ELSEVIER

On the use of stationary versus hidden Markov models to detect simple versus complex ecological dynamics[☆]

Brian C. Tucker, Madhur Anand*

Department of Biology, Laurentian University, 935 Ramsey Lake Road, Sudbury, Ont., Canada P3E 2C6

Abstract

The stationary Markov model (SMM) has been used to study simple ecological dynamics, such as classic Clementsian succession towards a climax. There has been considerable dissatisfaction among ecologists, however, because succession has been found to display complex dynamics. The application of hidden Markov models (HMM) is proposed for two reasons: (1) they can have multiple states with observations that need not converge on a stable configuration and (2) the hidden states allow for the detection of underlying ecological processes. A comparative analysis is made between the well-known SMM and the lesser known HMM using a range of hypothetical species response types with concentration on the prediction of ecological observation sequences and the detection of underlying ecological processes. The HMM provides similar predictive ability to that of the SMM in the case of simple dynamics but shows considerably improved performance for complex dynamics. The HMM also provides increased interpretive capabilities by suggesting where transitions in underlying hidden states can be identified, even when not apparent in the observable dynamics.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Hidden Markov; Stationary Markov; Species response; Log-likelihood; Hidden-state sequence; Ecological processes; Bayesian Information Criteria; Akaike's Information Criteria

1. Introduction

Since their introduction by Waggoner and Stephens (1970), Markov models have been widely applied in

ecology (e.g. Horn, 1975; Hulst, 1979; Binkley, 1980; Legg, 1980; Culver, 1981; Usher, 1981; Lippe et al., 1985; McAuliffe, 1988; Orlóci et al., 1993; Li, 1995; Valverde and Silvertown, 1997; Balzter, 2000; Logofet and Lesnaya, 2000; Benabdellah et al., 2003; Korotkov et al., 2001; Yemshanov and Perera, 2002; Tucker and Anand, 2003). The most common type of Markov model used has been the stationary Markov model (SMM). This is a relatively simple model in which the probability distribution of model states stabilizes on a fixed (or stable) distribution after a given num-

---

ber of model steps. It is this property that made the model ideal for modelling classic Clementsian succession (Clements, 1928) in which the distribution of model states represents species relative abundance in the community and the stable state represents the "climax". As such, it has been useful in some cases of modelling succession (Horn, 1975), recovery (Orlóci et al., 1993), and strong spatial environmental gradients (Tucker and Anand, 2003). The SMM is limited, however, in terms of its ability to describe dynamics other than monotonic increases or decreases in variables over space or time towards a stable state.

Most often in ecology, monotonic dynamics that stabilize on a fixed point are the exception rather than the rule (Anand, 2000). Ecological data often vary considerably from one observation to the next, even in the case of obvious successional patterns and strong gradients. This variability can mask underlying changes in the community making it difficult to distinguish between community states. For example, the succession of a forest may be masked by small perturbations, such as light wind damage or insect infestations. If these small perturbations in an otherwise "Clementsian" successional pathway occur with sufficient intensity or frequency, they may obscure the underlying succession process. That is, short-term variability may mask the long-term changes in community states. Childress et al. (1998) explored several stationary to semi-stationary Markov models and found them to perform poorly in primary succession when there was substantial species turnover. Korotkov et al. (2001) have applied SMM's to mixed boreal forest and found that the subclimax-to-climax transition does not occur in a mixed boreal forest in Central Russia and that a non-Markov model is required to describe the behaviour. Blatt et al. (2001) have shown how even small herbivore impacts can direct succession and result in a non-orderly approach to climax. In cases such as these, the SMM will not prove an effective descriptor of the dynamics because it requires that the ecological observations are represented directly by the model states and thus assumes that the observations are direct indicators of the community state. The model is therefore limited in terms of its ability to represent more abstract community states, such as successional stage or perturbation intensity. Benabdellah et al. (2003) defined "successional states" as model components that directly represent the forest community composition. However, we may wish to more objec-

tively define the underlying states, because we simply do not have enough information or knowledge to define them a priori. This approach may lead to the discovery of previously undetected community states.

The hidden Markov model (HMM) provides a framework for dealing with the shortcomings of the stationary Markov model. In contrast to the Markov and other single-matrix models, the discrete states of an HMM are partitioned such that there is a hierarchical coupling of two probabilistic matrices. HMMs are called hidden because the sequence of observations (ecological data) is the result of a stochastic process operating on top of a sequence of hidden states generated by a Markov process. At each state in the hidden sequence, an observable state is emitted according to the probability distribution for that state. It is likely that in many cases ecological observation sequences are the result of underlying hidden states with very different governing rules than simple analysis of the observations may suggest. For example, ecological processes (e.g. disturbance regimes) that are not easily observed directly may be changing in time or space, but must be determined from observable sequences of community states. This property allows the HMM to model more complex dynamics (e.g. non-monotonic responses), which are often present in ecological processes, than the SMM and offers increased flexibility to the modeller because it allows for the description of hierarchical, interacting systems or hidden processes underlying the observed ecological dynamics.

A similar approach would be that of Boudjema and Chau (1996) who use neural networks to look for attractors underlying noisy ecological dynamics. Borsuk et al. (2004) also apply probablistic models (in the form of Bayesian networks) to eutrification of a river estuary. While hidden Markov models can be considered a form of Bayesian network Ghahramani (1998), Borsuk et al. (2004) use the Bayesian network as an organising structure for a set of separate models. A second method of application might be to study the residuals derived from the examination of data by considering known factors that influence the dynamics. This would allow for the investigation and description, via the HMM, of the components of the ecological observations that cannot be easily explained by the known characteristics of a given ecological system. Our approach focuses on using the HMM as a predictor and descriptor of ecological

dynamics in order to investigate underlying community states.

The theory of hidden Markov models is not new (Baum and Petrie, 1966; Baum and Egon, 1967; Baum and Sell, 1968; Baum et al., 1970; Baum, 1972). Several tutorials on HMMs (Rabiner, 1989; Eddy, 1996; Ghahramani, 2001; Visser et al., 2002) have triggered their application in a wide range of disciplines, such as psychology, protein sequencing, and hydrology (Krogh et al., 1994; Krogh, 1998; Karplus et al., 1997; Thyer and Kuczera, 2003a,b; Aas et al., 1999; Kneale et al., 2001; Visser et al., 2002; Wong et al., 2001; Browning and Browning, 2002; Elliot and van der Hoek, 1997; Thomas et al., 2002). Hidden Markov models are flexible and relatively easy to implement since efficient methods exist to estimate the model parameters (Ghahramani, 1997), however, despite the wide appeal of these models for use in complex time series analysis and pattern detection in diverse fields, there have been few attempts to apply them to ecological problems. Viovy and Saint (1994) have applied HMMs to vegetation dynamics in satellite remote sensing time series, Hoef and Cressie (1997) to transect data in order to detect change-points in grassland vegetation, and Dale et al. (2002) to separate and assess the state and the underlying processes of a coastal wetland. While they all found HMMs to be adequate models for their cases, these authors did not link the methodology to or compare the results with stationary Markov models and no generalizations could be made regarding the types of species responses which were better modelled using the HMM.

We present here a comparative analysis of SMMs versus HMMs using a spectrum of hypothetical species response types. This paper also serves as a mini-review of HMMs with a focus on their application to ecological problems, specifically their ability to detect ecological processes and predict ecological observation sequences. The well-known SMM will serve as a familiar foundation for ecologists on which an understanding of the lesser known HMM can be built. Although there are several varieties of HMMs (Rabiner, 1989; Ghahramani, 1997; Brand, 1997; Fine et al., 1998) we will concentrate on the discrete, finite-state hidden Markov model. The discrete, finite states offer great flexibility to the model as hidden and observable states can represent either quantitative or qualitative properties of the system, thus, greatly expanding the field of

application of the model. Such an approach has been taken by Yoon and Korvin (2001) in fuzzy analysis of forest succession and Dambacher et al. (2003) who examine qualitative predictions from matrix models, however, their models lack the double embedded structure of HMMs.

2. Stationary Markov models

We begin by briefly describing the SMM. The SMM is completely defined by:

(i) N, the number of states in the model. Usually, the states reflect a direct observation in the system. The individual states are denoted as $S = S_1, S_2, \ldots, S_N$, and the state at time t is q_t.

