Beyond Time-Homogeneity for Continuous-Time Multistate Markov Models

Emmett B. Kendall, Jonathan P. Williams, Gudmund H. Hermansen, Frederic Bois & Vo Hong Thanh

Taylor & Francis
Taylor & Francis Group

Check for updates

Beyond Time-Homogeneity for Continuous-Time Multistate Markov Models

Emmett B. Kendall[a] ⓘ, Jonathan P. Williams[a,b] ⓘ, Gudmund H. Hermansen[b,c,d], Frederic Bois[e] ⓘ, and Vo Hong Thanh[e] ⓘ

[a]Department of Statistics, North Carolina State University, Raleigh, NC; [b]Centre for Advanced Study, Norwegian Academy of Science and Letters, Oslo, Norway; [c]University of Oslo, Oslo, Norway; [d]Peace Research Institute Oslo (PRIO), Oslo, Norway; [e]CERTARA UK Limited, Simcyp Division, Level 2-Acero, Sheffield, UK

ABSTRACT

Multistate Markov models are a canonical parametric approach for data modeling of observed or latent stochastic processes supported on a finite state space. Continuous-time Markov processes describe data that are observed irregularly over time, as is often the case in longitudinal medical data, for example. Assuming that a continuous-time Markov process is time-homogeneous, a closed-form likelihood function can be derived from the Kolmogorov forward equations—a system of differential equations with a well-known matrix-exponential solution. Unfortunately, however, the forward equations do not admit an analytical solution for continuous-time, time-*inhomogeneous* Markov processes, and so researchers and practitioners often make the simplifying assumption that the process is piecewise time-homogeneous. In this article, we provide intuitions and illustrations of the potential biases for parameter estimation that may ensue in the more realistic scenario that the piecewise-homogeneous assumption is violated, and we advocate for a solution for likelihood computation in a truly time-inhomogeneous fashion. Particular focus is afforded to the context of multistate Markov models that allow for state label misclassifications, which applies more broadly to hidden Markov models (HMMs), and Bayesian computations bypass the necessity for computationally demanding numerical gradient approximations for obtaining maximum likelihood estimates (MLEs). Supplemental materials are available online.

1. Introduction

Given a physical process that evolves over time, it is not always reasonable to expect that the *rate of evolution* remains constant over time. An obvious example of such a process is the progression of dementia as it develops with age. The rate at which young individuals (e.g., less than 50 years of age) develop dementia is virtually null compared to the rate observed for individuals beyond 70 years of age (i.e., all else constant, the probability of developing dementia within one year for an individual of age 40 is most certainly different than that probability for an individual of age 80; Williams et al. 2020). Fitting a continuous-time multistate Markov model or HMM in a time-inhomogeneous fashion, however, (i.e., where the transition rate matrix is a continuous function of time) has been a crux of numerous methodological approaches in the statistics community for many decades (e.g., Faddy 1976; Kay 1986; Jackson et al. 2003; Titman 2011; Jack et al. 2016; Williams et al. 2020; Siméon et al. 2021).

It is well-established that a continuous-time Markov process is characterized as the solution to the system of differential equations commonly referred to as the Kolmogorov forward equations (see, e.g., Karlin and Taylor 1981). While in the time-homogeneous case a closed-form matrix-exponential solution to the Kolmogorov forward equations exists, no known analytical solution is available in the time-inhomogeneous case.

Accordingly, an analytical expression for the likelihood function of a time-inhomogeneous multistate Markov model is unavailable, and so a variety of approximations have been proposed. For example, as in Williams et al. (2020), if it is assumed that the continuous-time Markov process transition rates are constant between integer ages, then the closed-form matrix-exponential solution to the forward equations exists for every integer age. Thus, in the piecewise time-homogeneous case, the full likelihood can be expressed with the transition probability matrices computed as products of multiple matrix-exponential solutions indexed by integer age. This solution facilitates the identifiability of time-inhomogeneous transition rates insofar as they are strictly piecewise-constant functions of time, which is surely an unrealistic, albeit pragmatic assumption about the biological process. Although pragmatic, approximating transition rates that are continuous functions of time, by rates that are piecewise-constant functions of time, of finer and finer resolution (e.g., yearly, monthly, weekly, etc.), quickly becomes computationally infeasible due to the rapidly growing number of matrix-exponential calculations required at finer resolutions.

Our contributions are: first, to provide intuitions and illustrations of the potential estimation biases that might result from modeling transition rates as piecewise-constant functions of time, when they are truly continuous functions of time;

CONTACT Emmett B. Kendall ✉ ebkendal@ncsu.edu 🖥 Department of Statistics, North Carolina State University, 5109 SAS Hall 2311 Katharine Stinson Dr. Raleigh, NC 27695-8203.

Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JCGS*.

second, to advocate for numerical ordinary differential equation (ODE) solutions to the time-inhomogeneous Kolmogorov forward equations that correctly specifies the transition rates as continuous functions of time; and third, to demonstrate, with both real and synthetic datasets, instances of when differences may and may not arise in estimated parameter values between the continuous versus piecewise-constant rates approaches—indicating the practical significance of numerical ODE solutions for fitting continuous-time, time-inhomogeneous Markov processes (HMMs or otherwise). Moreover, due to the computational complexity of matrix-exponential calculations, we find that a numerical ODE solution—which we implement via the R package deSolve (Soetaert, Petzoldt, and Setzer 2010)—leads to reduced computational time, even for time-homogeneous processes. The numerical ODE solver approach allows for computing the likelihood, but it does not yield a closed-form expression of the likelihood function. As such, gradient-based algorithms for finding MLEs of model parameters may be computationally prohibitive because the gradient might need to be approximated numerically. In any case, we propose Markov chain Monte Carlo (MCMC) computations within a fully Bayesian paradigm, avoiding the need to compute first derivatives of the likelihood function.

Markov models and HMMs play a fundamental role in many recent applications from modeling disease progression (Williams et al. 2020) and seizure detection (Furui, Akiyama, and Tsuji 2021), to assisting in volcanic eruption risk evaluation (Aspinall et al. 2006), as well as studying the efficacy of ceasefires on conflict dynamics (Williams et al. 2024). Applying numerical ODE solvers to fit continuous-time Markov models (HMMs or otherwise) in a truly time-inhomogeneous fashion has broad implications, particularly for studying longitudinal medical and biological datasets where the underlying phenomena exhibit continuously time-evolving rates of progression.

An earlier approach contributed by Titman (2011) proposed numerical methods for solving the Kolmogorov forward equations in the context of continuous-time, time-inhomogeneous multistate Markov processes, but does not consider HMMs. In Titman (2011), the transition rates are assumed to be smooth functions of time, and with the canonical strategy of using B-splines to model the transition rates as functions of time, Fisher scoring is used to compute MLEs of the model parameters. Derivatives with respect to the transition rate parameters are calculated numerically by extending the system of forward equations. While there exists a thorough investigation into the computational efficiency of applying numerical approaches, little is said in Titman (2011) about the inferential consequences of using these numerical methods as opposed to more common, approximation-based estimation procedures in the multistate Markov modeling literature. Other ideas to circumvent the piecewise-homogeneous approximation include time transformation techniques applied to the transition rates based on a common, smooth function of time that makes calculating MLEs more feasible (Hubbard, Inoue, and Fann 2008), and penalized splines of individual rates in a forward-only model (Joly et al. 2002). Further, there exist alternatives to the HMM framework for specification and estimation of time-dependent rates of progression, such as stochastic simulation algorithms (Thanh and Priami 2015; Marchetti, Priami, and Thanh 2017).

More recently, the methods of Titman (2011) have been extended to accommodate the HMM framework in the R package nhm (Titman 2019), and for some model specifications first derivatives of the likelihood function can be computed analytically. Bayesian implementations are not yet available in the nhm package. In addition to investigating a Bayesian approach, our article provides further clarity, intuition, and empirical studies to illustrate the consequences of a piecewise-homogeneous approximation when transition rates may be continuous functions of time. Hence, our article goes beyond the scope of what is presented in Titman (2011) by facilitating broader awareness for implementing these numerical ODE solvers within the context of continuous-time, time-inhomogeneous Markov models, as well as applying them to more complex data applications that go beyond the purview of what is capable of nhm. Additionally, Siméon et al. (2021) points to the lack of these numerical-integration techniques being applied in the statistics literature for correctly fitting multistate Markov models (HMMs or otherwise). Furthermore, Siméon et al. (2021) argues that piecewise-homogeneous approximations are slow and inaccurate relative to numerical methods.

Another alternative to solving the time dependent Kolmogorov forward equations is to empirically estimate the transition probability matrix via the product-integral solution to the Kolmogorov forward equations (see Borgan 1997; Aalen, Borgan, and Gjessing 2008; Andersen et al. 2012 for more detail). An extension of the Kaplan-Meier estimator (Kaplan and Meier 1958), the Aalen-Johansen estimator (Aalen and Johansen 1978) is an empirical product limit estimator for the transition probability matrix of a finite state, inhomogeneous Markov process. Although the Aalen-Johansen estimator offers a nonparametric approach to estimating the transition probability matrix, the empirical calculation of the estimator assumes that all state transitions are observed and observation times correspond to state transition times (Borgan 1997; Aalen, Borgan, and Gjessing 2008). Hence, similar to the piecewise-homogeneous approach, the Aalen-Johansen estimator relies on making assumptions about the evolution of the state process which, in most applications, are likely not met and can thus result in estimation bias (as shown in Sections 3.3 and 3.4.1). That being stated, an appeal of this estimator is that it can serve as a nonparametric MLE for the true transition probability matrix, and under certain conditions, is a uniformly consistent estimator (Andersen et al. 2012). The Aalen-Johansen estimator is a function of the Nelson-Aalen estimator (Nelson 1969, 1972), a nonparametric estimator of the cumulative transition intensity of the time-inhomogeneous Markov process. The Aalen-Johansen and Nelson-Aalen estimators are frequently used for assessing model goodness-of-fit when implementing other approaches to fitting multistate Markov models (Bureau, Shiboski, and Hughes 2003; Titman and Sharples 2008, 2010; Soper et al. 2020). Extending beyond Markov processes in particular, much work has been done to explore asymptotic properties of the Aalen-Johansen estimator in non-Markov settings (e.g., Datta and Satten 2001, 2002; Glidden 2002; Gunnes, Borgan, and Aalen 2007).

The last alternative to solving the time dependent Kolmogorov forward equations we consider is a discretization of time, altogether. A detailed comparison of discrete-time versus

continuous-time, time-homogeneous models can be found in Wan et al. (2016). They find that if inter-observation times are similar, a discrete-time model is as effective as a continuous-time model, as well as computationally more efficient. See also, Aalen et al. (1997) and Liu et al. (2015) for discretized approaches for fitting continuous-time Markov models. We present evidence, in Section 3.4.2, suggesting the reasonableness of discrete-time approximations for time-inhomogeneous HMMs—when regularly-spaced observation times are available at a high resolution—based on our analysis of real electrocorticography (ECoG) data recorded from a mice sleep study (the data are from Bojarskaite et al. 2020). Nonetheless, the importance of our investigation of continuous-time processes is built on the fact that many datasets do *not* exhibit observation times near enough a regular spacing that a discrete-time approximation would be effective.

The remainder of the article is structured as follows. In Section 2, the multistate Markov model and HMM frameworks are introduced, and mathematical intuition is provided for the potential biases of a piecewise-homogeneous approximation for continuously evolving time-inhomogeneous transition rates in the context of a simple linear regression illustration. Section 3 covers all details for the data analyses beginning with data descriptions (Section 3.1) and modeling approaches (Section 3.2). Then, Section 3.3 presents a simulation study of synthetic data generated from a continuous-time, time-inhomogeneous Markov process, to compare fitting an HMM with piecewise-homogeneous approximations, empirical approximations via the Aalen-Johansen and Nelson-Aalen estimators, as well as the time-inhomogeneous numerical ODE solution. We consider comparisons of these approaches on real datasets in Section 3.4. The article ends with concluding remarks in Section 4. The R code along with a workflow script file for reproducing all numerical results presented in this article is available at *https://ebkendall.github.io/research.html*.

