DISSERTATION

TRACKING OF MULTIPLE MERGING AND SPLITTING TARGETS WITH APPLICATION TO CONVECTIVE SYSTEMS

Submitted by

Curtis B. Storlie

Department of Statistics

In partial fullfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2005

COLORADO STATE UNIVERSITY

September 10, 2005

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY CURTIS STORLIE ENTITLED TRACKING OF MULTIPLE MERGING AND SPLITTING TARGETS WITH APPLICATION TO CONVECTIVE SYSTEMS BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Aubrey Poore

Douglas Nychka

Thomas Lee (Co-Adviser)

Jan Hannig (Co-Adviser)

F. Jay Breidt (Department Head)

ABSTRACT OF **DISSERTATION**

TRACKING OF MULTIPLE MERGING AND SPLITTING TARGETS WITH APPLICATION TO CONVECTIVE SYSTEMS

A statistical approach to multiple target tracking is presented which allows for birth, death, splitting and merging of targets. Targets are also allowed to go undetected for several frames. The splitting and merging of targets is a novel addition for a statistically based tracking algorithm. This addition is essential for the tracking of storms, which is the motivation for this work. The utility of this tracker extends well beyond the tracking of storms however. It can be valuable in other tracking applications that have splitting or merging, such as vortexes, radar/sonar signals, or groups of people. The method assumes that the location of a target behaves like a Gaussian Process when it is observable. A Markov Chain State Model decides when the birth, death, splitting, or merging of targets takes place. The tracking estimate is achieved by an algorithm that finds the tracks that maximize the conditional density of the unknown variables given the data. The problem of how to quantify the confidence in a tracking estimate is addressed as well. Some theoretical properties of this tracking estimate are also developed such as sufficient conditions for consistency. The properties of the proposed method will be demonstrated on simulated data. Finally, the method is applied to the problem for which it was designed, tracking storms from radar reflectivity data.

Curtis Byron Storlie
Statistics Department
Colorado State University
Fort Collins, CO 80523
Fall 2005

ACKNOWLEDGMENTS

I would like to express deep gratitude to my advisors Professor Thomas Lee and Professor Jan Hannig for their steadfast patience and guidance over the years. When they did not have time to meet with me, they made time. They went way out of their way on many occasions to help me in this pursuit. I cannot express in words how grateful I am for them.

Many thanks to Doug Nychka and NCAR for supporting me financially throughout the last two years. This support was essentially to work on the problem that eventually became the heart of this thesis. Doug not only provided me with a thesis problem but also played a crucial role in the development of the ideas used to solve this problem. I also want to mention Chris Davis at NCAR who helped me better understand the atmospheric science behind convective systems. Tim Hoar was also a very important resource for his help with the computational aspects of this problem.

I would also like to thank the other members of my committee, Professor Jay Breidt and Professor Aubrey Poore, for devoting their valuable time to oversee my research. The entire Colorado State University Statistics Department also deserves my thanks for giving me the opportunity to grow as a Statistician over the years.

I most certainly would not have gotten this far without the loving support of my wife Andrea. Her patience and understanding over the last several months has been supernatural. Most importantly, I would like to thank my Lord and Savior, Jesus Christ, for being with me always through all of my tribulations and giving me peace during this stressful time in my life.

Contents

List of Figures

Chapter 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. The approach developed in this thesis is motivated by the applications of storm tracking and tracking vortexes in turbulence fields. In these applications, the splitting and merging of targets is quite common. The term target will be used repeatedly throughout to denote a generic object of interest. In this thesis, we develop a tracking method that can effectively handle splitting and merging targets. The utility of the ideas developed here extend well beyond the tracking of storms and vortexes however. The method can be valuable in other tracking applications that have splitting or merging targets, such as radar/sonar signals or group tracking. In particular, the ideas developed here with respect to confidence in a tracking estimate and consistency of an estimate apply to any tracking application.

1.1 Two Scientific Problems

This section highlights two scientific problems for which the tracking of targets is an important step to their solutions.

1.1.1 Convective Systems

The first problem concerns the study of the evolution of storm/rainfall systems captured by satellite radar imaging techniques. Figure 1.1 shows radar reflectivity images evolving over time on July 14, 1996 from 1:00am to 3:30am. Radar reflectivity is highly correlated to rainfall intensity so that we can attribute the variation of color in these images to different rainfall intensities. In these images blue would indicate 0 inches/hour of rainfall increasing on a log scale to bright yellow indicating ≥ 2 inches/hour. The images are separated by 30 minutes.

The goal here is to track the larger convective systems in these images. For our purposes, a convective system is defined to be a rainfall system that is larger than 100 km in length (approximately 1° of latitude or longitude).

This problem has been studied previously in [20, 57]. The very short term behavior (less than 1 hour) of such systems are reasonably well-known, but the moderately short term (1 to 6 hours) and long term (1 to 2 days) behaviors are still largely unknown. It is certainly desirable if such longer term behaviors are better understood, and a useful tool that would help in this direction is a procedure that automatically monitors the movements and the interactions (e.g., merging and splitting) of all the structures in the overall system.

The merging of two systems that are located at the corner of South Dakota, Wyoming, and Nebraska can be seen in the first few images. These are clearly separate systems until