(ii) X_{1_i}, the initial distribution of states at $t = 1$, where

$$X_{1_i} = P[q_1 = S_i], \quad 1 \le i \le N. \tag{1}$$

(iii) **A**, the state transition probability matrix whose elements (transitions between states i, j) are

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \quad 1 \le i, j \le N. \tag{2}$$

The state distribution, X_t, for model steps $t + 1, t + 2, \ldots, T$ are generated by:

$$X_t = X_{t-1} \times A. \tag{3}$$

The model components might be defined by any number of community attributes, but in our case the elements of the initial state distribution, X_1, represent the initial species abundances in the community ($t = 1$) and the elements of the state transition matrix, **A**, represent the gain or loss of cover abundance from one species to all other species (species replacement) at each model step. The distribution of the model states, X_t represents the relative abundance of each species at any time t.

To illustrate further the SMM, Fig. 1 shows a conceptual representation of a two-state SMM. This model is fully specified by the initial condition (4) and the transition matrix (5). Performing an eigenanalysis on the transition matrix provides the final stable-state distribution (6). Fig. 2 shows the model output from (4) and (5) over six model steps. This could represent relative cover abundance of two species over 6 years. Since the transition matrix (5) strongly favours transitions into
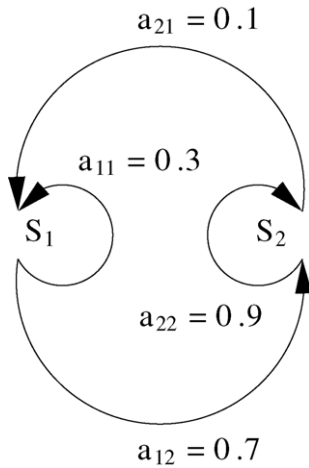
Fig. 1. A stationary Markov model with two states ($N = 2$), S_1 and S_2. Arrows represent possible state transitions at each model step according to the state transition matrix, $\mathbf{A} = a_{ij}$ (5).

state 2 (a_{12} and a_{12}), the state distribution stabilizes on (6) after only three model steps.

$$\mathbf{X}_1 = [\,0.5\ 0.5\,] \tag{4}$$

$$\mathbf{A} = \begin{bmatrix} 0.3\ 0.7 \\ 0.1\ 0.9 \end{bmatrix} \tag{5}$$

$$\mathbf{S}_{\text{stable}} = [\,0.125\ 0.875\,] \tag{6}$$

Clearly, the problem that must be considered prior to application of SMMs is the estimation of the parameters X_1 and A, given an observation sequence $O = O_1, O_2, \ldots, O_T$. Often X_1, the initial state distribution, is determined by the point at which observation of the system being modelled begins (e.g. the species abundances in a community immediately after perturbation in a recovery study). A can be parameterized through manipulative experiments or by detailed, fine-scale observation (Wootton, 2001 and Lippe et al., 1985). However, these data are not usually available because actual species transitions cannot often feasibly be observed in many ecological systems of interest (Hulst, 1979). The problem of estimating transitions indirectly from multivariate ecological time series (coenosere) was solved by Orlóci et al. (1993). The gain or loss in species abundances (to other species or bare ground) over each time step are used and recorded in a transition matrix unique to that time, t. The transition matrices are then averaged over time to produce a transition matrix that represents the overall observation sequence (and thus remains stationary).

3. Hidden Markov Models
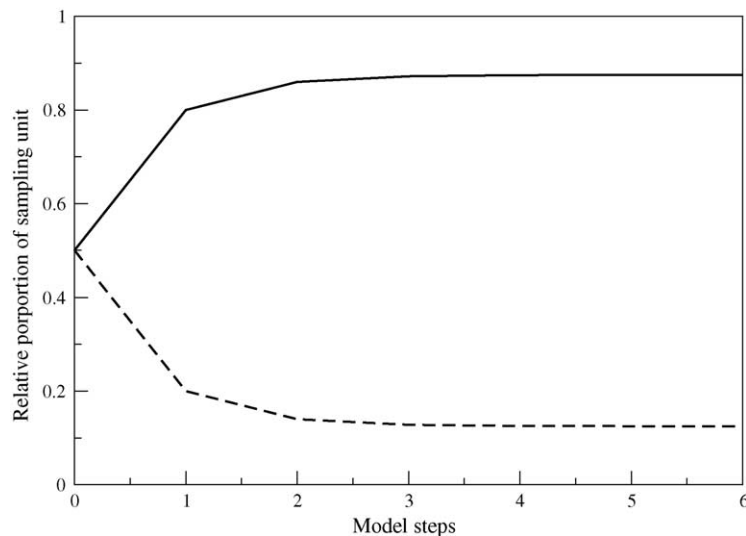
The HMM is characterized by the following:

Fig. 2. Dynamics of a two-state stationary Markov model over six model steps.

i. N, the number of hidden states in the model. The individual states are $S = S_1, S_2, \ldots, S_N$, and the state at time t is q_t.

ii. X_{1_i}, the initial distribution of states at $t = 1$, where

$$X_{1_i} = P[q_1 = S_i], \quad 1 \le i \le N. \tag{7}$$

iii. $\mathbf{A} = a_{ij}$, the state transition probability matrix, where

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \quad 1 \le i, j \le N. \tag{8}$$

iv. M, the number of discrete observable states for each hidden state. The observable states correspond to the observations made on the system. The individual states are $V = v_1, v_2, \ldots, v_M$.

v. $B = b_j(k)$, the probability distribution of observable states in state j, where

$$b_j(k) = P[v_k \text{ at } t | q_t = S_j],$$
$$1 \le j \le N, 1 \le k \le M. \tag{9}$$

Note that these first three components are similar to the SMM, but the states are hidden. Components (iv) and (v) link the hidden states to the observed.

The state of the model at each timestep changes according to the transition probabilities, a_{ij}, and emits an observable state from q_t according to $b_j(k)$ that represents the species cover abundance at time t.

Fig. 3 shows a discrete two-state HMM with six possible observations. The observable states $V = v_1, v_2, \ldots, v_M$ are $V = 1, 2, 3, 4, 5, 6$ and represent a cover abundance scale that could be used to sam-

ple relative species abundance in ecological communities. Of course, V can be of any size and type (e.g. $V = 1, 2, \ldots, 100$ for cover abundance estimated to the nearest percent) provided that the observations are discrete. The model is fully specified by (10)–(12).

$$\mathbf{X}_1 = [\, 0.5 \; 0.5 \,] \tag{10}$$

$$\mathbf{A} = \begin{bmatrix} 0.3 \; 0.7 \\ 0.1 \; 0.9 \end{bmatrix} \tag{11}$$

$$\mathbf{B} = \begin{bmatrix} 0.4 \; 0.2 \; 0.2 \; 0.1 \; 0.1 \; 0.1 \\ 0.1 \; 0.1 \; 0.1 \; 0.1 \; 0.2 \; 0.4 \end{bmatrix} \tag{12}$$

Note that $\mathbf{A}$, the state transition matrix, is the same as (5) in the SMM previously discussed and will have the identical stable state distribution of (6). Like the SMM, the underlying hidden-state distribution will stabilize, but the overlying observable states, $\mathbf{B}$, allow the model to capture the stochastic behaviour of observations. Examination of the probability distribution of observable states ($\mathbf{B}$) reveals that the much more probable S2 favours observable states 5 and 6 (higher cover abundance classes). The general behaviour of this model over many model steps would be a sequence of hidden states dominated by S2. As a result, the observation sequence would be comprised of a larger proportion of higher cover abundance observations. Fig. 4(A) shows one sample of the model over 100 steps. Fig. 4(B) shows the mean model behaviour of 1000 samples from the model over 100 steps. The initial lower
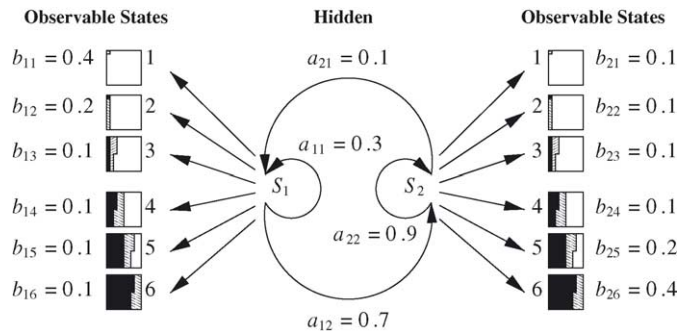

