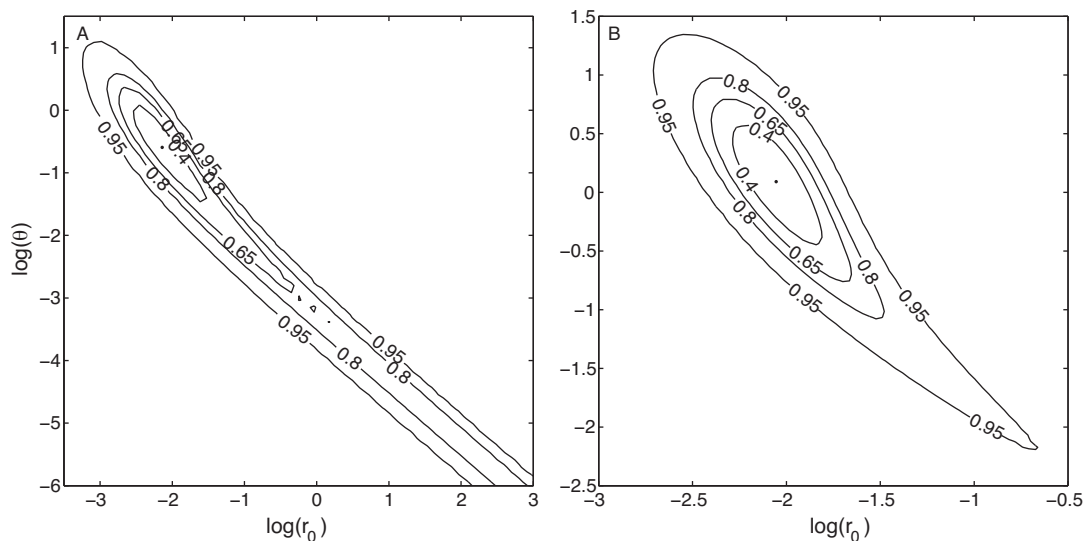
Fig. 4. Joint profile likelihood surfaces for two simulated data sets of the theta-logistic model (see also Table 3). Panel A: parameters used for simulation $\lambda_1 = (\theta = 0.5,\ r_0 = 0.1,$ $K = 900,\ Q = 0.01,\ R = 0.04)$. Panel B: $\lambda_2 = (\theta = 1.5,\ r_0 = 0.1,\ K = 900,\ Q = 0.01,\ R = 0.04)$. Following (Polansky et al., 2009) the joint profile log-likelihood surfaces have been transformed to confidence contours via a χ^2-distribution of the profiled models versus the model where all five parameters are estimated. Dots indicate the minima of the transformed surfaces equivalent to the maximum likelihood (ML) point. Both surfaces have elliptically shaped contours in proximity to the ML point in which case a quadratic approximation as used by HMM and ADMB is appropriate. While the surface for λ_2 (panel B) is close to quadratic even at the 95% level, the surface for λ_1 (panel A) departs from the quadratic shape at the 65% level.

cated by the profile likelihood surface. A possible explanation for this difference is that a substantial Monte Carlo sample size may be required to fully explore the posterior distribution when two parameters are highly correlated (Gamerman, 1997). If complications with parameter identifiability as illustrated in Fig. 4, panel A, are encountered in practical situations it is recommended to switch to a simpler model with fewer parameters e.g. by setting $\theta = 1$ (Clark et al., 2010).

For the data set generated with $\theta = 1.5$, the joint profile likelihood surface for $\log\theta$ and $\log r_0$ was well approximated by a quadratic function (Fig. 4, panel B). Thus, log-transforming θ and r_0 in the theta-logistic model avoids a boomerang-shaped likelihood surface (see e.g. Figure 2 in Polansky et al., 2009), which deviates considerably from a quadratic function. Thus, the CIs computed for HMM and ADMB in the log-transformed parameter space (Table 3, case λ_2) corresponded well to the confidence contours in Fig. 4, panel B. For BUGS credible intervals the conclusion was the same.

