MOVING-RESTING PROCESS WITH MEASUREMENT ERROR IN ANIMAL MOVEMENT MODELING

CHAORAN HU[1]*, L. MARK ELBROCH[2], THOMAS MEYER[3], VLADIMIR POZDNYAKOV[1], AND JUN YAN[1]

ABSTRACT. 1. Statistical modeling of animal movement is of critical importance. The continuous trajectory of an animal's movements is only observed at discrete, often irregularly spaced time points. Most existing models do not allow inactivity periods such as resting or sleeping.
2. The recently proposed moving-resting (MR) model is a Brownian motion governed by a telegraph process, which allows periods of inactivity in one state of the telegraph process. The MR model, like any other continuous-time model, naturally handles unequal sampling intervals.
3. The MR model shows promise in modeling the movements of predators with long inactive periods, such as many felids, but the lack of accommodation of measurement errors seriously prohibits its application in practice. Here we incorporate measurement errors in the MR model and derive basic properties of the model. Inferences are based on a composite likelihood using the Markov property of the chain composed by every other observed increments.
4. The performance of the method is validated in finite sample simulation studies. Application to the movement data of a mountain lion in Wyoming illustrates the utility of the method.

KEYWORDS: Composite likelihood; Dynamic programming; Markov process

1. INTRODUCTION

Statistical modeling of animal movement is of great importance in addressing fundamental questions about space use, movement, resource selection, and behavior in animal ecology (Hooten et al., 2017). The explosion of telemetric data on animal movement from the recent advancements in tracking and observation technologies presents countless opportunities and challenges (Cagnacci et al., 2010; Patterson et al., 2017). Telemetry devices, like Global Positioning System (GPS) receivers, can only determine an animal's position at discrete moments in time so a continuous trajectory is never available. Programming the device to record at a very fast fix rate could approximate a continuous trajectory, but this is seldom done due to battery-life limitations: it is very expensive to capture and collar an animal, so a long episodic time record is usually preferable to a highly detailed record, especially for animals that spend a great deal of time not moving around. GPS receivers often produce fixes at irregularly spaced time points, even if researchers program for regular intervals, due

1. Department of Statistics, University of Connecticut, 215 Glenbrook Rd., U-4120, Storrs, CT 06269, U.S.A.
2. Panthera, 8 West 40th Street, 18th Floor, NY 10018, U.S.A.
3. Department of Natural Resources & the Environment, University of Connecticut, 1376 Storrs Road, Unit 4087 Storrs, CT 06269, U.S.A.
* E-mail: chaoran.hu@uconn.edu.

to environmental factors such as satellite communication issues or sky occlusion. As a result, discrete-time models such as the state space model (e.g., Jonsen et al., 2005; Patterson et al., 2008; McClintock et al., 2012) are not realistic. Continuous-time models based on stochastic differential equations (SDE) (e.g., Preisler et al., 2004; Horne et al., 2007; Brillinger, 2010; McClintock, 2017; Scharf et al., 2017; Hanks et al., 2017; Parton and Blackwell, 2017) can handle the irregular spacing naturally. Nonetheless, the papers cited above assume perpetual motion and do not accommodate periods of inactivity. In contrast, on the time scale of most telemetry data, most animals alternate between periods of movement (foraging) and periods of inactivity such as prey handling or rest (Mashanova et al., 2010; Ueno et al., 2012; Jeschke, 2007). Realistic continuous-time models that accommodate inactive periods are needed.

The recently proposed moving-resting (MR) process (Yan et al., 2014) is a promising model that allows both irregularly time-spaced observations and inactive time periods. The MR process is a Brownian motion governed by a telegraph or on-off process (e.g., Zacks, 2004). Specifically, it allows an animal to alternate between a moving state, during which it moves in a Brownian motion (BM), and a resting state, during which it remains motionless. The switch between the two states is governed by a telegraph process, where the holding time (or duration) of each state is assumed to follow an exponential distribution. As an example, consider the movement of a mountain lion in the Gros Venture mountain range, Wyoming, where it spent a lot of time sleeping, hunting, or handling a kill. The top-left panel of Figure 1 shows the observed trajectory in Universal Transverse Mercator (UTM) eastings and northings relative to a fixed location for a two-month period in spring 2011. The top-right panel shows a sample path from the MR model. Those two trajectories are indeed visually similar, supporting the utility of the model. Note that the flat pieces of easting and northing coordinates are tied together over the same time periods for both real data and the simulated one.

A major limitation of the MR model is that it does not accommodate the measurement error or noise of telemetric devices. Adding measurement error to a BM model is not crucial as long as the noise is small in comparison to the total standard deviation of the increments of the Brownian motion between two consecutive time points (Pozdnyakov et al., 2014). In such cases, discarding the noise would not produce significant bias. The impact of the noise on inferences about MR processes, however, is much greater. For a given sequence of hidden states, the likelihood is the product of both densities *and* probabilities. With perfect instrumentation, a sequence of observed locations over time that are exactly the same, that is, when there is no change in neither the easting or northing coordinates result in a "flat" piece of trajectory; then the animal is known to be motionless over the time period spanned by the sequence. The likelihood contribution is the probability of staying in the motionless state instead of the density of the increment at zero. Adding even a tiny bit of noise would remove those flat pieces and, hence, cause drastic bias in the likelihood estimator of the parameters. One possible remedy is to round the observed coordinates, which enforces flat pieces. The number of such pieces, however, depends greatly on the rounding level, and there are no obvious rules to aid researchers in choosing best levels. A detailed illustration of the issue is given in Section 2.1 and 3.3.

Dealing with added noise in an MR process is challenging because it invalidates the Markov property of the joint location-state process. The transition density from one time point to the next can in principle be obtained from convolving the results for the MR process (Yan et al., 2014) with normally distributed noise, although that is computationally very

intensive. A lack of the Markov property of the joint location-state process means that the likelihood cannot be easily formed by multiplying these transition densities. Because the measurement errors are continuous, the dynamic programming tools of HMM based on a finite number of hidden states (Cappé et al., 2005) are not applicable. Further complications include that our observations are irregularly spaced in time, and that both the current hidden state and the current location observation depend on the previous hidden state. Therefore, the Baum-Welch algorithm (e.g., Zucchini et al., 2016) cannot be constructed. The generic simulation-based inferences such as iterated filtering (Ionides et al., 2011, 2015) or particle Markov chain Monte Carlo (Andrieu et al., 2010), available in R package `pomp` (King et al., 2016), are not applicable to our investigation due to the complexity of the MR process with measurement error. Alternatives to strict stillness in the MR model can be a BM with a much smaller infinitesimal variance parameter or an Ornstein–Uhlenbeck (OU) process (Uhlenbeck and Ornstein, 1930). Such formulations avoid directly modeling the measurement error. Nonetheless, they are still motions and conceptually not as preferable as strict resting for modeling animals staying at a particular location (sleeping, hunting, or handling). In addition, they introduce extra parameters which make estimation challenging. Such models are employed in a sensitivity analysis in Section 3.2.

Our contribution in this paper is a toolbox for applying the MR process with measurement error to animal movement modeling. First, we show that discarding the measurement error, even tiny ones, causes severe bias in estimation, and that rounding does not provide any satisfactory solution. To make inferences for MR process with measurement error, we establish that, after thinning every other observation, the remaining observations are location-state Markov. This facilitates a composite likelihood which contains two true likelihood components, one based on odd-numbered observations and the other based on even-numbered observations. The true likelihood of each component is computed with dynamic programming. The key for efficient computation is a small number (just two) of hidden states. The variance of the maximum composite likelihood estimator can be estimated through parametric bootstrap. The validity of the approach is confirmed in a simulation study. We then apply the approach to model the movement data of a mountain lion in Wyoming, whose trajectory is known to have long inactive periods. Our methods are publicly available in an R package `smam` (Hu et al., 2020b) with efficient C++ code.

2. METHODS

2.1. **Moving-Resting Process.** The MR process is a Brownian motion with an infinitesimal variance that is governed by an alternating renewal process with two different holding times. More specifically, the true location at time t is modeled by a two-component Markov process $(X(t), S(t))$. Here $S(t)$ is an unobservable continuous-time Markov chain with two states $\{0, 1\}$. The Markov chain starts in 0 or 1 according to an initial distribution $\nu(\cdot)$. It stays in state $i = 0, 1$ for an exponentially distributed holding time with rate λ_i, and then it switches to the other state with probability 1. We will assume that the initial distribution is stationary, that is,

$$\nu(1) = \Pr(S(0) = 1) = p_1 = \frac{\lambda_0}{\lambda_0 + \lambda_1},$$

and $\nu(0) = \Pr(S(0) = 0) = p_2 = 1 - p_1$. The $S(t) = 1$ means that the animal is moving at time t, $S(t) = 0$ means that the animal is in a motionless state.

An MR process $X(t)$, $t \geq 0$ that models the location of the animal is defined by a stochastic differential equation

$$\mathrm{d}X(t) = \begin{cases} \sigma \mathrm{d}B(t) & \text{if } S(t) = 1, \\ 0 & \text{if } S(t) = 0, \end{cases} \tag{1}$$

where $B(t)$ is the standard Brownian motion independent of $S(t)$, and $\sigma > 0$ is a parameter that determines the animal mobility when it moves. This SDE simply means that in state 1 the animal moves according to a Brownian motion. The location process $\{X(t)\}_{t\geq 0}$ itself is not Markov, but the location-state process $\{X(t), S(t)\}$ is a Markov process with stationary increments.

Using results on the telegraph process one can derive the marginal distribution of the increment $X(t) - X(0)$. This distribution forms the basis of the composite likelihood approach in Yan et al. (2014). The full maximum likelihood estimation based on dynamic programming was developed in Pozdnyakov et al. (2019). A more detailed description of the MR process and the derivation of the distribution of $X(t) - X(0)$ is given in Appendix A.

For actual animal movement data, we never observe the exact values of $X(t)$ but only $X(t)$ with added measurement errors. For an MR process, the probability of observing a zero increment is strictly positive. Adding noise makes this probability zero. Rounding can help, but it is not trivial to come up with an appropriate rounding level. Figure 1 (upper left) shows the easting/northing coordinates in the UTM coordinate system of a female mountain lion in spring 2012 in the Gros Ventre mountain range, Wyoming. The patterns of resting — places where both lines are more or less flat — and moving are readily apparent, which can hardly be captured by any existing model that assumes perpetual movements. The other three panels of Figure 1 show the coordinates of a simulated MR path without noise and with noise of two levels. The pattern is very similar to that in the upper left panel for the female mountain lion. The similarity is obvious, suggesting that an MR process might be a good model, but as shown next, ignoring the noise is disastrous in estimating the model parameters.