2. Multistate Markov Model Methods and Inference

We first formalize the common notations and specifications for multistate models. Let $s_i = \{s_{i,1}, \ldots, s_{i,n_i}\}$, for $i \in \{1, \ldots, n\}$ independent, represent an *observed* state sequence from a discrete, multistate process, $X(t)$, where for any time $t \geq 0$, $X(t) \in \{1, \ldots, m\}$, for some fixed, positive integer m. The inferential goal with regard to these n independent, multistate processes is to learn parameters characterizing the transitions between the m states over time. If $t_{i,1} < t_{i,2} < \cdots < t_{i,n_i}$, for $i \in \{1, \ldots, n\}$, represents the observed time points for each s_i, then the joint probability mass function is given by

$$f_{s_1, \ldots, s_n}(s_1, \ldots, s_n) = \prod_{i=1}^{n} P(X(t_{i,1}) = s_{i,1})$$
$$\times \prod_{k=2}^{n_i} P\left(X(t_{i,k}) = s_{i,k} \mid \{X(t_{i,j}) = s_{i,j}\}_{j=1}^{k-1}\right).$$

For the remainder of this article, however, we make the stochastic assumption that the evolution of $X(t)$ is defined by a Markov process. In particular, the Markov assumption simplifies the expression for $f_{s_1, \ldots, s_n}(s_1, \ldots, s_n)$ as follows. For any $t, h \geq 0$,

let $P(t, t + h)$ be the $m \times m$ transition probability matrix over the time period from t to $t + h$, for $X(t)$, where the rth row and s^{th} column of $P(t, t + h)$, denoted $P_{r,s}(t, t + h)$, represents the probability of transitioning from state r to state s over the time period t to $t + h$. The Markov assumption thus allows the joint probability mass function to be expressed as

$$f_{s_1, \ldots, s_n}(s_1, \ldots, s_n) = \prod_{i=1}^{n} P(X(t_{i,1}) = s_{i,1}) \cdot P_{s_{i,1}, s_{i,2}}(t_{i,1}, t_{i,2}) \cdots$$
$$\times P_{s_{i,n_i-1}, s_{i,n_i}}(t_{i,n_i-1}, t_{i,n_i}), \tag{1}$$

where

$$P_{s_{i,k-1}, s_{i,k}}(t_{i,k-1}, t_{i,k}) = P\{X(t_{i,k}) = s_{i,k} \mid X(t_{i,k-1}) = s_{i,k-1}\},$$

for $k = 2, 3, \ldots, n_i$. In what follows, we detail the various approaches for defining $P(t, t+h)$ depending on further stochastic assumptions made about the evolution of $X(t)$.

The mathematical expression for $P(t, t + h)$ is determined by whether time is measured as discrete or continuous. While the discrete-time approach leads to a simple expression (i.e., multiplying the one-time-point-ahead transition probability matrix h times, where t and h take values from a discrete grid of points such as the nonnegative integers), it is not adequate for modeling data that are observed irregularly in time. Thus, it is the necessary complexities of the continuous-time approach that we focus on in this article.

In the continuous-time, time-homogeneous setting, t and h take nonnegative, real values, and $P(t, t + h)$ is equivalent to $P(0, h)$. Further, so long as the transition probabilities $P_{r,s}(0, h)$, for $r, s \in \{1, \ldots, m\}$, are differentiable functions of h, $P(0, h)$ is defined by an $m \times m$ *infinitesimal generator* or *transition rate* matrix, labeled Q (i.e., the infinitesimal analogue of the one-time-point-ahead transition probability matrix arising in the discrete-time setting), via the system of differential equations

$$P'(0, h) = P(0, h) \cdot Q, \tag{2}$$

commonly referred to as the *Kolmogorov forward equations*. The $(r, s)^{\text{th}}$ component of Q, denoted $q_{r,s}$, can be interpreted as the instantaneous rate of transition from state r to state $s \neq r$:

$$q_{r,s} := \lim_{\delta \downarrow 0} \frac{P\{X(\delta) = s \mid X(0) = r)\}}{\delta}, \tag{3}$$

and $q_{r,r} := -\sum_{s:s \neq r} q_{r,s}$ for $r, s \in \{1, \ldots, m\}$ (see, e.g., Karlin and Taylor 1981, for comprehensive mathematical details). By standard matrix calculus arguments, it is readily verifiable that (2) has the analytical matrix-exponential solution

$$P(0, h) = e^{h \cdot Q}.$$

This closed-form solution makes it uncomplicated to express (1) as

$$f_{s_1, \ldots, s_n}(s_1, \ldots, s_n \mid Q) = \prod_{i=1}^{n} P(s_{i,1}) \cdot P_{s_{i,1}, s_{i,2}}(0, t_{i,2} - t_{i,1}) \cdots$$
$$\times P_{s_{i,n_i-1}, s_{i,n_i}}(0, t_{i,n_i} - t_{i,n_i-1})$$
$$= \prod_{i=1}^{n} P(s_{i,1}) \cdot \{e^{(t_{i,2} - t_{i,1}) \cdot Q}\}_{s_{i,1}, s_{i,2}} \cdots$$
$$\times \{e^{(t_{i,n_i} - t_{i,n_i-1}) \cdot Q}\}_{s_{i,n_i-1}, s_{i,n_i}},$$

where $P(s_{i,1})$ is the initial state probability, and $\{\cdot\}_{r,s}$ denotes the $(r, s)^{\text{th}}$ component of the matrix argument. Moreover, this likelihood expression admits a closed-form gradient expression with respect to the components of $\boldsymbol{Q}$, leading to efficient optimization algorithms for computing the MLEs, and the tractability of both the likelihood and its gradient are easily extended to the HMM framework for dealing with latent state sequences.

Statistical inferences are typically motivated by characterizing the components of $\boldsymbol{Q}$ as a function of covariates; for example, it may be of interest to determine if transition rates over a discretized progression of a disease (i.e., over m states of severity) are affected by baseline covariates such as sex, race, biomarkers, socioeconomic status, etc. It is not possible, however, to consider non-baseline covariates—such as age—that may be important for describing key features of the data (e.g., when studying dementia progression or other aging-related processes). This is due to the fact that the matrix-exponential solution to the Kolmogorov forward equations (2) relies on the assumption that $\boldsymbol{Q}$ is constant over time; this assumption is also why it suffices to take $t = 0$ in (2). Nonetheless, it remains paramount in many scientific and machine learning applications to develop extensions for estimating the effects of time-varying covariates on the transition rates in a continuous-time, time-inhomogeneous multistate Markov process defined by

$$\boldsymbol{P}'(t, t+h) = \boldsymbol{P}(t, t+h) \cdot \boldsymbol{Q}(t+h), \tag{4}$$

where $\boldsymbol{Q}(t+h)$ is the transition rate matrix defined as a function of time, with $h > 0$. The components of $\boldsymbol{Q}(t+h)$ are reminiscent

of (3), except are now functions of time, given by

$$q_{r,s}(t + h) := \lim_{\delta \downarrow 0} \frac{P\{X(t + h + \delta) = s \mid X(t + h) = r)\}}{\delta},$$

and $q_{r,r}(t + h) := -\sum_{s:s \neq r} q_{r,s}(t + h)$ for $r, s \in \{1, \ldots, m\}$; see also Karlin and Taylor (1981); Titman and Sharples (2010) and Titman (2011).

Unfortunately, there exists no known analytical solution to the time-inhomogeneous forward equations (4)—at least to the best of our knowledge—and rather than relying on numerical ODE solutions for likelihood and gradient computations, a common strategy observed throughout the statistical literature is to approximate a time-inhomogeneous Markov process by one that is piecewise time-homogeneous. In that case, the closed-form matrix-exponential solution applies in a piecewise-constant fashion, so that the Kolmogorov forward equations can be expressed as

$$\boldsymbol{P}'(t, t+h) = \boldsymbol{P}(t, t+h) \cdot \boldsymbol{Q}(g_d(t)), \tag{5}$$

where $g_d(t) := t - (t \mod d)$ for some fixed $d > 0$. In the case $d = 1$ it follows that $g_1(t) = \lfloor t \rfloor$, that is, the *floor* function taking the greatest integer less than or equal to t as its value. So, for example, if $d = 1$ and t is time as measured by the age of an individual, then $g_1(t) = \lfloor t \rfloor$ is the integer-valued age, and so defining $\boldsymbol{Q}$ as a function of integer age means that the transition rates only update with every birthday of the individual. Accordingly, in this example, for any $t \geq 0$ and $h \geq 0$ (5) has the closed-form solution

$$\boldsymbol{P}(t, t+h) = \begin{cases} e^{(\lfloor t+1 \rfloor - t) \cdot \boldsymbol{Q}(\lfloor t \rfloor)} \cdot e^{1 \cdot \boldsymbol{Q}(\lfloor t+1 \rfloor)} \cdot e^{1 \cdot \boldsymbol{Q}(\lfloor t+2 \rfloor)} \cdots e^{(t+h-\lfloor t+h \rfloor) \cdot \boldsymbol{Q}(\lfloor t+h \rfloor)} & \text{if } \lfloor t \rfloor \neq \lfloor t+h \rfloor \\ e^{h \cdot \boldsymbol{Q}(\lfloor t+h \rfloor)} & \text{else.} \end{cases}$$

In general, for arbitrary $d > 0$, the solution to (5) is

$$\boldsymbol{P}(t, t+h) = \begin{cases} e^{(g_d(t+d)-t) \cdot \boldsymbol{Q}(g_d(t))} \cdot e^{d \cdot \boldsymbol{Q}(g_d(t+d))} \cdots e^{(t+h-g_d(t+h)) \cdot \boldsymbol{Q}(g_d(t+h))} & \text{if } g_d(t) \neq g_d(t+h) \\ e^{h \cdot \boldsymbol{Q}(g_d(t+h))} & \text{else.} \end{cases}$$

It is seemingly a reasonable and pragmatic strategy to approximate the solution to the time-inhomogeneous forward equations (4) by choosing sufficiently small $d > 0$, but bias on the transition rate estimates may persist, nonetheless—if the transition rates of the underlying process are truly evolving in continuous-time—and the matrix-exponential computations involved in the piecewise solution can have a tremendous impact on computation time for small $d > 0$. Moreover, numerical ODE solutions exist for directly solving the system (4) and are readily available in standard software such as the R packages deSolve (Soetaert, Petzoldt, and Setzer 2010) and nhm (Titman 2019).

To provide insight and intuition on the biases on the transition rate parameters that may ensue from approximating a continuous-time process by a piecewise-constant function, we provide a simple illustration for how such biases arise in a linear model context next, in Section 2.1.