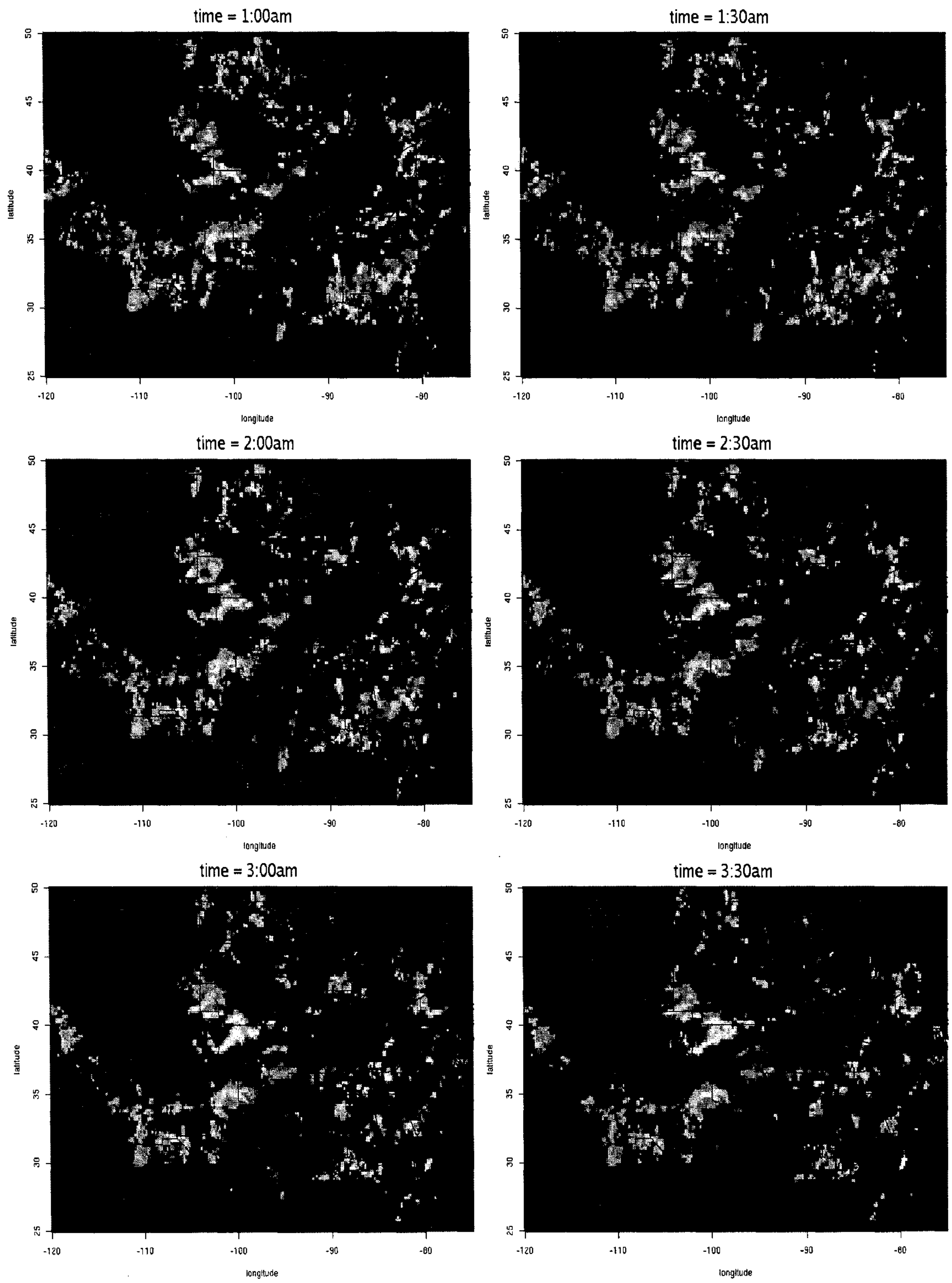

Figure 1.1: Radar Reflectivity Images Evolving During July 14, 1996

the third image, where they are now one system. There is also a splitting event in the panhandle of Florida as a fairly large system breaks apart into two smaller systems

Another aspect to this problem is the validation and improvement of the storm activity in Regional Climate Models. These are complex computer models for the state of the atmosphere. These models are used for prediction of weather and to otherwise generate data to answer scientific hypotheses. It is necessary for these models to maintain a certain degree of reality in terms of the storms that they produce. Comparing storm tracks from real data to those of the Regional Climate Model is one way to verify this.

1.1.2 2-D Turbulence

The second motivating problem focuses on a 2-dimensional turbulence simulation of freely decaying vortexes; see Figure 1.2. These images are of a 2-dimensional vorticity field with random initial conditions as it develops over time with no energy loss to the overall system. The white objects are centers of vorticity rotating in a clockwise direction, whereas the black vortexes have the opposite rotation.

Vortexes of the same spin will coalesce as they get close to each other. There is a good example of a merger on the left edge of the images a little bit below center. Also, vortexes of opposite spin have a tendency to parallel each other for a while before moving off in different directions. Two vortexes exhibiting this behavior are called dipoles. An example of dipoling can be seen by the two larger vortexes in the lower right of the images.

Recently such image sequences are a subject of much research [10, 45, 61], as it is (i) a paradigm for anisotropic geophysical and astrophysical turbulence, and (ii) it is also the most computationally accessible example of fluid turbulence. Automatic tracking of such vortexes is an important step to understanding the interactions amongst these structures.

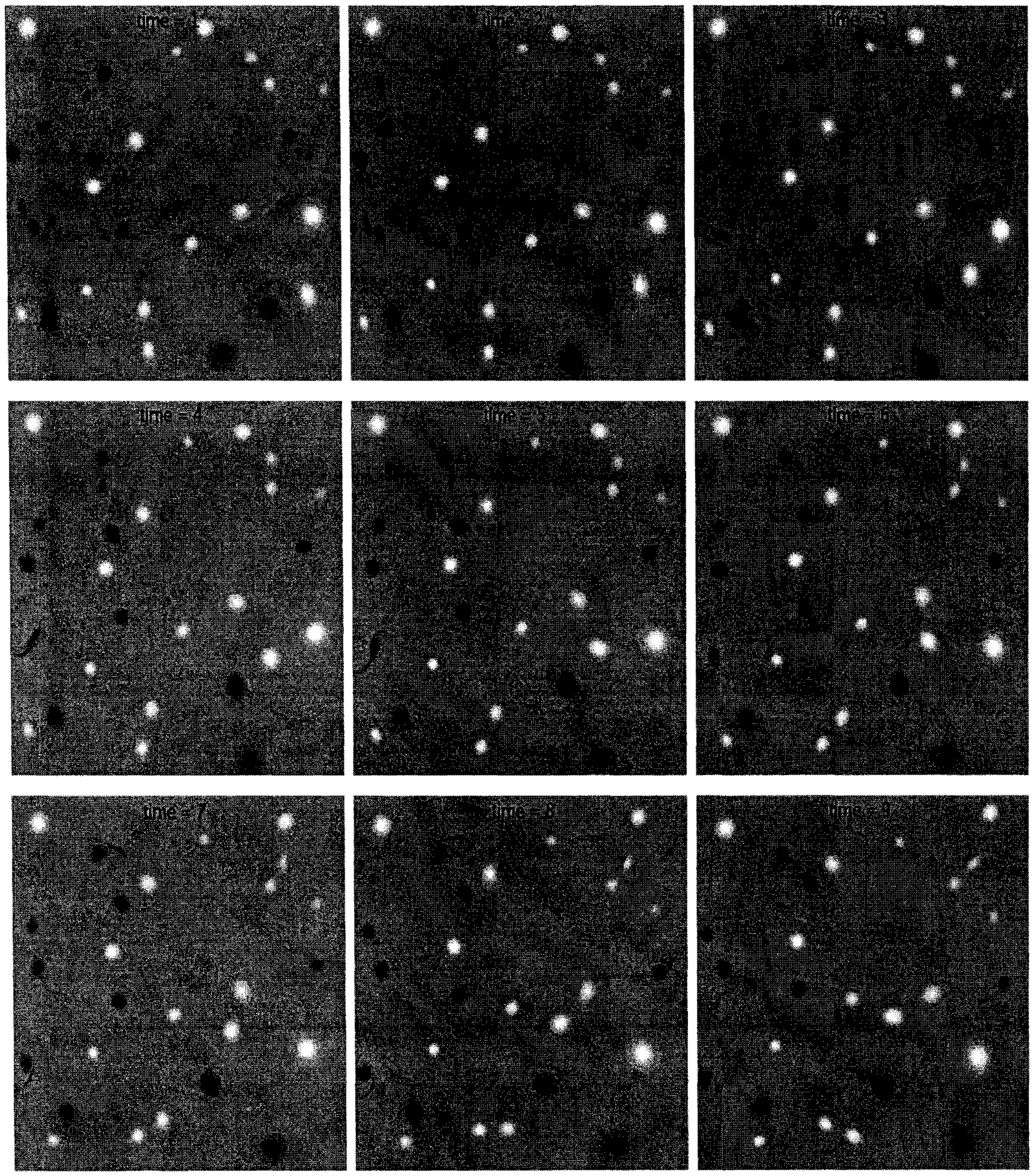

Figure 1.2: 2-D Turbulence Simulation

An important characteristic to these two tracking problems is that the targets exhibit merging, and for the former problem, the targets also exhibit splitting. Thus the tracking algorithm developed to monitor these systems must be able to identify this behavior.