We demonstrate the impact of noise on estimation by a simulation study. Consider a MR process with parameters $\lambda_1 = 1$ hour^{-1}, $\lambda_0 = 0.5$ hour^{-1}, $\sigma = 1$ km/hour$^{1/2}$. The measurement errors were independent Gaussian noise with standard deviation 0.05 km (1/20 of σ) and 0.01 km (1/100 of σ). The time interval between consecutive observations is 1 hour. We generated 100 datasets, each with sampling horizon 500 hours. The maximum likelihood estimates based on the MR process were obtained for datasets with and without noise, and for data with noise, three levels of rounding were considered, 10, 50, and 100 meters. That is, the coordinates in meters were rounded to the closest 10, 50, and 100 meters, respectively. Table 1 summarizes the parameter estimates based on the 100 replicates. When there was no noise, the point estimates were good, recovering the true parameters with high accuracy. For data with noise but no rounding, the optimization did not converge for most replications because of the Nelder-Mead simplex degeneracy (Nelder and Mead, 1965). With the help of various levels of rounding, the convergence percentage increases as the rounding becomes coarser, and the bias decreases but remain notable. This is true even for the cases with a noise standard deviation 0.01 km. It is indeed unclear how to choose an appropriate rounding level. Therefore, a practical model should handle the measurement errors directly.

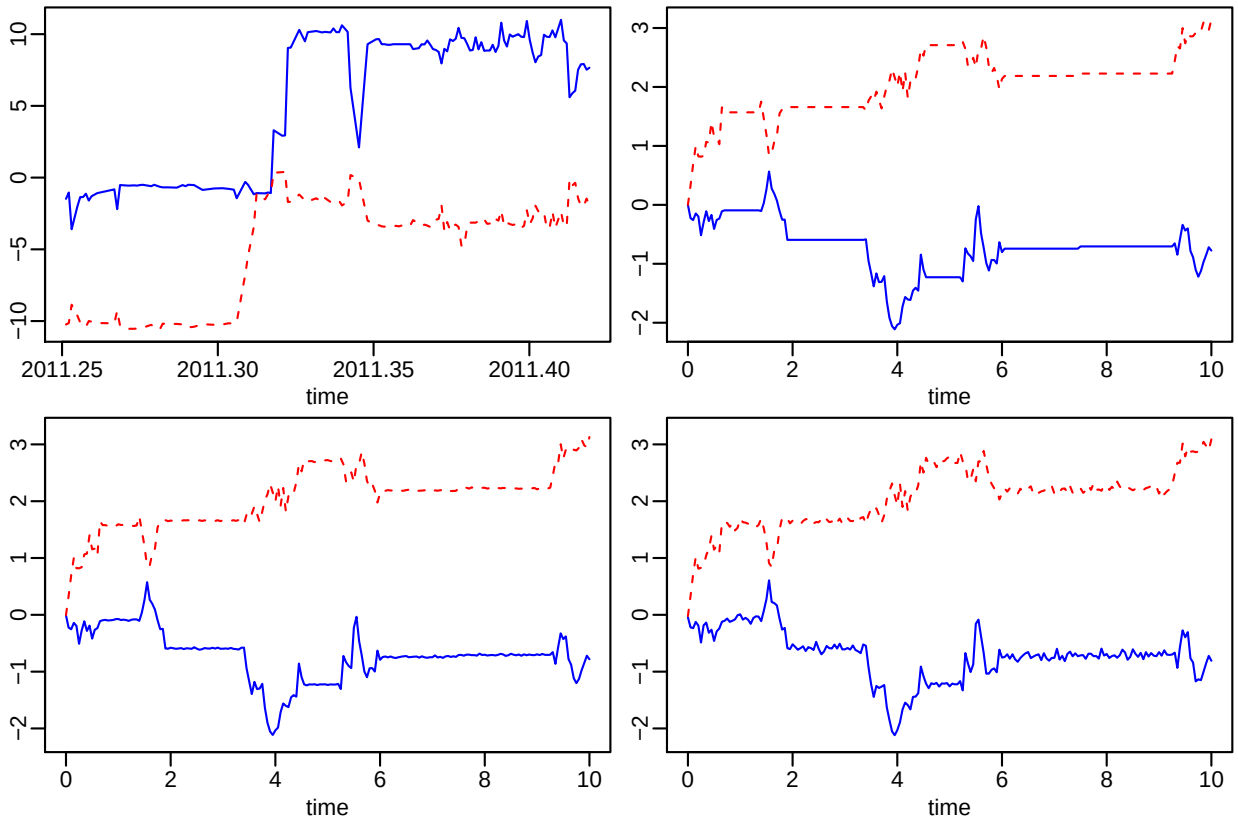

FIGURE 1. *Upper left*: Actual coordinates of a female mountain lion in a two-month period in 2012 in the Gros Ventre mountain range, Wyoming, with most observations separated by 8 hours. The y-axis is departure from the starting point. The solid blue line is UTM eastings (km) and the dashed red line is UTM northings (km). *Upper right*: Coordinates of a realization from a two-dimensional MR process. The two coordinates are dependent because the straight line segments representing resting periods are shared. *Bottom left*: Coordinates of the same realization as in upper right panel after adding Gaussian noise with standard deviation 0.01 km. *Bottom right*: Coordinates of the same realization as in upper right panel after adding Gaussian noise with standard deviation 0.05km.

2.2. **Moving-Resting Process with Measurement Error.** Suppose that the observations are recorded at times $t_0 = 0, t_1, \ldots, t_n$. Let $\{\epsilon_k\}_{k=0,\ldots,n}$ be independent and identically normally distributed random variables with mean 0 and variance σ_ϵ^2. A MR process with measurement error (MRME) $Z(t_k)$, $k = 0, 1, \ldots, n$, is the superimposition of a measurement error and the exact location:

$$Z(t_k) = X(t_k) + \epsilon_k, \tag{2}$$

where $X(\cdot)$ is a MR process in Equation (1), and ϵ_k's are independent $\mathrm{N}(0, \sigma_\epsilon^2)$ noises.

Let us mention here some important properties of the process $\{Z(t_k)\}_{k=0,\ldots,n}$. It is not Markov. Neither is the location-state process $\{Z(t_k), S(t_k)\}_{k=0,\ldots,n}$. To attain a Markov process, one might consider including the measurement errors. It is true that $\{Z(t_k), S(t_k), \epsilon_k\}_{k=0,\ldots,n}$ is Markov, but the cardinality of hidden states $(S(t_k), \epsilon_k)$ is a continuum in this case, which

TABLE 1. Summaries of the influence of measurement error on MR process parameter estimation based on 100 replicates. The true parameters of the MR process were $\lambda_1 = 1$ hour^{-1}, $\lambda_0 = 0.5$ hour^{-1}, $\sigma = 1$ km/hour$^{1/2}$. The measurement error was set as Gaussian noise with standard deviation (s.d.) 0.05 and 0.01. In each replicate, data were generated on a time horizon of 500 hours with sampling interval 1 hour. The mean and standard deviation of the point estimates, along with the convergence percentage of the optimizations under different setups are reported.

Gaussian noise	Rounding	Convergence	$\hat{\lambda}_1$		$\hat{\lambda}_0$		$\hat{\sigma}$	
s.d. (km)	(km)	percentage (%)	mean	s.d.	mean	s.d.	mean	s.d
—	—	100	1.00	0.17	0.50	0.05	1.00	0.05
0.05	—	7	348.08	121.53	2.82	0.08	5.42	0.98
	0.01	6	304.83	127.91	2.80	0.08	5.05	1.10
	0.05	14	240.54	135.93	2.54	0.11	4.73	1.34
	0.10	33	128.79	91.36	2.06	0.12	3.94	1.36
0.01	—	5	409.34	148.78	2.42	0.07	5.90	1.07
	0.01	10	346.62	165.24	2.25	0.09	5.64	1.46
	0.05	100	3.65	4.26	1.01	0.16	1.06	0.24
	0.10	99	1.33	0.28	0.67	0.07	0.95	0.05

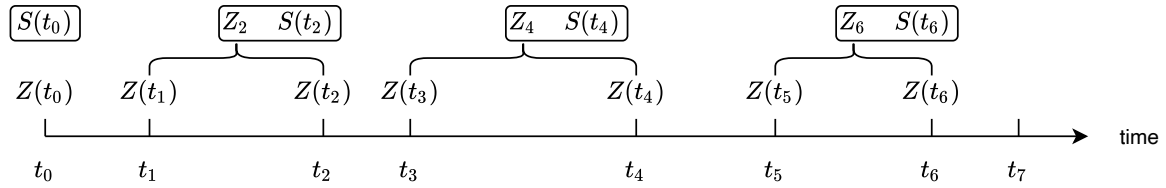

FIGURE 2. Increment-state sequence at even numbered time points are Markov.

makes the dynamic programming approach for computing likelihood infeasible. To address this difficulty we consider

$$\{Z(t_{2k}) - Z(t_{2k-1}), S(t_{2k})\}_{k=1,\ldots,\lfloor n/2 \rfloor}, \tag{3}$$

with initial state $S(t_0)$, where $\lfloor \cdot \rfloor$ is the floor function. This process (3) is Markov, because the increment of the Brownian motion between times t_{2k-1} and t_{2k} and measurement errors ϵ_{2k-1} and ϵ_{2k} are independent of observations collected by time t_{2k-2}; see Figure 2. Moreover, the process has a finite set of hidden states. This is an important property which we use to develop a forward algorithm for efficient computing of a composite likelihood in the next section. The conditional distribution of $(Z(t_{2k}) - Z(t_{2k-1}), S(t_{2k}))$ (given the observations up to time t_{2k-2}) depends only on state $S(t_{2k-2})$.

Because the increments of $\{Z(t_k)\}_{k=0,\ldots,n}$ are stationary, it is sufficient to derive the probability density function of $(Z(t) - Z(u), S(t))$ given $S(0)$ for any $0 < u < t$. This density can be computed in two steps. First, we need to find the conditional probability that the

animal is in state j at time u given that at time 0 it was in state i, that is,

$$\tau_{ij}(u) = P_i(S(u) = j),$$

where $P_i(\cdot) = \Pr(\cdot|S(0) = i)$.

Then we compute the joint distribution of the movement increment between times t and u and the state at time t, $(Z(t)-Z(u), S(t))$, given the state at time u, $S(u)$. Again, because of the stationarity, this distribution is the same as the distribution of $(Z(t-u)-Z(0), S(t-u))$ given $S(0)$. That is, the key (defective) density is

$$g_{ij}(z,t) = \frac{\partial}{\partial z} P_i(Z(t) - Z(0) \leq z, S(t) = j),$$

where $i, j = \{1, 0\}$ and z is a real number.