Fig. 3. A hidden Markov model with two states, S_1 and S_2, and six possible observations. Arrows in the hidden component indicate state transition probabilities of the transition matrix $\mathbf{A}$. Arrows connecting the hidden component to the observations of each state represent the observable state probabilities in $\mathbf{B}$. The boxes represent cover abundance classes where the hatch pattern indicates the cover abundance in a sampling unit (1 = 0–1%, 2 = 1–10%, 3 = 10–25%, 4 = 25–50%, 5 = 50–75%, and 6 = 75–100%).
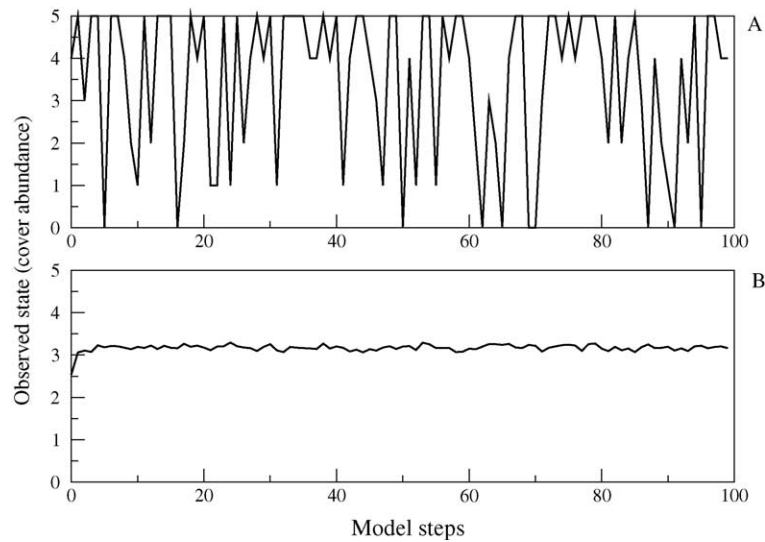
Fig. 4. Behaviour of HMM of Fig. 3 over 100 model steps. (A) One sample from the model; (B) average model output from 1000 samples from the model.

value of the observation mean is the result of the initial probability distribution where both states have equal probability of occurrence. The observation sequence quickly stabilizes, however, because of this model's hidden stable-state distribution (6).

While the SMM required the solution of only one problem prior to implementation, namely, estimation of the transition matrix, **A**, there are four problems that must be solved before applying HMMs to ecological data (Problems 1, 2, and 4 are described by Rabiner (1989) and Problem 3 is described by Visser et al. (2002)).

Problem 1. Find the probability of an observation sequence given a model, $P(O|\lambda)$, for any observation sequence, $O = O_1, O_2, \ldots, O_T$, and model, $\lambda = (A, B, X_1)$.

The likelihood of the observation sequence given the model, $P(O|\lambda)$, can be used as a relative measure of how well a particular model matches the data. The $P(O|\lambda)$ can be used for model selection from a pool of competing models of the same size. The model that results in the highest probability of generating the observation sequence is selected as the best model for the data. $P(O|\lambda)$ becomes very small in relatively short observation sequences and is often reported as the log-likelihood ($\log P(O|\lambda)$) because for sequences

over $T = 100$ the $P(O|\lambda)$ can exceed the numerical precision of the machine without scaling (Rabiner, 1989). The forward–backward procedure (Baum and Egon, 1967; Baum and Sell, 1968) can be used as an efficient method to find $P(O|\lambda)$ (see Appendix A.1 for more details).

Problem 2. Maximize the $P(O|\lambda)$ by adjusting the model parameters A, B, and X_1.

Parameters are estimated using the Baum–Welch (also the EM) algorithm as outlined by Rabiner (1989) and initially introduced by Baum et al. in a series of papers (Baum and Petrie, 1966; Baum and Egon, 1967; Baum and Sell, 1968; Baum et al., 1970; Baum, 1972). The Baum–Welch (EM) algorithm is a procedure that maximizes the probability of the observation sequence given the model ($P(O|\lambda)$) by iteratively estimating new model parameters until a local maximum of $P(O|\lambda)$ is reached; it finds the maximum likelihood estimate of the model parameters (see Appendix A.2 for more details).

Problem 3. Select the model that is most suitable for a given observation sequence, $O = O_1, O_2, \ldots, O_T$, from a pool of models having different N and/or M.

While the observable states are most likely fixed by the manner of data collection, the number of hidden

states of the model, N, in many cases can be varied. In fact, it may be of great interest to test many models having different numbers of states in an effort to understand how many underlying processes may be responsible for a particular ecological phenomenon. With this in mind, it is useful to find the maximum likelihood estimate of many models of different size when looking at an ecological data series. Obviously, by adding states to the model the $P(O|\lambda)$ will increase. When comparing models of a different size, there is a need for a method of model selection that takes into account not only the log-likelihood of the model but also the model complexity (number of model parameters). The solution to this problem may be particularly important to ecologists who are looking for possible relationships between the hidden states of the model and underlying ecological processes and/or interactions because it provides an answer to the question, "given the data and a series of models with differing numbers of states, how many states should the final model have?" This parallels the question, "given the data, how many important ecological processes might be operating?" Several methods exist for model selection, such as minimum description length criterion (MDL) (Rissanen, 1978), minimum message length (MML) (Wallace and Freeman, 1978), Bayesian Information Criterion (BIC) (Schwarz, 1978), and Akaike's Information Criterion (AIC) (Akaike, 1973). We use adjusted versions of BIC and AIC after Visser et al. (2002). The disagreement between the two information criteria with respect to the size of the models required for some observable sequences is not to be taken as a surprise. Visser et al. (2002) found the A-BIC criteria to be superior in computer simulations because it chose appropriately smaller models than A-AIC. Here we have used both methods in order to identify a method of selection that will provide better predictive power (A-AIC) and a method which provides a conservative lower bound for model state size (A-BIC) (see Appendix A.3 for more details).

Problem 4. Find the sequence of hidden states, $Q = q_1, q_2, \ldots, q_T$, that maximizes the probability of the observation sequence, $O = O_1, O_2, \ldots, O_T$, given the model, λ.

Finding the optimal sequence of hidden states is useful for assigning ecological significance to the hidden component of the model. If the hidden states are considered representative of hidden ecological states, the sequence of those states, $Q = q_1, q_2, \ldots, q_T$, that maximizes the likelihood of the observation sequence, $O = O_1, O_2, \ldots, O_T$, will reveal the most probable spatial or temporal pattern of ecological states responsible for the observed community states. While the transition matrix and the stable-state distribution will give an indication of the relative frequency of the states and is useful for a general description of the hidden process, the state sequence, Q, can be useful when looking at gradients or temporal dynamics where the particular order of the ecological states may indicate a shift in environmental conditions or ecological processes. It is possible to extract the optimal sequence of states from a discrete, finite-state Markov process using the Viterbi algorithm (see Viterbi (1967) for the original theory, Forney (1973) for a tutorial applied to HMMs, and Rabiner (1989) for the methodology used in this paper). The algorithm finds the maximum a posteriori probability (MAP) estimate of the state sequence of the Markov process (Forney, 1973).

