

Statistical consistency of the data association problem in multiple target tracking

Curtis B. Storlie

Los Alamos National Laboratory

e-mail: storlie@lanl.gov

Jan Hannig

University of North Carolina – Chapel Hill

e-mail: jan.hannig@unc.edu

and

Thomas C.M. Lee

University of California – Davis

e-mail: tcmllee@ucdavis.edu

Abstract: Simultaneous tracking of multiple moving objects extracted from an image sequence is an important problem which finds numerous applications in science and engineering. In this article we conduct an investigation of the theoretical properties a statistical model for tracking such moving objects, or targets. This tracking model allows for birth, death, splitting and merging of targets, and uses a Markov model to decide the times at which such events occur. This model also assumes that the track traveled by each target behaves like a Gaussian process. The estimated tracking solution is obtained via maximum likelihood. One of the contributions of this article is to establish the almost sure consistency to the data association problem by using these maximum likelihood tracking estimates. A major technical challenge for proving this consistency result is to identify the correct track (data association) amongst a group of similar (but incorrect) track proposals that are results of various combinations of target birth, death, splitting and/or merging. This consistency property of the tracking estimates is empirically verified by numerical experiments. To the best of our knowledge, this is the first time that a comprehensive study is performed for the large sample properties of a multiple target tracking method. In addition, the issue of how to quantify the confidence of a tracking estimate is also addressed.

AMS 2000 subject classifications: Primary 60G35; secondary 60G17, 60G15, 62M20.

Keywords and phrases: Data association, in-fill asymptotics, multiple hypothesis tracking, multiple target tracking, merging, splitting, Brownian motion.

Received July 2010.

Contents

1	Introduction	1228
1.1	A first description of the tracking problem	1229
1.2	Merging and splitting of targets	1231
2	The tracking model of Storlie, Lee, Hannig & Nychka	1233
2.1	Event Model	1233
2.2	Location Model	1235
2.2.1	Initial conditions for a target resulting from a birth event	1235
2.2.2	Initial conditions for a target resulting from a merging event	1236
2.2.3	Parent locations at the time of a merging event	1236
2.2.4	Initial conditions for a target resulting from a splitting event	1237
2.3	The tracking estimate	1237
2.4	Calculating $(\hat{U}, \hat{V}, \hat{P})$	1238
3	Main results: Asymptotic properties	1239
4	Simulated data results	1242
5	Conclusions & further work	1244
	Acknowledgments	1245
A	Model likelihood	1245
B	Derivation of event density (A2)	1245
C	Derivation of location density	1246
D	Technical Lemmas	1247
E	Propositions	1257
F	Proof of Theorem 1	1268
	References	1273

1. Introduction

Multiple target tracking has application to many scientific problems. It has importance in radar and signal processing, air traffic control, robot vision, GPS-based navigation, biomedical engineering, and video surveillance to name a few. Typically the objects of interest (i.e., the targets) are captured in the form of an image sequence. Each image frame contains the locations and perhaps other attribute (such as sizes and shapes) information of the targets. The goal of tracking is to recover this information and use it to help reconstruct the tracks that the targets traveled. Quite often the following two-step strategy would be adopted to solve this problem. The first step is to extract the locations and/or other attributes of the targets from each image frame. There is no unified solution for this first step, as different targets need different methods for extraction; e.g., human faces and missiles require very different target recognition methods to detect their appearances in an image. Once the target coordinates are located, the second step for solving the tracking problem is to link these coordinates together in such a way that coordinates of the same target detected at

different image frames are connected together as a reconstruction of the path that this target traveled. This article assumes that the target coordinates have already been extracted from the image sequence and focuses on the second step of coordinate linking. This second step is also known as the *data association* problem.

The work discussed in this paper was motivated by the practical need for tracking (i) storm activities captured in radar images and (ii) vortices generated in turbulence fields. In these two tracking applications, the splitting and merging of targets is quite common. A promising method is proposed Storlie, Lee, Hannig & Nychka (2009) for tracking of such, and other kinds of, merging and splitting targets. This tracking method models target locations with Gaussian processes, and estimates their tracks using a maximum likelihood approach. Movies presenting the results obtained by applying this tracking method to the storm activities and the vortices applications can be viewed at <http://www.stat.unm.edu/~storlie/tracking/>. In this article we complement the work of Storlie et al. (2009) by studying the asymptotic properties of their solution to the data association problem. We show that under certain regularity conditions this solution is strongly consistent. While this result is useful in and of itself, it is also our hope that the method of proof adopted here could be valuable for studying the large sample properties of solutions to other tracking problems as well. To the best of our knowledge, this is the first time that a comprehensive study is performed for the large sample properties of a multiple target tracking method.

1.1. A first description of the tracking problem

As mentioned before, in this article we assume that the coordinates (x, y) of the targets have been extracted from every frame of the image sequence. We do not know however which of the (x, y) locations corresponds to which target. The goal of a tracking method then, is to take the location data extracted over time and recover the track of each target. A track is defined to be the $(x(t), y(t))$ coordinates of a target at each time t during the image sequence.

To illustrate the idea further, consider the tracking problem depicted in assume that there are 4 targets at each time step as in Figure 1. The locations of the targets in this figure are simulated from the model to be described in Section 2. We are however ignoring the possibility of birth and death of targets as well as splitting and merging for the time being. The targets are free to change position from one time step (or image) to the next. The data association problem is to determine which temporal set of locations corresponds to one particular target. In other words, the goal is to sequentially connect the targets in the bottom left plot to form tracks. Since this is simulated data, the corresponding solution is known and the correct tracks are given in the bottom right plot. In many instances the observed locations of the targets include a non-negligible measurement error. In this case, it is typical to first solve the data association problem. The observations forming a track can then be smoothed to obtain the estimated paths of the solution.

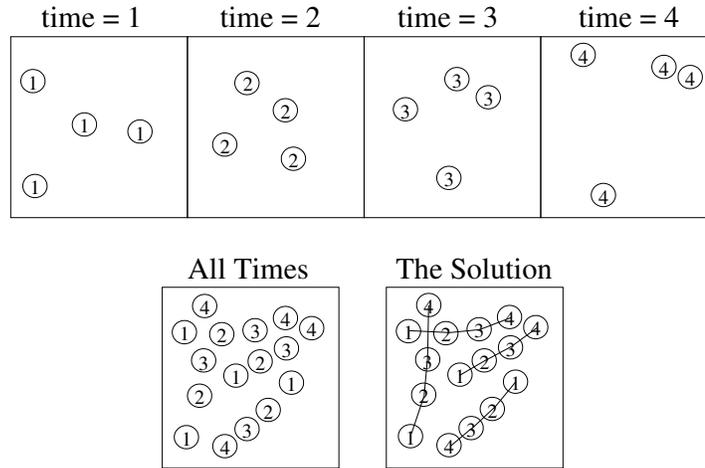


FIG 1. Illustration of the tracking problem.

Many statistical approaches to the target tracking problem have been studied in the engineering literature over the past thirty years; e.g., see the two very comprehensive books [Bar-Shalom, Li & Kirubarajan \(2001\)](#) and [Blackman & Popoli \(1999\)](#), and the references given therein. Most of these methods employ a statistical model to describe the motion of the targets to be tracked. Usually a Gaussian state space model is assumed, and the “best” estimates are defined as the set of tracks that maximizes the likelihood of the model. Note that we use the term likelihood to be the unconditional likelihood of the data as typical in statistics literature, as opposed to the conditional likelihood which is what is often meant in the tracking community. Virtually no attempts have been made to investigate the theoretical properties of these maximum likelihood estimates, although various practical algorithms have been proposed for computing them. The two most widely used algorithms are the Multiple Hypothesis Tracking algorithm of [Reid \(1979\)](#) and the Joint Probabilistic Data Association algorithm developed by [Fortmann, Bar-Shalom & Scheffe \(1983\)](#). See also [Mori, Chong, Tse & Wishner \(1986\)](#) for a good general mathematical framework for this problem. Most recently, a new class of filtering methods based on particle filtering or sequential Monte Carlo ([Gordon, Salmund & Smith \(1993\)](#), [Kitagawa \(1996\)](#), [Liu & Chen \(1998\)](#), [Doucet, Godsill & Andrieu \(2000\)](#) and [Doucet, de Freitas & Gordon \(2001\)](#)) have been developed. Still other methods such as the probability hypothesis density method ([Mahler \(2003\)](#) and [Vo, Singh & Doucet \(2005\)](#)) approximate the likelihood by propagating only the posterior expectation instead of the entire distribution through to subsequent times. Since the major focus of this article is to present a thorough theoretical analysis of the tracking estimates provided by the method of [Storlie et al. \(2009\)](#), the algorithmic issue of how to computing the estimates of target location will not be further discussed.

1.2. Merging and splitting of targets

The example shown in Figure 1 is often a simplistic representation of reality. Many applications will not have the same number of targets in each image. For examples, imperfect detection and occlusion will lead to missing targets in some images. There can also be false alarms or clutter. Furthermore, some targets may appear for the first time or disappear permanently in the middle of the image sequence. We will call these events birth and death respectively.

In addition, this article is motivated by the scientific need of tracking merging and splitting targets, such as storms or vortices. Figure 2 illustrates a more realistic example of an actual tracking problem motivating this work. These images are from a two-dimensional turbulence simulation of freely decaying vortices. The white objects are centers of vorticity rotating in a clockwise direction, whereas the black vortices have the opposite rotation. Vortices of the same spin will merge together as they move close to each other. There is a good example of a merger, between times 8 and 9, of two white vortices that are left of center and below center in the images.

In the example in Figure 2, there will be birth, death, and merging. In practice, the vortices also need to be identified with some target extraction procedure, and there is no *perfect* method for doing this. This leads to false alarm observations and missing observations in some time windows. The ability to deal with these issues is important for a practical tracking system. For the description of a tracking procedure that allows for birth, death, splitting, merging, false alarms, and missing observations, see [Storlie et al. \(2009\)](#). In this presentation, we are only interested in studying the asymptotic behavior of the estimator proposed in [Storlie et al. \(2009\)](#) as observations become more and more frequent. To this end, the asymptotic analysis is greatly simplified if false alarms are ignored. Studying the asymptotic properties under this simplification provides insight into the more complex situation when missing observations and false alarms are present. For example, false alarms can be thought of intuitively as targets that last only for a short time frame. If this is the case, the results presented here would still apply to the actual targets, even if false alarms are present.

Merging and splitting of targets can be common in radar applications as well, though in a slightly different context. That is, when two targets are close together, resolution limits may prevent them from being simultaneously detected. The detection method will then return only one (or even no) observation for these two targets. This certainly poses additional difficulties and challenges. Although this is perceived as a very important issue by [Daum \(1994\)](#) and [Blackman \(2004\)](#), we are unaware of any satisfactory solution to this. Most existing methods for tracking merging targets are not well defined in terms of an overall probabilistic model ([Trunk & Wilson, 1981](#); [Chang & Bar-Shalom, 1984](#); [Koch & van Keuk, 1997](#); [Genovesio & Olivo-Marin, 2004](#)).

Another limitation of most existing work on tracking methodology is the lack of theoretical understanding. Some notable exceptions are the work of [Cong & Hong \(1999\)](#) and [Li & Jing \(2003\)](#) which study the numerical convergence properties of their optimization algorithms, and the work of Chen,

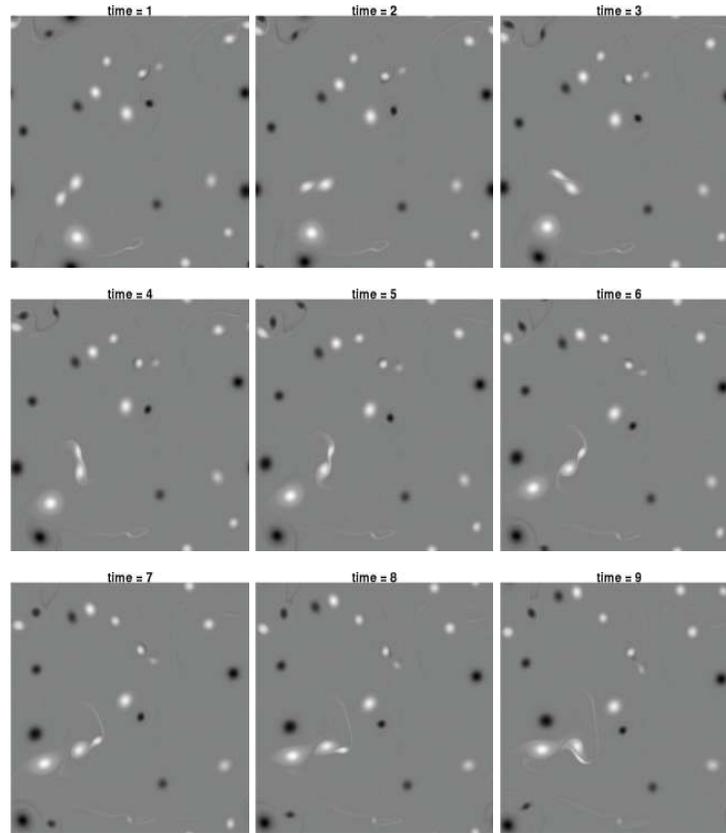


FIG 2. 2D Turbulence Simulation.

Li & Bar-Shalom (2004) which provides some theoretical justification for their method to choose the correct number of targets as the number of image frames goes to infinity. Also, for a somewhat different class of tracking problems, Hall & Rau (2000), Hall, Peng & Rau (2001) and Hall, Qiu & Rau (2007) provide tracking solutions with theoretical support. Despite of these various pieces of work, however, many important theoretical problems remain unsolved. For example, under what conditions can we obtain a consistent tracking estimate as the time increment between observations goes to zero? At what rate does this convergence take place? The novel contribution made in this article is to, via analyzing the method of Storlie et al. (2009), provide a first attempt at addressing these questions. While it is not surprising to expect the existence of a perfect tracking method when the number of observations increases to infinity, it is not obvious that the MLE of our tracking model converges almost surely to the correct solution. The current paper establishes this convergence result.

The rest of this article is organized as follows. Section 2 summarizes the major ingredients of the tracking method of Storlie et al. (2009). The almost

sure convergence of the maximum likelihood solution of this tracking method is established in Section 3. These properties are then illustrated via numerical experiments in Section 4. Concluding remarks and possible future work are offered in Section 5, while technical details are deferred to the appendix.

2. The tracking model of Storlie, Lee, Hannig & Nychka

In this section, for completeness, we summarize the multiple target tracking model proposed by Storlie et al. (2009), for which its theoretical properties are to be studied. This model has a continuous time stochastic component that describes (i) the events that occur and (ii) the locations of the targets to be tracked. The tracking estimate for the targets is obtained by using the model likelihood given in Appendix A.

Define a path, $(X(t), Y(t))$, to be the coordinates of a target at time $t > 0$. We observe the targets at discrete times $\mathbf{t} = (t_0, t_1, \dots, t_n)$. We assume a two-dimensional path, but the following could easily allow for paths in \mathbb{R}^3 . We wish to model the path of a target by a two-dimensional Gaussian process. The complication is that, due to the following reasons, we may not be able to observe the target at all times:

1. it will exist in the future, but does not exist yet (birth),
2. it no longer exists (death),
3. it broke off into 2 new targets (splitting), and
4. it combined with another target (merger).

The tracking model to be studied below has two parts which we will refer to as (i) the *Event Model* and (ii) the *Location Model*. The Event Model describes how and when targets come into existence and termination, while the Location Model describes how an existing target travels around.

Before describing the Event Model, again note that in many practical tracking problems, there may be missing observations and false alarms. Also, additional target attribute variables, such as size and shape, may be available. For the description of a more complete model that allows for the all of the above complications as well as attributes, false alarms, and missing observations, see Storlie et al. (2009). In the following we omit these features in order to simplify the asymptotic analysis. As mentioned, we can gain a lot of insight about the asymptotic behavior of the solution to the data association problem in this simplified case, and project this understanding to more complex scenarios (i.e., those with false alarms and missing observations).

2.1. Event Model

The Event Model is a continuous time Markov chain model that is very similar to a birth and death process. Four types of events can occur: births, deaths, splits, and mergers. The rate at which these events happen are λ_b , $N(t)\lambda_d$, $N(t)\lambda_s$, and $(N(t) - 1)\lambda_m$ respectively, where $N(t)$ is the number of targets

in existence at time t . It is assumed that the initial number of targets, $N_0 = N(0) \sim \text{Poisson}(\lambda_0)$.

The following notation will be used to describe the Event Model

$$\begin{aligned}
 U_{b,j} &= \text{number of births in the interval } [t_j, t_{j+1}) \\
 U_{d,j} &= \text{number of deaths in the interval } [t_j, t_{j+1}) \\
 U_{s,j} &= \text{number of splits in the interval } [t_j, t_{j+1}) \\
 U_{m,j} &= \text{number of mergers in the interval } [t_j, t_{j+1}). \tag{1}
 \end{aligned}$$

We will write $\mathbf{U}_b = (U_{b,1}, \dots, U_{b,n})$ and similarly for \mathbf{U}_d , \mathbf{U}_s , and \mathbf{U}_m . Also, denote the collection of N_0 and the \mathbf{U} 's by $\mathcal{U} = (N_0, \mathbf{U}_b, \mathbf{U}_d, \mathbf{U}_s, \mathbf{U}_m)$.

Each target, regardless of its status (e.g., alive or dead), will be uniquely identified by a positive integer starting from 1. We shall call such integers *indices*. The initial targets alive at time t_1 are arbitrarily labeled with indices 1 through N_0 . The following actions will be taken at the time whenever any one of the four possible events happens. When there is a birth the new target will be given the next available index. For example, if there are already 10 targets in the model (some currently alive, some could be dead), these targets would have been labeled uniquely with indices from 1 to 10, and the new target will be given an index of 11. When there is a death, all targets that are still alive are equally likely to be selected as the one that dies. When there is a split, all of the living targets are equally likely to be the parent, and the children will be given the next two available indices. Finally, for merging events all of the possible pairs of all living targets are equally likely to be the parents, and the child will be given the next available index.

Notice that the assumption that all targets are equally likely to be parents in a merger appears to be at odds with the principle that only close targets are eligible to merge together. We will rectify this issue in the Location Model to be described in Section 2.2. In short, locations of the parents of a merger are conditioned to be “close” to each other right before the merger. This shifts the burden of enforcing the property that “only close targets merge together” to the location model. This leads to an important simplification of the likelihood calculation since the location model depends on the Event Model but not vice-versa. On the other hand, this arrangement leads to complications in studying the theoretical properties of our tracking algorithm due to the loss of the Markov property.

We will specify which targets were involved in the events by

$$\begin{aligned}
 V_{b,j} &= \text{the collection of indices of targets that were born in the interval } [t_j, t_{j+1}) \\
 V_{d,j} &= \text{the collection of indices of targets that died in the interval } [t_j, t_{j+1}) \\
 V_{s,j} &= \text{the collection of triplets } (i_1, i_2, i_3) \text{ where } i_1 \text{ is the index of the parent} \\
 &\quad \text{and } i_2, i_3 \text{ are the children for every split in the interval } [t_j, t_{j+1}) \\
 V_{m,j} &= \text{the collection of triplets } (i_1, i_2, i_3) \text{ where } i_1, i_2 \text{ are the indices of the} \\
 &\quad \text{parents and } i_3 \text{ is the child for every merger in the interval } [t_j, t_{j+1}). \tag{2}
 \end{aligned}$$

Let $\mathbf{V}_b = (V_{b,1}, \dots, V_{b,n})$ and similarly for \mathbf{V}_d , \mathbf{V}_s , and \mathbf{V}_m . The collection of all the \mathbf{V} 's will be denoted as $\mathcal{V} = (\mathbf{V}_b, \mathbf{V}_d, \mathbf{V}_s, \mathbf{V}_m)$.

Lastly, it should be noted that this is a hidden Markov model in the sense that we do not actually observe the variables \mathcal{U} , and \mathcal{V} from the data. Predicting these variables is part of the tracking problem. This will be described further in Section 2.3.