From the Markov property of the location-state process $(X(t), S(t))$ and the independence of the added noise, the transition probability density function of $(Z(t) - Z(u), S(t))$ given $S(0)$ is

$$f(Z(t) - Z(u), S(t) = j \mid S(0) = i) = \sum_{l=0}^{1} \tau_{il}(u) g_{lj}(Z(t) - Z(u), t - u), \tag{4}$$

where $0 < u < t$ and $i, j \in \{0, 1\}$.

The formulas for $\tau_{ij}(t)$ and $g_{ij}(z,t)$ are presented in Appendix B. To ease the notation here we provide the description of one-dimensional MR process and one-dimensional MR process with measurement error. These formulas can be generalized to a multi-dimensional set up. The corresponding formulas are given in Appendix C. All our simulations and data analysis are performed using two-dimensional formulas.

2.3. **Composite Likelihood Estimation.** Since the full likelihood is unavailable, we resort to composite likelihood to estimate the parameters (Lindsay, 1988). A composite likelihood is a weighted product of likelihood segments

$$\mathrm{CL} = \prod_{k=1}^{K} L_k^{w_k},$$

where L_k is the true likelihood of the k-th data segment with a non-negative weight w_k, $k = 1, \ldots, K$, and K is the number of segments depending on the construction of the CL. The weights are useful, for example, in pairwise likelihood when some pairs with stronger dependence contribute more than other pairs. Here we employ a composite likelihood with equal, unit weights. Suppose that the location-state vectors are denoted as

$$\mathbf{Z} = (Z(t_0), Z(t_1), \ldots, Z(t_n))$$
$$\mathbf{S} = (S(t_0), S(t_1), \ldots, S(t_n)).$$

The observed data only contain $\mathbf{Z}$. We propose two ways to construct the composite likelihood.

2.3.1. *Two-Piece Composite Likelihood.* First, let us introduce two random vectors of observed increments and hidden states at even-numbered time points:

$$\mathbf{Z}_{even} = \left(Z_2, \ldots, Z_{2\lfloor n/2 \rfloor}\right),$$
$$\mathbf{S}_{even} = \left(S(t_0), S(t_2), \ldots, S(t_{2\lfloor n/2 \rfloor})\right),$$

where $Z_k = Z(t_k) - Z(t_{k-1})$. Then, because of the Markov property of sequence $\{Z_{2k}, S(t_{2k})\}$, the likelihood of increment-state observations is given by

$$L\left(\mathbf{Z}_{even}, \mathbf{S}_{even}; \boldsymbol{\theta}\right) = \nu(S(t_0)) \prod_{k=1}^{\lfloor n/2 \rfloor} f\left(Z_{2k}, S(t_{2k}) | S(t_{2k-2})\right),$$

in which $\boldsymbol{\theta} = (\lambda_1, \lambda_0, \sigma, \sigma_\epsilon)$ and $\nu(S(t_0))$ is the initial distribution of the state process $S(t)$ that is assumed to be stationary. The construction of the Markov Chain that corresponds to the even-numbered time points is illustrated by Figure 2.

Since $\mathbf{S}_{even}$ is not observed, we need to marginalize over all possible state trajectories:

$$L(\mathbf{Z}_{even}; \boldsymbol{\theta}) = \sum_{S(t_0), S(t_2), \ldots, S(t_{2\lfloor n/2 \rfloor}) \in \{0,1\}} L\left(\mathbf{Z}_{even}, \mathbf{S}_{even}; \boldsymbol{\theta}\right).$$

The cardinality of the set of the state trajectories is $2^{\lfloor n/2 \rfloor + 1}$, which makes the direct summation infeasible for even moderate n. It can, however, be tackled with the help of dynamic programming, specifically, by the normalized forward algorithm. A detailed description of the algorithm is given in Appendix D.

In a similar fashion, one can introduce the likelihood $L(\mathbf{Z}_{odd}; \boldsymbol{\theta})$ of the observed increments at the odd time points $\mathbf{Z}_{odd} = \left(Z_1, Z_3, \ldots, Z_{2\lfloor (n+1)/2 \rfloor - 1}\right)$. Adding two log-likelihoods together we get the following composite log-likelihood:

$$\mathrm{CL}(Z(t_0), \ldots, Z(t_n); \boldsymbol{\theta}) = \log L(\mathbf{Z}_{even}; \boldsymbol{\theta}) + \log L(\mathbf{Z}_{odd}; \boldsymbol{\theta}). \tag{5}$$

Each piece in (5) is a true log likelihood for about one half of the observations. The maximum composite likelihood estimator (MCLE) of $\boldsymbol{\theta}$ is the maximizer $\hat{\boldsymbol{\theta}}$ of (5).

2.3.2. *Marginal Composite Likelihood.* The second approach is to use the one-step transition density with the dependence between two consecutive increments ignored. If $\mathbf{S}$ were observed, for $i, j \in \{0, 1\}$, the likelihood of each pair of consecutive location-state observations

$$(\{Z(t_{k-1}), S(t_{k-1}) = i\}, \{Z(t_k), S(t_k) = j\})$$

is

$$\nu(i) g_{ij}(Z(t_k) - Z(t_{k-1}), t_k - t_{k-1}),$$

where $\nu(\cdot)$ is the stationary distribution of state process $\{S(t)\}_{t \geq 0}$. Since $\mathbf{S}$ is assumed to be unobserved, the likelihood of $(Z(t_{k-1}), Z(t_k))$ is

$$\sum_{j=0}^{1} \sum_{i=0}^{1} \nu(i) g_{ij}(Z(t_k) - Z(t_{k-1}), t_k - t_{k-1}).$$

The marginal composite log-likelihood is

$$\mathrm{CL}^*((Z(t_0), \ldots, Z(t_n)); \boldsymbol{\theta}) = \sum_{k=1}^{n} \log \left[\sum_{j=0}^{1} \sum_{i=0}^{1} \nu(i) g_{ij}(Z(t_k) - Z(t_{k-1}), t_k - t_{k-1}) \right].$$

Since the dependence among the increments is discarded, the resulting estimator is expected to be less efficient if the dependence is stronger.

Algorithm 1 Estimating standard error from parametric bootstrap.

input: Observed data; number of resampling M.
· Fit model to get the parameter estimates;
for $m = 1$ to M **do**
  · Generate a bootstrap sample on the observed time grids with the fitted parameters;
  · Fit model to the bootstrap sample to get bootstrap estimate;
end for
· Make inference based on the empirical distribution of M bootstrap estimates.

2.3.3. *Variance Estimation.* To make inferences about $\boldsymbol{\theta}$, we need the variance of $\hat{\boldsymbol{\theta}}$. it can be estimated easily by parametric bootstrap with the time points fixed because simulating from the MRME process with fitted parameters is straightforward. The general approach of parametric bootstrap is given as Algorithm 1.

Alternatively, we can estimate the variance by inverting the observed Godambe information matrix (Godambe, 1960). This approach, however, did not perform as well as the parametric bootstrap approach in our numerical studies.

3. SIMULATION AND REAL DATA APPLICATION

3.1. **Simulation Studies.** We ran three simulation studies to check the performance of the MCLE based on both the marginal composite likelihood and the two-piece composite likelihood. The objective of the studies were threefold: (1) to see if the procedures successfully recovered the model parameters, (2) to verify that standard errors could be obtained with the help of parametric bootstrap, and (3) to compare the performance of the marginal method to the two-piece one.

In Study 1, we generated movement data using the MRME model described by equation (2). The model parameters were set to be $\lambda_1 = 1$, $\lambda_0 = 0.5$, $\sigma = 1$, and $\sigma_\epsilon \in \{0.01, 0.05\}$. This is the same setup that was used for simulations in Section 2.1. For each configuration, we generated 200 two-dimensional datasets on a time grid from 0 to 1000, with sampling interval 5. The resulting data has length $n = 200$. Figure 3 presents the violin plots of the MCLE in Study 1 of the 200 replicates in comparison to the true values of the four parameters. Violin plots are similar to box plots with a rotated kernel density plot on each side. The horizontal bars in the panels are the true parameter values. For each parameter, the true value lies in the bulk part of the violin plot. This indicates that the true parameters are recovered well by both MCLE methods. The left and right panels represent different sized additional measurement errors. The variation of the estimates in the case of $\sigma_\epsilon = 0.01$ is smaller than that in the case of $\sigma_\epsilon = 0.05$, which is expected.

The second simulation study addresses the estimation of standard errors for both MCLE procedures via parametric bootstrap. The sampling horizon (the length of observation window) was set to two levels, 200 and 500 time units. The sampling intervals (the inverse sampling frequency) also had two levels, 1 and 5 time units. The parameters of MRME process were: $\lambda_1 = 1$, $\lambda_0 = 0.5$, $\sigma = 1$, and $\sigma_\epsilon = 0.01$. Table 2 (upper block) summarizes the results. Once again we can see that both the marginal method and the two-piece method recover the true parameters well. Their empirical standard errors are similar, suggesting that the two methods have comparable efficiency for these setups. Moreover, the standard