4. Data

A suite of test data representing hypothetical ecological observation sequences (dynamics) have been used to illustrate the behaviour of HMMs and to address the four problems of implementation. Fig. 5 shows theoretical data from six species response-types (A–E) where cover abundance in sampling units is measured using a scale of six possible observations (as in Fig. 3) over 20 sampling units (in either space or time). The data sets were simulated to illustrate possible species responses to various ecological scenarios. Response-type A shows a repetitive pattern of weak increase and decrease in abundance between each sampling period and could represent the annual cycling of a biennial or the cycling of vegetation cover as the result of wet and dry seasons. Response-type B and C illustrate responses to strong environmental gradients. Response-type D shows a more complex response where small short-term fluctuations are superimposed on a gradient that spans the sampling period. Response-type E shows a large initial increase and decrease followed by a repetitive pattern of weak increase and decrease. Such a response might be representative of the temporal dynam-
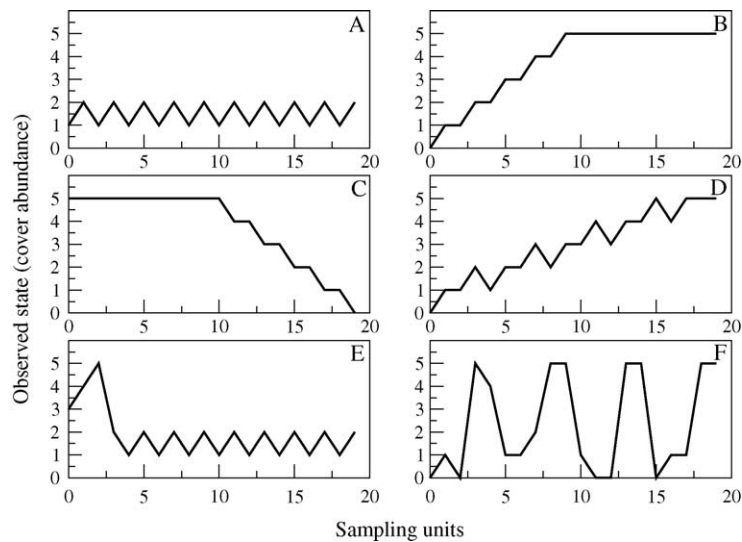
Fig. 5. Theoretical data of cover abundance of a sampling unit sampled using a scale (0–5) showing possible species responses to ecological scenarios: (A) recurring wet/dry period, (B and C) strong environmental gradient, (D) weak environmental gradient, (E) early succession of a colonizer after a major disturbance, (F) recurring major disturbance, such as extreme drought.

ics of an early successional stage species colonizing a fire-disturbed forest region, then declining as later stage successional species re-colonize. The repetitive pattern after sampling unit 4 represents responses to wet/dry periods as in response-type A. Response-type F shows a strong, repetitive pattern of species increase and decrease that might represent a recurring disturbance, such as extreme drought, that almost totally removes the species. Rapid increases in cover abundance would result from regeneration from seed during wet periods.

5. Results

5.1. Inadequacy of the stationary Markov model

A stationary Markov model was fitted to the matrix of species observation sequences in Fig. 5 according to the method of Orlóci et al. (1993). Although the stationary Markov model is normally fitted to a community matrix of several species, here it was applied individually to each response-type in order to make a direct comparison to the HMM projections and because each response-type represents a different ecological scenario. Fig. 6 shows the SMM projections of the single species after fitting of the model to each species observation sequence and probabilities

from randomisation testing for Markovity using the methodology of Orlóci et al. (1993). Response-types A, E, and F were not adequately fit by the SMM while response-types B, C, and D were. Response-type A projections stabilize quickly on a low cover abundance value. The projections of response-type B and D both capture the gradient response. However, B projections show a gradual increase in species abundance while D projections indicate a rapid increase in abundance followed by a stable period. These dynamics are opposite that of the actual species observation sequences. The model does show the gradient response of type C but cannot capture the long initial stable period prior to the gradient response or the complete disappearance of the species by the last sampling unit. The SMM does capture the initial higher cover abundance of response-type E in the early sampling units but fails to capture the much smaller, repetitive increases and decreases. The response of the model in the case of response-type F is similar to that of A in that it increases in abundance quickly before stabilizing.

5.2. Adequacy of hidden Markov models

For each theoretical response-type in Fig. 5, the log-likelihood of 500 trained models resulting from ran-
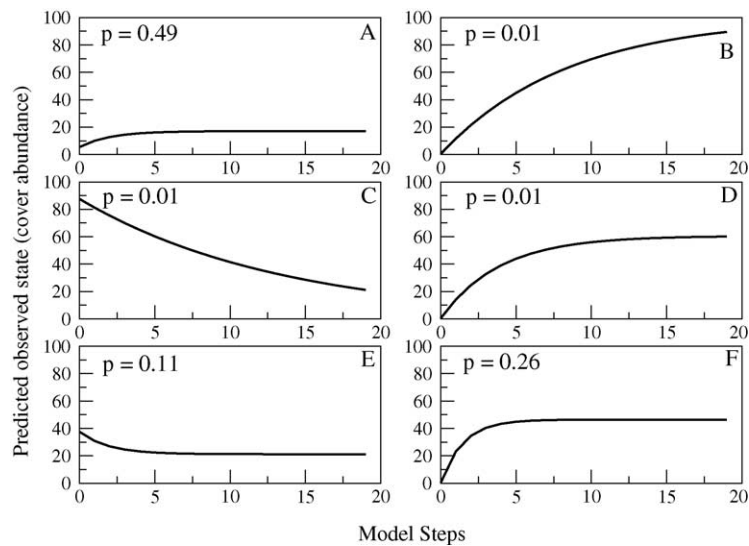
Fig. 6. The species abundances (% of sampling unit) from stationary Markov models fitted separately to each species and its accompanying barren ground variable (after Orlóci et al. (1993)). Only species projections are shown. Probabilities are from randomisation testing that the dynamics are not stationary Markov in nature. In the case of A, E, and F we reject the null hypothesis of a Markov process at $p < 0.05$.

dom a priori estimates of $\lambda = (A, B, X_1)$ was used as the selection criteria to choose the best model for the data given a fixed model size (N and M). This procedure was used to address the problem of the Baum–Welch algorithm finding only one local maximum of the likelihood estimate, while in fact the likelihood surface for the problem may have many local maxima

(Rabiner, 1989). By choosing many different initial values of λ, we attempt to sample much of the likelihood surface and ultimately choose the most likely model from many local maxima. The procedure was carried out separately for seven groups of models having $N = 2, 3, 4, 5, 6, 7, 8$ hidden states. Fig. 7 shows log $P(O|\lambda)$ for each response-type as the number of
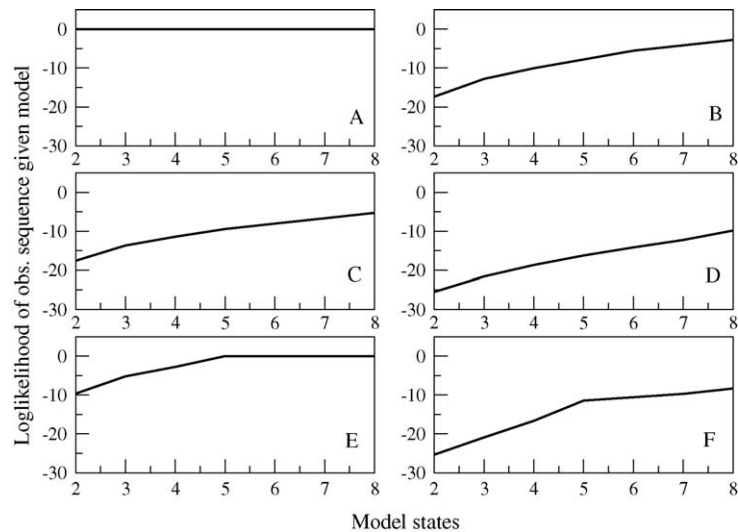

Fig. 7. The log-likelihood (log $P(O|\lambda)$) estimates for response-types A–F plotted against the number of hidden states in the model.

hidden states increases. The slope of the log-likelihood curve, as it approaches zero, indicates the improvement of $P(O|\lambda)$ by increasing the number of states. A log-likelihood of zero indicates a $P(O|\lambda) = 1$ where the model can produce exactly the observation sequence. The maximum likelihood estimate curves differ considerably between most of the species response-types. The curves for some species converge to log $P(O|\lambda) = 1$ at relatively low model orders (e.g. response-types A at two states, response-type E at five states) while others have yet to reach zero by eight states (e.g. response-types B, C, D, F). The log $P(O|\lambda)$ curve for response-type A is zero for all states, indicating that cyclic dynamics can be easily captured using an HMM having two states. The log $P(O|\lambda)$ curves for response-types B and C are almost the same at lower model sizes, but the curve for B increases with a greater slope than does that of C. Response-type E increases rapidly with increasing numbers of states, becoming zero at five states, while D and F both increase at a low slope and have the lowest log $P(O|\lambda)$ at eight states.