2.2. Location Model

When a target is determined to exist by the Event Model, the observed path of the i^{th} target $(X_i(t), Y_i(t))$ will be modeled by a Gaussian process. The Gaussian process is commonly used in tracking applications because it is mathematically straightforward to work with and yet it models the paths well in most applications. Target paths are assumed to run their course independently of other targets unless they are required to split or merge as determined by the Event Model. The dependency introduced by splitting and merging are described below.

The distribution of $X_i(t)$ will be defined below for the three cases of a target resulting from (i) birth, (ii) merger, and (iii) split. The distribution of $Y_i(t)$ will be similar with the obvious changes in notation and parameters and independent of $X_i(t)$ given the event variables \mathcal{U} and \mathcal{V} .

Let the x component of location of the i^{th} target at time t be denoted by $X_i(t)$. Also denote the time of initiation of the i^{th} target by ξ_i . If the i^{th} target exists at the first observation time t_1 , then it is assumed that $\xi_i = t_1$. Observed location is then governed by

$$X_i(t) = X_i(\xi_i) + H_i(t - \xi_i) + \sigma_i G_i(t - \xi_i), \quad (3)$$

where $H_i(t)$ is a smooth function corresponding to the target location, $G_i(t)$ is some continuous mean zero Gaussian process, describing random fluctuations such as errors of measurement; both $H_i(0) = G_i(0) = 0$. The initial position, $X_i(\xi_i)$ will depend on whether the target resulted from a birth, merger, or split. These are described in Sections 2.2.1 to 2.2.4 below.

The model in (3) is designed for $G_i(t)$ to be a Brownian motion. If we change the model for $G_i(t)$, we may wish to change (3) accordingly. For example, if we use integrated Brownian motion, we may want to add an initial velocity term instead of assuming that it is equal to zero.

2.2.1. Initial conditions for a target resulting from a birth event

Suppose that the i^{th} target resulted from a birth. It is assumed that the initial position is Gaussian. Specifically, $X_i(\xi_i) \sim \mathcal{N}(\mu_{X_0}, \sigma_{X_0}^2)$. For many problems, it may also seem reasonable to use a uniform distribution to model the initial location $X_i(\xi_i)$. However, very often the likelihood of a uniform distribution can be satisfactorily mimicked by sufficiently increasing the variance of a normal distribution. Thus we will keep the original Gaussian assumption for mathematical convenience.

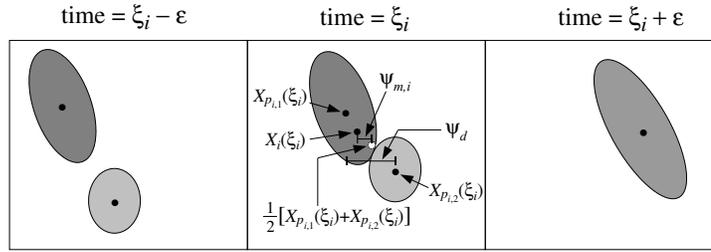


FIG 3. Physical Description of a Merger.

2.2.2. Initial conditions for a target resulting from a merging event

Now suppose that the i^{th} target is initiated from a merger. Let p_i be a vector containing the indices of the two targets that merge together to create the i^{th} target. If there is no size measurement made on the targets, which we are assuming here, then we can let the initial position of the child be the simple average of the positions of the parents at the time of merger plus a noise term $\psi_{m,i}$. We model $\psi_{m,i}$ as $\psi_{m,i} \sim \mathcal{N}(0, \sigma_{X_m}^2)$ with a small $\sigma_{X_m}^2$ so that the new target location is likely to be close to the average of the parents. Figure 3 displays a physical representation of this. Therefore we have

$$X_i(\xi_i) = \frac{1}{2} \{X_{p_{i,1}}(\xi_i) + X_{p_{i,2}}(\xi_i)\} + \psi_{m,i}. \tag{4}$$

The two parent targets are currently not required to be close to each other at the time of merger. To ensure that the parents locations are close to each other before merging, the difference between locations of the parents is conditioned to be small at the time of merger, ξ_i . This can be achieved as follows. Let d be a vector containing the three targets involved in the merger. The indices of the parents are d_1 and d_2 where d_3 is the index of the child. The difference in location between the two parents at the time of merger plus a random noise term is given by $D = X_{d_1}(\xi_{d_3}) - X_{d_2}(\xi_{d_3}) + \psi_d$, where $\psi_d \sim \mathcal{N}(0, \sigma_{X_d}^2)$. In a manner similar to a Brownian Bridge process, we then condition on the event $D = 0$. If σ_{X_d} is small, this will make it very likely that the two parent paths are close together right before the merger. Referring to Figure 3 once again, we see a merging event with a possible realization of ψ_d .

2.2.3. Parent locations at the time of a merging event

Notice that in our modeling so far, the two parent targets are not required to be close to each other at the time of a merging event. To ensure that the parents move close to each other before merging, the difference between locations of the parents at the time of merger is conditioned to be small. This is achieved as follows.

Let $d = (d_1, d_2, d_3)$ be a vector containing the indices of the three targets involved in a merging event where d_1 and d_2 are the parents while d_3 is the index of the child. Let D be the difference in location between the observed locations of the two parents at the time of merger plus a noise term,

$$D = X_{d_1}(\xi_{d_3}) - X_{d_2}(\xi_{d_3}) + \psi_d \quad (5)$$

where $\psi_d \sim \mathcal{N}(0, \sigma_{X_d}^2)$ and independent of the targets. If σ_{X_d} is small, then it is likely that ψ_d is small in absolute value. If we then condition the model for X_{d_1} and X_{d_2} on the event $D = 0$, this will ensure that the parents are only a small distance ψ_d apart at the time of the merging event. In Figure 3 once again, we see a merging event with a possible realization of ψ_d .

In general, there will be $N_m = \sum_{j=1}^n U_{m,j}$ merging events during the time window $[t_1, t_n]$. We will condition the target paths on all of these mergers in a manner similar to that above. This is described more precisely as follows. Let D_i be the D from (5) and $\psi_{d,i}$ be the corresponding ψ_d for the i^{th} merging event, $i = 1, \dots, N_m$. In a manner similar to a Brownian Bridge, the paths for (X_1, \dots, X_M) are then conditioned on the event $\{(D_1, \dots, D_{N_m}) = (0, \dots, 0)\}$, where M is the total number of targets that existed before time t_n .

2.2.4. Initial conditions for a target resulting from a splitting event

Suppose that the i^{th} target is initiated by a splitting event. To keep notation consistent with that for mergers, let p_i be a vector of length one that contains the index of the parent of the i^{th} target. The initial location of a target resulting from a split is given by

$$X_i(\xi_i) = X_{p_{i,1}}(\xi_i) + \psi_{s,i}, \quad (6)$$

where $\psi_{s,i} \sim \mathcal{N}(0, \sigma_{X_s}^2)$. Similar to the model for a merger, the initial position of a new target from a split is the same as that of the parent plus some error. It is assumed that $\sigma_{X_s}^2$ is small so that the new targets are likely to appear close to where the parent split.

2.3. The tracking estimate

Here we formally define our estimand and our method for estimating it. The setup for this estimation problem is as follows. We collect data at the following times t_1, \dots, t_n . At each time there are m_j observations. Let Z_{ij} be the i^{th} observation at time t_j , $i = 1, \dots, m_j$. Each Z_{ij} is a vector of the location values for a target. Further denote $\mathbf{Z}(t)$ as the collection of observations at time t , so that $\mathbf{Z}(t_j) = (Z_{1,j}, \dots, Z_{m_j,j})$, and let $\mathcal{Z} = \{Z_{ij} : j = 1, \dots, n, i = 1, \dots, m_j\}$ be the collection of observations at all observed times.

From our data, \mathcal{Z} , we need to decide which target to assign each observation to. Note that each observation can be assigned to only one target and each target can have only one observation assigned to it. We will create the variable

p_{ij} to be the index of the target that observation Z_{ij} originated from. Let $\mathcal{P} = \{p_{ij} : j = 1, \dots, n, i = 1, \dots, m_j\}$.

Now for a given \mathcal{Z} , the indices contained in \mathcal{P} will specify the tracks of each target. To completely specify a solution to the tracking problem, in addition to \mathcal{P} , we must also specify the events (births, deaths, splits, and mergers) that occurred with the variables \mathcal{U} and \mathcal{V} defined in (1) and (2). Thus the variables \mathcal{U} and \mathcal{V} together with \mathcal{P} denote a solution to the data association problem or the tracking solution. As mentioned before, the estimated paths can then be obtained by smoothing the observed locations of each track. We will denote our estimate of the tracking solution $(\mathcal{U}, \mathcal{V}, \mathcal{P})$ as $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$.

2.4. Calculating $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$

Assume for the moment that the parameters in the model described in Section 2 are known quantities. We will consider the estimation of these parameters later. We adopt the *Gelfand* style for density notation (Gelfand, 1990), and let $[X]$ denote the probability density function of the random variable X , $[X](x)$ to denote $[X]$ evaluated at x , and $[X | Y]$ to denote the conditional density of X given Y . Notice that the evaluation $[X | Y](x)$ is a random variable (i.e., a function of the random variable Y). Finally, we will use $[X | Y](x | y)$ to denote the evaluation of the conditional density for some observed value of the random variable $Y = y$. To achieve our tracking estimate, we will compute an approximation to the conditional density of $(\mathcal{U}, \mathcal{V}, \mathcal{P})$ given the data $\mathcal{Z} = z$,

$$[\mathcal{U}, \mathcal{V}, \mathcal{P} | \mathcal{Z}](u, v, p | z). \quad (7)$$

Note that this is also a probability mass function since the variables $(\mathcal{U}, \mathcal{V}, \mathcal{P})$ are discrete. From this it is natural to define our tracking estimate as

$$(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}}) = \arg \max_{u, v, p} [\mathcal{U}, \mathcal{V}, \mathcal{P} | \mathcal{Z}](u, v, p | z). \quad (8)$$

We can also interpret $[\mathcal{U}, \mathcal{V}, \mathcal{P} | \mathcal{Z}](\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}} | z)$ as the probability that $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$ is the correct solution given the data $\mathcal{Z} = z$.

We now consider the calculation of the density in (7). With the one-to-one mapping $g : (\mathcal{P}, \mathcal{Z}) \rightarrow (\mathcal{X}, \mathcal{Y}, \mathcal{Z})$, for a given \mathcal{Z} , the information contained in \mathcal{P} and $(\mathcal{X}, \mathcal{Y})$ is the same. Let $g^* : (\mathcal{P}, \mathcal{Z}) \rightarrow (\mathcal{X}, \mathcal{Y})$ be the function g without the last variable in its output. In Storlie et al. (2009) it is shown that (7) can be written as

$$[\mathcal{U}, \mathcal{V}, \mathcal{P} | \mathcal{Z}](u, v, p | z) = \frac{[\mathcal{U}, \mathcal{V}, \mathcal{X}, \mathcal{Y}](u, v, g^*(p, z))}{\sum_{j=1}^{\infty} [\mathcal{U}, \mathcal{V}, \mathcal{X}, \mathcal{Y}](u_j, v_j, g^*(p_j, z))}, \quad (9)$$

where $\{(u_j, v_j, p_j) : j = 1, 2, \dots\}$ is an enumeration of the possible tracking solutions. Of course in practice it may not be possible to exhaust all possible enumerations, but the solution (u, v, p) that maximizes (9) is the same one that maximizes the numerator.

In Appendix A a closed form approximation is derived for the model likelihood $[\mathcal{U}, \mathcal{V}, \mathcal{X}, \mathcal{Y}]$. It is important to note that the estimator defined in (8) uses (9) with this approximation for $[\mathcal{U}, \mathcal{V}, \mathcal{X}, \mathcal{Y}]$ rather than the actual density $[\mathcal{U}, \mathcal{V}, \mathcal{X}, \mathcal{Y}]$. In fact, the actual density would not be tractable to work with. Thus, all of the main results in Section 3 assume that $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$ is calculated with the approximate likelihood, which is explicitly presented in Equations (A1)-(A5). There are several optimization methods available to search for this maximum; see Blackman & Popoli (1999) for example.

Confidence Sets: We can also use (9) to get an approximation to the distribution of the possible tracking solutions given the data \mathcal{Z} . This is much more informative than just the estimate $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$. For example, we could make a confidence set of solutions such that the probability that the correct solution was any of the solutions in that set was $100(1 - \alpha)\%$. We could also calculate the probability that a given observation belongs to a certain target or the probability that two targets merged, etc.

Up until this point, we have ignored the issue of estimating the model parameters. If the tracking application requires the estimation of such parameters, one could do so with maximum likelihood or other suitable estimation method. The problem is that, before any tracking solution $(\mathcal{U}, \mathcal{V}, \mathcal{P})$ is specified, we would not know the values of the variables that should go into these estimates. We allow each of the solutions that we consider in (8) to have its own parameter estimates. This will necessarily make incorrect solutions have an overly optimistic likelihood and bias the distribution given in (9). We can limit the amount of this bias by setting reasonable bounds for the parameter estimates. This can be done very effectively in many practical problems, since prior information like a possible range for the numbers of targets and their velocities are typically available to the researcher. In the next section we show that as the sampling rate approaches infinity, the estimate (8) is identical to the correct solution eventually, even when the model parameters are estimated.

3. Main results: Asymptotic properties

In this section we show the strong consistency of the estimator $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$ defined by (8). First we highlight that our estimator assumes the observations are produced from the model described in Section 2, and that the estimator itself is constructed with the likelihood approximation given in (A1)-(A5).

We begin with the following notation. For each $k = 1, 2, \dots$ we collect from the process the observations $\mathcal{Z}_k = (\mathbf{Z}(t_1^k) \dots \mathbf{Z}(t_{n_k}^k))$ at times $0 \leq t_1^k < t_2^k < \dots < t_{n_k}^k \leq T$. Let $\Delta t_j^k = t_{j+1}^k - t_j^k$. Further let $t_{i,j}^k, j = 1, \dots, n_i^k$ denote the j^{th} time that the i^{th} target is observed, where n_i^k is the number of times the i^{th} target is observed. Also let $\Delta t_{i,j}^k = t_{i,j+1}^k - t_{i,j}^k$. At times it will be convenient to write $t_{i,j} = t_{i,j}^k, \Delta t_{i,j} = \Delta t_{i,j}^k$ and $n_i = n_i^k$, keeping in mind that these are still a function of k .

We will assume that we are using a Brownian motion model for the error component $G_i(t)$ in (3), (4) and (6). In addition, we will assume that the variance

scalers $\sigma_i^2 = \sigma^2$ for all i . The estimator we will use for σ^2 is defined as

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^{n_i} \frac{(X_i(t_{i,j+1}) - X_i(t_{i,j}))^2}{\Delta t_{i,j}} I_{E_{i,j}}, \quad (10)$$

where $N = \sum_{i=1}^m (n_i - 1)$ is the total number of consecutive differences from all tracks, I_A is the 0,1 indicator function of the event A , and the event $E_{i,j}$ is given by

$$\left\{ \frac{(X_i(t_{i,j+1}) - X_i(t_{i,j}))^2 + (Y_i(t_{i,j+1}) - Y_i(t_{i,j}))^2}{\Delta t_{i,j}} \leq K_1 \sqrt{\log \log N} \right\}$$

for some constant $K_1 > 0$. The indicator in (10) is to make the estimator more robust. Essentially this will eliminate extreme observations from biasing the variance estimate if the tracking estimate has incorrectly connected tracks. It will also be important to exclude extreme observations from the estimator even when the tracks are correctly specified as we will see. Following arguments similar to that of the proof of the Strong Law of Large Numbers it is straight forward to show that (10) is a consistent estimator of σ^2 when the tracks are correctly specified; see Lemma 8.

The conditions needed for Theorem 1 are as follows.

Condition 1. $\sup_j \{\Delta t_j^k\} = O(k^{-1})$ as $k \rightarrow \infty$.

Condition 2. The events of birth, death, splitting, and merging are distributed according to the Event Model of Section 2.1.

Condition 3. The location component, $H_i(t)$, of the observed location model in Section 2.2, satisfies $H_i(t) = \int_0^t h_i(s) ds$, where $h_i \in L_2^{loc}$ is locally square integrable for all targets.

Condition 4. The error component, $G_i(t)$, of the observed location model in Section 2.2, is a Brownian motion for all targets.

Condition 5. The likelihood $[\mathcal{U}, \mathcal{V}, \mathcal{X}, \mathcal{Y}]$ is calculated according to the approximation given in (A1)–(A5) which treats $H_i(t) + \sigma G_i(t)$ as a scaled Brownian motion for all targets.

Condition 6. The parameter estimates are confined to a compact set such that $\lambda_0, \lambda_b, \lambda_d, \lambda_s$, and λ_m are greater than zero and all the variance components of the location model are greater than 0.

Condition 7. The variance parameters for the random process components of $X_i(t)$ and $Y_i(t)$ which are σ_i^2 and η_i^2 respectively are such that $\sigma_i^2 = \sigma^2$ and $\eta_i^2 = \eta^2$ for all i . The estimates for σ^2 and η^2 are given by (10).

Condition 8. The estimate $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$ given in (8) is restricted to those with less than $M < \infty$ targets with $M > m$, where m is the number of targets in the correct solution $(\mathcal{U}, \mathcal{V}, \mathcal{P})$. The estimate $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$ is further restricted so that consecutive observations in a track must be such that

$$\{X_i(t_{i,j+1}) - X_i(t_{i,j})\}^2 + \{Y_i(t_{i,j+1}) - Y_i(t_{i,j})\}^2 \leq K_2 (\log k)^{-1}$$

for some positive constant K_2 .

The following theorem uses the propositions in the appendix to show that the tracking solution is estimated correctly in the limit.

Theorem 1. *Assume Conditions 1-8. Let $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$ be the sequence of correct tracking solutions. Let $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ be given by (8) restricted by Condition 8 for each k . Then there exists a K s.t. $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k = (\mathcal{U}, \mathcal{V}, \mathcal{P})_k$ for all $k > K$ almost surely (a.s.).*

The conclusion in Theorem 1 implies that for a given ω in a set with probability 1, there exists a $K(\omega)$ such that the estimate will equal the correct solution for all $k > K(\omega)$. This also implies that the parameters in the model can be estimated consistently as well, provided that the estimators used are consistent for their respective parameters when the tracks are correctly specified. Some possible estimators are given in Storlie et al. (2009).

Theorem 1 is proved in Appendix F by computing the likelihood ratio of a possible sequence of alternatives (i.e., any sequence of solutions that is not equal to $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$ for each k), $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$, to that of the correct solution sequence, $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$. We will then show that the supremum of this ratio converges to zero.

Remark 1. In practical terms, the consistent solution to the data association problem implies that we know without error the correct number of targets, as well as the number and time of the event for all births, deaths, splits, and mergers, respectively. The location of each target at each time is still not known *exactly*, but it can be inferred with standard smoothing techniques since it is known exactly which observations correspond to which target.

Remark 2. When $k = \infty$ it is clear that the target to track associations can be resolved since $(X_i(t), Y_i(t))$ is a continuous process. However, for finite k it is **not** always clear how to achieve a good estimate $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$, in a way such that this estimate is also consistent. These results show that $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ is one such estimate.

Remark 3. It is *only* for the purpose of deriving a likelihood $[\mathcal{U}, \mathcal{V}, \mathcal{X}, \mathcal{Y}]$ for the estimator in (8), that we assume a scaled Brownian motion model for $H_i(t) + \sigma G_i(t)$ from (3), or equivalently that $H_i(t) = 0$, as stated in Condition 5. Theorem 1 still allows for the *actual* motion of the target to be any $H_i(t)$ satisfying Condition 3. This is due to the Cameron Martin theorem, Theorem 18.22 of Kallenberg (2002), which under the assumptions of Conditions 3 and 4, implies that the distribution of $H_i(t) + \sigma G_i(t)$ and the distribution of $\sigma G_i(t)$ are mutually absolutely continuous. Since in the proof of the main Theorem, we are only interested in events of probability one, this allows us to just prove the result under the assumption that $H_i(t) = 0$. See Appendix F for more details.