1.2 Description of the Tracking Problem

The basic premise of the multiple target tracking problem is that we wish to follow targets through a sequence of images or scans. Each image contains information about the locations (and perhaps other attributes) of the targets. We wish to take this information and recover the tracks of the targets.

Tracking algorithms are usually dependent on an identification algorithm. The identification algorithm will record the location of each target that it finds in each image. In some cases, it may record other information about the target as well, such as size, shape, intensity, etc. The key user-defined parameter in an identification algorithm is the level of thresholding. If the threshold is too low, we will have too many many spurious observations or false alarms (often called clutter). If it is set too high, then we will have too many missed detections of actual targets.

Some methods use adaptive thresholding to aid in the identification process [62, 63, 29, 12]. In these approaches, the thresholding is set lower in regions that are likely to contain observations from actual targets (see discussion of gating in Section 1.3.2). This way we are more likely to detect an actual target if it still exists, but we will not pick up excessive false alarms in the other regions.

There are some algorithms that avoid the thresholding step altogether and identify and track the targets all in one step [30, 1, 58]. These methods are known as track before detect algorithms in the literature even though the tracking and detection are actually performed simultaneously. They have been shown to be very effective at tracking dim targets in the presence of clutter.

We are assuming the more traditional problem here however, that the tracking is performed on the data obtained after image processing. Adaptive thresholding can always be used in conjunction with this approach. The goal of the tracking algorithm then, is to take the location data recorded over time and recover the track of each target. A track is defined to be the $(x(t), y(t))$ coordinates of each target at each time during the image sequence.

To illustrate the idea further, assume there are 4 targets at each time step as in Figure 1.3. The locations of the targets in this figure are simulated from the model to be described in Chapter 2. We are however ignoring the possibility of birth and death of targets as well as missing and false alarm observations for the time being.

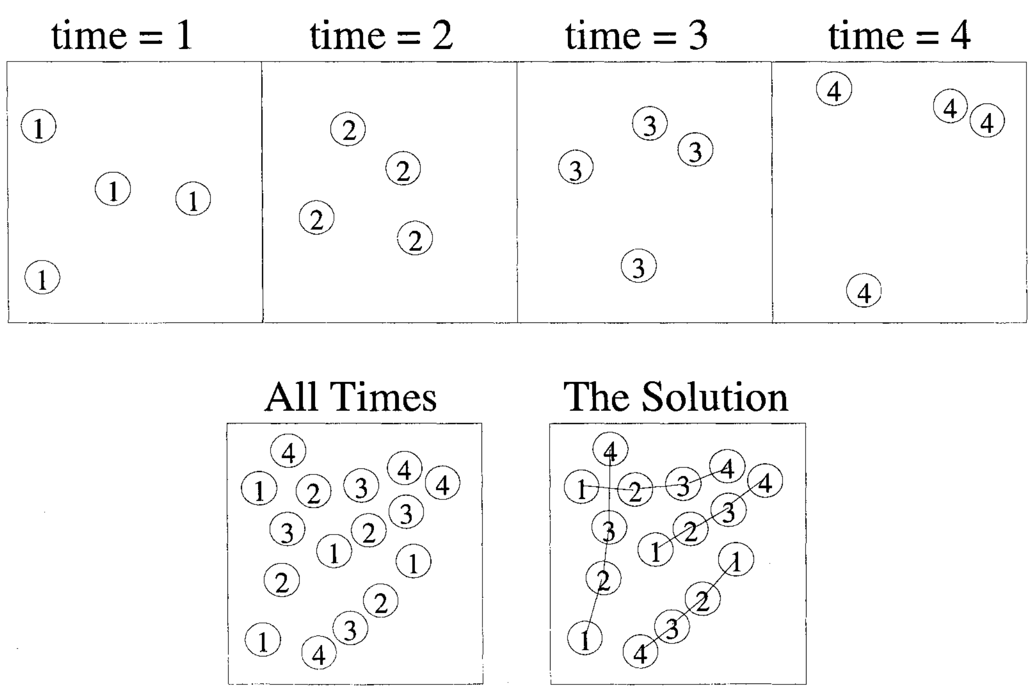

Figure 1.3: Illustration of the tracking problem.

The targets in the figure are labeled according to time so that time is distinguishable when all times are included in the plots at the bottom of the figure. The targets are free to change position from one time step (image) to the next. The data association problem is to determine which temporal set of locations corresponds to one particular target. Another way to think of this is how to sequentially connect the targets in the plot in the bottom left to form tracks. Since this is simulated data, the solution to this is known and the correct tracks are given in the bottom right plot. Typically the position and/or velocity of the targets is estimated once this data association problem is solved. An exception to this is the JPDA approaches discussed in Section 1.3.3.

There is a very slight difference in the way we will use the terms paths and tracks. We will use tracks to emphasize the data association problem in terms of selecting several sequential observations to put into a collection that we will call a track. A path will be used more to denote the continuous location curve of a target in 2-D space. So in our definition, a track is the discrete counterpart of a path.

Of course Figure 1.3 is an overly simplistic representation of reality. Most applications will not have the same number of targets in each image. Imperfect detection and occlusion will lead to missing targets in some images. There can also be false alarms or clutter. As mentioned previously, there is usually a trade off in number of false alarms to the number of missed detections as determined by the amount of thresholding that is used in the detection algorithm.

In addition, some targets may appear for the first time or disappear for good in the middle of the image sequence. These events, which we will call birth and death, can happen when a target moves into or out of the field of vision. In the case of storm tracking, the targets can actually form or disappear right in the middle of the field of vision.

As already mentioned, in our problems we deal with situations where two targets combine into one larger target (merging) and one target breaks apart into two smaller targets (splitting). Figure 1.4 shows our simple example again but now with a birth, missing observation, and a split.