Table 1 shows results for assessing model size selection. Both selection measures (A-BIC and A-AIC) agree on the model size for response-type A (two states), B (three states), and C (two states), but disagree on the remaining response-types. The more conservative A-BIC criteria consistently selects for smaller

Table 1
Model size suggested by the information criteria A-BIC and A-AIC

Response-type	Adjusted information criteria	
	A-BIC	A-AIC
A	2	2
B	3	3
C	2	2
D	3	4
E	2	3
F	3	5

models than does A-AIC, providing a range of models for response-type D (three to four states), E (two to three states), and F (three to five states). The parameter estimates for the model sizes selected by A-AIC are shown in Appendix A. Fig. 8 shows the mean predicted observations from 1000 models of the size selected by the A-AIC. The HMM models selected by A-AIC provide good fit to the response dynamics with the exception of response-type C. The model fails to reflect the initial stable dynamics from sampling unit 0–10 (Fig. 8C) and does not effectively capture the decline of response-type C to extinction. Response-type A is completely specified by the two-state model and represents a very simple observation sequence because of its repeating pattern. Response-type F also has a repeating pattern, but does not have the perfect replication of
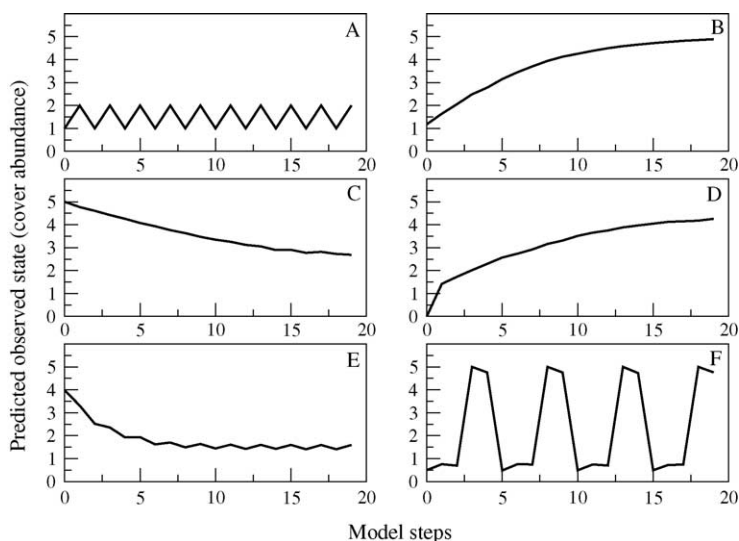

Fig. 8. The mean predicted observations of 1000 samples of the model size selected by A-AIC (Table 1) over 20 model steps. Models used in information criteria model selection provided the highest log $P(O|\lambda)$ from 500 models trained by the Baum–Welch algorithm.
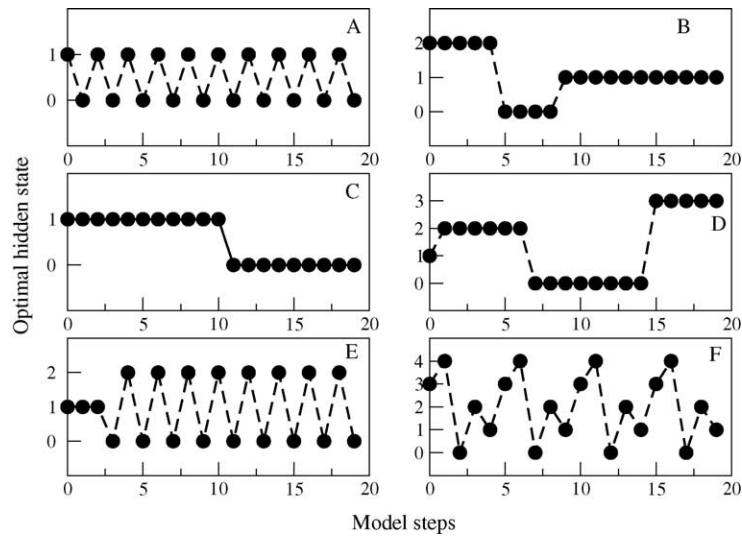
Fig. 9. The optimal state sequence, as determined by the Viterbi algorithm, that would maximize log $P(O|\lambda)$ given the model, $\lambda = (A, B, X_1)$. Circles represent the discrete states at each step.

observations as does Response-type A. The five-state HMM in Fig. 8F, however, does capture the strong repeating cycles of increases and decreases despite the original data lacking a perfectly repeating pattern as in response-type A. The gradient responses of B and D are both captured well by the HMM. The relatively steep gradient of response-type B compared to the gradual gradient response of D is also captured. The initial unimodal response of E is reflected in the HMM output as the period of slightly higher mean values (from model step 0–5) in Fig. 8. Even after capturing the initially high abundance of response-type E, the HMM is capable of capturing the fluctuations in relatively low levels of cover abundance from sampling unit 5 to 20. Although the SMM could capture the initial higher cover abundance values, it gives no indication of the repeating pattern of response-type E after sampling unit 5.

Fig. 9 shows the optimal sequence of hidden states of each species as determined by the Viterbi algorithm. Only the sequences for the model sizes selected by the A-AIC criteria (Table 1) are shown here. In some cases, the hidden state dynamics reflect the observed dynamics in Fig. 5, but in others they do not. The hidden-state sequence for response-type A is a simple oscillating pattern between the two states of the model. Response-types B, C and D show a gradual changing of states along the gradients but the rates at which they change

are different. Response-type B changes through two hidden states in the first seven model steps, coinciding closely with the steep gradient in Fig. 5B. It then remained in the third state. Response-type C does not exhibit a change in its hidden state until model step 11, which coincides with the gradient in Fig. 5C. Response-type D indicates a gradual progression through the four model states along the gradient. Response-type E indicates a period of hidden-state stability before an oscillating-state pattern emerges. Response-type F shows a repeating pattern of five states at the same frequency as the major increases in cover abundance in Fig. 5F.

6. Discussion

In the case of simple (monotonic) dynamics the HMM offers similar predictive power to that of the SMM, but the HMM provides improved predictive power in the case of more complex dynamics. The good performance of the SMM in describing the simple dynamics (response-type B & D) illustrates the model's well-known abilities and underscores why it has generated extensive interest as a succession model. In the case of response-type C, however, the HMM did not perform as well as the SMM, although neither model

captured the exact point of species decline. This behaviour is supported by the findings of Balzter (2000), who has shown that the stationary Markov model performs poorly in the case of increased variablility in later stages of the observation sequence. This is a consequence of the stationary transition matrix. Response-type A is completely specified by the two-state HMM and represents a trivial case for the HMM but not so for the SMM. The poor performance of the SMM in our example data supports the findings of Childress et al. (1998) who have found stationary and semi-stationary Markov models to perform poorly in situations with highly variable species responses during primary succession. Response-type F also has a repeating pattern, but does not have the perfect replication of observations as does response-type A. It is interesting, however, that the five-state HMM still captures the strong repeating cycles of species increases and decreases. Viovy and Saint (1994) have used this property to model seasonal vegetation dynamics in satellite remote sensing time series, where the states of the model were related to stages of vegetation growth in West African savannas, such as wet and dry seasons.

The dynamics of a HMM can be very noisy due to its doubly stochastic nature. Average model behaviour need not represent the dynamics of the individual observation sequences. When training the models to observation series, the stochastic nature of the observations are captured in the model parameters. For example, a sample observation sequence from an HMM could be used to train a second, equivalent model that would produce the same average behaviour. The mean values of many model samples represent the distribution dynamics. However, the ecologist may also be interested in samples of any particular model in order to appreciate the diversity of possible observation predictions. This is conceptually similar to the methods in MacFarlane et al. (2000) who use probability distributions of model parameters in a stand-level model of forest dynamics to deal with uncertainty in the physiological and morphological parameter estimates.