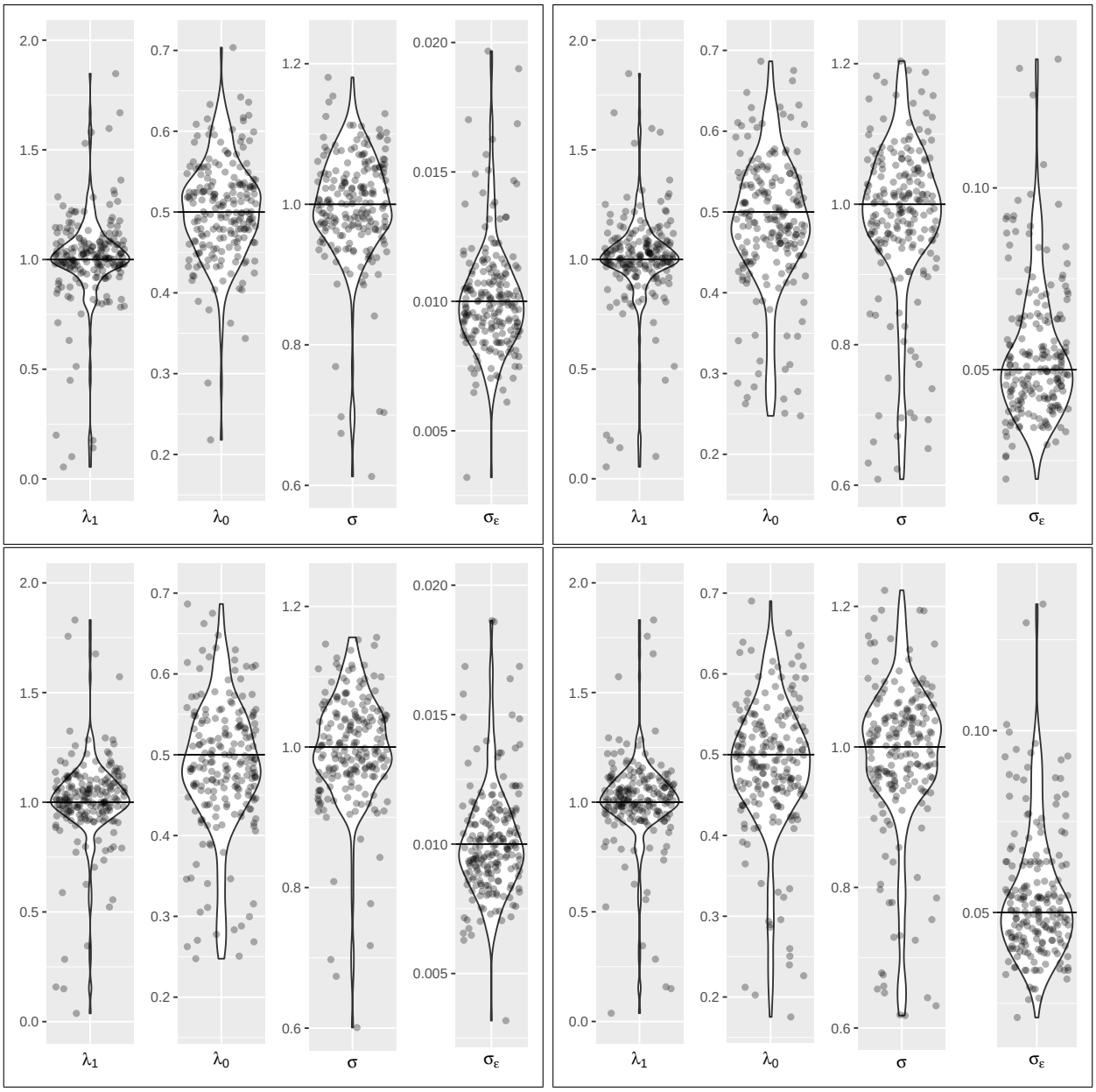

FIGURE 3. Violin plots of the MCLE in Study 1 with two-piece method (*top*) and marginal method (*bottom*) from 200 replicates. The horizontal bar in each panel is the true parameter value $\lambda_1 = 1$, $\lambda_0 = 0.5$, $\sigma = 1$, and $\sigma_\epsilon = 0.01$ (*left*) and 0.05 (*right*).

errors were estimated by the parametric bootstrap procedure with 50 replications. The estimated standard errors are reasonably close to empirical ones. The coverage rates of the 95% bootstrap confidence intervals are as low as 81% for λ_1 for the case with sampling interval 5 and sampling horizon 200. As the sampling interval decreases and the sampling horizon increases, the coverage rates get reasonably close to the nominal level.

Let us make a few remarks on the influence of sampling horizon and sampling interval on the efficiency of estimation. When the sampling interval is held fixed but the sampling horizon is 2.5 times longer, the ESE (empirical standard error) seems to be $\sqrt{2.5}(= 1.58)$ times smaller for most parameter estimates. The longer sampling horizon covers more moving-resting cycles $(M_j + R_j)$, and provides more information on both mobility and measurement error parameters. If we fix the sampling horizon and increase sampling frequency by reducing the sampling interval, however, the ESE of only $\hat{\sigma}$ reduces in proportion to the square root of the number of observations. Theoretically, if one can take observations nearly continuously, the mobility parameter σ can be estimated with absolute accuracy. Increasing sampling frequency also improves the estimation of λs to a certain degree, but improves the estimation of σ_ϵ drastically. The results show the difference between the domain expansion asymptotics (e.g., increasing sampling horizon) and the in-fill asymptotics (e.g., increasing sampling frequency).

Finally, let us note that the performance of the two methods is similar in both simulation studies described above. That was surprising because the marginal method basically ignores the dependence of the MRME process and treats increments as if they are independent. One possible explanation is that when the distance between two consecutive observations is relatively long, then the dependence between two consecutive increments of the MRME process is weaker. To illustrate this point, one can calculate the correlation of absolute values of consecutive increments via simulation by employing the auto-correlation function with lag 1 (ACF(1)) and lag 2 (ACF(2)). For example, for the same parameter set as in the above simulations and a long time horizon 100,000, both ACF(1) and ACF(2) for the sampling interval 5 are very close to 0 from a Monte Carlo study. For the sampling interval 0.1, however, they are 0.46 and 0.40, respectively. Our third simulation study was based on this design. We also considered sampling interval 0.8 in the simulation because it has a relatively large ACF(1), 0.23, and a significantly smaller ACF(2), 0.07. The results of Study 3 with these small sampling intervals are presented in Table 2 (lower block). It is clear that the two-piece procedure is preferable for datasets with shorter sampling intervals (more frequent observations).

3.2. **Sensitivity Analysis.** Simulation was conducted to explore the sensitivity of MRME model to misspecification. The MRME model assumes that the animal is strictly motionless during resting period. Here we simulate data points using two alternative approaches to strict stillness in the resting/encamping state: (1) a BM with a much smaller infinitesimal variance parameter; and (2) an OU process. The first process (with two alternating BMs) mathematically is very similar to the MR process, and we will call it a *Moving-Moving (MM) process*. In our sensitivity analysis, the BM model for the movement during resting period was set to have infinitesimal variance $\sigma_r \in \{0.01, 0.05\}$. The OU process is characterized by a stochastic differential equation

$$\mathrm{d}G(t) = \alpha(b - G(t))\mathrm{d}t + \sigma_{\mathrm{OU}}\mathrm{d}B(t),$$

where α and σ_{OU} are positive parameters. In our sensitivity analysis, we set $\alpha = 0.5$, $\sigma_{\mathrm{OU}} \in \{0.01, 0.05\}$. The sampling horizon, sampling interval, and the MRME parameters that are shared by all three models, $(\lambda_1, \lambda_0, \sigma, \sigma_\epsilon)$, were kept the same as those in Study 2. For each configuration, 200 replicates were generated.

Table 3 summarizes the means and standard errors of the MRME parameters from the 200 replicates. For $\lambda_1, \lambda_0, \sigma$, the performance of point estimators (including standard error)

TABLE 2. Summaries of Studies 2 and 3: average estimator (EST), empirical standard error (ESE), average parametric bootstrap standard error (ASE), and coverage rate (CR) of 95% large-sample confidence interval of MCLE with the two-piece method and the marginal method. The number of replications is 200.

Sampling horizon	Sampling interval	Parameter	True value	Two-piece method				Marginal method			
				EST	ESE	ASE	CR	EST	ESE	ASE	CR
				Study 2							
200	5	λ_1	1.0	0.961	0.546	0.508	0.80	0.982	0.570	0.516	0.81
		λ_0	0.5	0.493	0.169	0.240	0.93	0.485	0.211	0.251	0.89
		σ	1.0	0.966	0.189	0.163	0.80	0.970	0.194	0.166	0.82
		$\sigma_\epsilon(\times 10^{-2})$	1.0	1.145	0.710	0.651	0.91	1.136	0.665	0.648	0.92
	1	λ_1	1.0	1.104	0.394	0.386	0.94	1.057	0.410	0.414	0.88
		λ_0	0.5	0.502	0.093	0.092	0.96	0.488	0.109	0.109	0.92
		σ	1.0	1.011	0.084	0.088	0.93	1.001	0.089	0.097	0.92
		$\sigma_\epsilon(\times 10^{-2})$	1.0	1.002	0.070	0.068	0.94	1.002	0.069	0.068	0.93
500	5	λ_1	1.0	1.020	0.362	0.342	0.92	1.009	0.354	0.335	0.91
		λ_0	0.5	0.512	0.101	0.111	0.95	0.509	0.106	0.115	0.95
		σ	1.0	0.982	0.122	0.114	0.91	0.978	0.119	0.112	0.92
		$\sigma_\epsilon(\times 10^{-2})$	1.0	1.046	0.356	0.376	0.92	1.045	0.362	0.357	0.90
	1	λ_1	1.0	1.036	0.240	0.224	0.93	0.983	0.224	0.223	0.88
		λ_0	0.5	0.508	0.060	0.058	0.96	0.495	0.064	0.062	0.92
		σ	1.0	1.008	0.060	0.055	0.92	0.997	0.067	0.058	0.90
		$\sigma_\epsilon(\times 10^{-2})$	1.0	0.998	0.044	0.042	0.94	0.998	0.044	0.042	0.93
				Study 3							
160	0.8	λ_1	1.0	1.139	0.440			1.069	0.549		
		λ_0	0.1	0.106	0.030			0.096	0.039		
		σ	1.0	1.013	0.119			0.991	0.140		
		$\sigma_\epsilon(\times 10^{-2})$	1.0	0.998	0.045			0.998	0.045		
200	0.1	λ_1	1.0	1.031	0.238			1.132	0.329		
		λ_0	0.1	0.102	0.024			0.109	0.026		
		σ	1.0	1.002	0.040			1.007	0.043		
		$\sigma_\epsilon(\times 10^{-2})$	1.0	0.999	0.015			0.999	0.015		

is robust even if the fitted model is misspecified. The estimation of σ_ϵ is slightly inflated because of the extra variation during resting periods. However, its standard error is not affected by the misspecification. In conclusion, the MRME estimation procedure provides accurate estimates for animal mobility parameter σ and average moving and resting times even if the resting data point are generated from the two models that may be considered as alternatives to the MRME model.

TABLE 3. Summaries of Sensitivity Analysis: average estimator (EST), empirical standard error (ESE) of MCLE with the two-piece method and the marginal method. Fitted data are randomly generated from MRME where resting state is embedded with a BM or OU process. The number of replications is 200.