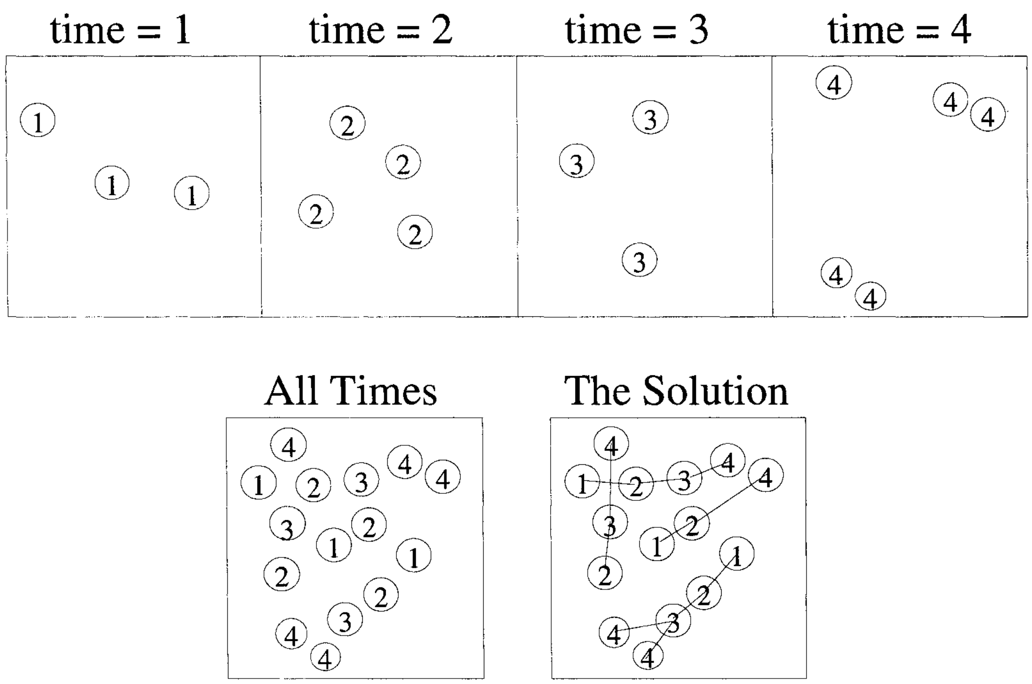

Figure 1.4: Illustration of the tracking problem.

1.3 Literature Review

In the following, we will give a summary of the work relevant to our problem that has been described in the tracking literature. For a more detailed review of tracking techniques, the reader is referred to two very comprehensive books by Bar-Shalom [2] and Blackman [6].

The target tracking problem has been studied extensively in the engineering literature over the past thirty years. The approaches to this problem can be separated into statistical and non-statistical groups. The non-statistical approaches typically use either image differencing techniques for consecutive images [46] or some sort of heuristic penalty score approach based on a smoothness of motion assumption[55, 52]. In either case an optimization problem is solved by shuffling the labels of the observations around to create tracks that optimize some criteria.

These methods have been used with some success [28] and even in the area of tracking storms [32, 33, 15, 11, 25]. These approaches are attractive because of their speed and simplicity, but their weakness is in the inability to adequately initiate tracks (birth) and terminate them (death). In addition there is no way to assess the confidence in the tracking estimate. Because of this, the focus recently in the tracking literature has been in the area of the more powerful statistical approaches. These methods employ a statistical model to describe the motion of the targets to be tracked. Usually a Gaussian state space model is assumed, the most popular of which is the nearly constant velocity model which assumes that the acceleration of the target is a white noise. This is equivalent to an Integrated Brownian Motion (IBM) model. From here the Multiple Hypothesis approach to the problem is to find the hypothesis (collection of tracks) that results in the maximum likelihood of the model. Mori [43] describes a good general mathematical framework for this problem.

The use of hypothesis here means a particular labeling of the observations to create a collection of tracks, such as in Figures 1.3 and 1.4. It is not to be confused with a statistical hypothesis, but there is some similarity. For example, if we considered the collection of tracks to be a parameter, then we may wish to decide if a particular parameter is correct versus an alternative.

There are many different ways to label the targets to create tracks. Each set of possible tracks is a collection of permutations of the labels of the observations in each image. Suppose again for the moment that there is no birth, death or false alarms and refer to Figure 1.3.

In this simple example, there are 4 targets in each image and 4 images, so there are $4!^4 = 331,776$ possible ways to create tracks. This number becomes prohibitively large quite quickly when the problem is more realistic in terms of more targets and more time steps, not to mention allowing for birth, death, missing values, and false alarms. It is not feasible to go through all possible combinations to find the global solution. Therefore, an important part of developing a tracking algorithm is not only posing an optimization problem to solve, but also solving that problem in an efficient way. In the following, we will describe the three of the most widely used approaches to solving this problem.

1.3.1 Nearest Neighbor (NN, GNN)

The simplest approach to approximate the global maximum to the likelihood is to use the Nearest Neighbor (NN) in terms of the Mahalanobis Distance [17]. Here the previous locations in a track are used to determine which target in the next image is the 'most likely' to be part of that track. The 'most likely' target is determined to be the one that maximizes conditional likelihood of its location given the previous locations in the track. This most likely target is the same as the one that minimizes the Mahalanobis distance. Each path is assumed to follow some Gaussian state space model independent of the other paths, so the conditional likelihood or Mahalanobis distance is calculated efficiently via the Kalman Filter. The problem with the Nearest Neighbor approach is that one observation can be assigned to multiple targets. The Global Nearest Neighbor (GNN) approach is the same as the Nearest Neighbor with the restriction that all observations must belong to only one track. For more information on the Nearest Neighbor approach, the reader is referred to [17] or [23].

1.3.2 Multiple Hypothesis Tracking (MHT)

A major advance that generated much interest in statistical approaches to tracking was made with the Multiple Hypothesis Tracking Algorithm (MHT) of Reid [48]. This method allows for target birth and death in addition to classifying false alarms separately. The optimization algorithm used is similar in flavor to GNN, but it delays final decisions about correspondence until several images later when more information is available.

The MHT logic works as follows. Consider a collection of possible tracks after n images to be a hypothesis, H_n. Suppose after time step (image) n, we know the K_n most likely hypotheses $H_{n,1} \ldots H_{n,K_n}$ as determined by their likelihood. Now a new set of observations are received at the next time step, $n+1$. These new observations are used in conjunction with $H_{n,1}$ to form all "feasible" new hypotheses assuming that $H_{n,1}$ was correct. A feasible hypotheses is one that is formed from compatible tracks. Compatible tracks are those that share no observations. The same thing is done then for the rest of the hypotheses $H_{n,2} \ldots H_{n,K_n}$. The collection of the K_{n+1} most likely hypotheses are kept from all of these newly formed hypotheses and the rest are discarded. This process is then repeated at time step $n+2$.

The process is started by considering all possible hypotheses after two time steps. From this collection, the K_2 most likely are kept and the rest are discarded and the algorithm is off and running.

To make the algorithm quite a bit more efficient, a procedure called gating is used to screen out very unlikely tracks before forming a hypotheses. Consider a given track T which is part of a particular hypothesis, H_n, at time step n. At time step $n+1$ it is stated above that all new observations are considered for inclusion into track T to form new hypotheses.

However, we can limit the number of new observations to be considered for inclusion into track T by the following strategy. We can form a prediction region for the (x, y) location of the target from track T at time step $n+1$ using the previous observations in the track. Only new observations within this prediction region (or gate) are considered for inclusion into track T.

The advantage of MHT over GNN is that decisions about correspondences or false alarm vs. target at time step n can wait until several time steps later to be resolved. Of course, it is still possible for the optimal solution to be deleted at one of the iterations if too few hypotheses are kept at that time step. It is particularly important to hold on to a lot of hypotheses in the early time steps until enough data is available to establish tracks and hence make hypothesis decisions more accurate.

Many improvements on speed and efficiency of the original MHT algorithm have been made. In Reid's original implementation, many unlikely hypotheses must be evaluated before they are deleted. However, Cox [18] extended Murty's method [44] for finding the N-best solutions to an assignment problem to the MHT. This can be used to inhibit the formation and evaluation of many unlikely hypotheses.