Remark 4. Condition 8 contains a mild regularity condition to ensure that the number of targets is not allowed to increase without bound in the estimation process. That is, some finite bound on the number of targets needs to be specified a priori, which has no effect on the result, provided this bound is greater than

the true number of targets present in the actual event. In practice, we can make this bound some large number that we are sure is larger than the number of targets present in the case at hand. The second part of Condition 8 is basically a form of screening or “gating” as it is called in the tracking literature. It simply prevents us from entertaining very unlikely matches.

Remark 5. Recall again that the estimate $(\hat{U}, \hat{V}, \hat{P})$ defined by (8) is constructed with the approximation to the model likelihood given in (A1)–(A5). Thus the results above hold for this estimator and not necessarily for the hypothetical estimator constructed using the true likelihood. For the following two reasons, we do not study the estimator that is based on the true likelihood. First, the form of this estimator is not tractable, making it very difficult to study analytically. Second, from a practical perspective, the consistent estimate $(\hat{U}, \hat{V}, \hat{P})$ that is being studied here can be calculated in a reasonable amount of time, while for the true likelihood based estimator, a large number of likelihood evaluations would need to be numerically approximated, and hence making its use practically infeasible.

Remark 6. The result of Theorem 1 does not apply directly to situations with missed detections and false alarms. However, if we assume an independent over time spatial Poisson process for false alarms, and an independent in time Bernoulli process for missed detections, as is often done in the tracking literature, it would not seem to affect the result (at least intuitively). Therefore, we conjecture that the same result holds even in this more complex case. The addition of these features, however, would add substantial complication to the proof, so we leave this result for a future work.

4. Simulated data results

This section presents simulation results from numerical experiments that were conducted to give a numerical example of Theorem 1 from Section 3. For all of these simulations, the data, \mathcal{Z} , is assumed to come from the model given in Section 2. The parameters used to simulate the different cases are given as follows: mean initial X location ($\mu_{X_0} = -113$), variance of initial X location ($\sigma_{X_0}^2 = 100$), variance scalar for BM’s in X location ($\sigma_i^2 = 0.1$ for all i), variance of difference in X location between parent and child after a split ($\sigma_{X_s}^2 = .5$), variance of difference in X location between parent and child after a merger ($\sigma_{X_m}^2 = .125$), variance of difference in X location between parents at time of merger ($\sigma_{X_d}^2 = 1$), mean initial Y location ($\mu_{Y_0} = 37.5$), variance of initial Y location ($\sigma_{Y_0}^2 = 100$), variance scalar for BM’s in Y location ($\eta_i^2 = .1$ for all i), variance of difference in Y location between parent and child after a split ($\sigma_{Y_s}^2 = .5$), variance of difference in Y location between parent and child after a merger ($\sigma_{Y_m}^2 = .125$), and variance of difference in Y location between parents at time of merger ($\sigma_{Y_d}^2 = 1$), where $\mu_{X_0}, \sigma_{X_0}^2, \dots, \sigma_{Y_d}^2$ are more precisely defined in Section 2.2. The event model parameters were set as mean number of initial targets ($\lambda_0 = 4$), rate of birth ($\lambda_b = 0.1$), rate of death ($\lambda_d = .02$), rate of split ($\lambda_s = 0.06$), and rate of merger ($\lambda_m = .08$).

For the parameter estimation, we restricted the parameter values to the following sets $\lambda_0 \in [0, 25]$, $\lambda_b \in [0.001, .25]$, $\lambda_d \in [0.001, .15]$, $\lambda_s \in [0.001, .15]$, $\lambda_m \in [0.001, .15]$, $\mu_{X_0} \in [-120, -85]$, $\sigma_{X_0}^2 \in [50, 1000]$, $\sigma_i^2 \in [0.001, 10.0]$, $\sigma_{X_s}^2 \in [0.001, 1.5]$, $\sigma_{X_m}^2 \in [0.001, 0.5]$, $\sigma_{X_d}^2 \in [0.001, 5.0]$, $\mu_{Y_0} \in [25, 50]$, $\sigma_{Y_0}^2 \in [50, 1000]$, $\eta_i^2 \in [.001, 10.0]$, $\sigma_{Y_s}^2 \in [0.001, 1.5]$, $\sigma_{Y_m}^2 \in [0.001, 0.5]$, and $\sigma_{Y_d}^2 \in [0.001, 5.0]$.

For these simulations the random location component $H_i(t)$ is an integrated Brownian motion, and the error component $G_i(t)$ is a Brownian motion for all targets. The three different cases studied here use three different time increments, $\Delta t = 1.0$, $\Delta t = 0.5$, and $\Delta t = 0.1$. For $\Delta t = 1.0$ we collect observations at times $0.0, 1.0, \dots, 9.0$, for $\Delta t = 0.5$ we collect observations at times $0, 0.5, \dots, 9.0$, etc. For each of the three cases we obtained the results of $N = 100$ realizations.

Recall that the aim of this numerical exercise is to empirically demonstrate the consistency of the tracking estimate $(\hat{U}, \hat{V}, \hat{P})$. Therefore we are interested in comparing the correct solution with other possible solutions that have high likelihood values. We use the following idea to find a set of high likelihood solutions that also contains the correct solution. First, we start with the correct solution and then consider making different changes to it. These changes are the same as those mentioned in the proof of Theorem 1; i.e., breaking apart a split into two deaths and a birth, breaking a track apart to form a birth and a death, etc. It is demonstrated in the proof of Theorem 1 that any other solution can be obtained by making a sequence of changes of this type.

The algorithm provided below is not intended as an algorithm to estimate unknown tracking solutions as it requires knowledge of the correct solution. However, it is sufficient for our purposes here, which is to show that eventually the correct solution has a high likelihood relative to other possible solutions. For an algorithm to use to solve the data association problem in practice see Storlie et al. (2009). *Algorithm 1* proceeds as follows.

Algorithm 1.

Step I: begin with the correct solution

Step II: consider all possible changes of the types below to the solution

1. *breaking a track into two track segments.*
2. *labeling a merger as two deaths and a birth.*
3. *labeling a split as a death and two births.*
4. *connecting a death with a birth to make one track.*
5. *labeling two deaths and a birth as a merger.*
6. *labeling a death and two births as a split.*

Step III: of the solutions resulting from these changes keep only solutions $(\tilde{U}, \tilde{V}, \tilde{P})$ such that $L(\tilde{U}, \tilde{V}, \tilde{P}) > L(U, V, P)/M$ where $L(\cdot) = [U, V, P | \mathcal{Z}](\cdot | z)$ is the conditional density given the data.

Step IV: repeat Steps II-III on the solutions obtained from Step III until the top K solutions remain unchanged.

TABLE 1

Results of 100 Realizations from a Brownian Motion Model. % **Est Correct** is the percentage of times that $(\hat{U}, \hat{V}, \hat{P})$ was equal to the correct solution (U, V, P) . % **Births Correct**, % **Deaths Correct**, % **Splits Correct**, % **Mergers Correct** are the percentages of all birth, death, splitting, and merging events, respectively, in the simulation that were labeled correctly by the estimate. % **Falling in 95% CS** is the percentage of times that the 95% confidence set (formed for each of the 100 data realizations) contained the correct solution

	$\Delta t=1.0$	$\Delta t=0.5$	$\Delta t=0.1$
% Est Correct	51.0	84.0	92.0
% Births Correct	90.1	97.9	99.4
% Deaths Correct	86.7	94.0	95.9
% Splits Correct	58.3	95.7	98.3
% Mergers Correct	86.4	98.3	99.5
% Falling in 95% CS	90.0	93.0	99.0

When Algorithm 1 converges, we have a set of solutions with high relative likelihood, and we are also assured that the correct solution is in this set. The estimate $(\hat{U}, \hat{V}, \hat{P})$ is the solution from this set with highest likelihood. Notice that Algorithm 1 does not directly consider switching observations between tracks, but this possibility is considered by breaking two tracks apart at the same time, then reconnecting them via Steps II-1 and II-4.

The simulation results for each of the time increments are given as the columns of Table 1. From Table 1 we can see that the estimation does improve substantially as Δt gets smaller. We see a dramatic improvement in the number of correct estimates. The percentage goes from 51.0% for the $\Delta t = 1.0$ case, to 84.0% for the $\Delta t = 0.5$ case, to 92.0% for the $\Delta t = 0.1$ case. The 95% confidence sets also have reasonable coverage for all sample sizes. It is worth noting that these results are actually illustrating a convergence in probability to the correct solution; i.e., $P((\hat{U}, \hat{V}, \hat{P}) = (U, V, P)) \rightarrow 1$, which is guaranteed by the stronger result of Theorem 1. From Theorem 1 we know that the 8 realizations in the $\Delta t = 0.1$ simulation that still have incorrect estimates could eventually be estimated correctly if we made Δt small enough.

5. Conclusions & further work

In this paper we have provided theoretical justification for the multiple target tracking method developed in Storlie et al. (2009). We have given sufficient conditions under which the estimate will converge to the correct solution almost surely. Our theoretical analysis revealed the importance of using a robust estimate of the variance component for the random process $G_i(t)$. The theoretical results were then demonstrated by simulation.

One important direction for future work is to generalize these results to more complicated multiple target tracking problems. For example one could examine the asymptotic properties of the estimate when using a more complicated (and more realistic) model such as integrated Brownian motion in the likelihood calculation. Also, it would be of much practical interest to investigate the prop-

erties of a model that allows for false alarms and missing observations such as that given in [Storlie et al. \(2009\)](#).

Acknowledgments

The authors are grateful to the reviewers and an associate editor, for their most constructive comments, most of which are incorporated in the current version of this article. The authors' work was supported in part by the National Science Foundation (NSF) under Grants 0707037, 1007543, 1016441, and 1007520.

Appendix A: Model likelihood

Here we present a closed form approximation to the likelihood of the model described in [Section 2](#). We denote the times at which the i^{th} target is observed as $\mathbf{t}_i = (t_{i,1}, \dots, t_{i,n_i})$. Also let $\mathbf{X}_i = (X_i(t_{i,1}), \dots, X_i(t_{i,n_i}))$ and let $\mathcal{X} = (\mathbf{X}_1, \dots, \mathbf{X}_m)$ be the collection of all observed x -coordinates of all targets. Define similarly for $\{\mathbf{Y}_i\}_{i=1}^m$ and \mathcal{Y} . Lastly recall the definition of \mathcal{U} and \mathcal{V} from [Section 2.1](#). Our goal is to obtain an approximate expression for the density, or likelihood, for the collection of random variables $\{\mathcal{U}, \mathcal{V}, \mathcal{X}, \mathcal{Y}\}$ that correspond to the targets. Since \mathcal{X} and \mathcal{Y} are independent given $\{\mathcal{U}, \mathcal{V}\}$, we can write the corresponding density as

$$[\mathcal{U}, \mathcal{V}, \mathcal{X}, \mathcal{Y}] = [\mathcal{U}, \mathcal{V}] \cdot [\mathcal{X} | \mathcal{U}, \mathcal{V}] \cdot [\mathcal{Y} | \mathcal{U}, \mathcal{V}]. \quad (\text{A1})$$

We shall call $[\mathcal{U}, \mathcal{V}]$ the *Event Density* and $\{[\mathcal{X} | \mathcal{U}, \mathcal{V}], [\mathcal{Y} | \mathcal{U}, \mathcal{V}]\}$ the *Location Density*. It is discussed in [Appendix B](#) that the Event Density $[\mathcal{U}, \mathcal{V}]$ can be approximated in closed form by

$$[\mathcal{U}, \mathcal{V}] \approx [N_0] \prod_{j=1}^n [U_{b,j}] \cdot [V_{b,j} | U_{b,j}] \cdot [U_{d,j}] \cdot [V_{d,j} | U_{d,j}] \cdot [U_{s,j}] \cdot [V_{s,j} | U_{s,j}] \cdot [U_{m,j}] \cdot [V_{m,j} | U_{m,j}]. \quad (\text{A2})$$

Depending on the choice of $G_i(t)$ in [\(3\)](#) (e.g., Brownian motion), the Location Density $\{[\mathcal{X} | \mathcal{U}, \mathcal{V}], [\mathcal{Y} | \mathcal{U}, \mathcal{V}]\}$ can also be approximated by a closed form expression; see [Appendix C](#).

Appendix B: Derivation of event density [\(A2\)](#)

Since the Event Model has independent increments, the target Event Density can be written as

$$[\mathcal{U}, \mathcal{V}] = [N_0] \prod_{j=1}^n [U_{b,j}, U_{d,j}, U_{s,j}, U_{m,j} | N(t_j)] \cdot [V_{b,j}, V_{d,j}, V_{s,j}, V_{m,j} | N(t_j), U_{b,j}, U_{d,j}, U_{s,j}, U_{m,j}], \quad (\text{A3})$$

where recall that $N(t)$ is the number of targets that exist at time t . Also, N_0 is the initial number of targets and is Poisson with parameter λ_0 so $[N_0](k) =$

$\lambda_0^k e^{-\lambda_0} / k!$. It is difficult to calculate the exact density for $(U_{b,j}, U_{d,j}, U_{s,j}, U_{m,j} \mid N(t_j))$, as they are dependent on each other. The rate of death $\lambda_d N(t)$, for example, changes when there is a birth, death, split or merger. Suppose $U_j = U_{b,j} + U_{d,j} + U_{s,j} + U_{m,j}$. The exact distribution of $(U_{b,j}, U_{d,j}, U_{s,j}, U_{m,j})$ would require us to sum over all the permutations of the order that the U_j events could happen in the interval $[t_j, t_{j+1})$. For each of these permutations, we would have to calculate the probability that the sum of U_j independent exponential random variables with respective rates (which are generally different) would be less than $\Delta t_j = t_{j+1} - t_j$.

Instead, we will approximate this probability by assuming that the rate of the occurrence of events stays constant during the interval $[t_j, t_{j+1})$. Specifically, we assume that the rate of each of the events during the interval is $\bar{\lambda}_{b,j} = \lambda_b$, $\bar{\lambda}_{d,j} = \lambda_d N(t_j)$, $\bar{\lambda}_{s,j} = \lambda_s N(t_j)$ and $\bar{\lambda}_{m,j} = \lambda_m N(t_j)$ for birth, death, splitting, and merging respectively.

With this approximation, the variables $(U_{b,j}, U_{d,j}, U_{s,j}, U_{m,j})$ are independent and $P(U_{d,j} = u)$ for example is the probability that the sum of u iid exponential random variables with rate $\bar{\lambda}_{d,j}$ are less than Δt_j . This is the same as the Poisson density with parameter $\bar{\lambda}_{d,j} \Delta t_j$ evaluated at u . Hence we have

$$[U_{b,j} \mid N(t_j)](u) \approx (\lambda_b \Delta t_j)^u e^{-\lambda_b \Delta t_j} / u!,$$

$$[U_{d,j} \mid N(t_j)](u) \approx (\bar{\lambda}_{d,j} \Delta t_j)^u e^{-\bar{\lambda}_{d,j} \Delta t_j} / u!,$$

$$[U_{s,j} \mid N(t_j)](u) \approx (\bar{\lambda}_{s,j} \Delta t_j)^u e^{-\bar{\lambda}_{s,j} \Delta t_j} / u!,$$

and

$$[U_{m,j} \mid N(t_j)](u) \approx (\bar{\lambda}_{m,j} \Delta t_j)^u e^{-\bar{\lambda}_{m,j} \Delta t_j} / u!.$$

Now consider the variables $(V_{b,i}, V_{d,i}, V_{s,i}, V_{m,i})$. Under the same approximation that $N(t)$ is constant during the interval $[t_j, t_{j+1})$, we have

$$[V_{b,j} \mid N(t_j), U_{b,j}](v) \approx 1,$$

$$[V_{d,j} \mid N(t_j), U_{d,j}](v) \approx (1/N(t_j))^{U_{d,j}},$$

$$[V_{s,j} \mid N(t_j), U_{s,j}](v) \approx (1/N(t_j))^{U_{s,j}},$$

and

$$[V_{m,j} \mid N(t_j), U_{m,j}](v) \approx \left(1 / \binom{N(t_j)}{2} \right)^{U_{m,j}}.$$

Thus we approximate (A3) as (A2).

Appendix C: Derivation of location density

Recall, that *only* for the purpose of evaluating the likelihood in (7), we use the assumption that $H_i(t) + \sigma G_i(t)$ is a Brownian motion, or equivalently that $H_i(t) = 0$, as stated in Condition 5. In this case, $X_i(t)$ is normally distributed for all t , and the observed location of all targets at all time points, has a multivariate

normal distribution. Recall that the times at which the i^{th} target is observed are denoted by $\mathbf{t}_i = (t_{i,1}, \dots, t_{i,n_i})$, $\mathbf{X}_i = (X_i(t_{i,1}), \dots, X_i(t_{i,n_i}))$ and $\mathcal{X} = (\mathbf{X}_1, \dots, \mathbf{X}_m)$. Then, $\mathcal{X} \sim N(\mu_X, \Sigma_X)$.

Recall from Section 2.2 that this mean and covariance will depend on the time of initiation, ξ , of the targets. The event variables \mathcal{U} and \mathcal{V} do not specify the exact values of ξ_i and ζ_i . They do however specify the interval between observations they are in. For the following, if it is known that ξ_i is in the interval (t_j, t_{j+1}) , we will set $\xi_i = t_j + \Delta t_j/2$.

Since μ_X and Σ_X depend on the exact values of ξ , this will be an approximation to the true density. In order to get the exact density, we would have to integrate out on the joint distribution of \mathcal{X} and ξ , given that the ξ_i 's are in their respective intervals. This would have to be done numerically and would not be feasible in practice. If the Δt_j are sufficiently small though, this approximation will be quite close to the true density.

Also recall from Section 2.2 that we need to then condition \mathcal{X} on the random variable D and evaluate this density at $D = 0$. Let d_i be the vector d defined in the last paragraph of Section 2.2.2 for the i^{th} merging event, $i = 1, \dots, N_d$. Then let

$$D_i = X_{d_{i,1}}(\xi_{d_{i,3}}) - X_{d_{i,2}}(\xi_{d_{i,3}}) + \psi_i \quad (\text{A4})$$

be the difference, D , for the i^{th} merging event. The random variable $\mathbf{D} = (D_1, \dots, D_{N_d})$ is also normally distributed, $\mathbf{D} \sim N(\mu_D, \Sigma_D)$. For the collection of both \mathcal{X} and \mathbf{D} we have $(\mathcal{X}, \mathbf{D})' \sim N(\mu, \Sigma)$, where

$$\mu = \begin{pmatrix} \mu_X \\ \mu_D \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_X & \Sigma_{X,D} \\ \Sigma'_{X,D} & \Sigma_D \end{pmatrix}. \quad (\text{A5})$$

The conditional distribution of \mathcal{X} given $\mathbf{D} = 0$, which we will just call the distribution of \mathcal{X} from this point forward, is given by the density $[\mathcal{X} | D = 0](x) = \phi(x; \mu^*, \Sigma^*)$, where $\phi(x; \mu^*, \Sigma^*)$ is the multivariate normal density with $\mu^* = \mu_X - \Sigma_{X,D} \Sigma_D^{-1} \mu_D$ and $\Sigma^* = \Sigma_X - \Sigma_{X,D} \Sigma_D^{-1} \Sigma'_{X,D}$.

In Section 3 we assume a Brownian motion for the random process $G_i(t)$ in (3). Using the Markov property the means and the covariances in (A5) can be well approximated by (A8) and (A9) below. If $G_i(t)$ is modeled by an integrated Brownian motion, the corresponding approximations for the means and covariances are given in Storlie et al. (2009).

Appendix D: Technical Lemmas

Here we present the Lemmas used in the proofs of Appendix E. All of the results in this section assume the Brownian motion model for $G_i(t)$ in (3). We begin with the Law of Iterated Logarithm for two-dimensional Brownian motion. The proof of Lemma 1 for Brownian motion in d dimensions can be found in Kallenberg (2002). Let $\|\cdot\|$ denote the Euclidean norm.