The log-likelihood estimates, found in the solution to Problem 1, provide a method of clustering ecological observation sequences by a model using some lower bound of $P(O|\lambda)$. For example, the observation sequences of response-type X could be compared to that of response-type Y by calculating $P(O_Y|\lambda_X)$ and $P(O_X|\lambda_Y)$. By using some lower bound of $P(O|\lambda)$,

it would be possible to determine if similar processes were responsible for the ecological observation sequences. This concept can be extended to a probabilistic distance measure between two HMMs (Juang and Rabiner, 1985 and Falkhausen et al., 1995) that can be considered analogous to the well-known Euclidean distance used to summarize ecological data sets. Given an N species community matrix and the corresponding N models, it would be possible to form a diagonal matrix of distance measures between the species HMMs that can then be summarized using classic ordination techniques.

In the case of response-type A, a two-state HMM provides perfect predictive power for the observation sequence ($P(O|\lambda) = 1$). In the other response-types, we see that increasing the number of states increases the log-likelihood by varying degrees. In the case of the gradient examples (response-types B, C, and D), the gradual increase in the log-likelihood as states are added is indicative of only moderate increases in the predictive ability of these models. The increases in power that can be achieved by adding states are subject to the limits of the underlying Markov process and the number of sampling units (model steps) between successive occurrences of a repeating phenomena. In the case of the former, adding states to improve model results when data are strongly non-linear (e.g. unimodal responses over many sampling units), will be largely ineffective until very large models are considered. In the case of the latter, the HMM must have at least as many hidden states as the length (in sampling units) of the repeating phenomena. The HMM's ability to capture responses approaching a stable state (response-type B) and its inability to capture responses leaving a stable condition (response-type C) illustrates a consequence of the model's underlying Markov properties. The HMM, because of its relation to the SMM, the state process, albeit hidden, is still stationary Markov and that the hidden-state distribution will stabilize after a fixed number of model steps. A more complex model (more states) is required to obtain the same log $P(O|\lambda)$ in response-type C as in B. The reason for this lies in the hidden component of the model. As it is a stationary Markov process, and therefore is sensitive to initial conditions, it does not perform well when there are dramatic changes late in the observation sequence. This illustrates the uni-directional nature of the SMM and, to a lesser extent, the HMM.

Even though the HMM, like the SMM, poorly predicted the gradient of response-type C, analysis of the hidden-state dynamics provides additional insight into the response that the SMM does not. While there are methods to evaluate the short-term, transient dynamics of matrix population models (Fox and Gurevitch, 2000; Yearsley, 2004), these approaches do not focus on association of the dynamics to different ecological process states. Favier et al. (2004) model the succession of a forest-savannah by linking vegetation states to key ecological processes, such as fire and establishment of trees. Such approaches, however, require knowledge regarding the active ecological states influencing species or community dynamics which may not be fully known for the system of interest. In such cases, an exploratory approach, such as the one described in this paper, may shed light on the linkages between vegetation states and ecological processes, as well as indicate the number of important processes in the system. The optimal sequence for the HMM of response-type C indicates that there is a change of hidden state when the species begins to respond to the environmental gradient. The model effectively partitioned the dynamics of response-type C into two main processes (species stability and species decline) and indicated where the process change occurs. The optimal state sequence for response-type E also indicates a major change of process that separates the initial high abundance values from the repeating increases and decreases of lower cover abundance.

The general trend of the hidden-state sequences mirroring major changes in the observation sequences is illustrative of the HMM's ability to capture observation-state linkages. The fact that the model states of the optimal sequence reflect the dynamics of the observation sequence indicates that the underlying states can provide meaningful information regarding potential processes responsible for the observation sequences. Knowledge of such linkages between the ecosystem components is required to make complex, linked ecosystem models (He et al., 1999). Often, however, data describing the linkages are missing, in which case the approach described here can be used to infer the linkages between ecological states from the data at hand. In some cases, (e.g. response-type B), a hidden-state change may not have been detected by the observed sequence. This is where the HMM is most useful. The hidden states, themselves, could represent abstract ecological states representing complex, multivariate processes and interactions (e.g. perturbation intensity or successional stage), but could also point to simpler structural changes (Hoef and Cressie, 1997; Dale et al., 2002).

When the hidden states are considered analogous to abstract ecological processes, such as successional stage, having a range of possible numbers of states may actually provide insight into the underlying system at hand. Phenomena that were initially considered separate processes in the ecological description of the system may be reconstituted into another set of equivalent processes that is larger (or smaller) than the first. Unlike Boudjema and Chau (1996), who use relatively large data sets to examine underlying components, we show that the HMM shows promise in giving insight to ecological process even with relatively short observation sequences. This promises to serve as an important characterisation and forcasting tool for ecological problems where data is limited or in the early states of collection.

7. Conclusions

The stationary Markov model (SMM) is a practical and effective model of simple ecological dynamics. The use of HMMs in ecological applications should not replace the SMM, but serve as a natural extension to the simpler model's capabilities. The HMM not only provides similar predictive ability to that of the SMM for simple dynamics, it shows considerably improved performance for cyclic dynamics and other non-monotonic responses and has the ability to summarize highly variable observation sequences. In addition to this, the HMM provides increased interpretive capabilities by representing 'hidden' processes by the underlying states. The optimal state sequence, given a model of a specific observation sequence, provides insight into the ecological processes underlying the observed dynamics. This property holds even when the average dynamics of the model are not exceptionally good predictors of the original data. Future work should consider the coupling of HMMs together to represent interactions between different response-types.

Acknowledgements

Appendix A

A.1. Problem 1: Probability of observation sequence

The forward–backward procedure (Baum and Egon, 1967 and Baum and Sell, 1968) can be used as an efficient method to find $P(O|\lambda)$. Below, we present a more simplified description than Rabiner (1989). The forward variable of the procedure is denoted as:

$$\alpha_t(i) = P(O_1, O_2, \ldots, O_t, q_t = S_i|\lambda). \tag{A.1}$$

To find $\alpha_t(i)$ via induction, we initialize the forward variable for state i as the product of the initial state probability, X_{1_i}, and the probability of the observation at $t = 1$ given state i.

$$\alpha_1(i) = X_{1_i} b_i(O_1), \quad 1 \le i \le N. \tag{A.2}$$

For each subsequent t and j, we find,

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N}(\alpha_t(i)a_{ij})\right] \times b_j(O_{t+1}),$$

$$1 \le t \le T - 1, \quad 1 \le j \le N. \tag{A.3}$$

The probability of the observation sequence is then found by summing all the forward variables across states.

$$P(O|\lambda) = \sum_{i=1}^{N}(\alpha_T(i)) \tag{A.4}$$

The backward variable is required for the estimation procedure and is similar to the forward variable.

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, \ldots, O_T, q_t = S_i|\lambda) \tag{A.5}$$

To find $\beta_t(i)$ via induction we initialize the backward variable.

$$\beta_T(i) = 1, \quad 1 \le i \le N. \tag{A.6}$$

For each subsequent t and j, we find,

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \times \beta_{t+1}(j),$$

$$t = T - 1, T - 2, \ldots, 1, \quad 1 \le j \le N. \tag{A.7}$$

For observation sequences longer than 100 the forward and backward variables must be scaled to keep them within the precision of the the computer (Rabiner, 1989).