2.1. Simple Illustration of Piecewise-Constant Approximation Bias

To understand how estimation biases on the transition rate parameters $\boldsymbol{Q} = \boldsymbol{Q}(t)$ may result from solving the system (5) with $\boldsymbol{Q} = \boldsymbol{Q}(g_d(t))$, it is insightful to understand how biases will result from fitting the parameters of a simple linear model depending on continuous-time t by a linear model depending on piecewise-constant-time $g_d(t) = t - (t \mod d)$ for some fixed $d > 0$. Accordingly, suppose that $y_1, \ldots, y_n$ are real-valued observations on the line

$$y_i = \beta_0 + \beta_1 t_i,$$

where t_i denotes an instance of a value in continuous-time, for $i \in \{1, \ldots, n\}$, and β_0 and β_1 are fixed scalars. Then regressing $y_1, \ldots, y_n$ on the piecewise-constant-time instances $g_d(t_1), \ldots g_d(t_n)$ yields the least squares projection $b_{0,n} + b_{1,n} g_d(t_i)$, for $i \in \{1, \ldots, n\}$, where
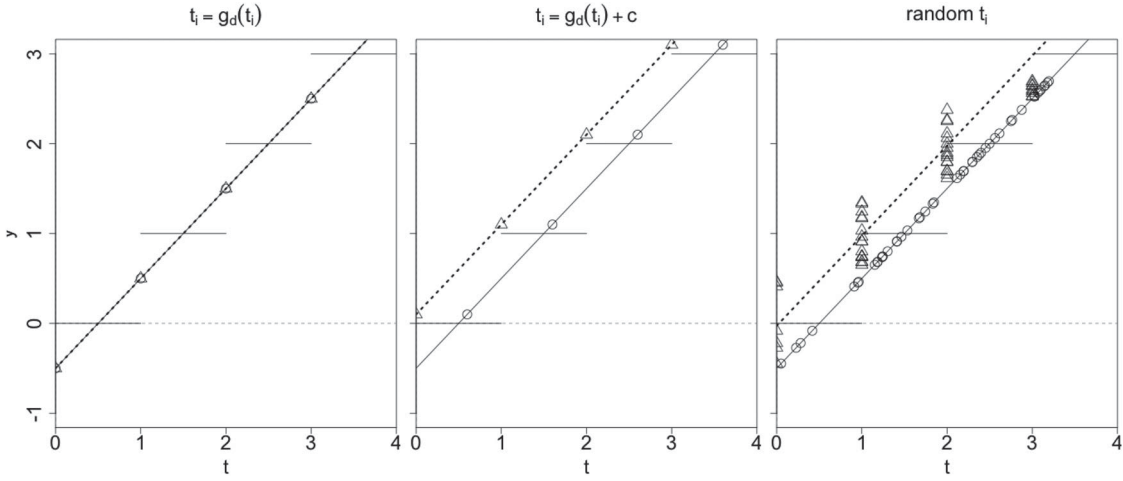
Figure 1. Illustration of Lemmas 1 and 2 with $d = 1$, that is, $g_d(t) = \lfloor t \rfloor$, for synthetic values $y_i = -\frac{1}{2} + t_i$, for $i \in \{1, \ldots, n\}$, for time instances $t_1, \ldots, t_n$. The left panel corresponds to integer-valued time instances $t_i = \lfloor t_i \rfloor$ (i.e., $c = 0$, as in Lemma 2), the middle panel corresponds to shifted time instances $t_i = \lfloor t_i \rfloor + c$ (with $c = 0.6$, as in Lemma 2), and the right panel corresponds to randomly generated time instances t_i. The solid black lines are plots of the line $-\frac{1}{2} + t$, and the dashed black lines are plots of the fitted least squares projection lines $b_{0,n} + b_{1,n}t$. The circles denote scatterplots of the observed points $\{(t_i, y_i)\}_{i=1}^n$, whereas the triangles denote scatterplots of the points $\{(g_d(t_i), y_i)\}_{i=1}^n = \{(\lfloor t_i \rfloor, y_i)\}_{i=1}^n$ – used to fit the dashed projection lines.

$$b_{0,n} = \beta_0 + \frac{\bar{t}_n\{\sum_{i=1}^n g_d^2(t_i)\} - \frac{1}{n}\sum_{i,j} t_i \cdot g_d(t_i)g_d(t_j)}{A_n}\beta_1$$

$$b_{1,n} = \beta_1 + \left(\frac{\sum_{i=1}^n (t_i - \bar{t}_n)\{g_d(t_i) - \frac{1}{n}\sum_{k=1}^n g_d(t_k)\}}{A_n} - 1\right)\beta_1,$$

(6)

$\bar{t}_n := \frac{1}{n}\sum_{i=1}^n t_i$, and assuming $A_n := \sum_{i=1}^n \{g_d(t_i) - \frac{1}{n}\sum_{k=1}^n g_d(t_k)\}^2 > 0$.

For fixed $d > 0$, so long as sufficient variation is observed in the $g_d(t_1), \ldots g_d(t_n)$ (i.e., there is an unbounded and heavy-tailed distribution of available time instances $t_1, \ldots, t_n$, for large n), the bias $\beta_1 - b_{1,n}$ for the slope parameter will vanish for large n, as stated in Lemma 1. The bias $\beta_0 - b_{0,n}$ for the baseline parameter, however, need not vanish for large n, as illustrated by Lemma 2. Figure 1 presents a graphical representation of these two lemmas.

Lemma 1. For fixed $d > 0$ and for every positive integer n, the least squares projection coefficients $b_{0,n}$ and $b_{1,n}$ can be expressed as in (6), and if $A_n > 0$, then

$$|\beta_1 - b_{1,n}| \leq \frac{d \cdot |\beta_1|}{(\frac{1}{n}A_n)^{\frac{1}{2}}}.$$

In particular, if $\lim_{n\to\infty} \frac{1}{n}A_n = \infty$, then $\lim_{n\to\infty} b_{1,n} = \beta_1$.

Proof. See Section A.1. □

Lemma 2. For fixed $d > 0$, suppose that $t_i = g_d(t_i) + c$, for some $c \geq 0$, for all $i \in \{1, \ldots, n\}$. Then, for every positive integer n with $A_n > 0$,

$$|\beta_0 - b_{0,n}| = c \cdot |\beta_1|.$$

In particular, in the case $\beta_1 \neq 0$, it follows that $b_{0,n} = \beta_0$ if and only if the observed time instances $t_1, \ldots, t_n$ happen to coincide precisely at discontinuity points of the function $g_d(\cdot)$.

Proof. See Section A.1. □

As illustrated by the panels in Figure 1, projecting points from a line that is continuous in time onto a piecewise-constant function of time can result in substantial bias in fitting the baseline parameter. Although the transition probability matrix for a continuous-time, time-inhomogeneous Markov process is more complicated than a simple linear function of time, the simple illustrations presented in this section provide the intuition for how fitting the rate parameters based on a piecewise-homogeneous rate matrix approximation $\boldsymbol{Q}(g_d(t))$ will likely result in biased estimation of the baseline rates of $\boldsymbol{Q}(t)$.

In Section 3.3, we provide empirical evidence to demonstrate such baseline rate bias using synthetic data generated from continuous-time, time-inhomogeneous Markov processes. Furthermore, we demonstrate, in Section 3.4.1, indications that the baseline rate bias can manifest from real data by comparing rate parameter estimates from fitting a continuously evolving time-inhomogeneous Markov model—based on a numerical ODE solution to (4)—versus a piecewise-homogeneous Markov model—based on the matrix-exponential-based solution to (5). We also present a second real data analysis in Section 3.4.2 suggesting that the piecewise-homogeneous approximation may be reasonable *if* the observed time instances fall on a grid, as described by $c = 0$ in Lemma 2.

2.2. Hidden Markov Models

In order to evince the broadness of scope in which our analyses and conclusions apply, we further extend the continuous-time Markov model framework with the alternative assumption that the state sequences $s_i = \{s_{i,1}, \ldots, s_{i,n_i}\}$, for $i \in \{1, \ldots, n\}$, are *unobserved*. The state sequences s_i are still assumed to be defined by a continuous-time, time-inhomogeneous Markov process; however, now let $y_i = \{y_{i,1}, \ldots, y_{i,n_i}\}$ represent *observed* response data, where the standard HMM formulation assumes that their random variable analogues $Y_{i,1}, \ldots, Y_{i,n_i}$ are condi-

tionally independent given the unobserved $s_{i,1}, \ldots, s_{i,n_i}$, and that $Y_{i,k}$ only depends on $s_{i,k}$, for every $k \in \{1, \ldots, n_i\}$, for every $i \in \{1, \ldots, n\}$. The same issues and potential rate biases—as discussed for (non-latent) Markov models—apply for approximating a time-inhomogeneous HMM by a piecewise-homogeneous HMM, but the likelihood function must be modified to account for uncertainty in the latent state sequences and the observed response data.

Given the standard HMM formulation, the density for $\boldsymbol{y_i}$ can be derived by integrating over all possible unobserved state sequences:

$$f_{y_i}(\boldsymbol{y_i}) = \sum_{s_{i,1}=1}^{m} \cdots \sum_{s_{i,n_i}=1}^{m} f_{y_i|s_i}(\boldsymbol{y_i} \mid \boldsymbol{s_i}) \cdot f_{s_i}(\boldsymbol{s_i})$$

$$= \sum_{s_{i,1}=1}^{m} \cdots \sum_{s_{i,n_i}=1}^{m} f_{s_i}(\boldsymbol{s_i}) \cdot \prod_{k=1}^{n_i} f_{y_{i,k}|s_{i,k}}(y_{i,k} \mid s_{i,k}),$$

where, as defined by the continuous-time Markov process that solves the time-inhomogeneous forward equations (4),

$$f_{s_i}(\boldsymbol{s_i}) = P(s_{i,1}) \cdot \prod_{k=2}^{n_i} P_{s_{i,k-1},s_{i,k}}(t_{i,k-1}, t_{i,k}).$$

Accordingly,

$$f_{y_i}(\boldsymbol{y_i}) = \sum_{s_{i,1}=1}^{m} P(s_{i,1}) \cdot f_{y_{i,1}|s_{i,1}}(y_{i,1} \mid s_{i,1})$$

$$\times \sum_{s_{i,2}=1}^{m} P_{s_{i,1},s_{i,2}}(t_{i,1}, t_{i,2}) \cdot f_{y_{i,2}|s_{i,2}}(y_{i,2} \mid s_{i,2}) \cdots$$

$$\times \sum_{s_{i,n_i}=1}^{m} P_{s_{i,n_i-1},s_{i,n_i}}(t_{i,n_i-1}, t_{i,n_i})$$

$$\cdot f_{y_{i,n_i}|s_{i,n_i}}(y_{i,n_i} \mid s_{i,n_i}),$$

which yields a joint density function over independent sequences $\boldsymbol{y_1}, \ldots, \boldsymbol{y_n}$:

$$f_{y_1,\ldots,y_n}(\boldsymbol{y_1}, \ldots, \boldsymbol{y_n}) = \prod_{i=1}^{n} \boldsymbol{\pi}^T \boldsymbol{D}_{(i,1)} \cdot \boldsymbol{P}(t_{i,1}, t_{i,2}) \boldsymbol{D}_{(i,2)} \cdots$$

$$\times \boldsymbol{P}(t_{i,n_i-1}, t_{i,n_i}) \boldsymbol{D}_{(i,n_i)} \cdot \boldsymbol{1}_m, \quad (7)$$

where $\boldsymbol{\pi} := \{P(s_{i,1} = 1), \ldots, P(s_{i,1} = m)\}^T$,

$$\boldsymbol{D}_{(i,k)} := \begin{pmatrix} f_{y_{i,k}|s_{i,k}}(y_{i,k} \mid s_{i,k} = 1) & & \\ & \ddots & \\ & & f_{y_{i,k}|s_{i,k}}(y_{i,k} \mid s_{i,k} = m) \end{pmatrix},$$

and $\boldsymbol{1}_m$ is an $m \times 1$ vector with each component taking the value 1. As per the discussions in Section 2, this likelihood function can only be computed numerically if $\boldsymbol{P}(t, t + h)$ is defined by the time-inhomogeneous forward equations (4), but it admits a closed-form approximation (for some fixed $d > 0$) based on the piecewise-homogeneous forward equations (5). These are the likelihood expressions that we compare using synthetic and real data in the sections that follow.

3. Empirical Results

3.1. Data Descriptions

Three datasets are studied in Sections 3.3 and 3.4: real cardiac allograft vasculopathy (CAV) data (Section 3.1.1), synthetic CAV data (Section 3.1.2), and real ECoG mice sleep study data (Section 3.1.3). Each data application serves a unique purpose in illustrating the practicality of using numerical ODE solvers for evaluating the transition probability matrix versus relying upon the piecewise time-homogeneous approximation or the Aalen-Johansen estimator, in the context of continuous-time, time-inhomogeneous Markov models. The focus is on comparing the resulting inferences from using the aforementioned modeling approaches, and specifically how the estimation of the transition rate parameters differ.