Lemma 1 (Law of the Iterated Logarithm). *For B a Brownian motion in \mathfrak{R}^2 ,*

$$P \left\{ \overline{\lim}_{t \downarrow 0} \frac{\|B(t)\|}{(2t \log \log(1/t))^{1/2}} = 1 \right\} = 1. \tag{A6}$$

Lemma 2 is Levy’s Modulus of Continuity for two-dimensional Brownian motion. Its proof follows the same general argument as for the one dimensional case given in Revuz & Yor (1999).

Lemma 2 (Levy’s Modulus of Continuity). *For B a Brownian motion in \mathfrak{R}^2 ,*

$$P \left\{ \overline{\lim}_{\varepsilon \rightarrow 0} \left(\sup_{\substack{0 \leq t_1 \leq t_2 \leq 1 \\ t_2 - t_1 \leq \varepsilon}} \frac{\|B(t_2) - B(t_1)\|}{(2\varepsilon \log(1/\varepsilon))^{1/2}} \right) = 1 \right\} = 1. \tag{A7}$$

For the purposes of comparing likelihoods, we need a convenient form for the location density. This will require the following Lemma. To write the overall location density, each observation is conditioned on the previous observations and all of the D_j variables from the mergers. We then take the product of all these conditional densities. Let $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,n_i})$ and let $\mathcal{F}_{i,j} = (X_{i,1}, \dots, X_{i,j}, \mathbf{X}_{i-1}, \dots, \mathbf{X}_1, D_1, \dots, D_{N_m}, \mathcal{U}, \mathcal{V})$. As usual let $t_{i,j}$ be the j^{th} time at which the i^{th} target is observed. Notice that we can write the x component of the likelihood as

$$[\mathcal{X} \mid \mathcal{U}, \mathcal{V}, \mathcal{P}] = \prod_{i=1}^m \prod_{j=1}^{n_i} [X_{i,j} \mid \mathcal{F}_{i,j-1}]. \tag{A8}$$

So we will need to give a convenient expression for $[X_{i,j} \mid \mathcal{F}_{i,j-1}]$.

Lemma 3. *Under Conditions 1-4 of Theorem 1, and further that $H_i(t) = 0$, for $j = 2, \dots, n_i$, the distribution of $X_{i,j}$ given $\mathcal{F}_{i,j-1}$ is Gaussian with mean and variance given respectively by*

$$\mu = X_i(t_{i,j-1}) + O(k^{-1}) \quad \text{and} \quad \sigma^2 = \Delta t_{i,j} \sigma^2 + O(k^{-2}). \tag{A9}$$

Proof. We can write the density of $X_{i,j}$ given $\mathcal{F}_{i,j-1}$ as

$$[X_{i,j} \mid \mathcal{F}_{i,j-1}](x_{i,j} \mid f_{i,j-1}) = \int_{-\infty}^{\infty} [X_{i,j}, X_i(\zeta) \mid \mathcal{F}_{i,j-1}](x_{i,j}, x_{i,\zeta} \mid f_{i,j-1}) dx_{i,\zeta}, \tag{A10}$$

where ζ is the termination time of the i^{th} target. But

$$\begin{aligned} & [X_{i,j}, X_i(\zeta) \mid \mathcal{F}_{i,j-1}](x_{i,j}, x_{i,\zeta} \mid f_{i,j-1}) \\ &= [X_{i,j} \mid X_i(\zeta), \mathcal{F}_{i,j-1}](x_{i,j} \mid x_{i,\zeta}, f_{i,j-1}) \cdot [X_i(\zeta) \mid \mathcal{F}_{i,j-1}](x_{i,\zeta} \mid f_{i,j-1}) \\ &= [X_{i,j} \mid X_{i,j-1}, X_i(\zeta)](x_{i,j}, x_{i,j-1}, x_{i,\zeta}) \cdot [X_i(\zeta) \mid \mathcal{F}_{i,j-1}](x_{i,\zeta}, f_{i,j-1}). \end{aligned}$$

For convenience of notation, drop the subscript i on $t_{i,j}$ to let $t_j = t_{i,j}$. Now the conditional distribution of $X_{i,j}$ given $X_{i,j-1} = x_{i,j-1}$ and $X_i(\zeta) = x_{i,\zeta}$ can be

shown to be Gaussian with mean and variance given respectively by

$$\mu = \frac{\zeta - t_j}{\zeta - t_{j-1}} x_{i,j-1} + \frac{\Delta t_j}{\zeta - t_{j-1}} x_{i,\zeta} \quad \text{and} \quad \tau^2 = \Delta t_j \sigma^2 - \frac{(\Delta t_j)^2}{\zeta - t_{j-1}} \sigma^2.$$

Also the conditional distribution of $X_i(\zeta)$ given $\mathcal{F}_{i,j-1} = f_{i,j-1}$ is Gaussian with mean and variance denoted by μ_{x_ζ} and $\sigma_{x_\zeta}^2$. Notice that we never observe $X_i(\zeta)$ so this conditional distribution has a variance $\sigma_{x_\zeta}^2 > 0$. We can then write the integral in (A10) as

$$C \int_{-\infty}^{\infty} \exp \left\{ -\frac{\left(x_j - \frac{\zeta - t_j}{\zeta - t_{j-1}} x_{j-1} + \frac{\Delta t_j}{\zeta - t_{j-1}} x_\zeta \right)^2}{2 \left(\Delta t_j \sigma^2 - \frac{(\Delta t_j)^2}{\zeta - t_{j-1}} \sigma^2 \right)} - \frac{(x_\zeta - \mu_{x_\zeta})^2}{2 \sigma_{x_\zeta}^2} \right\} dx_\zeta.$$

Following some tedious algebra we can see that the above expression is equal to

$$C' \exp \left\{ -\frac{(x_j - x_{j-1} + O(\Delta t_j))^2}{2(\sigma^2 \Delta t_j - O(\Delta t_j)^2)} \right\}$$

which we can recognize as the density of a Normal distribution with mean $x_{j-1} + O(\Delta t_j)$ and variance $\sigma^2 \Delta t_j - O(\Delta t_j)^2$. Now using Condition 1 gives the desired result. \square

The following Lemma is a version of the main Theorem presented in Shepp (1966). We will say that two measures μ_1 and μ_2 are equivalent if they have the same sets of measure zero. We will denote this $\mu_1 \sim \mu_2$. This Theorem gives necessary and sufficient conditions for a measure μ_X imposed by a Gaussian process $X(t)$ on $[0, T]$ to be equivalent to the measure imposed by a Brownian motion on $[0, T]$, μ_B . Let $m(t) = EX(t)$ and $\gamma(s, t) = E(X(s) - \mu(s))(X(t) - \mu(t))$.

Lemma 4 (Shepp’s Theorem). *Assume that $(\partial/\partial s)\gamma(s, t)$ is continuous for $0 \leq s \leq T, s \neq t$. Then $\mu_X \sim \mu_B$ if and only if*

$$(\partial/\partial s)\gamma(s, s^+) - (\partial/\partial s)\gamma(s, s^-) \equiv 1, \quad \text{for } 0 < s < T \tag{A11}$$

and there exists a function $k \in L^2$ for which

$$m(t) = \int_0^t k(u) du. \tag{A12}$$

The function k is unique and is given by $k(t) = m'(t)$ for almost every t .

We will now use Shepp’s Theorem to show that the measure imposed by a Brownian motion conditioned on $\{D_j : 1 \leq j \leq N_m\}$ is equivalent to μ_B .

Lemma 5. *Assume that $G_i(t)$ is the Brownian motion associated with path i in (3). Consider the measure, $\tilde{\mu}$ imposed by G_i given $(D_1 = 0, \dots, D_{N_m} = 0)$ where D_j is the difference between parent paths at the time of merger plus an error as defined in (A4). Then $\tilde{\mu} \sim \mu_B$.*

Proof. Without loss of generality (WLOG) we assume that the birth and death times are $\xi_i = 0$ and $\zeta_i = T$ so that G_i is on $[0, T]$. Let $m(t)$ and $\gamma(s, t)$ be the mean and covariance functions corresponding to $\tilde{\mu}$. We need to first show that $(\partial/\partial s)\gamma(s, t)$ is continuous. Then show that conditions (A11) and (A12) hold. But the second condition is trivially satisfied since $m(t) = 0$. So we just need to show continuity and (A11).

Let $s \leq t$ and let Σ be the covariance matrix for the vector $(G_i(s), G_i(t), D_1, \dots, D_{N_m})'$. Also, let Σ_1 and Σ_2 , be the covariance matrices for $(G_i(s), G_i(t))'$ and $(D_1, \dots, D_{N_m})'$ respectively. Finally let Σ_{12} be the matrix that contains the pairwise covariances of the elements in $(G_i(s), G_i(t))'$ with those in $(D_1, \dots, D_{N_m})'$ as its elements so that

$$\Sigma = \left(\begin{array}{c|c} \Sigma_1 & \Sigma_{12} \\ \hline \Sigma_{12}' & \Sigma_2 \end{array} \right).$$

Notice that

$$\Sigma_1 = \begin{pmatrix} s & s \\ s & t \end{pmatrix} \quad \text{and} \quad \Sigma_{12} = \begin{pmatrix} f_1(s) & f_2(s) & \cdots & f_{N_m}(s) \\ f_1(t) & f_2(t) & \cdots & f_{N_m}(t) \end{pmatrix}$$

for functions $f_j(t) = \text{Cov}(G_i(t), D_j)$, $j = 1, \dots, N_m$. The matrix Σ_2 is positive definite so let $\mathbf{v}_1, \dots, \mathbf{v}_{N_m}$ and $\lambda_1, \dots, \lambda_{N_m}$ be its eigenvectors and eigenvalues respectively. We can write the covariance matrix of $(B(s), B(t))'$ conditional on $(D_1, \dots, D_{N_m})'$ as $\Sigma^* = \Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'$. The off diagonal element of the matrix $\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'$ is given by

$$\begin{aligned} & (f_1(s), f_2(s), \dots, f_{N_m}(s)) \Sigma_2^{-1} \begin{pmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_{N_m}(t) \end{pmatrix} \\ &= \sum_{j=1}^{N_m} \frac{1}{\lambda_j} (f_1(s), f_2(s), \dots, f_{N_m}(s)) \mathbf{v}_j \mathbf{v}_j' \begin{pmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_{N_m}(t) \end{pmatrix} \\ &= \sum_{j=1}^{N_m} \sum_{k=1}^{N_m} \sum_{l=1}^{N_m} \frac{1}{\lambda_j} v_{jk} v_{jl} f_k(s) f_l(t). \end{aligned} \tag{A13}$$

This gives $\gamma(s, t) = \text{Cov}(G_i(s), G_i(t) \mid (D_1, \dots, D_{N_m}))$ in general as

$$\gamma(s, t) = s \wedge t - \sum_{j=1}^{N_m} \sum_{k=1}^{N_m} \sum_{l=1}^{N_m} \frac{1}{\lambda_j} v_{jk} v_{jl} f_k(s) f_l(t).$$

Hence we have

$$\frac{\partial}{\partial s} \gamma(s, t) = I_{\{s \leq t\}} - \sum_{j=1}^{N_m} \sum_{k=1}^{N_m} \sum_{l=1}^{N_m} \frac{1}{\lambda_j} v_{jk} v_{jl} f_k'(s) f_l(t). \tag{A14}$$

We claim that $f_j(t) = c_j t$ for some constant c_j . To see this notice that $D_j = X_{d_{j,1}}(\zeta_{d_{j,1}}) - X_{d_{j,2}}(\zeta_{d_{j,2}}) + \psi_j$ where the vector d_j is defined in (A4) so that $X_{d_{j,1}}$ and $X_{d_{j,2}}$ are the parents of a merger and $X_{d_{j,3}}$ is the child. So we have $f_j(t) = \text{Cov}(G_i(t), X_{d_{j,1}}(\zeta_{d_{j,1}})) - \text{Cov}(G_i(t), X_{d_{j,2}}(\zeta_{d_{j,2}}))$.

Now consider a path X_j . If X_j did not result from a birth then we can write $X_j(t)$ as the position of its parent(s) at time ξ_j plus $G_j(t)$. We can continue breaking the parent paths up in this same way until we have $X_j(\zeta_j) = c_1 G_{j_1}(\zeta_{j_1}) + \dots + c_m G_{j_m}(\zeta_{j_m}) + Y_j$, where the c_k are constants and Y_j is the sum of random variables such as $\psi_{m,k}$'s, $\psi_{s,k}$'s, and $X_k(\xi_k)$'s which are initial locations of paths resulting from birth, all of which are independent of G_i . All of the G_{j_k} are also independent of G_i unless $j_{k'} = i$ for some k' . In that case we have $\text{Cov}(G_i(t), X_j(\zeta_j)) = \text{Cov}(G_i(t), c_{k'} G_i(\zeta_i)) = c_{k'}(t \wedge \zeta_i) = c_{k'} t$ since $t \leq T = \zeta_i$. If there is no such k' so that $j_{k'} = i$ then $\text{Cov}(G_i(t), X_j) = 0$. Hence $f_j(t) = c_j t$ for some constant c_j possibly equal to zero.

This means that $\frac{\partial}{\partial s} \gamma(s, t)$ is continuous unless $s = t$ and also

$$\frac{\partial}{\partial s} \gamma(s, s^+) = 1 - \sum_{j=1}^{N_m} \sum_{k=1}^{N_m} \sum_{l=1}^{N_m} \frac{1}{\lambda_i} v_{jk} v_{jl} f'_k(s) f_l(s)$$

and

$$\frac{\partial}{\partial s} \gamma(s, s^-) = - \sum_{j=1}^{N_m} \sum_{k=1}^{N_m} \sum_{l=1}^{N_m} \frac{1}{\lambda_i} v_{jk} v_{jl} f'_k(s) f_l(s)$$

which satisfies (A11) and completes the proof. □

The next Lemma is needed to compare the location densities of two different sequences of solutions $(\hat{U}, \hat{V}, \hat{P})_k$ and $(\tilde{U}, \tilde{V}, \tilde{P})_k$. It gives a bound for the ratio of two location densities for two track segments with different $\mathcal{F}_{i,j}$ variables. Let $\mathcal{F}_i(t) = \mathcal{F}_{i,j'}$ for $j' = \max\{j : t_{i,j} \leq t\}$.

Lemma 6. *Assume Conditions 1-4 of Theorem 1, and further that $H_i(t) = 0$. Let Θ be the set of all pairs of tracking solutions sequences, $\hat{\theta}_k = (\hat{U}, \hat{V}, \hat{P})_k$ and $\tilde{\theta}_k = (\tilde{U}, \tilde{V}, \tilde{P})_k$ that have one of the differences 1a, 1b, 2, 3, 5a, 5b or 6b from the propositions. Consider a track segment $(x_1, \dots, x_{n'_k})$ such that*

$$(\tilde{X}_{i_1}(t_{j'}), \dots, \tilde{X}_{i_1}(t_{j'+n'_k})) = (\hat{X}_{i_2}(t_{j'}), \dots, \hat{X}_{i_2}(t_{j'+n'_k})) = (x_1, \dots, x_{n'_k}) \text{ for all } k$$

for some i_1 and i_2 . If $\tilde{X}_{i_1}(t_{j'})$ and $\hat{X}_{i_2}(t_{j'})$ are both the first observation of a track in their respective solutions then

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[\tilde{X}_{i_1}(t_{j'}) \mid \tilde{\mathcal{F}}_{i_1}(t_{j'-1})](x_1)}{[\hat{X}_{i_2}(t_{j'}) \mid \hat{\mathcal{F}}_{i_2}(t_{j'-1})](x_1)} = O(1) \text{ as } k \rightarrow \infty \text{ a.s.} \tag{A15}$$

and

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{\prod_{j=2}^{n'_k} [\tilde{X}_{i_1}(t_{j'+j-1}) \mid \tilde{\mathcal{F}}_{i_1}(t_{j'+j-2})](x_j)}{\prod_{j=2}^{n'_k} [\hat{X}_{i_2}(t_{j'+j-1}) \mid \hat{\mathcal{F}}_{i_2}(t_{j'+j-2})](x_j)} \leq a(\log(k))^b \text{ as } k \rightarrow \infty \text{ a.s.} \tag{A16}$$

for some constants a and b which depend on ω .

If neither $\tilde{X}_{i_1}(t_{j'})$ and $\hat{X}_{i_2}(t_{j'})$ are the first observation of a track in their respective solutions then

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{\prod_{j=1}^{n'_k} [\tilde{X}_{i_1}(t_{j'+j-1}) \mid \tilde{\mathcal{F}}_{i_1}(t_{j'+j-2})](x_j)}{\prod_{j=1}^{n'_k} [\hat{X}_{i_2}(t_{j'+j-1}) \mid \hat{\mathcal{F}}_{i_1}(t_{j'+j-2})](x_j)} \leq a(\log(k))^b \text{ as } k \rightarrow \infty \text{ a.s.} \tag{A17}$$

for some constants a and b which depend on ω .

Proof. For (A15) consider the case where target i_1 is a birth. Then

$$[\tilde{X}_{i_1}(t_{j'}) \mid \tilde{\mathcal{F}}_{i_1}(t_{j'-1})](x_1) \leq \phi(0; 0, \sigma_{X_0}^2) \tag{A18}$$

which is the mode of the normal density for the initial position of a target resulting from birth. This is true for any θ . Similarly if target i_1 is the result of a split or a merger, then

$$[\tilde{X}_{i_1}(t_{j'}) \mid \tilde{\mathcal{F}}_{i_1}(t_{j'-1})](x_1) \leq \phi(0; 0, \sigma_{X_s}^2)$$

and

$$[\tilde{X}_{i_1}(t_{j'}) \mid \tilde{\mathcal{F}}_{i_1}(t_{j'-1})](x_1) \leq \phi(0; 0, \sigma_{X_m}^2), \tag{A19}$$

respectively. Let C_1 be the maximum of the three quantities in (A18) and (A19).

Now for any Brownian path on a finite interval $P(\sup_{0 \leq t \leq T} |B(t)| < \infty) = 1$. Hence for every ω in a set with probability 1, $\sup_{0 \leq t \leq T} |B(t)| \leq M(\omega) < \infty$. So we know that x_1 is less than a constant for all ω . Also, $[\hat{X}_{i_2}(t_{j'}) \mid \hat{\mathcal{F}}_{i_1}(t_{j'-1})](x) > 0$ for $-\infty < x < \infty$ since for any of the three cases (target i_2 resulting from birth, split, or merger) it has support over the entire real line. Thus

$$\inf_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} [\hat{X}_{i_2}(t_{j'}) \mid \hat{\mathcal{F}}_{i_1}(t_{j'-1})](x_j) \geq C_2(\omega) > 0$$

and finally

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[\tilde{X}_{i_1}(t_{j'}) \mid \tilde{\mathcal{F}}_{i_1}(t_{j'-1})](x_1)}{[\hat{X}_{i_2}(t_{j'}) \mid \hat{\mathcal{F}}_{i_1}(t_{j'-1})](x_1)} \leq \frac{C_1}{C_2}$$

which gives the first result.

Now consider the ratios in (A16) and (A17). We only need to show (A16) since they are equivalent if in (A17) we relabel x_1 as x_2 and so on. By Lemma 3 can write the ratio in (A16) as

$$\frac{\prod_{j=2}^{n'_k} [\tilde{X}_{i_1}(t_{j'+j-1}) \mid \tilde{\mathcal{F}}_{i_1}(t_{j'+j-2})](x_j)}{\prod_{j=2}^{n'_k} [\hat{X}_{i_2}(t_{j'+j-1}) \mid \hat{\mathcal{F}}_{i_1}(t_{j'+j-2})](x_j)} = \left(\prod_{j=2}^{n_k} A_j \right) \cdot B$$

where

$$A_j = \sqrt{\frac{2\pi\Delta t_{j'}^k \hat{\sigma}^2 + O(k^{-2})}{2\pi\Delta t_{j'}^k \tilde{\sigma}^2 + O(k^{-2})}}, \tag{A20}$$

$$B = \exp \left\{ \sum_{j=2}^{n_k} \left(-\frac{(x_j - x_{j-1} + O(k^{-1}))^2}{2\Delta t_j^k \tilde{\sigma}^2 + O(k^{-2})} + \frac{(x_j - x_{j-1} + O(k^{-1}))^2}{2\Delta t_{j'}^k \hat{\sigma}^2 + O(k^{-2})} \right) \right\}$$

and all of the $O(h)$ are possibly different functions that tend to zero like ch as $k \rightarrow \infty$.