				Brownian motion resting				*OU process resting*			
Sampling	Sampling	Parameter	True	Two-piece		Marginal		Two-piece		Marginal	
horizon	interval		value	EST	ESE	EST	ESE	EST	ESE	EST	ESE
				$\sigma_r = 0.01$				$\alpha = 0.5$, $\sigma_{OU} = 0.01$			
200	5	λ_1	1.0	1.053	0.595	0.994	0.618	1.066	0.543	1.033	0.554
		λ_0	0.5	0.536	0.228	0.517	0.246	0.525	0.212	0.507	0.222
		σ	1.0	0.975	0.203	0.952	0.205	0.980	0.193	0.969	0.197
		$\sigma_\epsilon(\times 10^{-2})$	1.0	1.760	0.721	1.771	0.764	1.406	0.662	1.391	0.612
	1	λ_1	1.0	1.127	0.386	1.086	0.434	1.081	0.379	1.033	0.359
		λ_0	0.5	0.513	0.096	0.497	0.116	0.512	0.093	0.501	0.102
		σ	1.0	1.017	0.087	1.007	0.098	1.011	0.093	1.000	0.085
		$\sigma_\epsilon(\times 10^{-2})$	1.0	1.221	0.077	1.221	0.077	1.162	0.088	1.161	0.089
500	5	λ_1	1.0	1.056	0.398	1.072	0.377	1.033	0.333	1.034	0.330
		λ_0	0.5	0.498	0.103	0.505	0.104	0.507	0.096	0.505	0.100
		σ	1.0	0.998	0.136	1.002	0.130	1.001	0.114	1.001	0.117
		$\sigma_\epsilon(\times 10^{-2})$	1.0	1.812	0.487	1.810	0.481	1.374	0.410	1.370	0.392
	1	λ_1	1.0	1.010	0.224	0.976	0.215	1.018	0.221	0.977	0.216
		λ_0	0.5	0.496	0.058	0.488	0.066	0.499	0.057	0.490	0.066
		σ	1.0	0.999	0.057	0.993	0.061	1.002	0.054	0.993	0.058
		$\sigma_\epsilon(\times 10^{-2})$	1.0	1.225	0.052	1.226	0.052	1.168	0.051	1.167	0.050
				$\sigma_r = 0.05$				$\alpha = 0.5$, $\sigma_{OU} = 0.05$			
200	5	λ_1	1.0	1.372	0.526	1.392	0.538	1.264	0.606	1.266	0.624
		λ_0	0.5	0.733	0.264	0.733	0.279	0.612	0.276	0.606	0.298
		σ	1.0	0.967	0.178	0.970	0.179	0.989	0.172	0.985	0.188
		$\sigma_\epsilon(\times 10^{-2})$	1.0	2.446	1.361	2.442	1.405	2.572	1.084	2.627	1.150
	1	λ_1	1.0	1.295	0.572	1.259	0.641	1.226	0.522	1.230	0.645
		λ_0	0.5	0.537	0.118	0.516	0.159	0.536	0.118	0.512	0.159
		σ	1.0	1.048	0.116	1.037	0.133	1.032	0.103	1.029	0.135
		$\sigma_\epsilon(\times 10^{-2})$	1.0	3.690	0.236	3.699	0.240	3.204	0.216	3.208	0.214
500	5	λ_1	1.0	1.397	0.389	1.385	0.358	1.149	0.462	1.160	0.441
		λ_0	0.5	0.674	0.150	0.674	0.145	0.554	0.158	0.556	0.161
		σ	1.0	1.011	0.130	1.010	0.126	0.996	0.149	1.001	0.142
		$\sigma_\epsilon(\times 10^{-2})$	1.0	2.640	1.146	2.615	1.091	2.924	0.949	2.942	0.967
	1	λ_1	1.0	1.133	0.360	1.246	0.546	1.127	0.380	1.213	0.521
		λ_0	0.5	0.516	0.077	0.520	0.113	0.521	0.084	0.523	0.116
		σ	1.0	1.021	0.076	1.039	0.111	1.018	0.080	1.030	0.104
		$\sigma_\epsilon(\times 10^{-2})$	1.0	3.672	0.162	3.671	0.165	3.173	0.136	3.173	0.135

3.3. **Fitting Moving-Moving Process.** This subsection was motivated by an interesting question posed by a referee. For animal movement data generated from a MRME process, is it possible to accurately estimate the parameters of interest (λ_0, λ_1 and σ) by fitting a misspecified, yet simpler MM model? Such fits are attractive because the true likelihood of the MM model can be computed efficiently with the normalized forward algorithm as detailed in Pozdnyakov et al. (2019). In contrast, recall that the main difficulty with the MRME process is the continuum cardinality of the measurement error as a hidden state, which does not allow dynamic programming directly; instead, we have to resort to the composite likelihood approach. Adding random noise to a moving stage does not have a great impact on estimation (Pozdnyakov et al., 2014). So, if we model the measurement errors over the resting stage with a BM with a small infinitesimal variance, we still might hope to get reasonable estimates for λ_0, λ_1 and σ.

To verify this interesting idea, we generated data from the MRME process with the same parameters, sampling horizon and sampling interval setup as in Study 2 and fitted the MM model. Table 4 summarizes the results based on 200 replicates. This method performed quite well when the observations were close to each other in time. The estimates for both the λ's and σ that corresponded to the moving stage look unbiased, and the empirical standard errors are slightly lower than the ESE of the composite likelihood method. When the sampling intervals are longer, however, the mean duration of the moving stage ($1/\lambda_1$) is severely underestimated, exhibiting much higher standard errors than those from the MRME fit as tabulated in Table 2. These results are not surprising. The added measurement errors in the MRME model are stationary, and they have the same variance across the entire length of a resting period. The BM with a smaller infinitesimal variance is still a nonstationary motion that drifts away. Its variance grows linearly, and, as a result, a typical deviation is much smaller in the beginning of a long resting stage than in the end. When fitting the MM model to data generated from MRME processes with long sampling intervals, it is likely that some of the moving periods are mistaken as slow moving periods.

3.4. **Movement of a Mountain Lion.** The MRME model was applied to aforementioned GPS data collected on a mature female mountain lion living in the Gros Ventre Mountain Range near Jackson, Wyoming. The data were collected by a code-only GPS wildlife tracking collar from 2009 to 2012. The collar was programmed to record locations every 8 hours, but the actual sampling intervals were irregular. In the Grand Teton and Gros Ventre mountains of Wyoming, deep winter snows and prey migrations ensure that mountain-lion movements differ across seasons (Elbroch et al., 2013). So, we fitted a MRME model to the summer data (from June 1, 2012 to August 31, 2012) and winter data (from December 1, 2011 to February 29, 2012), respectively. These two periods of data were plotted as Figure 4. The summer data had an average sampling interval of 5.46 hours with standard deviation 5.14 hours, ranging from 0.5 hours to 32 hours. The average sampling interval was 5.58 hours in the winter data, with standard deviation 4.09 hours and range from 0.5 hours to 25 hours. The summer data has 401 observations and winter data has 392 observations. Table 5 summarizes the fitted MRME model parameters and their standard errors based on 100 bootstrap samples. The results from the marginal method are similar to those from the two-piece method except that the estimate of the rate parameter λ_1 of the moving state is noticeably smaller. Based on the comparison between the two methods in the simulation study, we discuss the results from the two-piece method.

TABLE 4. Summaries of fitting MM process to MRME data: average estimator (EST), empirical standard error (ESE). The number of replications is 200.

Sampling horizon	Sampling interval	Parameter	True value	Moving-Moving process EST	ESE
200	5	λ_1	1.0	1.630	3.039
		λ_0	0.5	0.463	0.208
		σ_1	1.0	1.082	0.655
		$\sigma_0(\times 10^{-2})$		2.922	8.671
	1	λ_1	1.0	1.040	0.368
		λ_0	0.5	0.512	0.077
		σ_1	1.0	1.004	0.076
		$\sigma_0(\times 10^{-2})$		1.404	0.093
500	5	λ_1	1.0	1.799	2.413
		λ_0	0.5	0.492	0.117
		σ_1	1.0	1.112	0.456
		$\sigma_0(\times 10^{-2})$		0.903	2.888
	1	λ_1	1.0	1.023	0.178
		λ_0	0.5	0.508	0.051
		σ_1	1.0	1.005	0.047
		$\sigma_0(\times 10^{-2})$		1.416	0.062

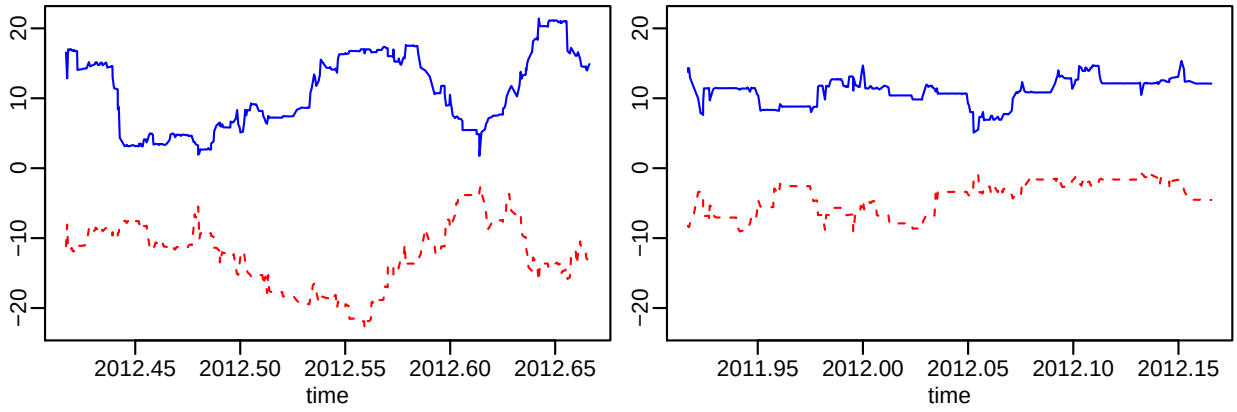

FIGURE 4. Actual coordinates of a female mountain lion in the Gros Ventre mountain range, Wyoming. The x-axis is time in years. The y-axis is departure from the starting point. The solid blue line is UTM easting (km) and the dashed red line is UTM northing (km). *Left:* Summer period data, from June 1, 2012 to August 31, 2012. *Right:* Winter period data, from December 1, 2011 to February 29, 2012.

TABLE 5. Analysis results for mountain lion movement data. Point estimates (EST) of MRME model from both two-piece method and marginal method and EST of Moving-Moving process are reported. Standard error of point estimates are evaluated by parametric bootstrap (SE).