There is also a track orientated approach to MHT in which the possible tracks are dealt with individually. All possible tracks are formed and maintained at each time step by confirming it or deleting it. A Sequential Likelihood Ratio Test (SQRT) is performed to determine if a track should be confirmed or deleted as described in Chapter 6 of [6]. Hypotheses are then reformed at each time step from compatible, likely tracks. The advantage here is that we can choose only the likely tracks, as deemed so by their likelihood, to be considered for hypothesis generation. The difficulty of course is that we need a way to keep straight which tracks are compatible and form only those hypotheses that contain compatible tracks. Alternatively, Poore [47] and Deb [21] have formed this problem as a multidimensional assignment problem and use Lagrangian Relaxation to achieve the compatible tracks constrained solution.

The difference between track oriented MHT and the traditional MHT is that in the former the actual tracks are maintained then hypotheses are formed as opposed to the latter where it forms and maintains the hypotheses directly. The question remains which approach is better, but it seems that both approaches give similar results in terms of accuracy (and speed when considering Cox's efficient implementation to the traditional MHT) [6].

A particular variation of the Multiple Hypothesis Algorithm used to solve the optimization problem we pose in Chapter 5 will be described in Section 5.2. For other applications of the MHT algorithm to a variety of scientific problems see [39, 64, 31, 34, 5]

1.3.3 Joint Probabilistic Data Association (JPDA)

Another approach that has received a lot of attention is the Joint Probabilistic Data Association (JPDA) algorithm developed by Fortmann, Bar-Shalom, and Scheffe [27]. The Probabilistic Data association approach is a different philosophy than the multiple hypothesis approach. It estimates the target state (position and velocity) by treating each observation as coming from a mixture distribution. For example, each new observation has a probability of originating from each target or a false alarm. The probability of a new observation originating from a particular target is computed conditionally based on the previous state of the target. The new state of the target is then estimated by a Gaussian approximation to what is really a Gaussian mixture using the Kalman Filter. Hence it uses the locations of all prospective targets at the current time to predict the state for each target at that time.

The JPDA has been shown to be very effectively for tracking targets in heavy clutter [40] and [3]. The downside for this method is that track initiation (birth) and termination (death) are more complicated [6] and [2]. Also, there is a coalescence problem for closely spaced targets [26]. For closely spaced targets the new observations produced by each target are close together. Hence, they both will have a high probability of coming from either target. In turn, the updated state estimates for each target will weight each observation highly. The result is that the state estimates will get closer and closer together. Research has been done on reducing the effect of this problem however [7].

1.3.4 Interacting Multiple Model (IMM)

The interacting multiple model (IMM) is a useful model for tracking maneuvering targets. The IMM was first proposed by Blom [8, 9] with a good review given in Chapters 1 - 3 of [2] and Chapter 4 of [6]. Some applications (i.e. tracking fighter jets) have targets that move for the most part in a straight line, but will suddenly alter their course (maneuver). The multiple model approach describes this behavior with separate models, one with a small motion variance component for the straight line motion and the other with a larger variance for the maneuvering. Of course, there can be more than two models for the behavior of the targets as well. In fact, applications for maneuvering targets typically assume three target motion models.

The target is allowed to switch from one model to the other at each time step according to a Markov chain. That is, at each time step, there is a probability $P_{i,j}$ of switching from model i to model j.

The IMM can be incorporated into any of the three optimization methods described above (GNN, MHT, or JPDA). For the storm tracking problem, this approach seems to be unnecessary since the motion of the storms tend to be fairly consistent. The IMM can be useful for the 2-D turbulence problem however, since the motion dynamics of the vortexes tend to change as vortexes get close to each other.

1.3.5 Comparison of GNN, MHT and JPDA

The GNN, JPDA, and MHT have been widely used for several decades. Unfortunately, there have been relatively few comparative studies performed. What has been done suggests that for tracking targets in cluttered environments, the JPDA and MHT are superior to the GNN. Also, the MHT seems to outperform the JPDA for more difficult problems [22, 60, 37]. The reason for this seems to be the delayed decision logic involved in the MHT. The JPDA uses only one step ahead logic to estimate the probability of an observation belonging to a target. Attempts have been made to extend the JPDA to incorporate a multiple frame logic [53, 24, 50]. Results from Salmond [53] for one of theses multi-scan JPDA methods indicate performance similar to that of MHT for a single target in clutter. In any case there is clear advantage to some form of delayed decision logic.

The MHT and JPDA methods have established a baseline for researchers to modify and improve these methods to fit there situation. Since the creation of these methods, contributions have typically been made by (1) developing better or more specific statistical models to describe the targets and/or (2) developing better/faster algorithms to approximate the corresponding optimization problem.

1.3.6 Merging and Splitting of Targets

Most recent approaches for target tracking are able to adequately handle track birth, death, missing values and clutter. But this paper is motivated by the problem of storm tracking from rainfall data. This problem comes with the issues above in addition to the merging and splitting of targets. Merging and splitting of targets has an important application to radar applications as well. When two targets are close together, resolution limits may prevent both targets from being detected. Hence, the scan returns one observation for both targets.

Although this is perceived as a very important problem by Daum [19] and Blackman [4], very little work has been done for merging and splitting of targets. What has been done for merging [59, 13, 38] and summarized in [6] pages 378-386, is not well defined in terms of an overall probabilistic model. It seems to be more of a patch for the problem than an accepted solution.

Also, the birth and death of targets for all of the methods described is dealt with in a hypothesis test approach. Recently however, Li [42] proposed the use of a Hidden Markov Model for birth and death of targets. This is more in line with the type of thinking that is needed here. We will extend this idea to include birth, death, splitting, and merging into an overall model for the targets. The overall model will then have a state component and a location (or motion) component. This way, birth, death, splitting, and merging can be predicted along with target state all in one step instead of deciding on birth, death, splitting and merging with a hypothesis test first, then predicting target state.

1.4 Contribution of this Thesis

The novel contribution made in this thesis is to address the concerns in Section 1.3.6 with a stochastic model that allows for splitting and merging of targets. The model we propose will incorporate birth, death, splitting, and merging of targets into the likelihood, as well as missed detections and false alarms. Predictions then can be made purely by considering the distribution of the relevant variables given the data. There is no need for hypothesis testing to initiate and terminate tracks as an intermediate step.

In addition to this, much can be added in the area of theoretical considerations for tracking methods. Convergence results to date deal with speed or convergence of the numerical algorithm used to solve the optimization problem such as in [41] or [16]. It is very necessary to have some theoretical justification for the optimization problem itself.

Chen [14] provides some theoretical justification for a method to choose the correct number of targets as the number of scans or images goes to infinity, but much more needs to be done. For example, under what conditions does the tracking estimate converge to the true values of the variables being estimated as the time increment between observations goes to zero? At what rate does this convergence take place? We also present a way to quantify the uncertainty in the tracking estimate which includes the uncertainty in birth, death, splitting and merging as well as the uncertainty in data association.

In the Chapter 2 we will describe the stochastic model that we use to explain the behavior of the targets to be tracked. Chapter 3 gives a more mathematical description of the model along with the likelihood. Estimation of the parameters in the model will be discussed in Chapter 4. Chapter 5 describes how to use the model to estimate the tracks of the targets and how to quantify the uncertainty involved in that estimate. Some theoretical justification for this tracking estimate in terms of its asymptotic properties is provided in Chapter 6. The methodology will then be illustrated on simulated data in Chapter 7. Lastly, we will look at the satellite reflectivity rainfall data and track the storms that are present in Chapter 8.