We will need to calculate a bound for how much different the estimates $\tilde{\sigma}^2$ and $\hat{\sigma}^2$ can be. Notice that in the conditions of any of the propositions, the sets of variables \tilde{D} and \hat{D} used to calculate $\tilde{\sigma}^2$ and $\hat{\sigma}^2$ can be different only by one variable.

Now all of the $\tilde{D}_{i,j} = \hat{D}_{i,j}$ except for possibly one, and $|\hat{N} - \tilde{N}| \leq 1$. Assume for the moment that $\tilde{N} = \hat{N} - 1$ and that $\hat{D}_{i',j'}$ does not appear in the estimate $\tilde{\sigma}^2$. Then

$$\begin{aligned} \hat{\sigma}^2 - \tilde{\sigma}^2 &= \left[(\hat{N} - 1) \sum_{i=1}^{\hat{m}} \sum_{j=1}^{\hat{n}_i} \hat{D}_{i,j}^2 - \tilde{N} \sum_{i=1}^{\tilde{m}} \sum_{j=1}^{\tilde{n}_i} \tilde{D}_{i,j}^2 \right] / (\hat{N}(\hat{N} - 1)) \\ &= - \left(\sum_{i=1}^{\tilde{m}} \sum_{j=1}^{\tilde{n}_i} \tilde{D}_{i,j}^2 \right) / (\hat{N}(\hat{N} - 1)) + \hat{D}_{i',j'}^2 / \hat{N} \\ &= -\tilde{\sigma}^2 / \hat{N} + \hat{D}_{i',j'}^2 / \hat{N} \end{aligned}$$

so that $-\tilde{\sigma}^2 / \hat{N} \leq \hat{\sigma}^2 - \tilde{\sigma}^2 \leq \hat{D}_{i',j'}^2 / \hat{N}$, but $\hat{D}_{i',j'}^2 \leq c\sqrt{\log \log \hat{N}}$ by Condition 7 and $\tilde{\sigma}^2 \leq K$ for some constant $K < \infty$ by Condition 6. So we have $K / \hat{N} \leq \hat{\sigma}^2 - \tilde{\sigma}^2 \leq c\sqrt{\log \log \hat{N}} / \hat{N}$. Then assuming that $\tilde{N} = \hat{N} + 1$ gives us the reverse inequality which gives us $|\hat{\sigma}^2 - \tilde{\sigma}^2| \leq c\sqrt{\log \log \hat{N}} / \hat{N}$.

Now consider the quantity A_j in (A20) we have

$$\begin{aligned} A_j &= \sqrt{\frac{\hat{\sigma}^2 + O(k^{-1})}{\tilde{\sigma}^2 + O(k^{-1})}} = \sqrt{1 + \frac{\hat{\sigma}^2 - \tilde{\sigma}^2 + O(k^{-1})}{\tilde{\sigma}^2 + O(k^{-1})}} \\ &\leq \sqrt{1 + c_2 \sqrt{\log \log \hat{N}} / \hat{N}} \leq \sqrt{1 + (c_2 \log \log \hat{N}) / \hat{N}} \end{aligned}$$

since $\hat{N} \leq ck$. But $\hat{N} \geq n'_k - 1$ so

$$\prod_{j=2}^{n'_k} A_j \leq \left(1 + (c_2 \log \log \hat{N}) / \hat{N} \right)^{\hat{N}/2}.$$

Now if $n_k \rightarrow \infty$ as $k \rightarrow \infty$, then

$$\frac{\left(1 + (c_2 \log \log \hat{N}) / \hat{N} \right)^{\hat{N}/2}}{(\log \hat{N})^{c_2/2}} \rightarrow 1$$

so that $\prod_{j=2}^{n_k} A_j \leq (\log \hat{N})^{c_2/2}$ eventually which means that

$$\prod_{j=2}^{n_k} A_j \leq c_3(\log k)^{c_4}. \tag{A21}$$

Of course if n'_k is bounded then $\prod_{j=2}^{n_k} A_j$ is also bounded.

Now for B notice that $(x_j - x_{j-1} + O(k^{-1}))^2 = (x_j - x_{j-1})^2 + (x_j - x_{j-1})O(k^{-1}) + O(k^{-2})$,

$$\frac{2\Delta t_j^k \hat{\sigma}^2}{2\Delta t_j^k \hat{\sigma}^2 + O(k^{-2})} = 1 + O(k^{-1}),$$

and

$$\frac{2\Delta t_j^k \tilde{\sigma}^2}{2\Delta t_j^k \tilde{\sigma}^2 + O(k^{-2})} = 1 + \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2 + O(k^{-1})} \leq 1 + \left(c_1 \sqrt{\log \log \hat{N}} \right) / \hat{N}.$$

So we can write B as

$$\begin{aligned} & \exp \left\{ \sum_{j=2}^{n_k} \left(-\frac{(x_j - x_{j-1})^2 + (x_j - x_{j-1})O(\frac{1}{k}) + O(\frac{1}{k^2})}{2\Delta t_j^k \hat{\sigma}^2} \left(1 + O\left(\frac{1}{k}\right) \right) \right. \right. \\ & \quad \left. \left. + \frac{(x_j - x_{j-1})^2 + (x_j - x_{j-1})O(\frac{1}{k}) + O(\frac{1}{k^2})}{2\Delta t_j^k \tilde{\sigma}^2} \left(1 + O\left(\sqrt{\log \log \hat{N}/\hat{N}}\right) \right) \right) \right\} \\ & = \exp \left\{ \sum_{j=2}^{n_k} \left(\frac{(x_j - x_{j-1})^2 O(\frac{1}{k}) + (x_j - x_{j-1})O(\frac{1}{k}) + (x_j - x_{j-1})^2 O(\sqrt{\log \log \hat{N}/\hat{N}}) + O(\frac{1}{k^2})}{2\Delta t_j^k \hat{\sigma}^2} \right) \right\}. \end{aligned} \tag{A22}$$

Now this track segment is from a solution from the conditions of one of Propositions 1a, 1b, 2, 3, 5a, 5b, 6a, or 6b so all of the tracks are correct tracks segments. Hence all of the (x_{j-1}, x_j) pairs are consecutive observations from a scaled Brownian path. Therefore

$$\frac{(x_j - x_{j-1})}{\sqrt{\Delta t_j^k}} \sim \mathcal{N}(0, \sigma^2).$$

Thus if $n'_k \rightarrow \infty$ and remember that $n'_k \leq ck$, then for any solutions $\tilde{\theta}$ and $\hat{\theta}$, B is no more than

$$\begin{aligned} & \exp \left\{ \left(\frac{1}{n'_k} \sum_{j=2}^{n'_k} \frac{(x_j - x_{j-1})^2}{\Delta t_j^k} \right) \frac{O(\frac{1}{k})n'_k}{2\hat{\sigma}^2} + \left(\sum_{j=2}^{n'_k} \frac{(x_j - x_{j-1})}{\sqrt{2\Delta t_j^k n'_k \log \log n'_k}} \right) \frac{O(\frac{1}{k})\sqrt{2n'_k \log \log n'_k}}{2\sqrt{\Delta t_j^k \hat{\sigma}^2}} \right. \\ & \quad \left. + \left(\sum_{j=2}^{n'_k} \frac{(x_j - x_{j-1})}{\sqrt{2\Delta t_j^k n'_k \log \log n'_k}} \right) \frac{O(\sqrt{\log \log \hat{N}/\hat{N}})\sqrt{2n'_k \log \log n'_k}}{2\sqrt{\Delta t_j^k \hat{\sigma}^2}} + \sum_{j=2}^{n'_k} O\left(\frac{1}{k}\right) \right\}. \end{aligned} \tag{A23}$$

The first term in the exponent is $O(1)$ a.s. by the strong law of large numbers. The second term is less than $(C_2 \log \log k)$ a.s. by the Law of the Iterated Logarithm and the fact that $n_k \leq ck$ for some constant c by Condition 1. The third term is less than $(C_3 \sqrt{k/\hat{N}} \log \log k)$ a.s. also by the Law of the Iterated Logarithm and the fact that $\hat{N} \leq ck$ and $n_k \leq ck$ for some constant c . The last term is $O(1)$ because we are adding up less than k terms which are all $O(k^{-1})$. Hence we are left with

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} B \leq \exp \left\{ C_1 + C_2 \log \log k + C_3 \sqrt{\frac{k}{\hat{N}}} \log \log k \right\}.$$

But in the propositions, we are restricting the solutions to have at most a total of M tracks. This means that for some constant c , $\hat{N} \geq ck$ for any solution, since at least one track must be accumulating observations as $k \rightarrow \infty$. Thus

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} B \leq \exp \{ C'_1 + C'_2 \log \log k \} = a(\log k)^b$$

for some constants a and b . Combining this with (A21) gives the desired result. □

Lemma 7 states that the probability that any of the two-dimensional Brownian motion paths will intersect at any time t in a finite interval is zero. This Lemma is an immediate consequence of Proposition 1.4.1 on page 353 of Khoshnevisan (2002).

Lemma 7. Consider any two (x, y) paths (X_1, Y_1) and (X_2, Y_2) from the location model in Section 2.2 where $G_1(t)$ and $G_2(t)$ are independent Brownian motions. Then,

$$P \left\{ \inf_{0 \leq t \leq 1} \|(X_1(t), Y_1(t)) - (X_2(t), Y_2(t))\| > 0 \right\} = 1. \tag{A24}$$

Lastly, Lemma 8 states that our robust estimate of the Brownian motion variance term is consistent.

Lemma 8. Assume Conditions 1-4 of Theorem 1, and further that $H_i(t) = 0$. The estimate $\hat{\sigma}^2$ given by (10) is strongly consistent as $k \rightarrow \infty$ when applied to the any sequence tracking solutions, $(\hat{U}, \hat{V}, \hat{P})_k$, that has less than M tracks and its tracks are made up entirely of correct track segments.

Proof. Recall

$$\hat{\sigma}_k^2 = \sum_{i=1}^m \sum_{j=1}^{n_i^k} \frac{(X_i(t_{i,j+1}^k) - X_i(t_{i,j}^k))^2}{\Delta t_j^k} I_{E_{i,j}}$$

where $N = \sum_{i=1}^m n_i^k$. Under the the assumption that all of the tracks are correct track segments, for each of the terms

$$\frac{X_i(t_{i,j+1}) - X_i(t_{i,j})}{\sqrt{\Delta t_j}} \sim \mathcal{N}(0, \sigma^2).$$

and they are independent. If we let $(Z_j^*)^2 = Z_j^2 I_{E_{i,j}}$ then we can write $\hat{\sigma}_k^2$ as

$$\frac{1}{N} S_N^* = \frac{1}{N} \sum_{j=1}^N Z_j^2 I_{E_{i,j}} = \frac{1}{N} \sum_{j=1}^N (Z_j^*)^2$$

where $Z_j \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$. There are a finite number of tracks in $(\hat{U}, \hat{V}, \hat{P})$ which means $N \geq c_1 k$ for some positive constant c_1 , so $N \rightarrow \infty$ as $k \rightarrow \infty$. Hence it is sufficient to show that $\frac{1}{N} S_N^* \rightarrow \sigma^2$ as $N \rightarrow \infty$ a.s.

Let $u_n = \lfloor \alpha^n \rfloor$, for $\alpha > 1$ which is fixed for now. We first need to show that

$$\sum_{n=1}^{\infty} P \left(\left| \frac{S_{u_n}^* - \mathbb{E} S_{u_n}^*}{u_n} \right| > \varepsilon \right) < \infty. \tag{A25}$$

But

$$\text{Var}(S_n^*) = \sum_{k=1}^n \text{Var}((Z_k^*)^2) \leq \sum_{k=1}^n \mathbb{E}((Z_k^*)^4) = n \mathbb{E}(Z_j^4 I_{E_{i,j}}) \leq n \mathbb{E}(Z_j^4) = 2n\sigma^2.$$

So it follows by Chebyshev's Inequality that the sum in (A25) is no more than

$$\sum_{n=1}^{\infty} \frac{\text{Var}(S_{u_n}^*)}{\varepsilon^2 u_n^2} \leq \sum_{n=1}^{\infty} \frac{2u_n \sigma^2}{\varepsilon^2 u_n^2}.$$

Now $y \leq 2\lfloor y \rfloor$ for $y \geq 1$ so we have that

$$\sum_{n=1}^{\infty} \frac{1}{u_n} \leq \sum_{n=1}^{\infty} \frac{1}{u_n} 2\alpha^{-n} = K < \infty$$

and finally

$$\sum_{n=1}^{\infty} P \left(\left| \frac{S_{u_n}^* - \mathbb{E} S_{u_n}^*}{u_n} \right| > \varepsilon \right) < \frac{2\sigma^2}{\varepsilon^2} K < \infty.$$

Hence by the first Borel-Cantelli Lemma, we have

$$P \left(\left| \frac{S_{u_n}^* - \mathbb{E} S_{u_n}^*}{u_n} \right| > \varepsilon \text{ eventually} \right) = 0$$

for all $\varepsilon > 0$. Taking a union over positive rational ε gives us that $(S_{u_n}^* - \mathbb{E} S_{u_n}^*)/u_n \rightarrow 0$ a.s. But $\mathbb{E} S_n^*/n = \mathbb{E}(Z_j^*)^2 \rightarrow \mathbb{E} Z_j^2 = \sigma^2$, so that $S_{u_n}^*/u_n \rightarrow \sigma^2$ a.s.

If $u_n \leq k \leq u_{n+1}$, then

$$\frac{u_n}{u_{n+1}} \frac{S_{u_n}^*}{u_n} \leq \frac{S_k^*}{k} \leq \frac{u_{n+1}}{u_n} \frac{S_{u_{n+1}}^*}{u_{n+1}}$$

but $u_{n+1}/u_n \rightarrow \alpha$ so

$$\frac{1}{\alpha} \sigma^2 \leq \underline{\lim} \frac{S_k^*}{k} \leq \overline{\lim} \frac{S_k^*}{k} \leq \alpha \sigma^2 \text{ a.s.}$$

This is true for all $\alpha > 1$. Now take $\alpha \downarrow 1$, rational, and we have

$$\lim_{k \rightarrow \infty} \hat{\sigma}_k^2 = \lim_{N \rightarrow \infty} \frac{S_N^*}{N} = \sigma^2 \text{ a.s.}$$

□

Corollary 1. *The estimate $\hat{\sigma}^2$ given by (10) is strongly consistent as $k \rightarrow \infty$ when applied to the correct sequence of tracking solutions, $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$.*

Proof. Immediate consequence of Lemma 8.

□

Appendix E: Propositions

In this section we introduce several propositions that are used to prove the main theorem of Section 3. As mentioned before, this is accomplished by computing the likelihood ratio of the best possible sequence of alternatives, $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$, to that of the correct solution sequence, $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$.

Before we can compare likelihoods, we must first consider how an alternative solution can differ from the correct one. To this end we will make use of the following notation. Recall that $t_{i,j}^k$, $j = 1, \dots, n_i^k$ is the j^{th} time that the i^{th} target is observed. At times it will be convenient to write $X_{i,j} = X_i(t_{i,j}^k)$, and $n_i = n_i^k$ keeping in mind these are still a function of k . For the tracking estimate $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ we will denote $X_{i,j}$ and n_i as $\hat{X}_{i,j}$ and \hat{n}_i respectively. For $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ we will denote them as $\tilde{X}_{i,j}$ and \tilde{n}_i . Adopt a similar notation for Y .

We will also make use of the following definition of a track and a track segment. A target track for target i is defined to be a sequence of observed locations $(X_{i,1}, X_{i,2}, \dots, X_{i,n_i})$. Define a track segment to be any subsequence of a track where consecutive elements are the same as those in the track, i.e. $(X_{i,j}, X_{i,j+1}, \dots, X_{i,l})$ where $1 \leq j \leq l \leq n_i$. Define a correct track to be a track in the correct solution $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$ and a correct track segment to be a track segment of a correct track.

We will consider the following six ways an alternative can differ from the correct solution. We will then argue that any incorrect solution may be obtained by applying these differences sequentially to the correct solution. These six alternatives are

1. incorrectly breaking a correct track into two track segments.
2. incorrectly labeling a merger as two deaths and a birth.

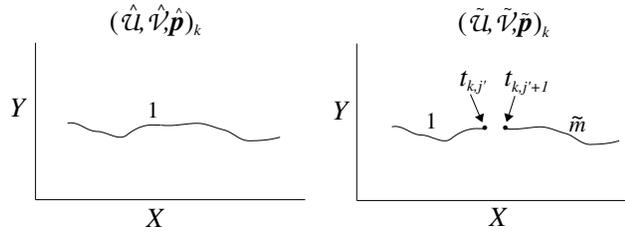


FIG 4. $(\tilde{U}, \tilde{V}, \tilde{P})_k$ incorrectly breaks target track 1 into two tracks.

3. incorrectly labeling a split as a death and two births.
4. incorrectly connecting a death with a birth to make one track.
5. incorrectly labeling two deaths and a birth as a merger.
6. incorrectly labeling a death and two births as a split.

The following propositions in this section deal with each of the six possible differences between solutions listed above. Each one assumes something about two sequences of tracking estimates $(\hat{U}, \hat{V}, \hat{P})_k$ and $(\tilde{U}, \tilde{V}, \tilde{P})_k$. Propositions 1a and 1b deal with the first difference listed above. They basically say that asymptotically it is not beneficial to break a correct target track into separate tracks. Proposition 1a considers breaking the track at a fixed time, while Proposition 1b considers breaking the track at an arbitrary time.

Proposition 1a. Assume Conditions 1-4, and further that $H_i(t) = 0$. Let Θ be the set of all pairs of tracking solutions sequences, $\hat{\theta}_k = (\hat{U}, \hat{V}, \hat{P})_k$ and $\tilde{\theta}_k = (\tilde{U}, \tilde{V}, \tilde{P})_k$ that have the following property. All of the tracks that make up $\hat{\theta}_k$ and $\tilde{\theta}_k$ are correct track segments and $\tilde{\theta}_k$ differs from $\hat{\theta}_k$ at every k only by breaking a correct target track labeled i_1 from $\hat{\theta}_k$ into two tracks by incorrectly specifying the death of target i_1 and the birth of target i_2 during a fixed time interval $[t_{j'}^k, t_{j'+1}^k)$ for each k . Then,

$$R_{1a} = \sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[(U, V, P)_k | \mathcal{Z}_k](\tilde{U}, \tilde{V}, \tilde{P})_k}{[(U, V, P)_k | \mathcal{Z}_k](\hat{U}, \hat{V}, \hat{P})_k} \leq O(k^{-3}(\log k)^c) \text{ as } k \rightarrow \infty \text{ a.s.}$$

for some positive constant c which depends on ω .