Season	Parameter	MRME				MM Model	
		Two-piece		Marginal			
		EST	SE	EST	SE	EST	SE
Summer	λ_1	2.841	0.459	1.090	0.280	1.205	0.450
	λ_0	0.179	0.014	0.158	0.015	0.141	0.014
	σ	1.335	0.106	0.999	0.104	1.015	0.127
	$\sigma_\epsilon(\times 10^2)$	1.854	0.087	1.879	0.078		
	$\sigma_0(\times 10^2)$					1.707	0.070
Winter	λ_1	6.225	0.825	4.720	0.711	5.280	2.306
	λ_0	0.118	0.010	0.114	0.009	0.112	0.010
	σ	1.506	0.095	1.454	0.089	1.448	0.231
	$\sigma_\epsilon(\times 10^2)$	0.908	0.036	0.934	0.043		
	$\sigma_0(\times 10^2)$					0.643	0.023

The estimates based on two-piece composite likelihood for summer data are $\hat{\lambda}_1 = 2.841$ hour^{-1}, $\hat{\lambda}_0 = 0.179$ hour^{-1}, $\hat{\sigma} = 1.335$ km/hour$^{1/2}$ and $\hat{\sigma_\epsilon} = 0.019$ km. On average, this lion was moving 0.352 hours for each 5.587 hours resting, and, if she kept moving for (exactly) one full hour, the average deviation from the initial position is 1.335 km in both directions (northing and easting). Compared to the summer data, the estimates for winter data are $\hat{\lambda}_1 = 6.225$ hour^{-1}, $\hat{\lambda}_0 = 0.118$ hour^{-1}, $\hat{\sigma} = 1.506$ km/hour$^{1/2}$ and $\hat{\sigma_\epsilon} = 0.009$ km, so, during the winter period, she spent 51.7% more time staying in place and 54.4% less time moving. This pattern reveals the general tendency for mountain lions to eat larger prey requiring longer handling times in winter than in summer in temperate systems (Knopff et al., 2010; Elbroch et al., 2013).

The estimate of σ_ϵ (standard deviation of Gaussian noise) indicates that the GPS tracking collar had about 10- to 20-meters of measurement error, and the error is twice as high in summer as in winter, which is consistent with the report by Owari et al. (2009). However, the variability observed by Owari et al. (2009) was primarily due to sky obstruction from broadleaf tree canopy, which cannot be the case for the conifer-dominated mountains in Wyoming. Instead, these differences in GPS error likely reflect the fact that mountain lions in the study area select thicker vegetation with shade and cover in which to bed in the summer and more open terrain with southern aspects in rugged terrain, which catch sun and provide thermoregulatory benefits in the winter (Kusler et al., 2017).

We also fit the MM process to the lion data with estimated parameters and standard errors from parametric bootstrapping reported in Table 5. The estimates of the shared parameters (λ_1, λ_0, and σ_1) from the MM fit are closer to the estimates based on the marginal composite likelihood than to those based on the two-piece composite likelihood. The standard errors, however, are larger than those from the MRME fit. Especially for the estimate of λ_1, the standard error from the MM fit is 0.02306, which is more than three times the standard error

from the marginal likelihood MRME fit (0.00711). This is the result from one data set, but it does echo the findings from the simulation study in Section 3.3.

4. DISCUSSION

Inactive periods and measurement errors are both necessary features of animal movement models, if we are to successfully reconstruct natural behaviors. Handling measurement errors in the MR process is especially critical because, if discarded, a microscopic amount of measurement error causes substantial bias in estimation. The MRME model incorporates measurement errors into the MR model for the first time. Because the full likelihood is intractable, two composite likelihood methods that are computationally feasible were used for estimating the parameters. The resulting estimators were shown to recover the truth in simulation studies, with uncertainty estimated reasonably well by parametric bootstrap. The model and methods form a notable step forward for animal movement modeling in practice.

For movement data from predators that are known to have long inactive periods (Jeschke, 2007), the MRME model has great potential for revealing insights for animal ecologists. For example, accurately differentiating resting from moving will aid researchers in refining energetic models with varied applications from determining kill rates (Laundré, 2005) to carnivore carrying capacity (Carbone and Gittleman, 2002), refining models estimating animal abundance dependent upon animal speed (Moeller et al., 2018), and in refining models that depend upon step length and direction (e.g. tortuosity) to indicate habitat preference, intraspecific interactions, and competition (Broekhuis et al., 2019). Model outputs will also support the interpretation of animal behavior at locations where they are not moving (generally called "GPS clusters" (Cristescu et al., 2015)), such as determining the locations of bed sites vital to the fitness of diverse species (Lima et al., 2005).

The MRME model can be extended to meet further practical needs. For some predators, an inactive period may have different purposes, such as resting, mitigating predation risk or food handling, and these behaviors are currently nearly inseparable without the addition of accelerometer data (Petroelje et al., 2020). Pozdnyakov et al. (2020) introduced the moving-resting-handling (MRH) process that allows two different types of motionless states. The moving period may also represent different behaviors (Benhamou, 2011; Kranstauber et al., 2012), and this can be accommodated by allowing the infinitesimal variance parameter to have multiple levels. More generally, animal behavior depends on a suite of intrinsic and extrinsic variables. Our MRME model provides an important benchmark for building these more realistic extensions.

Authors' Contributions

TM, VP, and JY conceived the model; CH, VP, and JY designed the inference methodology and led the project development; CH implemented the method, conducted the simulation studies and data analyses, and drafted the first version of the manuscript; LME collected the data; TM and LME provided ecological insights and interpreted the results; all authors contributed critically to the drafts and gave final approval for publication.

Data Availability

The mountain lion data is publicly available in the R package `smam`.

References

C. Andrieu, A. Doucet and R. Holenstein (2010). "Particle Markov chain Monte Carlo methods." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**, 269–342.

S. Benhamou (2011). "Dynamic approach to space and habitat use based on biased random bridges." *PLoS ONE* **6**, e14592.

D. R. Brillinger (2010). "Modeling spatial trajectories." In A. E. Gelfand, P. J. Diggle, M. Fuentes and P. Guttorp (eds.), *Handbook of Spatial Statistics*, pp. 463–476. Chapman & Hall/CRC Boca Raton, Florida, USA.

F. Broekhuis, E. K. Madsen, K. Keiwua and D. W. Macdonald (2019). "Using GPS collars to investigate the frequency and behavioural outcomes of intraspecific interactions among carnivores: A case study of male cheetahs in the Maasai Mara, Kenya." *PLoS ONE* **14**, e0213910.

F. Cagnacci, L. Boitani, R. A. Powell and M. S. Boyce (2010). "Animal ecology meets GPS-based radiotelemetry: A perfect storm of opportunities and challenges." *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 2157–2162.

O. Cappé, E. Moulines and T. Rydén (2005). *Inference in Hidden Markov Models*. Springer.

C. Carbone and J. L. Gittleman (2002). "A common rule for the scaling of carnivore density." *Science* **295**, 2273–2276.

B. Cristescu, G. B. Stenhouse and M. S. Boyce (2015). "Predicting multiple behaviors from gps radiocollar cluster data." *Behavioral Ecology* **26**, 452–464.

L. M. Elbroch, P. E. Lendrum, J. Newby, H. Quigley and D. Craighead (2013). "Seasonal foraging ecology of non-migratory cougars in a system with migrating prey." *PLOS ONE* **8**, e83375.

V. P. Godambe (1960). "An optimum property of regular maximum likelihood estimation." *Ann. Math. Statist.* **31**, 1208–1211.

E. Hanks, D. Johnson and M. Hooten (2017). "Reflected stochastic differential equation models for constrained animal movement." *Journal of Agricultural, Biological and Environmental Statistics* **22**, 353–372.

M. B. Hooten, D. S. Johnson, B. T. McClintock and J. M. Morales (2017). *Animal Movement: Statistical Models for Telemetry Data*. CRC Press.

J. S. Horne, E. O. Garton, S. M. Krone and S. Lewis, J (2007). "Analyzing animal movements using Brownian bridges." *Ecology* **88**, 2354–2363.

C. Hu, V. Pozdnyakov and J. Yan (2020a). "Density and distribution evaluation for convolution of independent gamma variables." *Computational Statistics* **35**, 327–342.

C. Hu, V. Pozdnyakov and J. Yan (2020b). *smam: Statistical Modeling of Animal Movements*. R package version 0.5.4. (https://CRAN.R-project.org/package=smam).

E. L. Ionides, A. Bhadra, Y. Atchadé and A. King (2011). "Iterated filtering." *The Annals of Statistics* **39**, 1776–1802.

E. L. Ionides, D. Nguyen, Y. Atchadé, S. Stoev and A. A. King (2015). "Inference for dynamic and latent variable models via iterated, perturbed Bayes maps." *Proceedings of the National Academy of Sciences* **112**, 719–724.

J. M. Jeschke (2007). "When carnivores are "full and lazy"." *Oecologia* **152**, 357–364.

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

A. A. King, D. Nguyen and E. L. Ionides (2016). "Statistical inference for partially observed Markov processes via the R package pomp." *Journal of Statistical Software* **69**, 1–43.

K. H. Knopff, A. A. Knopff, A. Kortello and M. S. Boyce (2010). "Cougar kill rate and prey composition in a multiprey system." *Journal of Wildlife Management* **74**, 1435–1447.

B. Kranstauber, R. Kays, S. D. LaPoint, M. Wikelski and K. Safi (2012). "A dynamic Brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement." *Journal of Animal Ecology* **81**, 738–746.

A. Kusler, L. M. Elbroch, H. Quigley and M. Grigione (2017). "Bed site selection by a subordinate predator: an example with the cougar (*Puma concolor*) in the greater yellowstone ecosystem." *PeerJ* **5**, e4010.

J. W. Laundré (2005). "Puma energetics: A recalculation." *Journal of Wildlife Management* **69**, 723–732.

S. L. Lima, N. C. Rattenborg, J. A. Lesku and C. J. Amlaner (2005). "Sleeping under the risk of predation." *Animal Behaviour* **70**, 723–736.

B. G. Lindsay (1988). "Composite likelihood methods." *Contemporary Mathematics* **80**, 221–239.

A. Mashanova, T. H. Oliver and V. A. Jansen (2010). "Evidence for intermittency and a truncated power law from highly resolved aphid movement data." *Journal of The Royal Society Interface* **7**, 199–208.

B. T. McClintock (2017). "Incorporating telemetry error into Hidden Markov Models of animal movement using multiple imputation." *Journal of Agricultural, Biological and Environmental Statistics* **22**, 249–269.

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

A. K. Moeller, P. M. Lukacs and J. S. Horne (2018). "Three novel methods to estimate abundance of unmarked animals using remote cameras." *Ecosphere* **9**, e02331.

J. A. Nelder and R. Mead (1965). "A simplex method for function minimization." *The Computer Journal* **7**, 308–313.

J. R. Norris (1998). *Markov Chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. Reprint of 1997 original.

T. Owari, H. Kasahara, N. Oikawa and S. Fukuoka (2009). "Seasonal variation of global positioning system (GPS) accuracy within the tokyo university forest in hokkaido." *Bulletin of Tokyo University Forests* **120**, 19–28.