Although common in the statistical literature to use the piecewise time-homogeneous assumption, if, for the real data analyses, the resulting parameter estimates are different from that of the numerical ODE approach, then the decision to use one method over the other is not to be overlooked. This is the motivation for the real data analyses because it is unknown which approach "truly" describes the process being modeled; therefore, if the inferential conclusions are different, practitioners must consciously decide if the piecewise-homogeneity assumption accurately represents the process, otherwise, the numerical ODE solver eliminates the need for that assumption. The same applies for implementing the Aalen-Johansen estimator and the assumption of observing all state transition times. More broadly, the choice to use the CAV data is both because it is well-studied in the multistate Markov model literature, as well as it offers data with unequal inter-observation times within subjects; alternatively, the ECoG mice sleep study provides an example where the inter-observation times within and between subjects is the same. We generate synthetic CAV data in the context of the simulation study to illustrate how fitting the rate parameters of a continuous-time, time-inhomogeneous Markov process, based on a piecewise-homogeneous approximation—with rate matrix $\boldsymbol{Q}(g_d(t))$—can lead to biased estimation of the baseline rates of $\boldsymbol{Q}(t)$.

3.1.1. Cardiac Allograft Vasculopathy Study Data

The first of the real data considered is the CAV data, available in the msm package for R (Jackson 2011), which comprises information and clinical assessments collected from 622 subjects who underwent heart transplant surgery. Data for each subject are available on roughly an annual basis for the years following surgery, and classification of the severity of deterioration of the subjects' arterial walls are recorded based on the clinical assessments (i.e., severity of CAV Jackson 2011). Associated with each clinical visit, for each subject, is the severity of CAV, classified as either 1: "no CAV"; 2: "mild/moderate CAV"; 3: "severe CAV"; or 4: "death." In the context of a multistate Markov process, these classifications define four states. Time must be treated as continuous—due to the irregular observation times of the clinical visits—and the observed state labels for CAV severity, associated with each clinical visit for each subject, must be assumed observed-with-error to account for diagnostic/clerical uncertainty.

3.1.2. Simulated Data

The synthetic data is generated from a time-inhomogeneous Markov process based on HMM parameters fit to the well-studied CAV dataset described in Section 3.1.1. In particular, the evolution of the Markov process rate matrix for this simulation is constructed to be inhomogeneous in continuous-time. Additionally, observations are made at irregular time instances, similar to the real CAV data, thus not guaranteeing all state transitions are observed, nor do the observation times necessarily correspond to transition times. Therefore, when the continuous-time, time-inhomogeneous HMM is fit to the synthetic data using either the piecewise-homogeneous approximation or the Aalen-Johansen estimator, compared to the numerical ODE solution approach, we are studying the model when it is both misspecified and correctly specified, respectively.

3.1.3. Mice Sleep Study Data

In the context of data collected from controlled experiments, especially when processing of the data is required, it is common that observations are recorded on a regularly-spaced grid of time instances. Thus, for the second real data application, ECoG data from a mice sleep study are used in which inter-observation times are equal. This study pertains to biological processes for which it is believed unrealistic to assume that transition rates are constant over time (Bojarskaite et al. 2020). ECoG and electromyography (EMG) measurements are recorded for each mouse at a frequency of 5000 observations per second, over hours. We construct an HMM to infer state sequences over the sleep-states 1: "intermediate state (IS)"; 2: "non-rapid eye movement (NREM)"; or 3: "rapid eye movement (REM)." Sleep researchers have manually annotated state labels that are provided in the dataset, but not at all observed time instances and we assume that misclassifications are possible. Hence, the theorized true state sequences are regarded as latent, and the observed state labels are understood as a response variable. We also use the observed ECoG time series as an additional response variable. In other sleep study applications of HMMs, researchers have considered the sleep-state labels as observed-without-error (e.g., Doroshenkov, Konyshev, and Selishchev 2007; Pan et al. 2012; Ghimatgar et al. 2019), or have attempted to label sleep-states in an unsupervised fashion (e.g., Flexerand et al. 2002).

The full details for the preprocessing of the raw data are provided in sec. 2 of Crainiceanu et al. (2009); however, instead of splitting the data into 30-sec epochs and applying the fast Fourier transform (FFT), we use 5-sec epochs. Additionally, we thinned the data to a frequency of approximately 556 observations per second instead of 5000. Once the FFT is applied on all epochs of the ECoG time series, what remains is a sequence of complex numbers corresponding to distinct frequencies. The modulus of FFT-mapped values determines the frequencies of waves that comprise the ECoG signal. In the context of studying sleep, the focus is on four separate frequency bands of brain waves and how the patterns they exhibit are signatures of the sleep-states. These wave bands and their respective frequency ranges (in Hz) are $\delta : [0.8, 4.0], \theta : [4.2, 8.0], \alpha : [8.2, 13.0]$, and $\beta : [13.2, 20.0]$ (Crainiceanu et al. 2009). For every 5-sec epoch, each of the four wave bands has an associated power, and we record their relative powers. Accordingly, the dataset we analyze

is discretized in 5-sec time increments, and to each time instance is an associated relative power of the δ, θ, α, and β bands, along with an observed sleep-state label, if available. If more than one observed sleep-state label resides in a single 5-sec epoch, for a given mouse, then the state label is considered missing (i.e., the likelihood function integrates over all states for that epoch, for that mouse).

3.2. Competing Methods

The five modeling approaches that will be compared are:

(A) : The numerical ODE solution (computed using the `deSolve` package (Soetaert, Petzoldt, and Setzer 2010) for R) to the correctly specified time-inhomogeneous forward equations (4), via a custom Metropolis-Hastings random walk MCMC algorithm.

(B) : The numerical ODE solution (computed using the `nhm` package (Titman 2019) for R) to the correctly specified time-inhomogeneous forward equations (4), via MLE.

(C) : The closed-form matrix-exponential-based solution to the time-homogeneous approximation (5), via a custom Metropolis-Hastings random walk MCMC algorithm.

(D) : The closed-form matrix-exponential-based solution to the time-homogeneous approximation (5), via MLE (computed using the `msm` package (Jackson 2011) for R).

(E) : The empirical approximation to the transition rate and transition probability matrices using the Nelson-Aalen and Aalen-Johansen estimators, respectively (computed using the `AalenJohansen` package (Bladt and Furrer 2023) for R).

Again, the primary interest for comparison is in how the inference differs depending on the approach used to evaluate the transition probability matrix for continuous-time, time-inhomogeneous Markov processes. Details of the properties of maximum likelihood estimation for continuous-time, multistate Markov processes are provided in Kalbfleisch and Lawless (1985) and Kay (1986), and others. A similar discussion but for Bayesian MCMC approaches is found in Zhao et al. (2016). See Moler and Van Loan (2003) for methods to evaluate the matrix exponential in (C) and (D). Lastly, for more theory related specifically to HMMs, Leroux (1992) details consistency results for maximum likelihood estimation and Vernet (2014) comments on posterior consistency.

3.2.1. Remarks on Computation

For the Bayesian approaches (A) and (C), the MCMC sampling is defined by a Metropolis-Hastings random walk. An adaptive proposal strategy is used during the burnin period to learn the proposal covariance for better mixing and quicker convergence. For approaches (B) and (D), maximum likelihood estimation is done via the BHHH (Titman 2019) and BFGS (Jackson 2011) algorithms, respectively. Both our custom numerical approach, (A), and the numerical approach in `nhm`, (B), implement the differential equation solvers from `deSolve` (Soetaert, Petzoldt, and Setzer 2010) which use the efficient and adaptive integration scheme developed in Petzold (1983). Note that when applying the `nhm` package, upper bounds on the parameter space are needed and successive observations with a time gap less than 0.001 are manually increased by 0.001 for numerical stability

purposes. Lastly, in all data applications in Sections 3.3 and 3.4, we assume a parametric form for the transition rate matrix. Note that the Aalen-Johansen and Nelson-Aalen estimators are not designed to be applied in the modeling setups of Sections 3.3 and 3.4 because the inferential focus for those analyses is to fit a given parametric form of the transition rates. Nonetheless, although the estimators for approach (E) are nonparametric, Section A.2 provides the details of a method we constructed for associating the transition rate parameters with the Nelson-Aalen estimator over all time points (a verification of the efficacy of this approach is provided in the supplemental materials). It is important to highlight that the `AalenJohansen` package offers the `prodint` function which numerically computes the product-integral representation of the transition probability matrix, analogous to `deSolve` with Riemann integration. Thus, a possible alternative to approach (A) would be to use `prodint` instead of `deSolve`; however, a more thorough application of approach (A) with `prodint` is not conducted because it would have been computationally infeasible given the MCMC structure and the size of the data used (for more information, see the supplemental materials). That being said, based on preliminary investigations, approach (A) with either `prodint` or `deSolve` should lead to similar results.

3.3. Simulation Study

Recall that the synthetic data is based on the CAV data from the `msm` package. Let the observed state label, $Y_{i,k}$, for the kth observation of the ith subject define an HMM response variable. Assuming that only state-adjacent misclassifications are possible, the probability mass function for $Y_{i,k}$ conditional on the true state $s_{i,k}$ is given by the rows of the table:

		$y_{i,k}$: observed state			
		1	2	3	4
	1	$1 - p_1$	p_1	0	0
$s_{i,k}$: true state	2	p_2	$1 - p_2 - p_3$	p_3	0
	3	0	p_4	$1 - p_4$	0
	4	0	0	0	1

where $p_1, p_2, p_3, p_4 \in [0, 1]$ are unknown parameters (see also, Sharples et al. 2003). Next, assuming that the progression of CAV is irreversible (e.g., as in Sharples et al. 2003; Jackson 2011), the transition rate matrix has the forward-model specification:

$$\boldsymbol{Q}(t) = \begin{pmatrix} -q_1 - q_2 & q_1 & 0 & q_2 \\ 0 & -q_3 - q_4 & q_3 & q_4 \\ 0 & 0 & -q_5 & q_5 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

and we model the rates as functions of covariates in a standard log-linear fashion (Jackson 2011; Williams et al. 2020):

$$q_j = \exp(\beta_{0,j} + \beta_{1,j} \cdot t + \beta_{2,j} \cdot \text{sex}), \tag{8}$$

for $j \in \{1, \ldots, 5\}$, where t is time in years since heart transplant surgery, "sex" is the biological sex of the subject, and $\beta_{0,j}$, $\beta_{1,j}$, and $\beta_{2,j}$ are unknown rate-specific coefficient parameters. This simple parametric model defines a continuous-time, time-inhomogeneous, multistate HMM that is fit to both the real and synthetic CAV datasets, and statistical inferences are made

on the transition rates for the progression of CAV in the years following heart transplant surgery.

We first fit the model to the CAV dataset by implementing a Metropolis-Hastings random walk MCMC algorithm on a posterior density kernel constructed with the likelihood function (7), where the transition probability matrices are computed numerically from the time-inhomogeneous forward equations (4) using the ODE solver in the `deSolve` package (Soetaert, Petzoldt, and Setzer 2010) in R. The prior densities for initial state probabilities $\boldsymbol{\pi}$ and the response function parameters p_1, p_2, p_3, p_4 are specified as logit-transformed Gaussian densities with diffuse scale parameters, and the transition rate parameters $\beta_{0,j}$, $\beta_{1,j}$, and $\beta_{2,j}$, for $j \in \{1, \ldots, 5\}$, are assigned Gaussian densities with diffuse scale parameters. The fitted posterior means are used to define the *true parameter values* of a continuous-time, time-inhomogeneous, multistate Markov process, from which we generate 100 synthetic datasets (though, we scale the *true* slope parameters $\beta_{1,j}$ by a factor of 3 for the purpose of illustration). Each of the 100 synthetic datasets consists of 2000 sampled subjects, and inter-observation times and sex covariate values are drawn consistent with the empirical frequencies observed in the real CAV dataset.