Proof. WLOG assume that $i_1 = 1$ and $i_2 = \tilde{m}$, where \tilde{m} is the number of tracks in $(\tilde{U}, \tilde{V}, \tilde{P})_k$; see Figure 4. Note that this implies $\hat{m} = \tilde{m} - 1$ where \hat{m} is the number of tracks in $(\hat{U}, \hat{V}, \hat{P})_k$. From (9) of Section 2.3 the ratio of the densities in the proposition can be written as

$$\frac{[(U, V, P)_k | \mathcal{Z}_k](\tilde{U}, \tilde{V}, \tilde{P})_k}{[(U, V, P)_k | \mathcal{Z}_k](\hat{U}, \hat{V}, \hat{P})_k} = \frac{[U_k, V_k](\tilde{U}_k, \tilde{V}_k)}{[U_k, V_k](\hat{U}_k, \hat{V}_k)} \cdot \frac{[\tilde{\mathcal{X}}_k | \tilde{U}_k, \tilde{V}_k](\tilde{\mathcal{X}}_k) \cdot [\tilde{\mathcal{Y}}_k | \tilde{U}_k, \tilde{V}_k](\tilde{\mathcal{Y}}_k)}{[\hat{\mathcal{X}}_k | \hat{U}_k, \hat{V}_k](\hat{\mathcal{X}}_k) \cdot [\hat{\mathcal{Y}}_k | \hat{U}_k, \hat{V}_k](\hat{\mathcal{Y}}_k)} \tag{A26}$$

Write $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,n_i})$ and

$$\mathcal{F}_{i,j} = (X_{i,1}, \dots, X_{i,j}, \mathbf{X}_{i-1}, \dots, \mathbf{X}_1, D_1, \dots, D_{N_m}),$$

with D_j being the difference variables resulting from merger as given in (A4). By convention let $\mathcal{F}_{i,0} = (\mathbf{X}_1, \dots, \mathbf{X}_{i-1}, D_1, \dots, D_{N_m})$. Then let $\hat{\mathcal{F}}_{i,j}$ and $\tilde{\mathcal{F}}_{i,j}$ denote the $\mathcal{F}_{i,j}$ variable for the solutions $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ and $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ respectively. Notice that we can write the x component of the likelihood for $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ given \mathcal{Z} as

$$[\hat{\mathcal{X}} | \hat{\mathcal{U}}, \hat{\mathcal{V}}](\hat{\mathcal{X}}) = \prod_{i=1}^{\tilde{m}} \prod_{j=1}^{\hat{n}_i} [\hat{X}_{i,j} | \hat{\mathcal{F}}_{i,j-1}](\hat{X}_{i,j})$$

for example. So we can write the ratio of the x location densities in (A26) by breaking it apart into tracks $2, \dots, \tilde{m} - 1$, which the two solutions have in common, then handle tracks 1 and \tilde{m} separately

$$\begin{aligned} \frac{[\tilde{\mathcal{X}}_k | \tilde{\mathcal{U}}_k, \tilde{\mathcal{V}}_k](\tilde{\mathcal{X}}_k)}{[\hat{\mathcal{X}}_k | \hat{\mathcal{U}}_k, \hat{\mathcal{V}}_k](\hat{\mathcal{X}}_k)} &= \frac{\prod_{i=1}^{\tilde{m}} \prod_{j=1}^{\tilde{n}_i} [\tilde{X}_{i,j} | \tilde{\mathcal{F}}_{i,j-1}](\tilde{X}_{i,j})}{\prod_{i=1}^{\tilde{m}-1} \prod_{j=1}^{\hat{n}_i} [\hat{X}_{i,j} | \hat{\mathcal{F}}_{i,j-1}](\hat{X}_{i,j})} \\ &= \frac{\prod_{i=2}^{\tilde{m}-1} \prod_{j=1}^{\tilde{n}_i} [\tilde{X}_{i,j} | \tilde{\mathcal{F}}_{i,j-1}](\tilde{X}_{i,j}) \cdot \prod_{j=1}^{\tilde{n}_1} [\tilde{X}_{1,j} | \tilde{\mathcal{F}}_{1,j-1}](\tilde{X}_{1,j})}{\prod_{i=2}^{\tilde{m}-1} \prod_{j=1}^{\hat{n}_i} [\hat{X}_{i,j} | \hat{\mathcal{F}}_{i,j-1}](\hat{X}_{i,j}) \cdot \prod_{j=1}^{\tilde{n}_1} [\hat{X}_{1,j} | \hat{\mathcal{F}}_{1,j-1}](\hat{X}_{1,j})} \\ &\quad \cdot \frac{[\tilde{X}_{\tilde{m},1} | \tilde{\mathcal{F}}_{\tilde{m},0}](\tilde{X}_{\tilde{m},1}) \cdot \prod_{j=1}^{\tilde{n}_{\tilde{m}}} [\tilde{X}_{\tilde{m},j} | \tilde{\mathcal{F}}_{\tilde{m},j-1}](\tilde{X}_{\tilde{m},j})}{[\hat{X}_{1,\tilde{n}_1+1} | \hat{\mathcal{F}}_{1,\tilde{n}_1}](\hat{X}_{1,\tilde{n}_1+1}) \cdot \prod_{j=\tilde{n}_1+2}^{\tilde{n}_1} [\hat{X}_{1,j} | \hat{\mathcal{F}}_{1,j-1}](\hat{X}_{1,j})}. \end{aligned} \quad (\text{A27})$$

But for $i = 2, \dots, \tilde{m} - 1$, and $j = 1, \dots, \hat{n}_i$, we have $\hat{X}_{i,j} = \tilde{X}_{i',j}$ for some i' in $1, \dots, \tilde{m} - 1$. The number of observations $\hat{n}_i = \tilde{n}_{i'}$ for all of these tracks are the same as well. Also $\hat{X}_{1,j} = \tilde{X}_{1,j}$ for $j = 1, \dots, \tilde{n}_1$ and $\hat{X}_{1,\tilde{n}_1+j} = \tilde{X}_{\tilde{m},j}$ for $j = 2, \dots, \tilde{n}_{\tilde{m}}$. So by Lemma 6 the first, second and third terms of (A27) are no more than $a^{\tilde{m}} (\log k)^{\tilde{m}b}$ for some uniform constants a and b so that

$$\frac{[\tilde{\mathcal{X}}_k | \tilde{\mathcal{U}}_k, \tilde{\mathcal{V}}_k](\tilde{\mathcal{X}}_k)}{[\hat{\mathcal{X}}_k | \hat{\mathcal{U}}_k, \hat{\mathcal{V}}_k](\hat{\mathcal{X}}_k)} \leq \frac{[\tilde{X}_{\tilde{m},1} | \tilde{\mathcal{F}}_{\tilde{m},0}](\tilde{X}_{\tilde{m},1})}{[\hat{X}_{1,\tilde{n}_1+1} | \hat{\mathcal{F}}_{1,\tilde{n}_1}](\hat{X}_{1,\tilde{n}_1+1})} \cdot a^{\tilde{m}} (\log k)^{\tilde{m}b} \quad \text{a.s.}$$

Since $\tilde{m} \leq M$ by Condition 8 we have

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[\tilde{\mathcal{X}}_k | \tilde{\mathcal{U}}_k, \tilde{\mathcal{V}}_k](\tilde{\mathcal{X}}_k)}{[\hat{\mathcal{X}}_k | \hat{\mathcal{U}}_k, \hat{\mathcal{V}}_k](\hat{\mathcal{X}}_k)} \leq \sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[\tilde{X}_{\tilde{m},1} | \tilde{\mathcal{F}}_{\tilde{m},0}](\tilde{X}_{\tilde{m},1})}{[\hat{X}_{1,\tilde{n}_1+1} | \hat{\mathcal{F}}_{1,\tilde{n}_1}](\hat{X}_{1,\tilde{n}_1+1})} a^M (\log k)^{Mb} \quad \text{a.s.} \quad (\text{A28})$$

Notice that the term in the numerator of (A28) is the density of the first observation of a new track and hence for all $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ must be smaller than the mode of the normal density with variance $\sigma_{X_0}^2$. Also, the denominator can be written out by utilizing Lemma 3. So we have,

$$\begin{aligned} &\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[\tilde{\mathcal{X}}_k | \tilde{\mathcal{U}}_k, \tilde{\mathcal{V}}_k](\tilde{\mathcal{X}}_k)}{[\hat{\mathcal{X}}_k | \hat{\mathcal{U}}_k, \hat{\mathcal{V}}_k](\hat{\mathcal{X}}_k)} \\ &\leq \frac{\phi(0; 0, \sigma_{X_0}^2) \cdot a^M (\log k)^{Mb}}{\inf_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \exp \left\{ -\frac{(\hat{X}_1(t_{j'+1}^k) - \hat{X}_1(t_j^k) + O(\frac{1}{k}))^2}{2\hat{\sigma}^2 \Delta t_{j'}^k + O(\frac{1}{k^2})} \right\} / \sqrt{2\pi \hat{\sigma}^2 \Delta t_{j'}^k + O(\frac{1}{k^2})}}. \end{aligned} \quad (\text{A29})$$

The same is true for \mathcal{Y} and it is independent of \mathcal{X} . Recall that the variance parameter of the Brownian motion for Y_i is denoted as η_i . Then we have

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[\tilde{\mathcal{X}}_k, \tilde{\mathcal{Y}}_k | \tilde{\mathcal{U}}_k, \tilde{\mathcal{V}}_k](\tilde{\mathcal{X}}_k, \tilde{\mathcal{Y}}_k)}{[\hat{\mathcal{X}}_k, \hat{\mathcal{Y}}_k | \hat{\mathcal{U}}_k, \hat{\mathcal{V}}_k](\hat{\mathcal{X}}_k, \hat{\mathcal{Y}}_k)} \leq \frac{\phi(0; 0, \sigma_{X_0}^2) \phi(0; 0, \sigma_{Y_0}^2) \cdot a^M (\log k)^{Mb}}{\inf_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \mathcal{D}} \tag{A30}$$

where

$$\mathcal{D} = \frac{\exp \left\{ - \left(\frac{(\hat{X}_1(t_{j'+1}^k) - \hat{X}_1(t_{j'}^k) + O(k^{-1}))^2}{2\hat{\sigma}^2 \Delta t_{j'}^k + O(k^{-2})} + \frac{(\hat{Y}_1(t_{j'+1}^k) - \hat{Y}_1(t_{j'}^k) + O(k^{-1}))^2}{2\hat{\eta}^2 \Delta t_{j'}^k + O(k^{-2})} \right) \right\}}{2\pi \hat{\sigma} \hat{\eta} \Delta t_{j'}^k + O(k^{-3/2})}.$$

Now

$$\frac{(\hat{X}_1(t_{j'+1}^k) - \hat{X}_1(t_{j'}^k))^2}{\hat{\sigma}^2} \stackrel{\mathcal{L}}{=} \frac{\sigma^2}{\hat{\sigma}^2} (B(t_{j'+1}^k) - B(t_{j'}^k)) \tag{A31}$$

where $B(t)$ is a Brownian motion or a Brownian motion conditional on the D_j 's of Lemma 5. In either case, because of Lemma 5, any path properties of Brownian motion will apply. Also realize that all of the track segments in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ are correct so by Lemma 8, $\hat{\sigma}^2 \rightarrow \sigma^2$. Using this fact along with (A31) on \mathcal{D} , we have

$$\begin{aligned} \inf_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \mathcal{D} &= \frac{\exp \left\{ - \frac{(1+o(1))(B(t_{j'+1}^k) - B(t_{j'}^k))^2 + (1+o(1))(\tilde{B}(t_{j'+1}^k) - \tilde{B}(t_{j'}^k))^2}{2\Delta t_{j'}^k + O(k^{-2})} \right\}}{2\pi \sigma_1 \eta_1 \Delta t_{j'}^k + O(k^{-3/2})} \\ &= \frac{\exp \left\{ - \frac{(B(t_{j'+1}^k) - B(t_{j'}^k))^2 + (\tilde{B}(t_{j'+1}^k) - \tilde{B}(t_{j'}^k))^2}{2\Delta t_{j'}^k \log \log(1/\Delta t_{j'}^k) + \log \log(1/\Delta t_{j'}^k) O(k^{-2})} (1 + o(1)) \log \log(1/\Delta t_{j'}^k) \right\}}{C \Delta t_{j'}^k + O(k^{-3/2})} \\ &= \frac{\exp \left\{ - \frac{(B(t_{j'+1}^k) - B(t_{j'}^k))^2 + (\tilde{B}(t_{j'+1}^k) - \tilde{B}(t_{j'}^k))^2}{2\Delta t_{j'}^k \log \log(1/\Delta t_{j'}^k)} (1 + o(1)) \log \log(1/\Delta t_{j'}^k) \right\}}{C \Delta t_{j'}^k + O(k^{-3/2})}, \tag{A32} \end{aligned}$$

where B and \tilde{B} are independent Brownian motions. Now remember that $[t_{j'}^k, t_{j'+1}^k]$ is a fixed interval for each k . Also assume that B is a Brownian motion (not conditioned on D_j 's) so that $B(t_{j'+1}^k) - B(t_{j'}^k) \stackrel{\mathcal{L}}{=} B^*(\Delta t_{j'}^k)$ where B^* is a Brownian motion. We can now apply Lemma 1 to \mathcal{D} which gives

$$\inf_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \mathcal{D} \geq \frac{\exp \left\{ - (1 + o(1)) \log \log(1/\Delta t_{j'}^k) \right\}}{C \Delta t_{j'}^k + O(k^{-3/2})}. \tag{A33}$$

Note that B^* may have been a different Brownian motion for each k , but there is a countable number of them so the sets with probability zero can be joined. If $B(t)$ was a Brownian motion conditioned on D_j 's then the result in (A33) is still true since the path set for $B(t)$ is the same as that of Brownian motion

excluding the paths in a set with probability zero. Hence we are left with

$$\inf_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \mathcal{D} = \frac{\log(1/\Delta t_{j'}^k)^{-1+o(1)}}{C\Delta t_{j'}^k + O(k^{-3/2})} \quad (\text{A34})$$

$$\geq C'k(\log k)^{-(1+\varepsilon)} \text{ eventually a.s.} \quad (\text{A35})$$

for some constant C' and an arbitrarily small $\varepsilon > 0$. We used Condition 1 to get from (A34) to (A35). This says that the location contribution to the density of target 1 at time $t_{j'}^k$ goes to ∞ like $k/(\log(k))^c$. This along with (A30) gives,

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[\tilde{\mathcal{X}}_k, \tilde{\mathcal{Y}}_k | \tilde{\mathcal{U}}_k, \tilde{\mathcal{V}}_k](\tilde{\mathcal{X}}_k, \tilde{\mathcal{Y}}_k)}{[\hat{\mathcal{X}}_k, \hat{\mathcal{Y}}_k | \hat{\mathcal{U}}_k, \hat{\mathcal{V}}_k](\hat{\mathcal{X}}_k, \hat{\mathcal{Y}}_k)} \leq O(k^{-1}(\log(k))^c) \text{ a.s.} \quad (\text{A36})$$

We also need to consider the event density contribution to the ratio of (A26). The event densities $[\mathcal{U}_k, \mathcal{V}_k](\hat{\mathcal{U}}_k, \hat{\mathcal{V}}_k)$ and $[\mathcal{U}_k, \mathcal{V}_k](\tilde{\mathcal{U}}_k, \tilde{\mathcal{V}}_k)$ will be different only in their contribution of the number of events during the interval $[t_{j'}^k, t_{j'+1}^k]$. For $(\tilde{\mathcal{U}}_k, \tilde{\mathcal{V}}_k)$, there is one more death and one more birth in the interval than for $(\hat{\mathcal{U}}_k, \hat{\mathcal{V}}_k)$, so from (A2) we have

$$\begin{aligned} & \frac{[\mathcal{U}_k, \mathcal{V}_k](\tilde{\mathcal{U}}_k, \tilde{\mathcal{V}}_k)}{[\mathcal{U}_k, \mathcal{V}_k](\hat{\mathcal{U}}_k, \hat{\mathcal{V}}_k)} \\ &= \frac{P(U_{b,j'} = \hat{U}_{b,j'} + 1)P(U_{d,j'} = \hat{U}_{d,j'} + 1)P(U_{s,j'} = \hat{U}_{s,j'})P(U_{m,j'} = \hat{U}_{b,j'})P(V_k = \tilde{V}_k)}{P(U_{b,j'} = \hat{U}_{b,j'})P(U_{d,j'} = \hat{U}_{d,j'})P(U_{s,j'} = \hat{U}_{s,j'})P(U_{m,j'} = \hat{U}_{b,j'})P(V_k = \hat{V}_k)} \\ &= \frac{\left(\frac{(\lambda_b \Delta t_{j'}^k)^{(\hat{U}_{b,j'} + 1)}}{(\hat{U}_{b,j'} + 1)!} e^{-\lambda_b \Delta t_{j'}^k}\right) \left(\frac{(\lambda_d \tilde{N}_{j'} \Delta t_{j'}^k)^{(\hat{U}_{d,j'} + 1)}}{(\hat{U}_{d,j'} + 1)!} e^{-\lambda_d \tilde{N}_{j'} \Delta t_{j'}^k}\right) P(V_k = \tilde{V}_k)}{\left(\frac{(\lambda_b \Delta t_{j'}^k)^{\hat{U}_{b,j'}}}{\hat{U}_{b,j'}!} e^{-\lambda_b \Delta t_{j'}^k}\right) \left(\frac{(\lambda_d \tilde{N}_{j'} \Delta t_{j'}^k)^{\hat{U}_{d,j'}}}{\hat{U}_{d,j'}!} e^{-\lambda_d \tilde{N}_{j'} \Delta t_{j'}^k}\right) P(V_k = \hat{V}_k)} \\ &= \frac{(\lambda_b \Delta t_{j'}^k)(\lambda_d \tilde{N}_{j'} \Delta t_{j'}^k)}{(\hat{U}_{b,j'} + 1)(\hat{U}_{d,j'} + 1)}. \end{aligned} \quad (\text{A37})$$

But notice that $\hat{U}_{b,j'} \geq 0$ and $\hat{U}_{d,j'} \geq 0$ while $\tilde{N}_{j'} \leq M$. Hence

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[\mathcal{U}_k, \mathcal{V}_k](\tilde{\mathcal{U}}_k, \tilde{\mathcal{V}}_k)}{[\mathcal{U}_k, \mathcal{V}_k](\hat{\mathcal{U}}_k, \hat{\mathcal{V}}_k)} = O(\Delta t_{j'}^k)^2 = O(k^{-2}) \text{ a.s.} \quad (\text{A38})$$

Putting this together with (A36) gives us,

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k | \mathcal{Z}_k](\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k}{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k | \mathcal{Z}_k](\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k} \leq O(k^{-3}(\log(k))^c) \text{ a.s.}$$

as we desired. \square

Proposition 1b. Assume Conditions 1-4, and further that $H_i(t) = 0$. Let Θ be the set of all pairs of tracking solutions sequences, $\hat{\theta}_k = (\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ and

$\tilde{\theta}_k = (\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ that have the following property. All of the tracks that make up $\hat{\theta}_k$ and $\tilde{\theta}_k$ are correct track segments and $\tilde{\theta}_k$ differs from $\hat{\theta}_k$ at every k only by breaking a correct target track labeled i_1 from $\hat{\theta}_k$ into two tracks by incorrectly specifying the death of i_1 and the birth of target i_2 during an arbitrary time interval $[t_{j'}^k, t_{j'+1}^k]$. Then,

$$R_{1b} = \sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}_k](\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k}{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}_k](\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k} \leq O(k^{-2+\varepsilon}) \text{ as } k \rightarrow \infty \text{ a.s.}$$

for some positive constant c which depends on ω and some $\varepsilon > 0$ which can be made arbitrarily small.

Proof. This proof follows the exact same logic as the previous proof of Proposition 1a. The only difference being that now $t_{j'}^k$ is an arbitrary time. Hence the only change will be to apply Lemma 2 instead of Lemma 1 to (A32). So for the \mathcal{D} of (A30) we end up with

$$\mathcal{D} \geq \frac{\exp \left\{ - \sup_j \frac{(B(t_{j+1}^k) - B(t_j^k))^2 + (\tilde{B}(t_{j+1}^k) - \tilde{B}(t_j^k))^2}{2\Delta t_j^k \log(1/\Delta t_j^k)} (1 + o(1)) \log(1/\Delta t_{j'}^k) \right\}}{C\Delta t_{j'}^k + O(k^{-3/2})}.$$

Now use Lemma 2 which results in

$$\begin{aligned} \inf_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \mathcal{D} &\geq \frac{\exp \{ - (1 + o(1)) \log(1/\Delta t_{j'}^k) \}}{C\Delta t_{j'}^k + O(k^{-3/2})} \\ &= \frac{\Delta(t_{j'}^k)^{1+o(1)}}{C\Delta t_{j'}^k + O(k^{-3/2})} \\ &\geq C'k^{-\varepsilon} \text{ eventually a.s.} \end{aligned} \tag{A39}$$

for some constant C' and an arbitrarily small $\varepsilon > 0$. This along with (A30) gives us

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[\mathcal{X}_k, \mathcal{Y}_k \mid \tilde{\mathcal{U}}_k, \tilde{\mathcal{V}}_k](\tilde{\mathcal{X}}_k, \tilde{\mathcal{Y}}_k)}{[\mathcal{X}_k, \mathcal{Y}_k \mid \hat{\mathcal{U}}_k, \hat{\mathcal{V}}_k](\hat{\mathcal{X}}_k, \hat{\mathcal{Y}}_k)} \leq a^M k^{-\varepsilon} (\log k)^{Mb} \leq C''k^{-\varepsilon} \text{ eventually a.s.} \tag{A40}$$

for some constant C'' and $\varepsilon > \epsilon$ which can still be taken to be arbitrarily small.