Similarly to Lambert et al. (2005), the frequentist properties of the three estimation methods were evaluated. To this end we used so-called violin-plots (Fig. 3), where the empirical distribution of 200 parameter estimates was compared with the average of the corresponding 200 95% interval bounds. In discussing our results it is important to stress that CIs provided by frequentist methods (HMM and ADMB) and credible intervals provided by Bayesian methods (BUGS) have fundamentally different interpretations. A 95% CI is an interval which contains the true parameter in 95% of a large number of repeated experiments. Conversely, a 95% credible interval is an interval which has a 95% posterior probability of containing the parameter for the experiment at hand. From Fig. 3 it was evident that the CIs were consistent with corresponding quantiles of the empirical distributions. This further confirms the validity of the quadratic approximation of the log-likelihood function. The empirical distributions of the BUGS parameter estimates under vague prior assumptions were largely identical to their HMM and ADMB counterparts. However, Fig. 3 showed that even when assigning vague priors it cannot be expected that credible intervals coincide with frequentist CIs, which by definition do not incorporate *a priori* knowledge. In addition, considerable differences in

credible intervals were present between the two BUGS analyses using different vague priors (Fig. 3). Thus, it is crucial, when employing Bayesian methods in the absence of *a priori* knowledge, to assess the sensitivity of credible intervals to the choice of distribution for the vague prior.

ADMB uses automatic differentiation to estimate the states and parameters of the model, which is the main reason for its computing time superiority (Table 2). This advantage will only increase further as models become more complex and the number of parameters grows. The main disadvantage of ADMB is, that the Laplace approximation for the density of the random effects (here equivalent to the latent states) must be reasonable. In our test cases the latent state estimation of ADMB was close to identical to the HMM and BUGS results (Fig. 2), which justifies using the Laplace approximation. If results from alternative methods are not available, the quality of the approximation can be assessed using the built-in importance sampling functionality (p. 35, ADMB-project, 2009b). Another possible complication of ADMB is that some programming experience in C++ is required. The HMM approach, on the other hand, has the advantage of being language independent, i.e. the method can be implemented in any programming language, for which a function optimizer is available. The programming background of the modeler is therefore of minor concern. The computing speed of the HMM approach is, at worst, proportional to the number of grid cells squared, a number which grows rapidly with increasing state dimension. Thus, HMMs are best suited for one or two-dimensional problems. BUGS depends less on state dimension because it is Monte Carlo based and it requires no density approximations nor differentiability. Consequently, BUGS is flexible and applicable to the widest variety of problems of the three methods we have examined. In addition, WinBUGS (Spiegelhalter et al., 2003) can be used to view and produce BUGS code graphically. This further increases the accessibility of the method.

BUGS and Monte Carlo based methods in general have tuning parameters that cannot be estimated from data and therefore require subjective input from the modeler. The tuning parameters include the number of samples, burn-in time, thinning rate, convergence assessment, and choice of prior distribution, all of which

influence the estimation results significantly. This fact is underlined in the BUGS manual (Spiegelhalter et al., 1996, p. 1), and it is emphasized that the modeler using BUGS must have a sound understanding of the Gibbs sampler. Our results supported this in that computing times (Table 2) and interval estimation (Fig. 3) were significantly influenced by the choice of prior. In contrast, ADMB has no tuning parameters as such, but it does have certain options that are more or less relevant depending on the type of problem, for example the SEPARABLE FUNCTION construct. HMM has two tuning parameters: the extent of the grid and the grid resolution. Limiting the state-space involves a risk of truncating the latent state path. To minimize this risk the approach of de Valpine and Hastings (2002) was followed, where bounds are chosen so wide that the probability of latent path truncation is negligible. Naturally, wider grid extents and higher grid resolution entail an increase in computation time. Thus, determining the value of these parameters is a trade-off between computing speed and accuracy of results. Generally, if one is uncertain about the grid specifications, we recommend to start with a wide and coarse grid to get preliminary results, and then adapt extents and refine the grid accordingly if needed. If the conclusion is unchanged on the adapted grid there is strong evidence that the latent path is enclosed and properly resolved by the discretization.

5. Conclusion

In summary, the three methods considered in this paper are all powerful approaches to nonlinear state-space modeling of ecological systems. ADMB is by far the fastest method owing to its use of the Laplace approximation and automatic differentiation. This limits ADMB to problems where the state distributions are unimodal, which, however, is the case in the majority of practical examples. In contrast, HMM and BUGS are more general and are able to handle arbitrary state distributions. HMM requires specification of a spatial grid and is limited to problems with low state dimensions, say below four. BUGS has fewest model restrictions, but requires specification of prior information and other subjective input from the modeler in the form of algorithmic tuning parameters.

State-space methods provide a natural paradigm for ecosystem modeling. Thus, it is imperative that the ecological community is alert to progress in other scientific fields where state-space models are used and developed. This paper evaluated the performance, with respect to estimation accuracy and speed, of three advanced methods for state-space analysis. The study showed that state and parameter estimation performance for all three methods was largely identical, however with BUGS providing overall wider credible intervals for parameters than HMM and ADMB confidence intervals.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ecolmodel.2011.01.007.