A. Parton and P. Blackwell (2017). "Bayesian inference for multistate 'step and turn' animal movement in continuous time." *Journal of Agricultural, Biological and Environmental Statistics* **22**, 373–392.

T. Patterson, L. Thomas, C. Wilcox, O. Ovaskainen and J. Matthiopoulos (2008). "State-space models of individual animal movement." *Trends in Ecology and Evolution* **23**, 87–94.

T. A. Patterson, A. Parton, R. Langrock, P. G. Blackwell, L. Thomas and R. King (2017). "Statistical modelling of individual animal movement: An overview of key methods and a discussion of practical challenges." *Advances in Statistical Analysis* **101**, 399–438.

T. R. Petroelje, J. L. Belant, D. E. Beyer and N. J. Svoboda (2020). "Identification of carnivore kill sites is improved by verified accelerometer data." *Animal Biotelemetry* **8**, 1–10.

V. Pozdnyakov, L. Elbroch, A. Labarga, T. Meyer and J. Yan (2019). "Discretely observed Brownian motion governed by telegraph process: Estimation." *Methodology and Computing in Applied Probability* **21**, 907–920.

V. Pozdnyakov, L. M. Elbroch, C. Hu, T. Meyer and J. Yan (2020). "On estimation for brownian motion governed by telegraph process with multiple off states." *Methodology and Computing in Applied Probability* **22**, 1275—1291.

V. Pozdnyakov, T. H. Meyer, Y.-B. Wang and J. Yan (2014). "On modeling animal movements using Brownian motion with measurement error." *Ecology* **95**, 247–253.

H. K. Preisler, A. A. Ager, B. K. Johnson and J. G. Kie (2004). "Modeling animal movements using stochastic differential equations." *Environmetrics* **15**, 643–657.

H. Scharf, M. Hooten and D. Johnson (2017). "Imputation approaches for animal movement modeling." *Journal of Agricultural, Biological and Environmental Statistics* **22**, 335–352.

T. Ueno, N. Masuda, S. Kume and K. Kume (2012). "Dopamine modulates the rest period length without perturbation of its power law distribution in Drosophila melanogaster." *PLoS One* **7**, e32007.

G. E. Uhlenbeck and L. S. Ornstein (1930). "On the theory of the brownian motion." *Phys. Rev.* **36**, 823–841.

J. Yan, Y.-W. Chen, K. Lawrence-Apfel, I. Ortega, V. Pozdnyakov, S. Williams and T. Meyer (2014). "A moving-resting process with an embedded Brownian motion for animal movements." *Population Ecology* **56**, 401–415.

S. Zacks (2004). "Generalized integrated telegraph processes and the distribution of related stopping times." *Journal of Applied Probability* **41**, 497–507.

W. Zucchini, MacDonald, I. L. and R. Langrock (2016). *Hidden Markov Models for Time Series: An Introduction Using R*. Chapman and Hall/CRC, 2 edition.

Appendix A. Details of Moving-Resting Process

The MR process is a Brownian motion with an infinitesimal variance that is governed by an alternating renewal process with two different holding times. Let random variables $\{M_i\}_{i\geq 1}$ be independent exponential variables with rate λ_1, and $\{R_i\}_{i\geq 1}$ be independent exponential variables with rate λ_0. These are the holding times. There are two possible alternating sequences of the holding times, $(M_1, R_1, M_2, R_2, \dots)$ or $(R_1, M_1, R_2, M_2, \dots)$. Which one represents a particular realization depends on an initial distribution. A continuous-time state process, $S(t)$, $t \geq 0$, takes only two values, 0 and 1, and it is defined by the holding times. In particular, for sequence $(M_1, R_1, M_2, R_2, \dots)$, if there exists $k \geq 0$ such that

$$\sum_{j=1}^{k}(M_j + R_j) < t \text{ but } \sum_{j=1}^{k}(M_j + R_j) + M_{k+1} \geq t,$$

then $S(t) = 1$; otherwise, $S(t) = 0$. For sequence $(R_1, M_1, R_2, M_2, \dots)$, if there exists $k \geq 0$ such that

$$\sum_{j=1}^{k}(R_j + M_j) < t \text{ but } \sum_{j=1}^{k}(R_j + M_j) + R_{k+1} \geq t,$$

then $S(t) = 0$, otherwise, $S(t) = 1$. The state process is stationary, if the initial probability of $\{S(0) = 1\}$ is set as

$$p_1 = \frac{\lambda_0}{\lambda_0 + \lambda_1},$$

and the initial probability of $\{S(0) = 0\}$ is set as $p_0 = 1 - p_1$ (e.g., Norris, 1998, p.120).

Properties and inferences of the MR process (1) have been studied in Yan et al. (2014) and Pozdnyakov et al. (2019). A key element is the distribution of occupation times, that is, the total time spent in the moving state by time t

$$M(t) = \int_0^t S(s)\mathrm{d}s,$$

and the total time spent in the resting state $R(t) = t - M(t)$. Let $P_i(\cdot)$ be the conditional probability $\Pr(\cdot|S(0) = i)$. Zacks (2004) derived computationally efficient formulas for the following (defective) densities for $0 < w < t$:

$$\begin{aligned}
p_{11}(w,t)\mathrm{d}w &= P_1(M(t) \in \mathrm{d}w, S(t) = 1),\\
p_{10}(w,t)\mathrm{d}w &= P_1(M(t) \in \mathrm{d}w, S(t) = 0),\\
p_{01}(w,t)\mathrm{d}w &= P_0(R(t) \in \mathrm{d}w, S(t) = 1),\\
p_{00}(w,t)\mathrm{d}w &= P_0(R(t) \in \mathrm{d}w, S(t) = 0).
\end{aligned}$$

Having this at hand, one can derive the marginal distribution of the increment $X(t) - X(0)$. Here we derive for the one-dimensional case; higher dimensional cases work similarly. Without loss of generality, let $X(0)$ to be 0, and $X(t)$ becomes the increment from time 0 to time t. Then, the joint distribution of the increment $X(t)$ and $S(t)$, $t > 0$, is

$$\begin{aligned}
P_1(X(t) \in \mathrm{d}x, S(t) = 1) &= h_{11}(x,t)\mathrm{d}x,\\
P_1(X(t) \in \mathrm{d}x, S(t) = 0) &= h_{10}(x,t)\mathrm{d}x,\\
P_0(X(t) \in \mathrm{d}x, S(t) = 0) &= h_{00}(x,t)\mathrm{d}x + e^{-\lambda_0 t}\delta_0(x),\\
P_0(X(t) \in \mathrm{d}x, S(t) = 1) &= h_{01}(x,t)\mathrm{d}x,
\end{aligned}$$

where $\delta_0(x)$ is the delta function with an atom at 0, $x \in \mathbb{R}$, and $h_{ij}(x,t)$, $i,j \in \{0,1\}$, are functions derived in Yan et al. (2014):

$$\begin{aligned}
h_{11}(x,t) &= e^{-\lambda_1 t}\phi(x;\sigma^2 t) + \int_0^t \phi(x;\sigma^2 w)p_{11}(w,t)\mathrm{d}w,\\
h_{10}(x,t) &= \int_0^t \phi(x;\sigma^2 w)p_{10}(w,t)\mathrm{d}w,\\
h_{00}(x,t) &= \int_0^t \phi(x;\sigma^2 (t-w))p_{00}(w,t)\mathrm{d}w,\\
h_{01}(x,t) &= \int_0^t \phi(x;\sigma^2 (t-w))p_{01}(w,t)\mathrm{d}w.
\end{aligned}$$

with $\phi(\cdot;\sigma^2)$ being the density function of normal distribution $\mathrm{N}(0,\sigma^2)$. The marginal distribution of increment $X(t)$ can be obtained by summing out $S(t)$ and $S(0)$, which forms the basis of the composite likelihood in Yan et al. (2014). The full maximum likelihood estimation based on dynamic programming was developed in Pozdnyakov et al. (2019).

Appendix B. Details of Moving-Resting Process with Measurement Error

Again, here we present our derivations for a one-dimensional case. The real-world animal movement data sets are two-dimensional. The formulas from below can be generalized to a d-dimensional case; see Appendix C.

First, let us calculate the marginal distribution of $Z(t) - Z(0)$, the increment of one-dimensional $\{Z(v)\}_{v\geq 0}$ from time 0 to time t. Consider

$$\Delta Z(t) = Z(t) - Z(0) = X(t) - X(0) + \xi(t),$$

where $\xi(t) \sim \mathrm{N}(0, 2\sigma_\epsilon^2)$ is independent of process $(X(\cdot), S(\cdot))$. Note that the observed increment between time points t_{2k+1} and t_{2k+2} is given by

$$Z(t_{2k+2}) - Z(t_{2k+1}) = X(t_{2k+2}) - X(t_{2k+1}) + \epsilon_{2k+2} - \epsilon_{2k+1}.$$

Because of independence the difference of the corresponding measurement errors has the following normal distribution: $(\epsilon_{2k+2} - \epsilon_{2k+1}) \sim \mathrm{N}(0, 2\sigma_\epsilon^2)$. Without loss of generality, we assume that $X(0) = 0$. Define

$$g_{ij}(z,t) = \frac{\partial}{\partial z} P_i(\Delta Z(t) \leq z, S(t) = j),$$

where $i, j \in \{1, 0\}$. Then, by conditioning on $\xi(t)$ we get that

$$\begin{aligned} g_{11}(z,t)\mathrm{d}z =& P_1(\Delta Z(t) \in \mathrm{d}z, S(t) = 1) \\ =& \int_{\mathbb{R}} P_1(X(t) + \xi(t) \in \mathrm{d}z, S(t) = 1, \xi(t) \in \mathrm{d}x) \\ =& \int_{\mathbb{R}} P_1(X(t) \in \mathrm{d}z - x, S(t) = 1)\phi(x; 2\sigma_\epsilon^2)\mathrm{d}x \\ =& \int_{\mathbb{R}} h_{11}(z - x, t)\phi(x; 2\sigma_\epsilon^2)\mathrm{d}x\mathrm{d}z, \end{aligned}$$

where, as before, $\phi(\cdot; \sigma^2)$ is the density function of $\mathrm{N}(0, \sigma^2)$. Similarly, one can get that

$$\begin{aligned} g_{10}(z,t) &= \int_{\mathbb{R}} h_{10}(z - x, t)\phi(x; 2\sigma_\epsilon^2)\mathrm{d}x, \\ g_{01}(z,t) &= \int_{\mathbb{R}} h_{01}(z - x, t)\phi(x; 2\sigma_\epsilon^2)\mathrm{d}x, \\ g_{00}(z,t) &= \int_{\mathbb{R}} h_{00}(z - x, t)\phi(x; 2\sigma_\epsilon^2)\mathrm{d}x + e^{-\lambda_0 t}\phi(z; 2\sigma_\epsilon^2). \end{aligned}$$