The event model is the same as previously in Proposition 1a,

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[\mathcal{U}_k, \mathcal{V}_k](\tilde{\mathcal{U}}_k, \tilde{\mathcal{V}}_k)}{[\mathcal{U}_k, \mathcal{V}_k](\hat{\mathcal{U}}_k, \hat{\mathcal{V}}_k)} = O(k^{-2}) \text{ a.s.}$$

This along with (A40) gives us,

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}_k](\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k}{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}_k](\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k} \leq O(k^{-2+\varepsilon}) \text{ a.s.}$$

□

Propositions 2 and 3 deal with the second and third differences listed earlier. These propositions say that it is not beneficial asymptotically to break apart a correctly specified merging or splitting event.

Proposition 2. *Assume Conditions 1-4, and further that $H_i(t) = 0$. Let Θ be the set of all pairs of tracking solution sequences, $\hat{\theta}_k = (\hat{U}, \hat{V}, \hat{P})_k$ and $\tilde{\theta}_k = (\tilde{U}, \tilde{V}, \tilde{P})_k$ that have the following property. All of the track segments that make up $\hat{\theta}_k$ and $\tilde{\theta}_k$ are correct track segments and $\tilde{\theta}_k$ differs from $\hat{\theta}_k$ at every k only by relabeling a correct merging event of targets i_1 and i_2 into target i_3 as the deaths of targets i_1 and i_2 and the birth of target i_3 . Then*

$$R_2 = \sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[(U, V, P)_k | Z_k](\tilde{U}, \tilde{V}, \tilde{P})_k}{[(U, V, P)_k | Z_k](\hat{U}, \hat{V}, \hat{P})_k} \leq O(k^{-2}(\log k)^c) \text{ as } k \rightarrow \infty \text{ a.s.}$$

for some positive constant c which depends on ω .

Proof. WLOG assume that $i_1 = 1, i_2 = 2$ and $i_3 = 3$. The relabeling described in the proposition is portrayed in Figure 5. Here we have that for $i = 1, \dots, m$, and $j = 1, \dots, \hat{n}_i$, the locations $\hat{X}_{i,j} = \tilde{X}_{i',j}$ for some i' . The number of observations $\hat{n}_i = \tilde{n}_{i'}$ for all of these tracks are the same as well. Hence by Lemma 6 the ratio of the x location densities in (A26) is no more than $a^{\tilde{m}}(\log k)^{\tilde{m}b}$ for some constants a and b ,

$$\frac{[\tilde{\mathcal{X}}_k | (\tilde{U}, \tilde{V}, \tilde{P})_k](\tilde{\mathcal{X}})}{[\hat{\mathcal{X}}_k | (\hat{U}, \hat{V}, \hat{P})_k](\hat{\mathcal{X}})} \leq a^{\tilde{m}}(\log k)^{\tilde{m}b} \text{ a.s.}$$

and since there are less than M tracks the solutions $(\hat{U}, \hat{V}, \hat{P})_k$ and $(\tilde{U}, \tilde{V}, \tilde{P})_k$, we have

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[\tilde{\mathcal{X}}_k | (\tilde{U}, \tilde{V}, \tilde{P})_k](\tilde{\mathcal{X}})}{[\hat{\mathcal{X}}_k | (\hat{U}, \hat{V}, \hat{P})_k](\hat{\mathcal{X}})} \leq a^M(\log k)^{Mb} \text{ a.s.}$$

Of course the same is true for \mathcal{Y} as well so we have

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[\tilde{\mathcal{X}}_k, \hat{\mathcal{Y}}_k | \tilde{U}_k, \tilde{V}_k](\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}_k)}{[\hat{\mathcal{X}}_k, \hat{\mathcal{Y}}_k | \hat{U}_k, \hat{V}_k](\hat{\mathcal{X}}, \hat{\mathcal{Y}}_k)} \leq a^{2M}(\log k)^{2Mb} \text{ a.s.} \tag{A41}$$

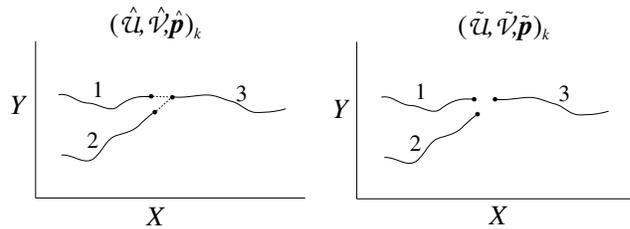


FIG 5. $(\tilde{U}, \tilde{V}, \tilde{P})_k$ incorrectly breaks apart a merger into two deaths and a birth.

For the event model contribution, the event model for $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ has 2 more events in $[t_{j'}^k, t_{j'+1}^k)$, than $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$. As in (A38) then we have

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[\mathcal{U}_k, \mathcal{V}_k](\tilde{\mathcal{U}}_k, \tilde{\mathcal{V}}_k)}{[\mathcal{U}_k, \mathcal{V}_k](\hat{\mathcal{U}}_k, \hat{\mathcal{V}}_k)} \leq O(k^{-2}) \text{ a.s.}$$

This along with (A41) gives us the desired result. □

Proposition 3. *Assume Conditions 1-4, and further that $H_i(t) = 0$. Let Θ be the set of all pairs of tracking solution sequences, $\hat{\theta}_k = (\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ and $\tilde{\theta}_k = (\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ that have the following property. All of the tracks that make up $\hat{\theta}_k$ and $\tilde{\theta}_k$ are correct track segments and $\tilde{\theta}_k$ differs from $\hat{\theta}_k$ at every k only by relabeling a correct splitting event of target i_1 into targets i_2 and i_3 as the death of target i_1 and the birth of targets i_2 and i_3 . Then*

$$R_3 = \sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}_k](\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k}{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}_k](\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k} \leq O(k^{-2}(\log k)^c) \text{ as } k \rightarrow \infty \text{ a.s.}$$

for some positive constant c which depends on ω .

Proof. This is symmetric with respect to the difference between solutions in Proposition 2. The proof will therefore be identical. □

We will say that two targets labeled i_1 and i_2 at times t_1 and t_2 are distinct if their labels are different in the correct solution $(\mathcal{U}, \mathcal{V}, \mathcal{P})$. That is to say that they are not the same physical target. Proposition 4 deals with the fourth difference listed earlier in the chapter. It assumes that $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$ has the death of a target labeled i_1 and the birth of a target labeled i_2 in the same interval $[t_{j'}^k, t_{j'+1}^k)$. These two events are not necessarily labeled correctly. However target i_1 at time $t_{j'}^k$ is assumed to be distinct from target i_2 at time $t_{j'+1}^k$. The solution $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})$ then connects these two track segments which is not consistent with the correct solution; see Figure 6. Proposition 4 then says that there can be no differences of this type eventually.

Proposition 4. *Assume Conditions 1-4, and further that $H_i(t) = 0$. Let Θ be the set of all pairs of tracking solution sequences, $\hat{\theta}_k = (\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ and $\tilde{\theta}_k =$*

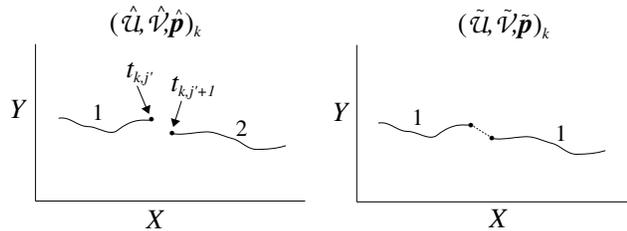


FIG 6. $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ incorrectly joins target tracks 1 and 2 into one track.

$(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ that have the following property. The sequences $\tilde{\theta}_k$ and $\hat{\theta}_k$ differ at every k only by joining the birth and death of two targets i_1 and i_2 which are distinct into one target track. Then Θ is the empty set.

Proof. WLOG assume that $i_1 = 1$ and $i_2 = 2$. The observations that get incorrectly joined together in a track in the solution $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})$ are $(\hat{X}_1(t_{j'}^k), \hat{Y}_1(t_{j'}^k))$ and $(\hat{X}_2(t_{j'+1}^k), \hat{Y}_2(t_{j'+1}^k))$. Now these are observations from distinct targets in the correct solution so by Lemma 7 and the continuity of Brownian paths, we know that

$$\inf_j \left\{ \left(\hat{X}_2(t_{j'+1}^k) - \hat{X}_1(t_{j'}^k) \right)^2 + \left(\hat{Y}_2(t_{j'+1}^k) - \hat{Y}_1(t_{j'}^k) \right)^2 \right\} \\ = \inf_j \left\| \left(\hat{X}_1(t_{j'}^k), \hat{Y}_1(t_{j'}^k) \right) - \left(\hat{X}_2(t_{j'+1}^k), \hat{Y}_2(t_{j'+1}^k) \right) \right\|^2 \geq C_1 > 0 \text{ a.s.}$$

So by Condition 8 there cannot be a solution that connects these two observations in the same track eventually. Hence such a sequence $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ does not exist. \square

Note that this is the only place where we use the second part of Condition 8. This is a very reasonable assumption to make, since it only prevents us from forming discontinuous paths. It seems however, that the likelihood should prevent us from doing this anyway. We do indeed believe that this is the case, but need to develop tighter bounds in formulation of Lemma 6 before we can remove the second part of Condition 8.

The next proposition deals with the fifth difference listed in the beginning of the chapter. It basically says that it is not advantageous asymptotically to take actual deaths and a birth and merge them together.

We will say that an event in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ corresponds to an event in $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$ if they happen to the same target in the same interval. For example if target 1 dies in the interval $[t_{j'}^k, t_{j'+1}^k]$ in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ and in $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$ target 1 merges with target 2 in that interval, then the death of target 1 in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ corresponds to the merger of target 1 with target 2 in $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$. Also two events in a solution are distinct if they are not the same event.

Proposition 5a. *Assume Conditions 1-4, and further that $H_i(t) = 0$. Let Θ be the set of all pairs of tracking solution sequences, $\hat{\theta}_k = (\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ and $\tilde{\theta}_k = (\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ that have the following property. All of the tracks that make up $\hat{\theta}_k$ and $\tilde{\theta}_k$ are correct track segments and $\tilde{\theta}_k$ differs from $\hat{\theta}_k$ at every k only by declaring a merging event in place of two deaths and a birth for two targets i_1 and i_2 that died in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ with a target i_3 that was born in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$. It is further assumed that the deaths of targets i_1 and i_2 and the birth of target i_3 may be incorrectly specified in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$, but at least two of these three events must correspond to two distinct events in the correct solution $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$. Then,*

$$R_{5a} = \sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}_k](\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k}{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}_k](\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k} = 0 \text{ eventually as } k \rightarrow \infty \text{ a.s.}$$

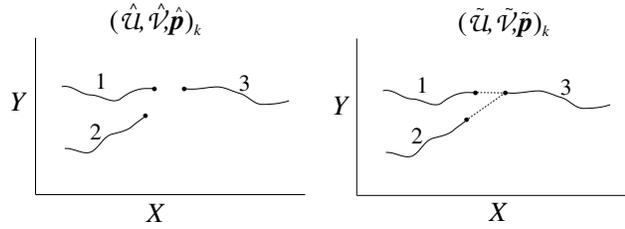


FIG 7. $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ incorrectly merges targets 1 and 2 into target 3.

Proof. WLOG let $i_1 = 1$, $i_2 = 2$, and $i_3 = 3$. The relabeling described in the proposition are portrayed in Figure 7. Let the unknown times at which the two distinct events occur be denoted τ_1 and τ_2 and assume WLOG that $\tau_1 \leq \tau_2$. Since these are the times of actual events from the event model, $\tau_1 < \tau_2$ with probability 1. Hence by condition 1 there is a K such that for all $k > K$ there will be a sample time, $t_{j'}^k$, in the interval (τ_1, τ_2) . For this sample time, one of the targets involved in the proposed merging event in $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ will be missing unless targets 1 and 2 merge before τ_1 . But missing observations are not allowed and both targets still exists in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ at time $t_{j'}^k$, so this would violate the hypothesis. Thus $[\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}}]_k, \mathcal{Z}_k((\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k, \mathcal{Z}_k) = 0$ eventually. \square

Proposition 5b is very similar to proposition 5a, but now there is no restriction that any of the two deaths and a birth in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ correspond to events in $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$. Thus they can be at arbitrary times. In this case there may be an advantage to combine these three events into a merger by switching to the alternative $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$. This will not be a problem however as we will see that there would have to be too many other negative differences before $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ could make use out of any possible advantage it may gain from the difference in Proposition 5b.

Proposition 5b. *Assume Conditions 1-4, and further that $H_i(t) = 0$. Let Θ be the set of all pairs of tracking solution sequences, $\hat{\theta}_k = (\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ and $\tilde{\theta}_k = (\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ that have the following property. All of the tracks that make up $\hat{\theta}_k$ and $\tilde{\theta}_k$ are correct track segments and $\tilde{\theta}_k$ differs from $\hat{\theta}_k$ at every k only by declaring a merging event in place of two deaths and a birth for two targets i_1 and i_2 that died in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ with a target i_3 that was born in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$. Then*

$$R_{5b} = \sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k | \mathcal{Z}_k]((\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k)}{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k | \mathcal{Z}_k]((\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k)} \leq bk^2(\log k)^c \text{ as } k \rightarrow \infty \text{ a.s.}$$

for some positive constants b and c which depend on ω .

Proof. This differs from the previous proposition since, the deaths of targets i_1 and i_2 and the birth of target i_3 in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ do not necessarily correspond to events in the correct solution. WLOG assume that $i_1 = 1$, $i_2 = 2$ and $i_3 = 3$. Here we have the reverse case of Proposition 2, so again we have that for

$i = 1, \dots, m$ and $j = 1, \dots, \hat{n}_i$, the observations $\hat{X}_{i,j} = \tilde{X}_{i',j}$ for some i' . The number of observations $\hat{n}_i = \tilde{n}_{i'}$ for all of these tracks are the same as well. The same arguments as in the proof of Proposition 2 lead us to

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[\tilde{\mathcal{X}}_k, \hat{\mathcal{Y}}_k | \tilde{\mathcal{U}}_k, \tilde{\mathcal{V}}_k](\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}_k)}{[\hat{\mathcal{X}}_k, \hat{\mathcal{Y}}_k | \hat{\mathcal{U}}_k, \hat{\mathcal{V}}_k](\hat{\mathcal{X}}, \hat{\mathcal{Y}}_k)} \leq a^{2M} (\log k)^{2Mb} \quad \text{a.s.} \quad (\text{A42})$$

For the event model contribution, the two deaths and a birth $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ do not necessarily correspond to events in the correct solution, so it is possible that they are all in the same interval $[t_{j'}^k, t_{j'+1}^k)$. This would be the worst case, since if they were not, by arguments of the previous proof, $[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k, \mathcal{Z}_k](\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k, \mathcal{Z}_k) = 0$ eventually. So assume that all three events are in the same interval for all k . In which case, the event density for $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ and $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ will differ only in their contributions during the interval $[t_{j'}^k, t_{j'+1}^k)$. The event model for $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ has two more events in $[t_{j'}^k, t_{j'+1}^k)$ than does $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$. In the same manner we derived (A37) we can see that

$$\frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k](\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k}{[(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k](\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k} = \frac{(\lambda_d(\tilde{N}_{j'} - 1)\Delta t_{j'}^k)(\tilde{U}_{b,j'} + 1)(\tilde{U}_{d,j'} + 1)(\tilde{U}_{m,j'} + 2)}{(\lambda_b\Delta t_{j'}^k)(\lambda_d\tilde{N}_{j'}\Delta t_{j'}^k)^2(\tilde{U}_{m,j'} + 1)}.$$

After taking into account that $\tilde{U}_{b,j'}$, $\tilde{U}_{d,j'}$, and $\tilde{N}_{j'}$ are no more than M while $\hat{U}_{m,j'} \geq 0$, we have

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k](\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k}{[(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k](\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k} \leq ck^2 \quad \text{a.s.}$$

Combining this with (A42) gives the desired result. \square

Proposition 6a is the counterpart of Proposition 5a for splitting instead of merging.

Proposition 6a. *Assume Conditions 1-4, and further that $H_i(t) = 0$. Let Θ be the set of all pairs of tracking solution sequences, $\hat{\theta}_k = (\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ and $\tilde{\theta}_k = (\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ that have the following property. All of the tracks that make up $\hat{\theta}_k$ and $\tilde{\theta}_k$ are correct track segments and $\tilde{\theta}_k$ differs from $\hat{\theta}_k$ at every k only by declaring a splitting event in place of two births and a death for a target i_1 that died in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ with two targets i_2 and i_3 that were born in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$. It is further assumed that the death of target i_1 and the birth of targets i_2 and i_3 may be incorrectly specified in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$, but at least two of these three events must correspond to two distinct events in the correct solution $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$. Then,*

$$R_{6a} = \sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k | \mathcal{Z}_k](\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k}{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k | \mathcal{Z}_k](\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k} = 0 \quad \text{eventually as } k \rightarrow \infty \text{ a.s.}$$

Proof. This follows the exact same arguments as the proof of Proposition 5a. \square

Proposition 6b is also the counterpart of Proposition 5b for splitting instead of merging. Again, when the events in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ that we join together into a splitting event in $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ do not correspond to actual events in the correct solution, there may be an advantage to switch to $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$. This will need to be dealt with in the proof of Theorem 1.

Proposition 6b. *Assume Conditions 1-4, and further that $H_i(t) = 0$. Let Θ be the set of all pairs of tracking solution sequences, $\hat{\theta}_k = (\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ and $\tilde{\theta}_k = (\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ that have the following property. All of the track segments that make up $\hat{\theta}_k$ and $\tilde{\theta}_k$ are correct track segments and $\tilde{\theta}_k$ differs from $\hat{\theta}_k$ at every k only by declaring a splitting event in place of two births and a death for a target i_1 that died in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ with two targets i_2 and i_3 that were born in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$. Then*

$$R_{6b} = \sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}_k][(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k]}{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}_k][(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k]} \leq bk^2(\log k)^c \text{ as } k \rightarrow \infty \text{ a.s.}$$

for some positive constants b and c which depend on ω .

Proof. Because of the symmetry of splitting and merging, this is identical to the proof of Proposition 5b. \square

Appendix F: Proof of Theorem 1

Proof of Theorem 1. Due to the Cameron Martin theorem, Theorem 18.22 of [Kallenberg \(2002\)](#), under the assumptions of Conditions 3 and 4, the distribution of $H_i(t) + \sigma G_i(t)$ and the distribution of $\sigma G_i(t)$ are mutually absolutely continuous. Because we are only interested in events of probability one, such as almost sure convergence, we can without loss of generality assume that $H_i(t) = 0$ for all targets.

Let $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ be an arbitrary incorrect solution sequence. The solution $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ has incorrect tracks, and/or incorrectly labeled events. Realize, however that the tracks of any incorrect solution are made up of correct track segments which are only joined together incorrectly. We can define \mathcal{T} to be the minimal set of correct track segments that makes up $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$. Since there is a finite number of possible track segments for each k , this minimal set exists for all k .