Chapter 2

The Stochastic Model

In this chapter we define a continuous time stochastic model that we assume defines the motion and state of the targets to be tracked. By using a continuous time model, irregularly spaced observations in time and fill in asymptotics are much easier to deal with versus a discrete time model. This model will then be used to estimate the tracks of the targets by using the model likelihood given in Chapter 3.

Define a path, $(X(t), Y(t))$, to be the coordinates of a target at time t, $t > 0$. We observe the targets at discrete times $\underline{t} = (t_0, t_1, \ldots, t_n)$. We are assuming a 2-dimensional path, but the following could allow for paths in $\Re^3$. We wish to model a path of a target (for example a storm) by a 2 dimensional Gaussian Process (GP).

The complication is that we may not observe the target if we look in on the process at a time t, because of several issues:

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)

4. it combined with another target (merger)

5. it may not be found by our detection procedure (missing).

In addition, to these five issues there is another problem:

6. there may be false alarms (clutter) that are found by the detection algorithm.

We would like to model these false alarms separately from the actual targets.

In the following we describe an overall model for the targets that has four components: State Model, Missing State Model, Location Model, and Attribute Models. The State Model describes how and when targets come into existence and terminate. The Missing State Model describes when an existing target is observable or missing. The Location Model describes how an existing target moves and the Attribute Models describe other characteristics of the targets, such as size, orientation, or intensity.

We will also allow for false alarms to appear in this model. There are essentially two types of false alarms. Those that exist for an extended period of time are called persistent false alarms or persistent clutter. Those that can be treated as a one time occurrence that lasts an infinitesimal length of time will just be called false alarms or clutter. We are only concerned with the latter here since persistent clutter will be modeled as an actual target as suggested by Blackman [6]. The model for false alarms will be described separately in Section 2.5.

2.1 Target State Model

To deal with problems 1-5 above we propose a continuous time Markov chain model that determines how and when a birth, death, split, or merger occurs. This *State Model* is very similar to a birth and death process. There are four types of events that can occur in this process: 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 Markov Chain assumption implies that the times between successive events (birth, death, split, or merger) are independent exponentially distributed random variables. However, if this assumption is not reasonable, another positive continuous distribution can be used. The Markov property would then be lost however, which would make the likelihood calculation in Section 3.1.1 more difficult.

The following notation will be used to describe the state model

$$\begin{aligned} U_{b,j} &= \text{number of births in the interval } [t_j, t_{j+1}) \\ U_{d,j} &= \text{number of deaths in the interval } [t_j, t_{j+1}) \\ U_{s,j} &= \text{number of splits in the interval } [t_j, t_{j+1}) \\ U_{m,j} &= \text{number of mergers in the interval } [t_j, t_{j+1}). \end{aligned} \tag{2.1}$$

We will also let $\boldsymbol{U}_b = (U_{b,1}, \ldots, U_{b,n})$ and similarly for $\boldsymbol{U}_d$, $\boldsymbol{U}_s$, and $\boldsymbol{U}_m$. Then denote the collection of N_0 and the $\boldsymbol{U}$'s by $\mathcal{U} = (N_0, \boldsymbol{U}_b, \boldsymbol{U}_d, \boldsymbol{U}_s, \boldsymbol{U}_m)$. We will also need to specify which targets were involved in the events. Let,

$$
\begin{aligned}
V_{b,j} &= \text{the collection of indices of targets that were born in the interval } [t_j, t_{j+1}) \\
V_{d,j} &= \text{the collection of indices of targets that died in the interval } [t_j, t_{j+1}) \\
V_{s,j} &= \text{the collection of triplets } (i_1, i_2, i_3) \text{ where } i_1 \text{ is the index of the parent and} \\
&\quad\ i_2, i_3 \text{ are the children for every split interval } [t_j, t_{j+1}) \\
V_{m,j} &= \text{the collection of triplets } (i_1, i_2, i_3) \text{ where } i_1, i_2 \text{ are the indices of the parents and} \\
&\quad\ i_3 \text{ is the child for every merger in the interval } [t_j, t_{j+1}).
\end{aligned} \tag{2.2}
$$

It is assumed that when a target is born it takes on the next available index. When there is a death, each target that is alive up until the time of death is equally likely to be the one that dies. When there is a split, all of the targets alive up until that time are equally likely to be the parent and the children take on the next two available indexes. Finally, for a merger all of the N choose 2 pairs of targets for the N targets alive up until that time are equally likely to be the parents and the child takes on the next available index.

Let $\boldsymbol{V}_b = (V_{b,1}, \ldots, V_{b,n})$ and similarly for $\boldsymbol{V}_d$, $\boldsymbol{V}_s$, and $\boldsymbol{V}_m$. The collection of all the $\boldsymbol{V}$'s will be denoted $\mathcal{V} = (\boldsymbol{V}_b, \boldsymbol{V}_d, \boldsymbol{V}_s, \boldsymbol{V}_m)$. We will say that the variables $\mathcal{U}$, and $\mathcal{V}$, describe the target state model. This is a hidden Markov model in that we do not actually get to observe these states from the data. Predicting the state variables, which targets merged with which etc., is part of the tracking problem. This will be described further in Chapter 5.

We will say that a target is initiated if it is a new target as a result of a birth, split, or merger. Likewise, a target is terminated if it no longer exists because it died, split or merged. When a target exists, it may or may not be observable. We will describe how to model this behavior next.

2.2 Missing State Model

The State Model defines in a probabilistic way how birth, death, splitting, and merging occur. We now define how an existing target may become unobservable or missing. It is certainly not ideal to simply allow this target to die and start a new path when it appears again. Missing observations are usually handled in the tracking literature by a white noise. That is, if a target exists, it has probability P_d of being detected and producing an observation at a given time t and this is *iid* over time.

If missing observations have a tendency to be correlated over time, an alternative to the white noise model would be to use a continuous time Markov Chain that simply changes between two states, missing and observable. While in the observable state, the target goes missing with rate ν_1 and when missing, the target becomes observable with rate ν_0. In either case, we will let $W_i(t)$ be the missing state variable that represents the observability of the i^{th} target at time t. Suppose that $W_i(t) = 1$ if the i^{th} target is observable at time t and $W_i(t) = 0$ if it is missing.

It is worth mentioning that if we have attribute information such as size or intensity, it may be beneficial to allow P_d to be proportional to the mean size or intensity of the target. This would allow targets with intensity near the threshold value for intensity to go missing more often for example. This could also be incorporated into the rates ν_0 and ν_1 if we use the Markov model.

2.3 Target Location Model

As mentioned previously, when a target is determined to exist by the State Model, the path of the i^{th} target, $(X_i(t), Y_i(t))$ will behave like a Gaussian Process. Target paths are assumed to run their course independently of other targets unless they are required to split or merge as determined by the State Model. The dependency introduced by splitting and merging will be worked into the following.

The distribution of $X_i(t)$ will be defined below for the three cases of a target resulting from (1) birth, (2) merger, and (3) split. The distribution of $Y_i(t)$ will be similar with the obvious changes in notation and parameters and independent of $X_i(t)$.