For the simulation study, we compare the five approaches to fitting the HMM as outlined in Section 3.2. Furthermore, the matrix-exponential-based solutions (C) and (D) are fitted—for each of the 100 synthetic datasets—for each choice of $d \in \{1/6, 1, 2\}$, as in (5), representing time-homogeneity at the resolution of bi-monthly, yearly, and bi-yearly, respectively. The solutions (C) and (D) become computationally prohibitive at the monthly resolution, that is, $d = 1/12$. A summary of the results is displayed in Figures 2 and 3; violin plots of the posterior means or MLEs for the baseline coefficient parameters for each transition rate are shown, along with the coverage of 0.95 level credible or confidence intervals. Note that solutions (C) and (D) represent the model fit when it is misspecified because the data are simulated with the rate matrix updating continuously as a function of time, not piecewise-constant with respect to time. Next, solution (E) represents a model fit when it is misspecified for the following three reasons: states are observed with error, not all transitions between states are necessarily observed, and observation times do not necessarily correspond to transition times. Only solutions (A) and (B) are fit using a correctly specified model. Violin plots for all HMM parameters are in the supplemental materials.

Consistent with the intuition provided in Section 2.1, it is evident from Figures 2 and 3 that substantial estimation bias for the baseline rate parameters (i.e., $\beta_{0,j}$ for $j \in \{1, \ldots, 5\}$) is exhibited by the HMMs fitted with the time-homogeneous approximation (5), for each choice of $d \in \{1/6, 1, 2\}$, whereas the coefficient estimates based on the numerical ODE solver approaches to fitting the time-inhomogeneous HMM (4) are well-concentrated on the true coefficient values. Additionally, a similar bias is observed with regards to approach (E). Note that despite approach (B) leading to unbiased estimates, the coverage probabilities for $\beta_{0,1}$, $\beta_{0,4}$, and $\beta_{0,5}$ are markedly less than the nominal level of 0.95, in contrast to those of (A). Moreover, less time is required for computing the likelihood function based on the numerical ODE solution than is required for the matrix-exponential-based computations for $d = 1/6$, as shown in
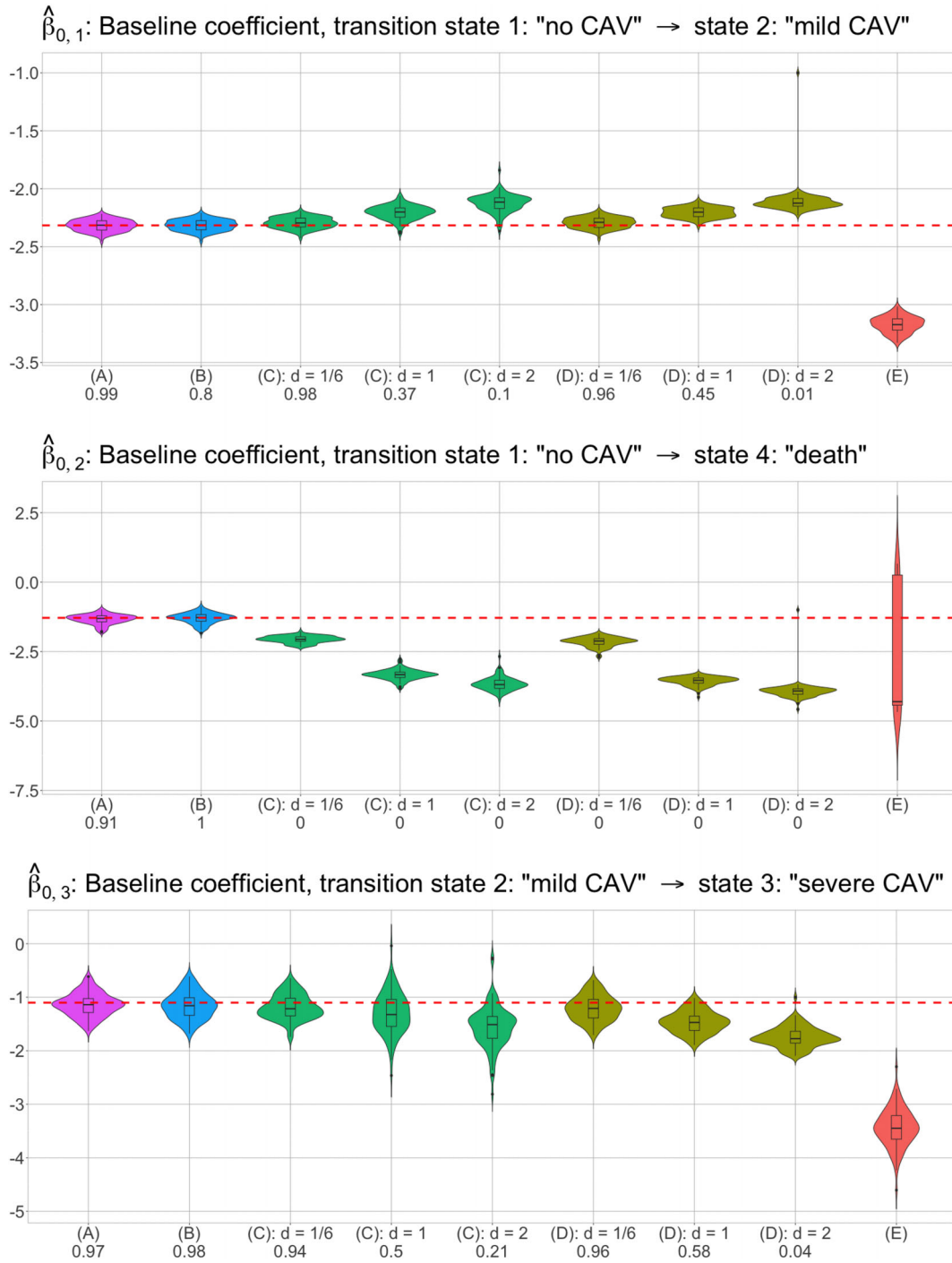
Figure 2. For approaches (A) and (C), presented are violin plots of the posterior means corresponding to the simulation study of 100 synthetic datasets. The violin plots associated with approaches (B) and (D) are of the MLEs. The violin plots for approach (E) are from the empirical estimation process outlined in Section A.2. The baseline coefficient parameters for the transition rates are as specified in (8). Where available, the numbers below each label report coverages of 0.95 level credible or confidence intervals.

Table 1. There are observed, however, reduced computation times for the matrix-exponential-based solutions for $d \in \{1, 2\}$, but these come at the cost of more substantial estimation bias.

3.4. Real Data Results

3.4.1. Cardiac Allograft Vasculopathy Study Data
As described in the simulation study in Section 3.3, the continuously time-evolving transition rates were constructed as a

feature of the synthetic or *artificial* CAV data (in order to assess the estimation performance observed when fitting the HMMs to data that are actually generated from a Markov process with this feature). Instead, in this section, we model the real CAV dataset in the same manner as Section 3.3 and compare the five approaches (A), (B), (C), (D), and (E) except now, time-evolving transition rates *may not* be a feature exhibited by these data, but in that case the slope coefficients for t (i.e., $\beta_{1,j}$ for $j \in \{1, \dots, 5\}$) would be estimated as zero and the five approaches should

$\hat{\beta}_{0,4}$: Baseline coefficient, transition state 2: "mild CAV" → state 4: "death"

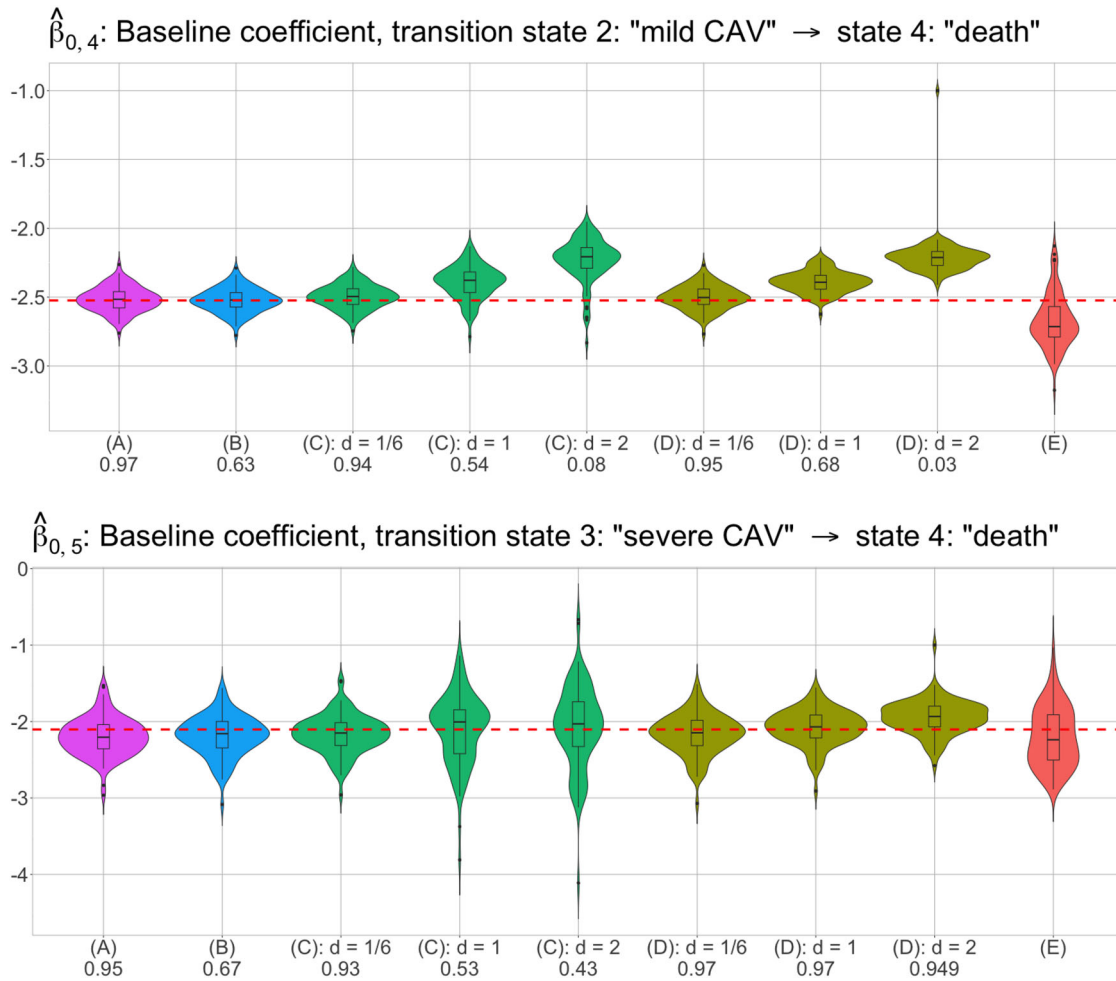$\hat{\beta}_{0,5}$: Baseline coefficient, transition state 3: "severe CAV" → state 4: "death"

Figure 3. For approaches (A) and (C), presented are violin plots of the posterior means corresponding to the simulation study of 100 synthetic datasets. The violin plots associated with approaches (B) and (D) are of the MLEs. The violin plots for approach (E) are from the empirical estimation process outlined in Section A.2. The baseline coefficient parameters for the transition rates are as specified in (8). Where available, the numbers below each label report coverages of 0.95 level credible or confidence intervals.

Table 1. Displayed are the times to compute the likelihood function evaluated at the true parameter values, averaged over the 100 synthetic datasets.