Let the difference between solutions described in Propositions 1a, 1b, ... 6b be referred to as difference 1a, difference 1b, ..., difference 6b. We can apply differences 1a, 1b, 2, and 3 sequentially to break $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$ into the track segments in \mathcal{T} . The diagrams in Figure 8 illustrate this process. We can then connect these track segments together using differences 4, 5a, 5b, 6a, and 6b sequentially to form the tracks of $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$; See the last transition in Figure 8. Furthermore, we can do this by applying any of the differences 5a, 5b, 6a, and 6b that we may need before applying any difference 4's. So all of the difference 1a, 1b, 2, 3, 5a, 5b, 6a are applied to solutions with tracks that are correct track

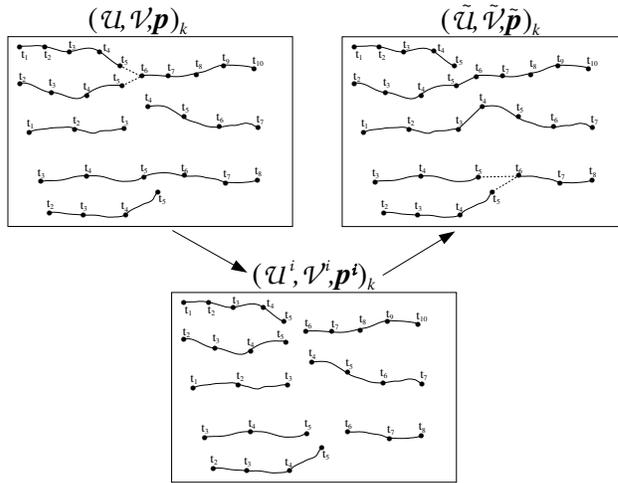


FIG 8. Break $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$ into track segments and connect them to get $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$.

segments and hence fit into the conditions of their respective propositions. In actuality, there can be no difference 4's eventually by Proposition 4 anyway.

So there is a sequence of solutions that starts with $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$ and passes through several incorrect solutions to arrive at $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$. Each element of this sequence has one and only one of the differences described above from the previous element. We shall write this sequence as

$$(\mathcal{U}, \mathcal{V}, \mathcal{P})_k, (\mathcal{U}^1, \mathcal{V}^1, \mathcal{P}^1)_k, (\mathcal{U}^2, \mathcal{V}^2, \mathcal{P}^2)_k, \dots, (\mathcal{U}^l, \mathcal{V}^l, \mathcal{P}^l)_k, (\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k.$$

We can write the likelihood ratio of any incorrect solution $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ to that of the correct solution $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$ as

$$\begin{aligned} & \frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}_k](\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k}{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}_k](\mathcal{U}, \mathcal{V}, \mathcal{P})_k} \\ &= \frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}_k](\mathcal{U}^1, \mathcal{V}^1, \mathcal{P}^1)_k}{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}_k](\mathcal{U}, \mathcal{V}, \mathcal{P})_k} \cdot \frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}_k](\mathcal{U}^2, \mathcal{V}^2, \mathcal{P}^2)_k}{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}_k](\mathcal{U}^1, \mathcal{V}^1, \mathcal{P}^1)_k} \\ & \dots \frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}_k](\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k}{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}_k](\mathcal{U}^l, \mathcal{V}^l, \mathcal{P}^l)_k}. \end{aligned} \tag{A43}$$

Let Θ be the set of all tracking solution sequences satisfying Condition 8 eventually. That is they have no more than M tracks and they restrict the distance between consecutive observations in a track to be less than $(c \log k^{-1})$. We claim that the correct sequence of solutions $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$ is in this set.

Obviously $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$ has no more than M tracks. And for the difference between consecutive observations, in any of the tracks we have

$$\begin{aligned}
& \sup_{\substack{0 \leq t_1 \leq t_2 \leq T \\ t_2 - t_1 \leq ck^{-1}}} ((X(t_2) - X(t_1))^2 + (Y(t_2) - Y(t_1))^2)^{1/2} \\
& \leq \sigma \vee \eta \sup_{\substack{0 \leq t_1 \leq t_2 \leq T \\ t_2 - t_1 \leq ck^{-1}}} \left(\frac{(X(t_2) - X(t_1))^2 + (Y(t_2) - Y(t_1))^2}{\sigma^2 \vee \eta^2} \right)^{1/2} \\
& \leq \sigma \vee \eta \sup_{\substack{0 \leq t_1 \leq t_2 \leq T \\ t_2 - t_1 \leq ck^{-1}}} \left(\frac{(X(t_2) - X(t_1))^2}{\sigma^2} + \frac{(Y(t_2) - Y(t_1))^2}{\eta^2} \right)^{1/2} \\
& = \sigma \vee \eta \sup_{\substack{0 \leq t_1 \leq t_2 \leq T \\ t_2 - t_1 \leq ck^{-1}}} \|B(t_2) - B(t_1)\| \\
& \leq (2c_1 k^{-1} \log(c_1^{-1} k))^{1/2} \text{ eventually as } k \rightarrow \infty
\end{aligned}$$

by Lemma 2. Hence all of the consecutive differences in the correct solution will eventually be smaller than $c \log k^{-1}$. Also note that by Proposition 4, there can be no sequence $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k \in \Theta$ that has any difference 4's from the correct solution.

Let $\Theta' = \Theta \setminus \{(\mathcal{U}, \mathcal{V}, \mathcal{P})_k\}$. We need to show that the supremum over $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k \in \Theta'$ of the ratio in (A43) converges to 0 as $k \rightarrow \infty$. By Propositions 1a, 1b, 2, 3, 5a, and 6a, any of the ratios in (A43) that have differences 1a, 1b, 2, 3, 5a, and 6a, are $\leq O(k^{-2}(\log k)^c)$. However, if any of the terms in (A43) have differences 5b or 6b, they can be as big as $ck^2(\log k)^c$. If there are two many of these differences then the ratio may not converge to 0. We will then have to consider how these differences could be applied to obtain $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$. We will first consider how there can be one difference 5b or 5a in an interval $[t_{j'}^k, t_{j'+1}^k]$, then consider multiple differences.

Suppose exactly one difference 5b was applied to $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$ during the interval $[t_{j'}^k, t_{j'+1}^k]$ to obtain $(\mathcal{U}^{i+1}, \mathcal{V}^{i+1}, \mathcal{P}^{i+1})_k$. For difference 5b, we must merge together two deaths and a birth, of which no two of these three events can correspond to distinct actual events in $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$. Otherwise this would be difference 5a. The three events must also be in the same interval $[t_{j'}^k, t_{j'+1}^k]$, otherwise in a manner similar to that of Proposition 5a, the ratio

$$\frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}](\mathcal{U}^{i+1}, \mathcal{V}^{i+1}, \mathcal{P}^{i+1})_k}{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}](\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k} = 0$$

eventually.

So before we can apply difference 5b, we must first use differences 1a, 1b, 2, or 3, to create at least two of the three events (two deaths and a birth) in $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$. Notice that we cannot use differences 2 and 3 together to create these events since then two of the three events would correspond to distinct events in $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$.

So there are exactly five ways difference 5b can be applied.

1. We could use a correctly labeled death in $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$. We would then still need a birth and a death. This would require using at least one difference 1a in the interval $[t_{j'}^k, t_{j'+1}^k]$ previously in our sequence of solutions to get

to $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$. It is difference 1a not difference 1b since the interval that contains the correctly labeled death is fixed. The overall contribution then of difference 5b to the ratio in (A43) is no more than

$$R_{1a}R_{5b} \leq O(k^{-3}(\log k)^c) \cdot bk^2(\log k)^c = O(k^{-1}(\log k)^{2c}).$$

2. We could use a correctly labeled birth in $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$. We would then still need two deaths. This would require using at least two applications of difference 1a in the interval $[t_{j'}^k, t_{j'+1}^k)$ previously in our sequence of solutions.
3. We could use difference 2 to get two deaths and a birth previous to $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$, but only two of these three events can be used in the difference 5b, otherwise we would be reconstructing a correct merger. So we would still need to apply difference 1a previous to $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$ at least once to get another birth or death.
4. We could use difference 3 to get a deaths and two births previous to $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$, but we would still need at least one more death to apply difference 5b. Hence we would need to apply difference 1a previous to $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$ at least once to get the other death.
5. Lastly, we could apply difference 5b in an arbitrary time interval, but this would require us to create two deaths and a birth in the interval $[t_{j'}^k, t_{j'+1}^k)$ previous to $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$. This would require at least two applications of difference 1b since now the interval is arbitrary.

Notice for cases 2-5 that there must be at least two applications of differences 1a, 1b, 2, or 3. The largest resulting ratio from any of these is $O(k^{-2+\varepsilon})$ for an arbitrarily small ε . Hence the overall contribution of difference 5b to the ratio of (A43) for cases 2-5 is no more than $O(k^{-2+2\varepsilon})$.

Therefore if there is one difference 5b applied in any interval then the overall contribution of that difference to the ratio in (A43) for any of the five cases is no more than $O(k^{-1}(\log k)^{2c})$.

Now Suppose exactly one difference 6b was applied to $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$ during the interval $[t_{j'}^k, t_{j'+1}^k)$ to obtain $(\mathcal{U}^{i+1}, \mathcal{V}^{i+1}, \mathcal{P}^{i+1})_k$. Because of the symmetry of the problem the logic is identical to the five cases above and the overall contribution of that difference to the ratio of (A43) is no more than $O(k^{-1}(\log k)^{2c})$.

We now consider N^* differences 5b and/or 6b applied in a single interval $[t_{j'}^k, t_{j'+1}^k)$, where $N^* > 1$. Again let $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$ be the element of the solution sequence just before we apply the N^* differences 5b and/or 6b. Note that applying differences 5b and/or 6b N^* times requires at least N^* deaths and at least N^* births in $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$ during the interval $[t_{j'}^k, t_{j'+1}^k)$. Note that this N^* combinations of differences 5b and/or 6b can only use one of an actual death, an actual birth, an actual merger with difference 2, or an actual split with difference 3 since eventually only one of these events will be in the interval $[t_{j'}^k, t_{j'+1}^k)$. There are again the same five cases to consider:

1. We could use a correctly labeled death in $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$. We would then still need at least N^* more births and this would require N^* applications

of difference 1a in the interval $[t_{j'}^k, t_{j'+1}^k)$ previously in our sequence of solutions. So the contribution of the N^* differences 5b and/or 6b to the ratio of (A43) is no more than

$$(R_{1a})^{N^*} (R_{5b})^{N^*} = O(k^{-N^*} (\log k)^{2N^*c}).$$

2. We could use a correctly labeled birth in $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$. We would then still need at least N^* deaths. This would require using at least N^* applications of difference 1a previously. So the contribution of differences 5b and/or 6b to the ratio is again no more than

$$(R_{1a})^{N^*} (R_{5b})^{N^*} = O(k^{-N^*} (\log k)^{2N^*c}).$$

3. We could use difference 2 to get two deaths and a birth previous to $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$, but we would still need $N^* - 1$ more births which would require $N^* - 1$ applications of difference 1a previously. This makes the contribution of differences 5b and/or 6b to the ratio no more than

$$R_2 \cdot (R_{1a})^{N^*-1} (R_{5b})^{N^*} = O(k^{-(N^*-1)} (\log k)^{2N^*c}).$$

4. We could use difference 3 to get a death and two births previous to $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$, but we would still need at least $N^* - 1$ more deaths. Hence we would need to apply difference 1a at least $N^* - 1$ times previous to $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$ get the other deaths. This makes the contribution to the ratio no more than

$$R_3 \cdot (R_{1a})^{N^*-1} (R_{5b})^{N^*} = O(k^{-(N^*-1)} (\log k)^{2N^*c}).$$

5. Lastly, we could apply the differences 5b and/or 6b in an arbitrary time interval. Let N_5 and N_6 be the number of applications of difference 5b and difference 6b respectively so that $N_5 + N_6 = N^*$. Let N_b and N_d be the minimum number of births and deaths needed respectively. Notice that $N_b = N_5 + 2N_6$ and $N_d = 2N_5 + N_6$. The minimum number of applications of difference 1b that we would need is

$$\min_{\substack{N_5, N_6 \\ N_5 + N_6 = N^*}} \{N_b \vee N_d\} = \min_{\substack{N_5, N_6 \\ N_5 + N_6 = N^*}} \{(N_5 + 2N_6) \vee (2N_5 + N_6)\}.$$

This minimum is achieved when $N_5 = \lceil \frac{N^*}{2} \rceil$ and $N_6 = \lfloor \frac{N^*}{2} \rfloor$ or vice versa. In either case this makes

$$\min_{\substack{N_5, N_6 \\ N_5 + N_6 = N^*}} \{N_b \vee N_d\} = 2 \left\lceil \frac{N^*}{2} \right\rceil + \left\lfloor \frac{N^*}{2} \right\rfloor \geq N^* + \left\lceil \frac{N^*}{4} \right\rceil \quad \forall N^* \geq 1.$$

This means we need at least $\lceil (5/4)N^* \rceil$ applications of difference 1b previous to $(\mathcal{U}^i, \mathcal{V}^i, \mathcal{P}^i)_k$, which means the contribution of the N^* differences 5b and/or 6b to the ratio is no more than

$$(R_{1b})^{(5/4)N^*} (R_{5b})^{N^*} \leq Ck^{-(1/2+\epsilon)N^*} (\log k)^{N^*c} = O\left(k^{-(1/2+\epsilon)N^*}\right)$$

for an arbitrarily small ϵ and $\epsilon > \epsilon$. This along with the case of $N^* = 1$ tells us that for any number of differences 5b and/or 6b applied in any one interval then the overall contribution of that difference to the ratio of (A43) is no more than $O(k^{-1+\epsilon})$ which happens when $N^* = 2$.

So if we apply any number of differences 5b or 6b to the correct solution $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$ to obtain an incorrect solution $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ we have

$$\sup_{(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k \in \Theta} \frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k | \mathcal{Z}](\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k}{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k | \mathcal{Z}](\mathcal{U}, \mathcal{V}, \mathcal{P})_k} \leq O(k^{-1+\epsilon}). \quad (\text{A44})$$

Of course if we don't apply any differences 5b or 6b then we had to apply at least one of the differences 1a, 1b, 2, 3, 5a, or 6a. This would make the ratio

$$\sup_{(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k \in \Theta} \frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k | \mathcal{Z}](\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k}{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k | \mathcal{Z}](\mathcal{U}, \mathcal{V}, \mathcal{P})_k} \leq O(k^{-2+\epsilon}).$$

Hence in general we have (A44). This means for a given ω there exists a $K(\omega)$ s.t. for all $k > K(\omega)$,

$$\sup_{(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k \in \Theta} \frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k | \mathcal{Z}_k](\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k}{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k | \mathcal{Z}_k](\mathcal{U}, \mathcal{V}, \mathcal{P})_k} < 1.$$

Hence, our estimate $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k = (\mathcal{U}, \mathcal{V}, \mathcal{P})_k$ for all $k > K$. \square

References

- BAR-SHALOM, Y., LI, X. & KIRUBARAJAN, T. (2001), *Estimation with Applications to Tracking and Navigation*, John Wiley & Sons, Inc., New York, NY.
- BLACKMAN, S. (2004), 'Multiple hypothesis tracking for multiple target tracking', *IEEE A&E Systems Magazine* **19**, 5–18.
- BLACKMAN, S. & POPOLI, R. (1999), *Design and Analysis of Modern Tracking Systems*, Artech House, Boston.
- CHANG, K. & BAR-SHALOM, Y. (1984), 'Joint probabilistic data association for multitarget tracking with possibly unresolved measurements and maneuvers', *IEEE Transactions on Automatic Control* **29**, 585–594.
- CHEN, H., LI, X. & BAR-SHALOM, Y. (2004), 'On joint track initiation and parameter estimation under measurement origin uncertainty', *IEEE Transactions on Aerospace and Electronic Systems* **40**, 675–694.
- CONG, S. & HONG, L. (1999), 'Computational complexity analysis for multiple hypothesis tracking', *Mathematical and computer modeling* **29**, 1–16. [MR1698393](#)
- DAUM, F. E. (1994), 'The importance of resolution in multiple target tracking', *Signal and Data Processing of Small Targets 1994, SPIE Proceedings* **2235**, 329–338.

- DOUCET, A., DE FREITAS, N. & GORDON, N. (2001), *Sequential Monte Carlo Methods in Practice*, Springer Verlag, New York, NY. [MR1847783](#)
- DOUCET, A., GODSILL, S. & ANDRIEU, C. (2000), 'On sequential Monte Carlo sampling methods for Bayesian methods', *Statistics and Computing* **10**, 197–208.
- FORTMANN, T., BAR-SHALOM, Y. & SCHEFFE, M. (1983), 'Sonar tracking of multiple targets using joint probabilistic data association', *IEEE Journal of Oceanic Engineering* **8**, 173–184.
- GELFAND, A. (1990), 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association* **85**(410), 398–409. [MR1141740](#)
- GENOVESIO, A. & OLIVO-MARIN, J. (2004), 'Split and merge data association filter for dense multi-target tracking', *Proceedings of the 17th International Conference on Pattern Recognition* pp. 677–680.
- GORDON, N., SALMOND, D. & SMITH, A. (1993), 'Novel approach to nonlinear/non-Gaussian Bayesian state estimation', *IEE Proceedings-F* **140**, 107–113.
- HALL, P., PENG, L. & RAU, C. (2001), 'Local likelihood tracking of fault lines and boundaries', *Journal of the Royal Statistical Society B* **63**, 569–582. [MR1858403](#)
- HALL, P., QIU, P. & RAU, C. (2007), 'Tracking edges, corners and vertices in an image'. unpublished manuscript.
- HALL, P. & RAU, C. (2000), 'Tracking a smooth fault line in a response surface', *Annals of Statistics* **28**, 713–733. [MR1792784](#)
- KALLENBERG, O. (2002), *Foundations of Modern Probability*, 2nd edn, Springer-Verlag, New York, NY. [MR1876169](#)
- KHOSHNEVISAN, D. (2002), *Multiparameter Processes; An Introduction to Random Fields*, Springer-Verlag, New York, NY. [MR1914748](#)
- KITAGAWA, G. (1996), 'Monte Carlo filter and smoother for non-Gaussian nonlinear state models', *Journal of Computational and Graphical Statistics* **5**, 1–25. [MR1380850](#)
- KOCH, W. & VAN KEUK, G. (1997), 'Multiple hypothesis track maintenance with possibly unresolved measurements', *IEEE Transactions on Aerospace and Electronic Systems* **33**, 883–892.
- LI, J. & JING, Z. (2003), 'Practical system for tracking multiple maneuvering targets', *Optical Engineering* **42**, 2439–2451.
- LIU, J. & CHEN, R. (1998), 'Sequential monte carlo methods for dynamic systems', *Journal of the American Statistical Association* **93**, 1032–1044. [MR1649198](#)
- MAHLER, R. (2003), 'Multitarget Bayes filtering via first-order multitarget moments', *IEEE Transactions on Aerospace and Electronic Systems* **39**, 1152–1178.
- MORI, S., CHONG, C., TSE, E. & WISHNER, R. (1986), 'Tracking and classifying multiple targets without a priori identification', *IEEE Transactions on Automatic Control* **31**, 401–409.

- REID, D. (1979), 'An algorithm for tracking multiple targets', *IEEE Transactions on Automatic Control* **24**, 843–854.
- REVUZ, D. & YOR, M. (1999), *Continuous Martingales and Brownian Motion*, Springer-Verlag, New York, NY. [MR1725357](#)
- SHEPP, L. (1966), 'Radon nikodym derivatives of Gaussian measures', *Annals of Mathematical Statistics* **37**, 321–354. [MR0190999](#)
- STORLIE, C., LEE, T., HANNIG, J. & NYCHKA, D. (2009), 'Tracking of multiple merging and splitting targets: A statistical perspective (with discussion)', *Statistica Sinica* **19**, 1–52. [MR2487870](#)
- TRUNK, G. & WILSON, J. (1981), 'Track initiation of occasionally unresolved radar targets', *IEEE Transactions on Aerospace and Electronic Systems* **17**, 122–130.
- VO, B., SINGH, S. & DOUCET, A. (2005), 'Sequential Monte Carlo methods for Bayesian multitarget filtering with random finite sets', *IEEE Transactions on Aerospace and Electronic Systems* **41**, 1224–1245.