2.3.1 Birth

Suppose that the i^{th} target resulted from a birth. Let the x component of location and velocity of the i^{th} target at time t be denoted by $X_i(t)$ and $X_i'(t)$ respectively. 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 $\xi_i = t_1$. Then

$$X_i(t) = X_i(\xi_i) + X_i'(\xi_i)(t - \xi_i) + \sigma_i G_i(t - \xi_i) \tag{2.3}$$

where $G_i(t)$ is some continuous mean 0 Gaussian Process. It is also assumed that the initial position and velocity are Gaussian. Specifically, $X_i(\xi_i) \sim N\left(\mu_{X_0}, \sigma^2_{X_0}\right)$ and $X_i'(\xi_i) \sim N\left(\mu_{X_0'}, \sigma^2_{X_0'}\right)$.

As mentioned in the introduction, integrated Brownian Motion (IBM) is a popular model to use for the random motion component $G(t)$. This is given by,

$$G_i(t) = \int_0^t B_i(s)ds \tag{2.4}$$

where $B_i(t)$ is the Brownian Motion driving the i^{th} path.

The model in (2.3) is set up for $G_i(t)$ to be an IBM. If we change the model for $G_i(t)$, we may wish to change (2.3) accordingly. For example, if we use twice integrated Brownian Motion, we may want to add an initial acceleration term instead of assuming that it is 0.

To implement an Interacting Multiple Model here we would have several random motion models $G_{i,1} \ldots G_{i,K}$. To maintain the continuous time nature of our model, we would need a continuous time Markov Chain that would switch back and forth between motion models. With probability $P_{i,j}$ it would switch from model i to model j with the time between switches distributed as independent exponential random variables.

2.3.2 Merger

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. Then we have,

$$X_i(t) = \frac{1}{2}\left(X_{p_{i,1}}(\xi_i) + X_{p_{i,2}}(\xi_i)\right) + \psi_{m,i} + \left[\frac{1}{2}\left(X'_{p_{i,1}}(\xi_i) + X'_{p_{i,2}}(\xi_i)\right) + \psi'_{m,i}\right](t-\xi_i) + \sigma_i G_i(t-\xi_i) \tag{2.5}$$

where $\psi_{m,i} \sim N(0, \sigma^2_{X_m})$ and $\psi'_{m,i} \sim N(0, \sigma^2_{X'_m})$.

Notice that the initial position of the new target (child) is the average of the positions of the merged targets (parents) at the time of merger plus a little bit of noise $\psi_{m,i}$. Figure 2.1 displays a physical representation of this. Also, the initial velocity of the new target is the average of the velocities of the parents at the time of merger plus a little bit of noise $\psi'_{m,i}$. Presumably, $\sigma^2_{X_m}$ and $\sigma^2_{X'_m}$ are small so that the new target location and velocity are likely to be close to the averages of the parents.

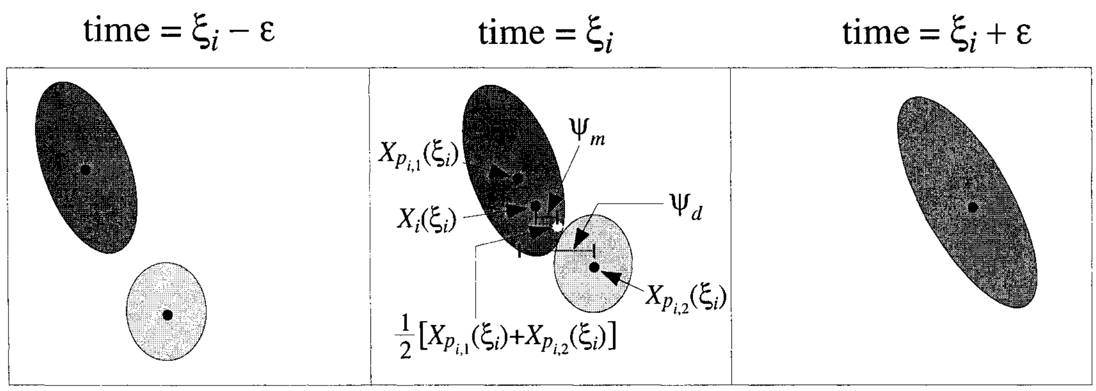

Figure 2.1: Physical Description of a Merger

The two parent targets are currently not required to be close to each other at the time of merger. To ensure that the parents get 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. A more precise explanation of this is 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

$$D = X_{d_1}(\xi_{d_3}) - X_{d_2}(\xi_{d_3}) + \psi_d \tag{2.6}$$

where $\psi_d \sim N(0, \sigma^2_{X_d})$. We will then condition the model for X on D and evaluate this model at $D = 0$. 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 2.1 once again, we see a merging event with a possible realization of ψ_d.

For the storms problem, Figure 2.1 is a realistic physical description of what is actually happening. That is, two targets really do become one target. Other applications of merging targets may not share this same physical representation. For the case of unresolved radar observations, there is obviously still two separate targets that just produce one observation because they are very close together. However, we may wish to track the centroid of these two targets, which is the best that we can do until the targets show up as separate observations again. In this framework then, we actually do have a new merged target that can be thought of as a phantom target in between the actual targets, which is waiting to split into the actual targets again.

2.3.3 Split

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. Then the location is given by

$$X_i(t) = X_{p_{i,1}}(\xi_i) + \psi_{s,i} + \left[X'_{p_{i,1}}(\xi_i) + \psi'_{s,i}\right](t - \xi_i) + \sigma_i G_i(t - \xi_i) \tag{2.7}$$

where $\psi_i \sim N(0, \sigma^2_{X_s})$ and $\psi'_i \sim N(0, \sigma^2_{X'_s})$. Similar to the model for a merger, the initial position and velocity of a new target from a split is the same as that of the parent plus some error. It is assumed that $\sigma^2_{X_s}$ is small so that the new targets are likely to appear close to where the parent split. Similarly, $\sigma^2_{X'_s}$ should be small so that the new targets have a velocity similar to that of their parent.

2.3.4 Measurement Error

In many cases, we do not get to observe the exact locations of the targets because of random error in our measuring device. Rather, at a time t, we get to observe $X_i(t) + \varepsilon$ where ε is an independent measurement error term. If these measurement errors are independent over time as well, then for $j = 1, \ldots, n$ we would observe

$$X_i^*(t_j) = X_i(t_j) + \varepsilon_{i,j} \tag{2.8}$$

where it is assumed that

$$\varepsilon_{i,j} \overset{iid}{\sim} \mathcal{N}(0, \sigma_{X,e}^2). \tag{2.9}$$

2.4 Target Attribute Models

In this section we will describe several models that can be used in conjunction with the location and state models if some auxiliary information about target attributes is available. We present a few special cases of attributes that are commonly available (Size, Orientation, Intensity). Other attributes may be handled in a similar manner.

2.4.1 Size

The size of the target, which is a measure of area for the 2-dimensional case, can be taken into account by the following. We will use the length of the minor and major axes, $R_1(t)$ and $R_2(t)$, of the best fitting ellipse to each target as shown in Figure 2.2.