Approach	(A)	(C): $d = 1/6$	(C): $d = 1$	(C): $d = 2$
Likelihood evaluation time (seconds)	2.53	2.73	0.93	0.75

largely yield the same inference on the baseline rates. Thus, any observed difference in the estimated transition rates is likely attributable to time-evolving rates being exhibited by the data, and it would be rather contrived to assume that the underlying rates of CAV progression actually evolve in a piecewise-constant fashion—let alone to identify the correct resolution of the discontinuity points. Similarly, there is no reason to believe that the progression of CAV occurs precisely at the observed time points. Hence, any discrepancies in the resulting inference from the five approaches is indicative that future practitioners applying continuous-time, time-inhomogeneous Markov models should make a conscious and justified decision for which method of evaluating the transition probability matrix is used.

In any case, the 0.95 level credible or confidence sets for the baseline rate parameters—fitted by all five approaches and for each $d \in \{1/6, 1, 2\}$—are presented in Figure 4.

Differences are observed in the intervals [across the estimation approaches] most prominently between the numerical ODE solver approaches versus the piecewise-homogeneous approximations for lower resolutions, d, and approach (E). Comparing numerical approaches (A) and (B), the confidence intervals of (B) for the baseline parameters tend to be narrower, with the exception of $\beta_{0,2}$. This wide confidence set phenomenon for approach (B) is also seen in the confidence sets for the initial state parameters (provided in the supplemental materials). For completeness, Table 2 outlines the computation speed for the likelihood evaluations using the different approaches. The same trend as in the simulation study of Section 3.3 manifests. Finally, there is one special case (that we are aware of) in which the piecewise-homogeneous approach (for some $d > 0$) may be a reasonable approximation to the numerical ODE solution, for fitting a time-inhomogeneous HMM with a continuously evolving rate matrix; this special case is discussed in the next section.

3.4.2. Mice Sleep Study Data

The final real data application studied is from the mice sleep study data described in Section 3.1.3. Unlike the irregular observation times in the CAV data, the mice sleep study data are
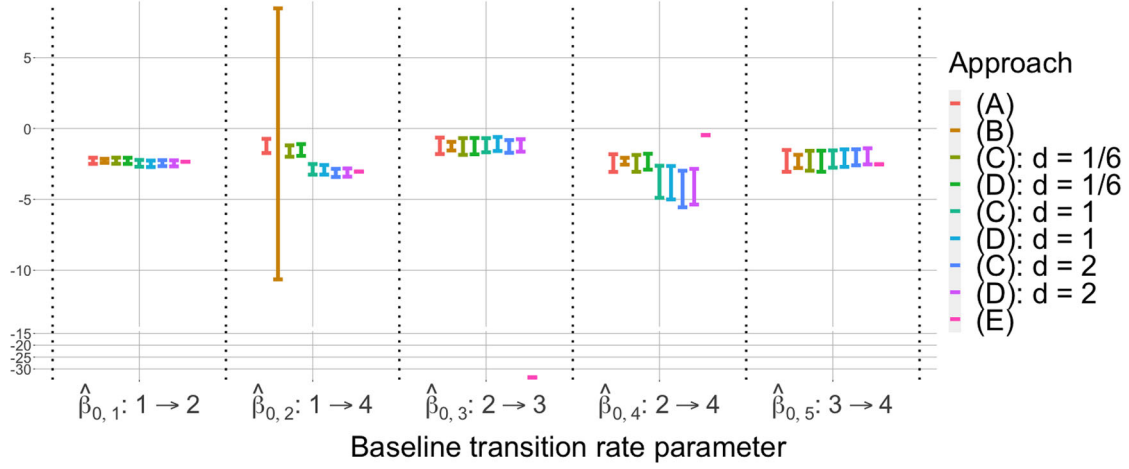
95% Credible and Confidence Sets

Figure 4. For each parameter (i.e., within the dotted vertical lines), the vertical intervals ordered left to right correspond to the "Approach" ordered top to bottom. For approaches (A) and (C), highest posterior density credible intervals are reported for the 0.95 level, for the fitted baseline coefficients, $\beta_{0,j}$, as in (8), for all transitions. For approaches (B) and (D), 0.95 level confidence intervals based on the Hessian are reported, as given by the nhm (Titman 2019) and msm packages (Jackson 2011), respectively. Due to the lack of availability of an approach for constructing a confidence interval for approach (E), only point estimates are provided.

Table 2. Displayed are the times to compute the likelihood function, for each approach evaluated at the posterior means associated with approach (A).

Approach	(A)	(C): $d = 1/6$	(C): $d = 1$	(C): $d = 2$
Likelihood evaluation time (seconds)	1.80	1.94	0.59	0.45

processed such that observation times are equally spaced. This is typically the type of scenario in which a discrete-time (and possibly time-inhomogeneous) Markov model would be appropriate, and that is exactly what the piecewise-homogeneous approach reduces to if the discontinuity points of $g_d(\cdot)$ align with the regularly-spaced grid of observed time instances. If it is suspected, however, that the transition rates evolve continuously in time, then a continuous-time Markov model becomes necessary for proper identification of the transition rates as continuous functions of time (e.g., as log-linear functions of a time-based covariate). Nonetheless, if the regularly-spaced grid of observed time instances is at a relatively high resolution (i.e., corresponding to relatively small $d > 0$) with respect to the rate of change of the transition rates over time, then—in the sense of a first-order Taylor approximation—the piecewise-

homogeneous approximation should yield similar inferences as that of the numerical ODE solution. This follows because the observed time-stamps $t_{i,k}$, for $k \in \{1, \ldots, n_i\}$ and $i \in \{1, \ldots, n\}$, in a dataset will *not* be distorted (i.e., rounded by some modulus) by the piecewise-homogeneous approach due to the existence of $d > 0$ such that $g_d(t_{i,k}) = t_{i,k}$, for all i, k, as illustrated by Lemma 2 with $c = 0$.

We compare—when $c = 0$ and $d > 0$ is small— the piecewise-homogeneous and numerical ODE solution approaches to time-inhomogeneous HMM-based inference on a real dataset from a mice sleep study. All state transitions are allowed, as specified by the rate matrix presented next. An additional complication, however, is that the states 1: "IS", 2: "NREM", and 3: "REM" are not an exhaustive list of sleep-states that have been identified by sleep researchers—they are just a simplified list of sleep-states of interest for this illustration. That being so, we must define an additional auxiliary state 4 as a catchall state to allow us to properly identify states 1, 2, and 3. For $t > 0$,

$$Q(t) = \begin{pmatrix} -q_1 - q_2 - q_3 & q_1 & q_2 & q_3 \\ q_4 & -q_4 - q_5 - q_6 & q_5 & q_6 \\ q_7 & q_8 & -q_7 - q_8 - q_9 & q_9 \\ q_{10} & q_{11} & q_{12} & -q_{10} - q_{11} - q_{12} \end{pmatrix}$$

where each q_j, for $j \in \{1, \ldots, 12\}$, is expressed as

$$q_j = \exp(\beta_{0,j} + \beta_{1,j} \cdot t), \tag{9}$$

where $\beta_{0,j}$ and $\beta_{1,j}$ are unknown rate-specific coefficient parameters, for $j \in \{1, \ldots, 12\}$.

The probability mass function for the observed state labels $Y_{i,k}$ conditional on the true state $s_{i,k}$ is given by the rows of the table:

		$y_{i,k}$: observed state			
		1	2	3	4
$s_{i,k}$: true state	1	$1 - p_1 - p_2$	p_1	p_2	0
	2	p_3	$1 - p_3 - p_4$	p_4	0
	3	p_5	p_6	$1 - p_5 - p_6$	0
	4	0	0	0	1

Note that the auxiliary state 4 is never actually observed, but is included in the probability mass function for completeness. Next, we adopt, from Langrock et al. (2013), a state-conditional Dirichlet density function for the observed relative powers,
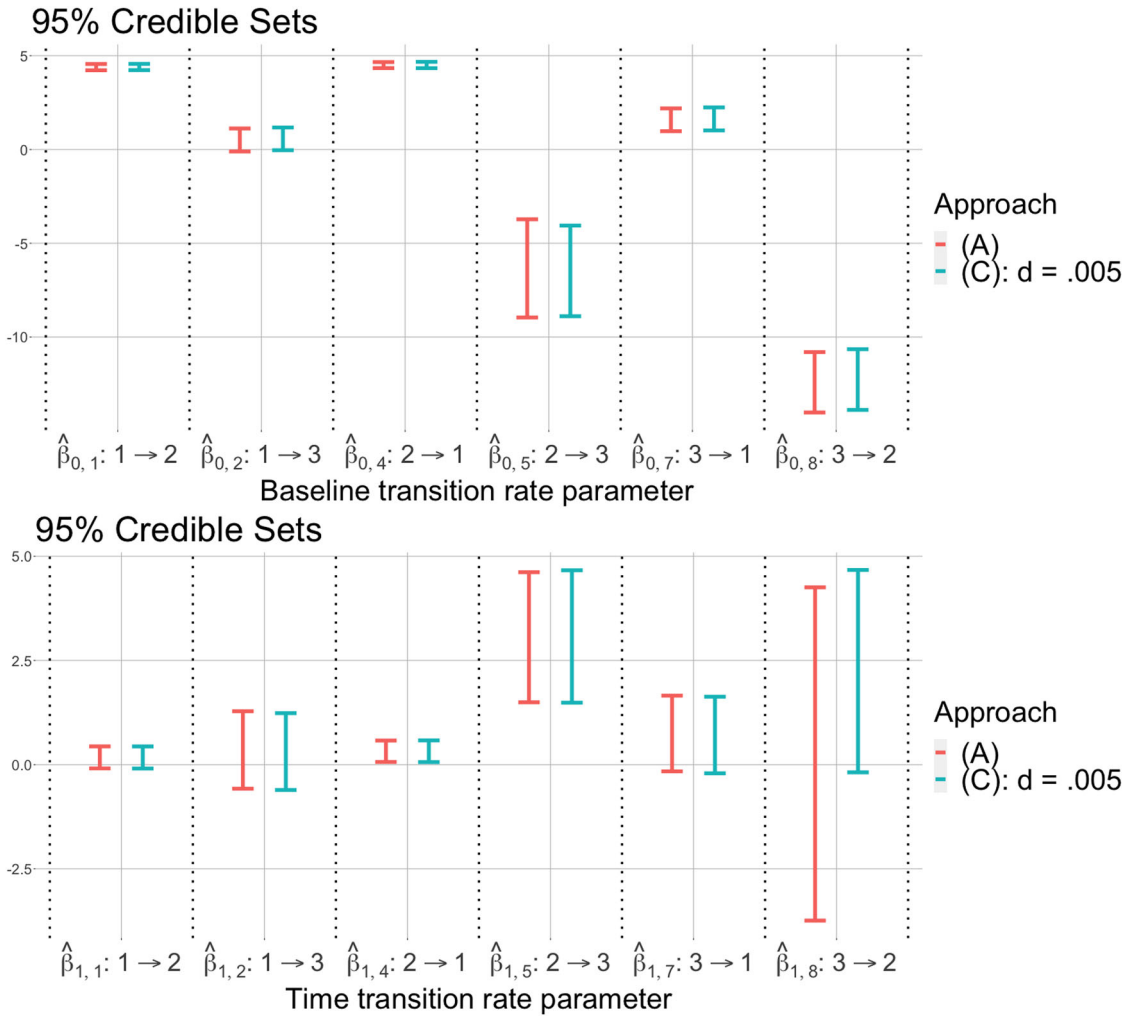
95% Credible Sets

$\hat{\beta}_{0,1}: 1 \to 2 \quad \hat{\beta}_{0,2}: 1 \to 3 \quad \hat{\beta}_{0,4}: 2 \to 1 \quad \hat{\beta}_{0,5}: 2 \to 3 \quad \hat{\beta}_{0,7}: 3 \to 1 \quad \hat{\beta}_{0,8}: 3 \to 2$