References

ADMB-project, 2009a. AD Model Builder: Automatic Differentiation Model Builder. Developed by David Fournier and freely available from admb-project.org.

ADMB-project, 2009b. Random Effects in AD Model Builder: ADMB-RE User Guide.

Cappé, O., Moulines, E., Ryden, T., 2005. Inference in Hidden Markov Models. Springer.

Casella, G., George, E., 1992. Explaining the Gibbs sampler. American Statistician 46 (3), 167–174.

Clark, F., Brook, B., Delean, S., Reşit Akçakaya, H., Bradshaw, C., 2010. The theta-logistic is unreliable for modelling most census data. Methods in Ecology and Evolution 1 (3), 253–262.

de Valpine, P., Hastings, A., 2002. Fitting population models incorporating process noise and observation error. Ecological Monographs 72, 57–76.

Gamerman, D., 1997. Sampling from the posterior distribution in generalized linear mixed models. Statistics and Computing 7 (1), 57–68.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1 (3), 515–533.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Barnardo, J., Berger, J., David, A., Smith, A. (Eds.), Bayesian Statistics, vol. 4. Clarendon Press, Oxford, UK, pp. 169–193.

Gimenez, O., Bonner, S., King, R., Parker, R., Brooks, S., Jamieson, L., Grosbois, V., Morgan, B., Thomas, L., 2008. WinBUGS for population ecologists: Bayesian modeling using Markov Chain Monte Carlo methods. In: Thomson, David L., C., E.G., Conroy, M.J. (Eds.), Modeling Demographic Processes In Marked Populations. Vol. 3 of Mathematics and Statistics. Springer, US, pp. 883–915.

Gimenez, O., Rossi, V., Choquet, R., Dehais, C., Doris, B., Varella, H., Vila, J., Pradel, R., 2007. State-space modelling of data on marked individuals. Ecological Modelling 206 (3–4), 431–438.

Jonsen, I.D., Flemming, J.M., Myers, R.A., 2005. Robust state-space modeling of animal movement data. Ecology 86 (11), 2874–2880.

Kitagawa, G., 1987. Non-Gaussian state-space modeling of nonstationary time series. Journal of the American Statistical Association 82 (400), 1032–1041.

Lambert, P., Sutton, A., Burton, P., Abrams, K., Jones, D., 2005. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. Statistics in Medicine 24 (15), 2401–2428.

Madsen, H., 2008. Time Series Analysis. Chapman & Hall/CRC London, London.

Madsen, H., Thyregod, P., 2010. Introduction to General and Generalized Linear Models. Chapman & Hall/CRC, London.

Morales, J., Haydon, D., Frair, J., Holsinger, K., Fryxell, J., 2004. Extracting more out of relocation data: building movement models as mixtures of random walks. Ecology 85 (9), 2436–2445.

Mortensen, S.B., 2009. Markov and mixed models with applications. Ph.D. Thesis, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark.

Oehlert, G.W., 1992. A note on the delta method. American Statistician 46 (1), 27–29.

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.

Polansky, L., De Valpine, P., Lloyd-Smith, J., Getz, W., 2009. Likelihood ridges and multimodality in population growth rate models. Ecology 90 (8), 2313–2320.

Schofield, M., Barker, R., MacKenzie, D., 2009. Flexible hierarchical mark-recapture modeling for open populations using WinBUGS. Environmental and Ecological Statistics 16 (3), 369–387.

Skaug, H., Fournier, D., 2006. Automatic approximation of the marginal likelihood in non-gaussian hierarchical models. Computational Statistics & Data Analysis 51 (2), 699–709.

Spiegelhalter, D., Thomas, A., Best, N., Gilks, W., 1996. Bayesian Inference Using Gibbs Sampling. Version 0.5 (Version ii). MRC Biostatistics Unit, Cambridge.

Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2003. WinBUGS User Manual. Version 1.4. MRC Biostatistics Unit, Cambridge, UK.

Vonesh, E.F., 1996. A note on the use of laplace's approximation for nonlinear mixed-effects models. Biometrika 83 (2), 447–452.

Wang, G., 2007. On the latent state estimation of nonlinear population dynamics using Bayesian and non-Bayesian state-space models. Ecological Modelling 200 (3–4), 521–528.

Wolfinger, R., Xihong, L., 1997. Two Taylor-series approximation methods for nonlinear mixed models. Computational Statistics & Data Analysis 25 (4), 465–490.

Zucchini, W., MacDonald, I., 2009. Hidden Markov Models for Time Series. Chapman & Hall/CRC, London.