Next, let us denote

$$\tau_{ij}(t) = P_i(S(t) = j).$$

One can show that

$$\begin{aligned} \tau_{01}(t) &= P_0(S(t) = 1) \\ &= \sum_{n=0}^{\infty} \left[P\left(\sum_{k=1}^{n+1} R_k + \sum_{k=1}^{n} M_k \leq t \right) - P\left(\sum_{k=1}^{n+1} R_k + \sum_{k=1}^{n+1} M_k \leq t \right) \right] \\ &= \sum_{n=0}^{\infty} H(t; n, \lambda_1, n+1, \lambda_0), \end{aligned}$$

where $\{M_i\}_{i\geq 1}$ and $\{R_i\}_{i\geq 1}$ are defined in Section 2.1, and a summation over an empty set is 0 and $H(t; a_1, b_1, a_2, b_2)$ is a special function involving convolutions of independent gamma variables studied by Hu et al. (2020a). Specifically, with $F(t; a_1, b_1, a_2, b_2)$ being the distribution function of the sum of two independent gamma variables with parameters (a_1, b_1) and (a_2, b_2), respectively, $H(t; a_1, b_1, a_2, b_2) = F(t; a_1, b_1, a_2, b_2) - F(t; a_1 + 1, b_1, a_2, b_2)$

can be computed efficiently (Hu et al., 2020a, Lemma 1). Using similar techniques, we obtain that

$$\tau_{10}(t) = \sum_{n=0}^{\infty} H(t; n, \lambda_0, n+1, \lambda_1),$$
$$\tau_{00}(t) = \sum_{n=0}^{\infty} H(t; n, \lambda_0, n, \lambda_1),$$
$$\tau_{11}(t) = \sum_{n=0}^{\infty} H(t; n, \lambda_1, n, \lambda_0).$$

Appendix C. Formulas for High Dimensions

Let $\mathbf{X}(t_k)$ and $\mathbf{Z}(t_k)$ be d-dimensional random vectors. Set $\mathbf{X}(t_k) = (X_1(t_k), \ldots, X_d(t_k))$, $\epsilon_{\mathbf{k}} = (\epsilon_{1k}, \ldots, \epsilon_{dk}) \sim \mathrm{MN}(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I})$, and $\epsilon_{\mathbf{i}}$ and $\epsilon_{\mathbf{j}}$ are independent for $i \neq j$. Then, $\mathbf{Z}(t_k) = (Z_1(t_k), \ldots, Z_d(t_k)) = (X_1(t_k) + \epsilon_{1k}, \ldots, X_d(t_k) + \epsilon_{dk})$. The density h_{ij} in the d-dimension case is given by

$$h_{10}(\mathbf{x}, t) = \int_0^t \prod_{i=1}^{d} \phi(x_i; \sigma^2 w) p_{10}(w, t) \mathrm{d}w.$$

Then we get that

$$\begin{aligned} g_{10}(\mathbf{z}, t) &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} h_{10}(\mathbf{z} - \mathbf{x}; \sigma^2 w) \prod_{i=1}^{d} \phi(x_i; 2\sigma_\epsilon^2) \mathrm{d}x_1 \ldots \mathrm{d}x_d \\ &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \left[\int_0^t \prod_{i=1}^{d} \phi(z_i - x_i; \sigma^2 w) p_{10}(w, t) \mathrm{d}w \right] \prod_{i=1}^{d} \phi(x_i; 2\sigma_\epsilon^2) \mathrm{d}x_1 \ldots \mathrm{d}x_d \\ &= \int_0^t \prod_{i=1}^{d} \left[\int_{\mathbb{R}} \phi(z_i - x_i; \sigma^2 w) \phi(x_i; 2\sigma_\epsilon^2) \mathrm{d}x_i \right] p_{10}(w, t) \mathrm{d}w. \end{aligned}$$

Similarly, we also have

$$g_{00}(\mathbf{z}, t) = \int_0^t \prod_{i=1}^{d} \left[\int_{\mathbb{R}} \phi(z_i - x_i; \sigma^2 (t-w)) \phi(x_i; 2\sigma_\epsilon^2) \mathrm{d}x_i \right] p_{00}(w, t) \mathrm{d}w + e^{-\lambda_0 t} \prod_{i=1}^{d} \phi(z_i, 2\sigma_\epsilon^2),$$
$$g_{01}(\mathbf{z}, t) = \int_0^t \prod_{i=1}^{d} \left[\int_{\mathbb{R}} \phi(z_i - x_i; \sigma^2 (t-w)) \phi(x_i; 2\sigma_\epsilon^2) \mathrm{d}x_i \right] p_{01}(w, t) \mathrm{d}w,$$
$$\begin{aligned} g_{11}(\mathbf{z}, t) = &\int_0^t \prod_{i=1}^{d} \left[\int_{\mathbb{R}} \phi(z_i - x_i; \sigma^2 w) \phi(x_i; 2\sigma_\epsilon^2) \mathrm{d}x_i \right] p_{11}(w, t) \mathrm{d}w \\ &+ e^{-\lambda_1 t} \prod_{i=1}^{d} \left[\int_{\mathbb{R}} \phi(z_i - x_i; \sigma^2 t) \phi(x_i; 2\sigma_\epsilon^2) \mathrm{d}x_i \right]. \end{aligned}$$

Let us mention that these formulas do not require numerical multiple integral evaluation and, as a consequence, are computationally efficient.

Appendix D. Normalized Forward Algorithm

First, let us define the forward variables by

$$\alpha(\mathbf{Z}_{even}(t_{2k}), S(t_{2k}), \boldsymbol{\theta}) = \sum_{S(t_0), S(t_2), \ldots, S(t_{2k-2}) \in \{0,1\}} \nu(S(t_0)) \times \prod_{j=1}^{k} f\left(Z(t_{2j}) - Z(t_{2j-1}), S(t_{2j}) \mid S(t_{2j-2})\right),$$

where $\mathbf{Z}_{even}(t_{2k}) = (Z(t_0), Z(t_2), \ldots, Z(t_{2k}))$, $k = 1, \ldots, \lfloor n/2 \rfloor$, and the initial forward variable $\alpha\left(\mathbf{Z}_{even}(t_0), S(t_0), \boldsymbol{\theta}\right) = \nu(S(t_0))$. Then, using the Markov property of (3) one can show that the forward variables satisfy the following recursive relationship:

$$\alpha\left(\mathbf{Z}_{even}(t_{2k+2}), S(t_{2k+2}), \boldsymbol{\theta}\right) = \sum_{S(t_{2k}) \in \{0,1\}} \alpha\left(\mathbf{Z}_{even}(t_{2k}), S(t_{2k}), \boldsymbol{\theta}\right) \times f\left(Z(t_{2k+2}) - Z(t_{2k+1}), S(t_{2k+2}) | S(t_{2k})\right).$$

Note that for every k we have two forward variables. To get one of the next two forward variables we need to calculate two transitional values in (4), multiply each kth forward variable by an appropriate transitional value, and finally sum up these two quantities. The bottom line is that the transition from $\alpha\left(\mathbf{Z}_{even}(t_{2k}), S(t_{2k}), \boldsymbol{\theta}\right)$ to $\alpha\left(\mathbf{Z}_{even}(t_{2k+2}), S(t_{2k+2}), \boldsymbol{\theta}\right)$ for each k requires a constant (independent of k) number of operations. This allows us to compute the likelihood in linear time with respect to n, because

$$L(\mathbf{Z}_{even}; \boldsymbol{\theta}) = \sum_{S(t_{2\lfloor n/2 \rfloor}) \in \{0,1\}} \alpha\left(\mathbf{Z}_{even}(t_{2\lfloor n/2 \rfloor}), S(t_{2\lfloor n/2 \rfloor}), \boldsymbol{\theta}\right).$$

Now, when the sample size n is large, the series of multiplications may cause underflow problems where some terms are too small to be distinguished from zero by a computer. A normalized forward algorithm addresses the underflow problem. More specifically, let us introduce the normalized forward variables as

$$\bar{\alpha}\left(\mathbf{Z}_{even}(t_{2k}), S(t_{2k}), \boldsymbol{\theta}\right) = \frac{\alpha\left(\mathbf{Z}_{even}(t_{2k}), S(t_{2k}), \boldsymbol{\theta}\right)}{L\left(\mathbf{Z}_{even}(t_{2k}); \boldsymbol{\theta}\right)},$$

and let

$$\rho\left(\mathbf{Z}_{even}(t_{2k+2}); \boldsymbol{\theta}\right) = \frac{L\left(\mathbf{Z}_{even}(t_{2k+2}); \boldsymbol{\theta}\right)}{L\left(\mathbf{Z}_{even}(t_{2k}); \boldsymbol{\theta}\right)}.$$

Then, the update formulas for normalized forward variable $\bar{\alpha}\left(\mathbf{Z}_{even}(t_{2k}), S(t_{2k}), \boldsymbol{\theta}\right)$ and $\rho\left(\mathbf{Z}_{even}(t_{2k+2}); \boldsymbol{\theta}\right)$ are

$$\bar{\alpha}\left(\mathbf{Z}_{even}(t_{2k+2}), S(t_{2k+2}), \boldsymbol{\theta}\right) = \frac{1}{\rho\left(\mathbf{Z}_{even}(t_{2k+2}); \boldsymbol{\theta}\right)} \times \sum_{S(t_{2k}) \in \{0,1\}} \bar{\alpha}\left(\mathbf{Z}_{even}(t_{2k}), S(t_{2k}), \boldsymbol{\theta}\right) \times f\left(Z(t_{2k+2}) - Z(t_{2k+1}), S(t_{2k+2}) | S(t_{2k})\right),$$

and

$$\rho\left(\mathbf{Z}_{even}(t_{2k+2}); \boldsymbol{\theta}\right) = \sum_{S(t_{2k+2}), S(t_{2k}) \in \{0,1\}} \bar{\alpha}\left(\mathbf{Z}_{even}(t_{2k}), S(t_{2k}), \boldsymbol{\theta}\right)$$

$$\times f\left(Z(t_{2k+2}) - Z(t_{2k+1}), S(t_{2k+2})|S(t_{2k})\right).$$

Finally, the likelihood function is given by

$$\log L\left(\mathbf{Z}_{even}(t_{2k}); \boldsymbol{\theta}\right) = \sum_{k=1}^{\lfloor n/2 \rfloor} \log \rho\left(\mathbf{Z}_{even}(t_{2k}); \boldsymbol{\theta}\right).$$