A.2. Problem 2: Estimating model parameters to maximize $P(O|\lambda)$

Parameters are estimated using the iterative Baum–Welch algorithm as outlined by Rabiner (1989) and initially introduced by Baum et al. (Baum and Petrie, 1966; Baum and Egon, 1967; Baum and Sell, 1968; Baum et al., 1970; Baum, 1972). Rabiner (1989) points out that for any finite sequence of observations used as training data, there is no optimal way of estimating the parameters of the model and that using the Baum–Welch algorithm allows for only the local maximization of $P(O|\lambda)$ for a chosen $\lambda = (A, B, X_1)$. Thus, for a given observation sequence, and several a priori estimates of λ, there can be several a posteriori maximum likelihood estimates of λ. This results in several competing models being found for a single observation sequence after several Baum procedures with different a priori estimates of λ. Here, we use random values for the a priori estimates of $\lambda = (A, B, X_1)$. The random values are standardized according to (A.8).

$$\sum_{i=1}^{N} a_{ij} = 1, \quad \sum_{i=1}^{N} b_{ij} = 1, \quad \sum_{i=1}^{N} X_{1_i} = 1. \tag{A.8}$$

Using the forward and backward algorithms, the initial estimates of A, B, and X_1 can be re-estimated to improve $P(O|\lambda)$. Below, we present a more simplified description than Rabiner (1989).

$$\xi_t(i, j) = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)} \tag{A.9}$$

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i, j) \tag{A.10}$$

The new estimates of λ are,

$$\bar{X}_i = \gamma_1(i), \tag{A.11}$$

$$\bar{a}_{ij} = \frac{\sum_{i=1}^{T-1} \xi_t(i, j)}{\sum_{i=1}^{T-1} \gamma_t(i)}, \tag{A.12}$$

$$\bar{b}j(k) = \frac{\sum_{t=1}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}. \tag{A.13}$$

A.3. Problem 3: Model size selection

The unadjusted AIC and BIC are largely unusable in model selection with these models because of the rapid increase in parameters as model orders increase. Instead, adjusted versions of the Bayesian and Akaike's information criteria, A-BIC and A-AIC, respectively, are used for model selection of HMMs (Visser et al., 2002). These methods give a score based on the model's likelihood (log $P(O|\lambda)$) but add a penalty component based on the model's complexity. Model complexity is quantified using the number of free parameters in the model. In the case of HMMs, however, the enumeration of free parameters need not include those parameters of the HMM with a value of zero because they have no effect on the stochastic processes of the model (provide no information regarding the model processes) (Visser et al., 2002). This is a particularly important consideration with hidden Markov models because of the rapid increase in the number of free parameters with increasing numbers of states. The model size that minimizes these functions is chosen as the best model size where there is an appropriate trade-off between log $P(O|\lambda)$ and model complexity.

The Bayesian and Akaike's information criteria are given below,

$$\text{BIC} = -2 \log L + 2n_p \tag{A.14}$$

$$\text{AIC} = -2 \log L + n_p \log(T) \tag{A.15}$$

where, n_p (number of free parameters) is $N(N-1) + N(M-1) + (N-1)$.

Adjusted versions of the criteria are given by,

$$A - \text{BIC} = -2 \log L + 2(n_p - n_{p=0} + n_{p=1}) \tag{A.16}$$

$$A - \text{AIC} = -2 \log L + 2(n_p - n_{p=0} + n_{p=1}) \log(T) \tag{A.17}$$

Where, $n_{p=0}$ is number of parameters equal to 0, $n_{p=1}$ is number of parameters equal to 1.

A.4. Response-type HMM parameter estimates

Response-type A

$$\mathbf{X}_1 = [\,0.00 \; 1.00\,] \tag{A.18}$$

$$\mathbf{A} = \begin{bmatrix} 0.00 \; 1.00 \\ 1.00 \; 0.00 \end{bmatrix} \tag{A.19}$$

$$\mathbf{B} = \begin{bmatrix} 0.00 \; 0.00 \; 1.00 \; 0.00 \; 0.00 \; 0.00 \\ 0.00 \; 1.00 \; 0.00 \; 0.00 \; 0.00 \; 0.00 \end{bmatrix} \tag{A.20}$$

Response-type B

$$\mathbf{X}_1 = [\,0.00 \; 0.00 \; 1.00\,] \tag{A.21}$$

$$\mathbf{A} = \begin{bmatrix} 0.75 \; 0.25 \; 0.00 \\ 0.00 \; 1.00 \; 0.00 \\ 0.20 \; 0.00 \; 0.80 \end{bmatrix} \tag{A.22}$$

$$\mathbf{B} = \begin{bmatrix} 0.00 \; 0.00 \; 0.00 \; 0.50 \; 0.50 \; 0.00 \\ 0.00 \; 0.00 \; 0.00 \; 0.00 \; 0.00 \; 1.00 \\ 0.20 \; 0.40 \; 0.40 \; 0.00 \; 0.00 \; 0.00 \end{bmatrix} \tag{A.23}$$

Response-type C

$$\mathbf{X}_1 = [\,0.00 \; 1.00\,] \tag{A.24}$$

$$\mathbf{A} = \begin{bmatrix} 1.00 \; 0.00 \\ 0.09 \; 0.91 \end{bmatrix} \tag{A.25}$$

$$\mathbf{B} = \begin{bmatrix} 0.11 \; 0.22 \; 0.22 \; 0.22 \; 0.22 \; 0.00 \\ 0.00 \; 0.00 \; 0.00 \; 0.00 \; 0.00 \; 1.00 \end{bmatrix} \tag{A.26}$$

Response-type D

$$\mathbf{X}_1 = [\,0.00 \; 1.00 \; 0.00 \; 0.00\,] \tag{A.27}$$

$$\mathbf{A} = \begin{bmatrix} 0.86 \; 0.00 \; 0.00 \; 0.14 \\ 0.00 \; 0.00 \; 1.00 \; 0.00 \\ 0.19 \; 0.00 \; 0.81 \; 0.00 \\ 0.00 \; 0.00 \; 0.00 \; 1.00 \end{bmatrix} \tag{A.28}$$

$$\mathbf{B} = \begin{bmatrix} 0.00 & 0.00 & 0.22 & 0.56 & 0.22 & 0.00 \\ 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.56 & 0.44 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.38 & 0.62 \end{bmatrix} \tag{A.29}$$

Response-type E

$$\mathbf{X}_1 = [\, 0.00 \ 1.00 \ 0.00 \,] \tag{A.30}$$

$$\mathbf{A} = \begin{bmatrix} 0.00 & 0.00 & 1.00 \\ 0.33 & 0.67 & 0.00 \\ 1.00 & 0.00 & 0.00 \end{bmatrix} \tag{A.31}$$

$$\mathbf{B} = \begin{bmatrix} 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.33 & 0.33 & 0.33 \\ 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix} \tag{A.32}$$

Response-type F

$$\mathbf{X}_1 = [\, 0.00 \ 0.00 \ 0.00 \ 1.00 \ 0.00 \,] \tag{A.33}$$

$$\mathbf{A} = \begin{bmatrix} 0.00 & 0.00 & 1.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.00 & 0.00 \\ 0.00 & 1.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \\ 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix} \tag{A.34}$$

$$\mathbf{B} = \begin{bmatrix} 0.50 & 0.25 & 0.25 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.25 & 0.75 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \\ 0.50 & 0.50 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.25 & 0.75 & 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix} \tag{A.35}$$

References

Aas, K., Eikvil, L., Huseby, R., 1999. Applications of hidden Markov chains in image analysis. Pattern Recognit. 32, 703–713.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B., Cáski, F. (Eds.), Second International Symposium on Information Theory, Budapest, Akademiai Kaidó. pp. 267–281.

Anand, M., 2000. Fundamentals of vegetation change: complexity rules. Acta Biotheor. 48, 1–14.

Balzter, H., 2000. Markov chain models for vegetation dynamics. Ecol. Model. 126, 139–154.

Baum, L., 1972. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. Inequalities 3, 1–8.

Baum, L., Egon, J., 1967. An inequality with applications to statistical estimation for probabilistic functions of a Markov process and to a model for ecology. Bull. Am. Math. Soc. 73, 360–363.

Baum, L., Petrie, T., 1966. Statistical inference for probabilistic functions of finite state Markov chains. Ann. Math. Stat. 37, 1554–1563.

Baum, L., Petrie, T., Soules, G., Weiss, N., 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Ann. Math. Stat. 41 (1), 164–171.

Baum, L., Sell, G., 1968. Growth functions for transformations on manifolds. Pac. J. Math. 27 (2), 211–227.

Benabdellah, B., Albrecht, K., Pomaz, V., Denisenko, E., Logofet, D., 2003. Markov chain models for forest successions in the Erzgebirge, Germany. Ecol. Model. 159, 145–160.

Binkley, C., 1980. Is succession in hardwood forests a stationary Markov process? Forest Sci. 26, 566–570.

Blatt, S., Janmaat, J., Harmsen, R., 2001. Modelling succession to include a herbivore effect. Ecol. Model. 139, 123–136.