Baseline transition rate parameter

95% Credible Sets

$\hat{\beta}_{1,1}: 1 \to 2 \quad \hat{\beta}_{1,2}: 1 \to 3 \quad \hat{\beta}_{1,4}: 2 \to 1 \quad \hat{\beta}_{1,5}: 2 \to 3 \quad \hat{\beta}_{1,7}: 3 \to 1 \quad \hat{\beta}_{1,8}: 3 \to 2$

Time transition rate parameter

Figure 5. For each parameter (i.e., within the dotted vertical lines), the vertical intervals ordered left to right correspond to the "Approach" ordered top to bottom. The highest posterior density credible intervals at the 0.95 level are reported for the fitted coefficients in (9), for transitions of interest (i.e., excluding transitions to/from the auxiliary state 4). The value $d = 0.005$ is chosen so that $g_d(\cdot)$ aligns with the 5-sec time increments in the dataset (i.e., $g_d(t) = t$).

denoted $X_{i,k} := (X_{\delta,i,k}, X_{\theta,i,k}, X_{\alpha,i,k}, X_{\beta,i,k})$, of the δ, θ, α, and β bands, respectively, for all $k \in \{1, \ldots, n_i\}$ and $i \in \{1, \ldots, n\}$:

$$f_\lambda(x_{i,k} \mid s_{i,k} = s) = \frac{\Gamma(\lambda_\delta^{(s)} + \lambda_\theta^{(s)} + \lambda_\alpha^{(s)} + \lambda_\beta^{(s)})}{\Gamma(\lambda_\delta^{(s)})\Gamma(\lambda_\theta^{(s)})\Gamma(\lambda_\alpha^{(s)})\Gamma(\lambda_\beta^{(s)})}$$

$$\cdot x_{\delta,i,k}^{\lambda_\delta^{(s)}-1} \cdot x_{\theta,i,k}^{\lambda_\theta^{(s)}-1} \cdot x_{\alpha,i,k}^{\lambda_\alpha^{(s)}-1} \cdot x_{\beta,i,k}^{\lambda_\beta^{(s)}-1},$$

where $s \in \{1,2,3,4\}$, $\lambda_\delta^{(s)}, \lambda_\theta^{(s)}, \lambda_\alpha^{(s)}, \lambda_\beta^{(s)} > 0$, $x_{\delta,i,k}, x_{\theta,i,k}, x_{\alpha,i,k}, x_{\beta,i,k} \in [0,1]$, and $x_{\delta,i,k} + x_{\theta,i,k} + x_{\alpha,i,k} + x_{\beta,i,k} = 1$.

Figure 5 presents the 0.95 level credible sets for the rate parameters fitted by approaches (A) and (C) specified in Section 3.2. For the piecewise-homogeneous approach, (C), we set $d = 0.005$ so that $g_d(\cdot)$ aligns with the 5-sec time increments in the dataset (i.e., $g_d(t) = t$) scaled by a factor of 1/1000. We find that the credible set intervals always overlap across the two approaches for both the baseline and time transition rate coefficients. This confirms the intuition that for sufficiently small $d > 0$, if $g_d(t) = t$, both the numerical ODE solution and matrix-exponential-based solution lead to the same 0.95 level credible sets. Additionally, we find that, consistent with the belief of sleep researchers, the NREM $\to$ REM transition is non-constant over

Table 3. Displayed are the times to compute the likelihood function, for each approach evaluated at the posterior means associated with approach (A).

Approach	(A)	(C): $d = .005$
Likelihood evaluation time (seconds)	9.40	2.70

time. Hence, the piecewise-homogeneous approach is likely a reasonable first-order approximation (based on modeling the rates as log-linear functions of time). Lastly, Table 3 provides the computation speed for the likelihood evaluations using the two approaches.

4. Concluding Remarks

Continuous-time multistate Markov models (HMMs or otherwise) are widely used in data analysis applications involving processes that evolve over time, and generally speaking it is artificial to assume that a naturally occurring process would either exhibit rates of evolution that change in an arbitrarily regular piecewise-constant fashion, or that all observation times correspond to state transition times. Although these assumptions motivate pragmatic approximations for fitting

a time-inhomogeneous, continuous-time Markov model, we demonstrate that such approximations may lead to non-negligible biases in estimation. Furthermore, we demonstrate that there exist exact numerical-integration solutions. Moreover, these solutions are computationally more efficient than the approximations, and the computational tools to implement them are readily and freely available. Given the comprehensive list of approaches tested in this analysis, we find that the numerical-integration approaches, based on either Bayesian (approach (A)) or Frequentist (approach (B)) computations, lead to the most consistent estimation of the standard parametric structures that we consider in the simulation study in Section 3.3 and real data analyses in Section 3.4. Lastly, we note that the numerical integration can be implemented via the Riemann integral (by way of `deSolve` and/or `nhm`) or the product integral (by way of `prodint`).

Appendices

A.1. Proofs

Proof of Lemma 1. Let $y := (y_1, \ldots, y_n)^T$ so that $y = W\beta$, where $\beta := (\beta_0, \beta_1)^T$, and

$$W := \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}.$$

Then the least squares projection coefficients in (6) are obtained from the expression

$$b_n = (X^T X)^{-1} X^T y = (X^T X)^{-1} X^T W \beta,$$

where $b_n := (b_{0,n}, b_{1,n})^T$, and

$$X := \begin{pmatrix} 1 & g_d(t_1) \\ \vdots & \vdots \\ 1 & g_d(t_n) \end{pmatrix}.$$

Next, observe that for any $i \in \{1, \ldots, n\}$,

$$t_i - \bar{t}_n = t_i - g_d(t_i) + g_d(t_i) - \frac{1}{n} \sum_{k=1}^n g_d(t_k) + \frac{1}{n} \sum_{k=1}^n \{g_d(t_k) - t_k\}.$$

Multiplying both sides by $\{g_d(t_i) - \frac{1}{n} \sum_{k=1}^n g_d(t_k)\}$, summing over $i \in \{1, \ldots, n\}$, and dividing by $A_n = \sum_{i=1}^n \{g_d(t_i) - \frac{1}{n} \sum_{k=1}^n g_d(t_k)\}^2 > 0$, for sufficiently large n (by assumption), gives

$$\frac{\sum_{i=1}^n (t_i - \bar{t}_n)\{g_d(t_i) - \frac{1}{n} \sum_{k=1}^n g_d(t_k)\}}{A_n}$$
$$= \frac{\sum_{i=1}^n \{t_i - g_d(t_i)\}\{g_d(t_i) - \frac{1}{n} \sum_{k=1}^n g_d(t_k)\}}{A_n} + 1 + 0.$$

Then,

$$\left| \frac{\sum_{i=1}^n (t_i - \bar{t}_n)\{g_d(t_i) - \frac{1}{n} \sum_{k=1}^n g_d(t_k)\}}{A_n} - 1 \right|$$
$$= \left| \frac{\sum_{i=1}^n \{t_i - g_d(t_i)\}\{g_d(t_i) - \frac{1}{n} \sum_{k=1}^n g_d(t_k)\}}{A_n} \right|$$
$$\leq \frac{\sum_{i=1}^n d \cdot |g_d(t_i) - \frac{1}{n} \sum_{k=1}^n g_d(t_k)|}{A_n}$$
$$= d \cdot \frac{\|v\|_1}{\|v\|_2^2}$$

$$\leq d \cdot \frac{\sqrt{n} \|v\|_2}{\|v\|_2^2}$$
$$\leq \frac{d}{(\frac{1}{n} A_n)^{\frac{1}{2}}},$$

where $v \in \mathbb{R}^n$ is defined by $v_i := g_d(t_i) - \frac{1}{n} \sum_{k=1}^n g_d(t_k)$ for $i \in \{1, \ldots, n\}$. Since $d > 0$ is fixed, the proof is completed by applying the assumption that $\lim_{n \to \infty} \frac{1}{n} A_n = \infty$. □

Proof of Lemma 2. From the first expression in (6) and since, by assumption, $t_i = g_d(t_i) + c$,

$$\frac{b_{0,n} - \beta_0}{\beta_1} = \frac{\bar{t}_n \{\sum_{i=1}^n g_d^2(t_i)\} - \frac{1}{n} \sum_{i,j} t_i \cdot g_d(t_i) g_d(t_j)}{A_n}$$
$$= \frac{\bar{t}_n \{\sum_{i=1}^n g_d^2(t_i)\} - \frac{1}{n} \sum_{i,j} \{g_d^2(t_i) + c \cdot g_d(t_i)\} g_d(t_j)}{A_n}$$
$$= \frac{\bar{t}_n \{\sum_{i=1}^n g_d^2(t_i)\} - \frac{1}{n} \{\sum_{j=1}^n g_d(t_j)\} \sum_{i=1}^n \{g_d^2(t_i) + c \cdot g_d(t_i)\}}{A_n}$$
$$= \frac{\{\bar{t}_n - \frac{1}{n} \sum_{j=1}^n g_d(t_j)\}\{\sum_{i=1}^n g_d^2(t_i)\} - \frac{c}{n}\{\sum_{i=1}^n g_d(t_i)\}^2}{A_n}$$
$$= c \cdot \frac{\{\sum_{i=1}^n g_d^2(t_i)\} - \frac{1}{n}\{\sum_{i=1}^n g_d(t_i)\}^2}{A_n}$$
$$= c.$$

□

A.2. Transition Rate Parameter Estimation via the Nelson-Aalen Estimator

Recall that the transition rate matrix, $Q(t)$, from Sections 3.3 and 3.4.1 has the form

$$Q(t) = \begin{pmatrix} -q_1 - q_2 & q_1 & 0 & q_2 \\ 0 & -q_3 - q_4 & q_3 & q_4 \\ 0 & 0 & -q_5 & q_5 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

where each $q_j, j = 1, 2, \ldots, 5$, has the following form

$$q_j = q_j(t \mid \text{sex}) = \exp(\beta_{0,j} + \beta_{1,j} \cdot t + \beta_{2,j} \cdot \text{sex}),$$

with $t \geq 0$ and sex $\in \{0, 1\}$. The Nelson-Aalen estimator estimates the *cumulative* transition rate function (Aalen and Johansen 1978; Borgan 1997). Let $A_j(t \mid \text{sex})$ be the cumulative transition rate function defined as

$$A_j(t \mid \text{sex}) := \int_0^t q_j(s \mid \text{sex}) \, ds,$$

and let $\hat{A}_j(t \mid \text{sex})$ denote the Nelson-Aalen estimator. The `aalen_johansen` function provides $\hat{A}_j(t \mid \text{sex})$, for each observed transition time, $t \geq 0$, and for $j = 1, 2, \ldots, 5$. Then, in order to retrieve estimates for $\beta_{0,j}, \beta_{1,j}$, and $\beta_{2,j}, \forall j$, define

$$g(\beta_{0,j}, \beta_{1,j}, \beta_{2,j} \mid t, \text{sex}) := \int_0^t q_j(s \mid \text{sex}) \, ds$$
$$= \frac{\exp(\beta_{0,j} + \beta_{1,j} \cdot t + \beta_{2,j} \cdot \text{sex}) - \exp(\beta_{0,j} + \beta_{2,j} \cdot \text{sex})}{\beta_{1,j}},$$

and use the ℓ_1 loss to formulate the parameter estimates as

$$\widehat{\beta}_j^T = \begin{pmatrix} \hat{\beta}_{0,j} \\ \hat{\beta}_{1,j} \\ \hat{\beta}_{2,j} \end{pmatrix} = \underset{\beta_{0,j}, \beta_{1,j}, \beta_{2,j}}{\text{argmin}} \sum_t \left| g(\beta_{0,j}, \beta_{1,j}, \beta_{2,j} \mid t, \text{sex}) - \hat{A}_j(t \mid \text{sex}) \right|.$$