Borsuk, M., Stow, C., Reckhow, K., 2004. A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. Ecol. Model. 173, 219–239.

Boudjema, G., Chau, N., 1996. Revealing dynamics of ecological systems from natural recordings. Ecol. Model. 91, 15–23.

Brand, M., 1997. Coupled hidden Markov models for modeling interacting processes. Tech. Rep. Learning and Common Sense Technical Report 405, Media Lab Perceptual Computing, MIT.

Browning, S., Browning, B., 2002. On reducing the statespace of hidden Markov models for the identity by decent process. Theor. Popul. Biol. 62, 1–8.

Childress, W., Crisafulli, C., Rykiel, E., 1998. Comparison of Markovian matrix models of a primary successional plant community. Ecol. Model. 107, 93–102.

Clements, F., 1928. Plant Succession and Indicators. Hafner Press, New York.

Culver, D., 1981. On using Horn's Markov succession model. Am. Nat. 117, 572–574.

Dale, M., Dale, P., Li, C., Biswas, G., 2002. Assessing impacts of small perturbations using a model-based approach. Ecol. Model. 156, 185–199.

Dambacher, J., Li, H., Rossignol, P., 2003. Qualitative predictions in model ecosystems. Ecol. Model. 161, 79–93.

Eddy, S., 1996. Hidden Markov models. Curr. Opin. Struct. Biol. 6, 361–365.

Elliot, R., van der Hoek, J., 1997. An application of hidden Markov models to asset allocation problems. Finance Stochastics 1, 229–238.

Falkhausen, M., Reininger, H., Wolf, D., 1995. Calculation of distance measures between hidden Markov models. Proceedings of the Eurospeech, Madrid.

Favier, C., Chave, J., Fabing, A., Schwartz, D., Dubois, M., 2004. Modelling forest-savanna mosaic dynamics in man-influenced environments: effects of fire, climate and soil heterogeneity. Ecol. Model. 171, 85–102.

Fine, S., Singer, Y., Tishby, N., 1998. The hierarchical hidden Markov model: analysis and applications. Machine Learn. 32, 41–62.

Forney, G., 1973. The Viterbi algorithm. Proc. IEEE 61, 268–278.

Fox, G., Gurevitch, J., 2000. Population numbers count: tools for near-term demographic analysis. Am. Nat. 156, 242–256.

Ghahramani, Z., 1997. Factorial hidden Markov models. Machine Learn. 29, 245–275.

Ghahramani, Z., 1998. Learning dynamic Bayesian networks. In: Adaptive Processing of Sequences and Data Structures. Springer-Verlag, Berlin, pp. 168–197.

Ghahramani, Z., 2001. An introduction to hidden Markov models and Bayesian networks. Int. J. Pattern Recognit. Artif. Intell. 15 (1), 9–42.

He, H., Mladenoff, D., Crow, T., 1999. Linking an ecosystem model and a landscape model to study forest species response to climate warming. Ecol. Model. 114, 213–233.

Hoef, J.V., Cressie, N., 1997. Using hidden Markov chains and empirical Bayes change-point estimation for transect data. Environ. Ecol. Stat. 4, 247–264.

Horn, H., 1975. Markovian properties of forest succession. In: Ecology and Evolution of Communities. Harvard University Press, Massachusetts, USA, pp. 196–211.

Hulst, R., 1979. On the dynamics of vegetation: Markov chains as models of succession. Vegetatio 40, 3–14.

Juang, B.-H., Rabiner, L., 1985. A probabilistic distance measure for hidden Markov models. Tech. Rep. 64 (2). AT&T Bell Laboratories.

Karplus, K., Sjolander, K., Barrett, C., Cline, M., Haussler, D., Hughey, R., Holm, L., Sander, C., 1997. Predicting protein structure using hidden Markov models. Proteins Struct. Funct. Genet. (Suppl. 1) 134–139.

Kneale, P., See, L., Smith, A., September 2001. Towards defining evaluation measures for neural network forecasting models. In: Pullar, D. (Ed.), Proceedings of the Sixth International Conference on GeoComputation. GeoComputation CD-ROM.

Korotkov, V., Logofet, D., Loreau, M., 2001. Succession in mixed boreal forest of Russia: Markov models and non-Markov effects. Ecol. Model. 142, 25–38.

Krogh, A., 1998. An introduction to hidden Markov models for biological sequences. In: Computational Methods in Molecular Biology. Elsevier Science, pp. 45–63.

Krogh, A., Brown, M., Mian, I., Haussler, D., 1994. Hidden Markov models in computational biology: applications to protein modeling. J. Mol. Biol. 235, 1501–1531.

Legg, C., 1980. A Markovian approach to the study of heath vegetation dynamics. Bull. Ecol. 11, 393–404.

Li, B.-L., 1995. Stability analysis of a nonhomogenous Markovian landscape model. Ecol. Model. 82, 247–256.

Lippe, E., DeSmidt, J., Glenn-Lewin, D., 1985. Markov models and succession: a test from a heathland in the Netherlands. J. Ecol. 73, 775–791.

Logofet, D., Lesnaya, E., 2000. The mathematics of Markov models: what Markov chains can really predict in forest successions. Ecol. Model. 126, 258–298.

MacFarlane, D., Green, E., Valentine, H., 2000. Incorporating uncertainty into parameters of a forest process model. Ecol. Model. 134, 27–40.

McAuliffe, J., 1988. Markovian dynamics of simple and complex desert plant communities. Am. Nat. 131, 459–490.

Orlóci, L., Anand, M., He, X., 1993. Markov chain: a realistic model for temporal coenosere? Biometrie-Praximetrie 33, 7–26.

Rabiner, L., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE 77 (2), 257–285.

Rissanen, J., 1978. Modeling by shortest data description. Automatica 14, 465–471.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6, 461–464.

Thomas, L., Allen, D., Morkel-Kingsbury, N., 2002. A hidden Markov chain model for the term structure of bond credit risk spreads. Int. Rev. Finan. Anal. 11, 311–329.

Thyer, M., Kuczera, G., 2003. A hidden Markov model for modelling long-term persistence in multi-site rainfall time series. Part 1. Model calibration using a Bayesian approach. J. Hydrol. 275, 12–26.

Thyer, M., Kuczera, G., 2003. A hidden Markov model for modelling long-term persistence in multi-site rainfall time series. Part 2. Real data analysis. J. Hydrol. 275, 27–48.

Tucker, B., Anand, M., 2003. The use of matrix models to detect natural and pollution-induced forest gradients. Community Ecol. 4 (1), 89–110.

Usher, M., 1981. Modelling ecological succession, with particular reference to Markovian models. Vegetatio 46, 11–18.

Valverde, T., Silvertown, J., 1997. Canopy closure rate and forest structure. Ecology 78, 1555–1562.

Viovy, N., Saint, G., 1994. Hidden Markov models applied to vegetation dynamics using satellite remote sensing. IEEE Transact. Geosci. Remote Sens. 32, 906–917.

Visser, I., Raijmakers, M., Molenaar, P., 2002. Fitting hidden Markov models to phycological data. Sci. Programming 10, 185–199.

Viterbi, A., 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Transact. Inf. Theory IT-13, 260–269.

Waggoner, P., Stephens, G., 1970. Transition probabilities for a forest. Nature 225, 1160–1161.

Wallace, C., Freeman, P., 1978. Estimation and inference by compact coding. J. Royal Stat. Soc. B 49, 241–252.

Wong, J., McDonald, K., Palazoglu, A., 2001. Classification of abnormal plant operation using multiple process variable trends. J. Process Control 11, 409–418.

Wootton, J., 2001. Prediction in complex communities: analysis of empirically derived markov models. Ecology 82, 580–598.

Yearsley, J., 2004. Transient population dynamics and short-term sensitivity analysis of matrix population models. Ecol. Model. 177, 245–258.

Yemshanov, D., Perera, A., 2002. A spatially explicit stochastic model to simulate boreal forest cover transitions: general structure and properties. Ecol. Model. 150, 189–209.

Yoon, J., Korvin, A., 2001. Evaluating a model of forest succession using fuzzy analysis. Stochastic Anal. Appl. 19, 893–901.