Supplementary Materials

Acknowledgments

Disclosure Statement

Funding

ORCID

Emmett B. Kendall http://orcid.org/0000-0002-7985-0205
Jonathan P. Williams http://orcid.org/0000-0001-5164-9430
Frederic Bois http://orcid.org/0000-0002-4154-0391
Vo Hang Thanh https://orcid.org/0000-0003-4599-0215

References

Aalen, O., Borgan, O., and Gjessing, H. (2008), *Survival and Event History Analysis: A Process Point of View*, New York: Springer. [2]

Aalen, O. O., Farewell, V. T., De Angelis, D., Day, N. E., and Nöel Gill, O. (1997), "A Markov Model for HIV Disease Progression Including the Effect of HIV Diagnosis and Treatment: Application to AIDS Prediction in England and Wales," *Statistics in Medicine*, 16, 2191–2210. [3]

Aalen, O. O., and Johansen, S. (1978), "An Empirical Transition Matrix for Non-homogeneous Markov Chains based on Censored Observations," *Scandinavian Journal of Statistics*, 5, 141–150. [2,13]

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (2012), *Statistical Models based on Counting Processes*, New York: Springer. [2]

Aspinall, W., Carniel, R., Jaquet, O., Woo, G., and Hincks, T. (2006), "Using Hidden Multi-State Markov Models with Multi-Parameter Volcanic Data to Provide Empirical Evidence for Alert Level Decision-Support," *Journal of Volcanology and Geothermal Research*, 153, 112–124. [2]

Bladt, M., and Furrer, C. (2023), "Conditional Aalen–Johansen Estimation," arXiv preprint arXiv:2303.02119. [7]

Bojarskaite, L., Bjørnstad, D. M., Pettersen, K. H., Cunen, C., Hermansen, G. H., Åbjørsbråten, K. S., Chambers, A. R., Sprengel, R., Vervaeke, K., Tang, W., et al. (2020), "Astrocytic ca2+ Signaling is Reduced During Sleep and is Involved in the Regulation of Slow Wave Sleep," *Nature Communications*, 11, 1–16. [3,7]

Borgan, Ø. (1997), "Three Contributions to the Encyclopedia of Biostatistics: The Nelson-Aalen, Kaplan-Meier, and Aalen-Johansen," Preprint series. Statistical Research Report *http://urn.nb.no/URN:NBN:no-23420*. [2,13]

Bureau, A., Shiboski, S., and Hughes, J. P. (2003), "Applications of Continuous Time Hidden Markov Models to the Study of Misclassified Disease Outcomes," *Statistics in Medicine*, 22, 441–462. [2]

Crainiceanu, C. M., Caffo, B. S., Di, C.-Z., and Punjabi, N. M. (2009), "Nonparametric Signal Extraction and Measurement Error in the Analysis of Electroencephalographic Activity During Sleep," *Journal of the American Statistical Association*, 104, 541–555. [7]

Datta, S., and Satten, G. A. (2001), "Validity of the Aalen–Johansen Estimators of Stage Occupation Probabilities and Nelson–Aalen Estimators of Integrated Transition Hazards for non-Markov Models," *Statistics & Probability Letters*, 55, 403–411. [2]

——— (2002), "Estimation of Integrated Transition Hazards and Stage Occupation Probabilities for Non-Markov Systems Under Dependent Censoring," *Biometrics*, 58, 792–802. [2]

Doroshenkov, L., Konyshev, V., and Selishchev, S. (2007), "Classification of Human Sleep Stages based on EEG Processing Using Hidden Markov Models," *Biomedical Engineering*, 41, 25–28. [7]

Faddy, M. (1976), "A Note on the General Time-Dependent Stochastic Compartmental Model," *Biometrics*, 32, 443–448. [1]

Flexerand, A., Dorffner, G., Sykacekand, P., and Rezek, I. (2002), "An Automatic, Continuous and Probabilistic Sleep Stager based on a Hidden Markov Model," *Applied Artificial Intelligence*, 16, 199–207. [7]

Furui, A., Akiyama, T., and Tsuji, T. (2021), "A Time-Series Scale Mixture Model of EEG with a Hidden Markov Structure for Epileptic Seizure Detection," arXiv:2111.06526, pp. 1–5. [2]

Ghimatgar, H., Kazemi, K., Helfroush, M. S., and Aarabi, A. (2019), "An Automatic Single-Channel EEG-based Sleep Stage Scoring Method based on Hidden Markov Model," *Journal of Neuroscience Methods*, 324, 108320. [7]

Glidden, D. V. (2002), "Robust Inference for Event Probabilities with Non-Markov Event Data," *Biometrics*, 58, 361–368. [2]

Gunnes, N., Borgan, Ø., and Aalen, O. O. (2007), "Estimating Stage Occupation Probabilities in Non-Markov Models," *Lifetime Data Analysis*, 13, 211–240. [2]

Hubbard, R. A., Inoue, L., and Fann, J. (2008), "Modeling Nonhomogeneous Markov Processes via Time Transformation," *Biometrics*, 64, 843–850. [2]

Jack, C. R., Therneau, T. M., Wiste, H. J., Weigand, S. D., Knopman, D. S., Lowe, V. J., Mielke, M. M., Vemuri, P., Roberts, R. O., Machulda, M. M., Senjem, M. L., Gunter, J. L., Rocca, W. A. and Petersen, R. C. (2016), "Transition Rates between Amyloid and Neurodegeneration Biomarker States and to Dementia: A Population-based, Longitudinal Cohort Study," *Lancet Neurol*, 15, 56–64. [1]

Jackson, C. H. (2011), "Multi-State Models for Panel Data: The msm package for R," *Journal of Statistical Software*, 38, 1–29. [6,7,8,11]

Jackson, C. H., Sharples, L. D., Thompson, S. G., Duffy, S. W., and Couto, E. (2003), "Multistate Markov Models for Disease Progression with Classification Error," *The Statistician*, 52, 193–209. [1]

Joly, P., Commenges, D., Helmer, C., and Letenneur, L. (2002), "A Penalized Likelihood Approach for An Illness–Death Model with Interval-Censored Data: Application to Age-Specific Incidence of Dementia," *Biostatistics*, 3, 433–443. [2]

Kalbfleisch, J., and Lawless, J. F. (1985), "The Analysis of Panel Data Under a Markov Assumption," *Journal of the American Statistical Association*, 80, 863–871. [7]

Kaplan, E. L., and Meier, P. (1958), "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 53, 457–481. [2]

Karlin, S., and Taylor, H. E. (1981), *A Second Course in Stochastic Processes*, New York: Elsevier. [1,3,4]

Kay, R. (1986), "A Markov Model for Analysing Cancer Markers and Disease States in Survival Studies," *Biometrics*, 42, 855–865. [1,7]

Langrock, R., Swihart, B. J., Caffo, B. S., Punjabi, N. M., and Crainiceanu, C. M. (2013), "Combining Hidden Markov Models for Comparing the Dynamics of Multiple Sleep Electroencephalograms," *Statistics in Medicine*, 32, 3342–3356. [11]

Leroux, B. G. (1992), "Maximum-Likelihood Estimation for Hidden Markov Models," *Stochastic Processes and their Applications*, 40, 127–143. [7]

Liu, Y.-Y., Li, S., Li, F., Song, L., and Rehg, J. M. (2015), "Efficient Learning of Continuous-Time Hidden Markov Models for Disease Progression," in *Advances in Neural Information Processing Systems* (Vol. 28), pp. 3599. [3]

Marchetti, L., Priami, C., and Thanh, V. H. (2017), *Simulation Algorithms for Computational Systems Biology*, Cham: Springer. [2]

Moler, C., and Van Loan, C. (2003), "Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later," *SIAM Review*, 45, 3–49. [7]

Nelson, W. (1969), "Hazard Plotting for Incomplete Failure Data," *Journal of Quality Technology*, 1, 27–52. [2]

——— (1972), "Theory and Applications of Hazard Plotting for Censored Failure Data," *Technometrics*, 14, 945–966. [2]

Pan, S.-T., Kuo, C.-E., Zeng, J.-H., and Liang, S.-F. (2012), "A Transition-Constrained Discrete Hidden Markov Model for Automatic Sleep Staging," *Biomedical Engineering Online*, 11, 1–19. [7]

Petzold, L. (1983), "Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations," *SIAM Journal on Scientific and Statistical Computing*, 4, 136–148. [7]

Sharples, L. D., Jackson, C. H., Parameshwar, J., Wallwork, J., and Large, S. R. (2003), "Diagnostic Accuracy of Coronary Angiography and Risk Factors for Post–Heart-Transplant Cardiac Allograft Vasculopathy," *Transplantation*, 76, 679–682. [8]

Siméon, S., Beaudouin, R., Brotzmann, K., Braunbeck, T., and Bois, F. Y. (2021), "Multistate Models of Developmental Toxicity: Application to Valproic Acid-Induced Malformations in the Zebrafish Embryo," *Toxicology and Applied Pharmacology*, 414, 115424. [1,2]

Soetaert, K., Petzoldt, T., and Setzer, R. W. (2010), "Solving Differential Equations in R: Package deSolve," *Journal of Statistical Software*, 33, 1–25. [2,4,7,8]

Soper, B. C., Nygård, M., Abdulla, G., Meng, R., and Nygård, J. F. (2020), "A Hidden Markov Model for Population-Level Cervical Cancer Screening Data," *Statistics in Medicine*, 39, 3569–3590. [2]

Thanh, V. H., and Priami, C. (2015), "Simulation of Biochemical Reactions with Time-Dependent Rates by the Rejection-based Algorithm," *The Journal of Chemical Physics*, 143, 08B601_1. [2]

Titman, A. (2019), *nhm: Non-Homogeneous Markov and Hidden Markov Multistate Models*, R package version 0.1.0. *https://CRAN.R-project.org/package=nhm* [2,4,7,11]

Titman, A. C. (2011), "Flexible Nonhomogeneous Markov Models for Panel Observed Data," *Biometrics*, 67, 780–787. [1,2,4]

Titman, A. C., and Sharples, L. D. (2008), "A General Goodness-of-Fit Test for Markov and Hidden Markov Models," *Statistics in Medicine*, 27, 2177–2195. [2]

——— (2010), "Model Diagnostics for Multi-State Models," *Statistical Methods in Medical Research*, 19, 621–651. [2,4]

Vernet, E. (2014), "Posterior Consistency for Nonparametric Hidden Markov Models with Finite State Space," arXiv:1311.3092. [7]

Wan, L., Lou, W., Abner, E., and Kryscio, R. J. (2016), "A Comparison of Time-Homogeneous Markov Chain and Markov Process Multi-State Models," *Communications in Statistics: Case Studies, Data Analysis and Applications*, 2, 92–100. [3]

Williams, J. P., Hermansen, G. H., Strand, H., Clayton, G., and Nygård, H. M. (2024), "Bayesian Hidden Markov Models for Latent Variable Labeling Assignments in Conflict Research: Application to the Role Ceasefires Play in Conflict Dynamics," *Annals of Applied Statistics*, 18, 2034–2061. [2]

Williams, J. P., Storlie, C. B., Therneau, T. M., Jack Jr, C. R., and Hannig, J. (2020), "A Bayesian Approach to Multistate Hidden Markov Models: Application to Dementia Progression," *Journal of the American Statistical Association*, 115, 16–31. [1,2,8]

Xu, S., Chen, M., Feng, T., Zhan, L., Zhou, L., and Yu, G. (2021), "Use ggbreak to Effectively Utilize Plotting Space to Deal with Large Datasets and Outliers," *Frontiers in Genetics*, 12, 774846.

Zhao, T., Wang, Z., Cumberworth, A., Gsponer, J., de Freitas, N., and Bouchard-Côté, A. (2016), "Bayesian Analysis of Continuous Time Markov Chains with Application to Phylogenetic Modelling," *Bayesian Analysis*, 11, 1203–1237. [7]