We are assuming here that a best fitting ellipse for each target is determined by standard imaging techniques such as in [51]. If the data is such that we cannot fit an ellipse, but

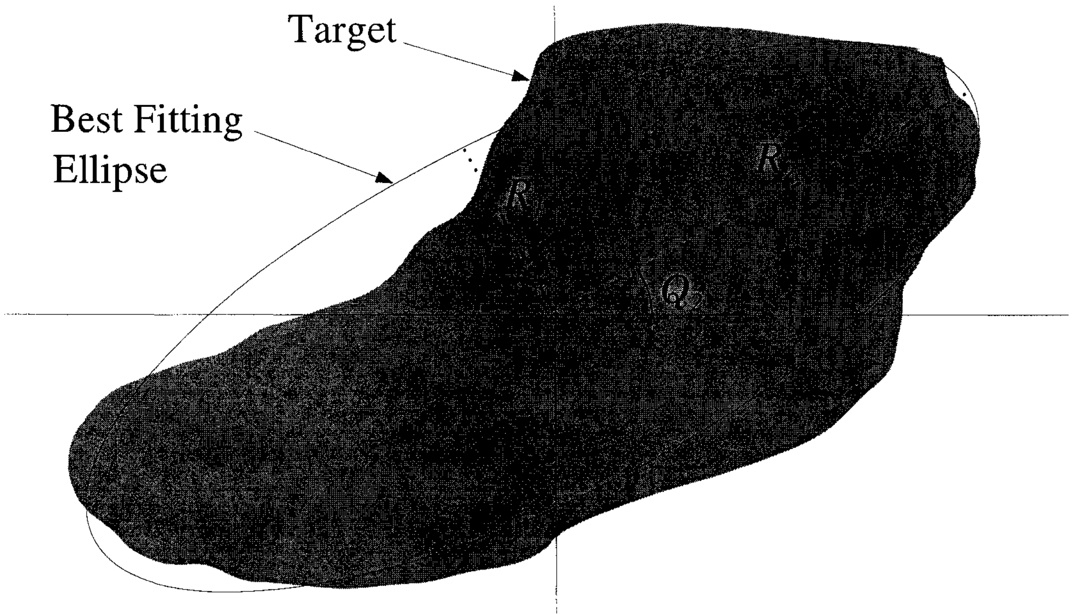

Figure 2.2: R_1, R_2, and Q_2

we do have the overall size, then we can use this same model for size, $S(t) = R_1(t)R_2(t)$, which ends up being less complicated than the discussion that is to follow. Note that we have excluded the factor of π in the target size definition for convenience of notation.

Let $R_{1,i}(t)$ and $R_{2,i}(t)$ be the radii of the minor and major axes respectively for the i^{th} target at time t. $R_{1,i}(t)$ and $R_{2,i}(t)$ will be treated as random variables from a log-normal distribution with parameters $(\mu_{R_{1,i}}, \sigma^2_{R_{1,i}})$ and $(\mu_{R_{2,i}}, \sigma^2_{R_{2,i}})$ respectively. These observations are assumed to be independent over time. In other words $R_1(t)$ and $R_2(t)$ are assumed to be log-normal white noise with their respective parameters which remain constant over time. This assumption may not be met in many applications, but should be approximately true unless the size of targets changes quite drastically. It is also certainly possible to make the mean size a parametric function of time so that it can grow in the beginning and get smaller over time like a storm system would tend to do. Or perhaps for radar applications, the mean size can remain constant, but the variance could be a function of distance from the radar to allow for more measurement variability the further a target is from the measurement source. In any case, the assumption of independence of size observations over time is necessary in order to adequately deal with splitting and merging of targets; see the discussion at the end of this section.

Since $R_{1,i}(t)$ and $R_{2,i}(t)$ are log-normal, $S_i(t)$ is also log-normal with parameters $(\mu_{R_{1,i}} + \mu_{R_{2,i}}, \sigma^2_{R_{1,i}} + \sigma^2_{R_{2,i}})$. Hence the mean value of $S_i(t)$ is

$$E[S_i(t)] = \exp\left\{\mu_{R_{1,i}} + \mu_{R_{2,i}} + \frac{1}{2}\left(\sigma^2_{R_{1,i}} + \sigma^2_{R_{2,i}}\right)\right\} \tag{2.10}$$

In the event of a split, the mean sizes of the two new targets will sum to the mean size of the parent target. For example, if $E[S_i]$ is the mean size of the i^{th} target which splits into targets j and k, then we have $E[S_i] = E[S_j] + E[S_k]$. The new parameters are otherwise free and will need to be specified or estimated. In the event of a merger, the same conservation of mean size rule will apply.

In this formulation, it is possible that at some times by random chance, $R_{1,i}(t)$ will be greater than $R_{2,i}(t)$. From the data we receive from the best fitting ellipse, we won't know when this is happening. So in reality, we only observe the order statistics $R_{(1)}(t) = R_1(t) \wedge R_2(t)$ and $R_{(2)}(t) = R_1(t) \vee R_2(t)$. Thus it is the likelihood of the collection of $R_{(1)}$ and $R_{(2)}$ variables that needs to be calculated in Chapter 3.

It seems as though it would be beneficial to allow the radii or size to change as positive, continuous processes. An exponential Brownian motion or a Bessel Process would be viable options. A problem comes in with merging and splitting though. For example, the size of a child after merger should be equal to the sum of the sizes of the parents. The sums of exponential Brownian motions are hard to work with. Also, in the event of a split, we need to somehow randomly distribute the size of the parent into the two children. This makes it very difficult, if not impossible, to construct process that models this behavior with a tractable likelihood. We may be able to use an approximate likelihood however. This problem is potentially an area of further research.

2.4.2 Orientation

The orientation of a target can be measured by the angle, Q_2 of the major axis corresponding to R_2 as shown in Figure 2.2. We will use the usual notation $Q_{2,i}(t)$ to represent the angle for the i^{th} target at time t. We will assume that $Q_{2,i}(t)$ comes from a VonMises distribution on $[0, \pi]$ with parameters α_i and β_i. Like R_1 and R_2, $Q_2(t)$ is assumed to be *iid* over time.

Also, since we only get to observe $R_{(2)}(t)$ we won't know which axis is the major axis, so we can't observe Q_2. We do however get to observe the angle of $R_{(2)}(t)$ which we will call $Q_{(2)}(t)$. The likelihood for the collection of $Q_{(2)}$ variables will be given in Chapter 3.

2.4.3 Intensity

The intensity $I(t)$ of a target can be defined for the storms application to be the maximum rainfall rate of the pixels that make up the storm, the average of rainfall rates for all of the pixels, or some combination of the two. For the 2-D turbulence problem, we can use vorticity in place of rainfall rate. For other applications the intensity of a target may be otherwise defined in a suitable way.

Once we have decided on a definition for intensity of a target, it can be modeled in a manner similar to the other attributes. We can assume some appropriate distribution and let the observations be *iid* over time. For example, we could assume a log-normal distribution,

$$I(t) \sim logN(\mu_{I_i}, \sigma_{I_i}^2) \tag{2.11}$$

For a merging and splitting, it may be reasonable that the mean intensity is conserved in a manner similar to size. For example, if $E[I_i]$ is the mean intensity of the i^{th} target which splits into targets j and k, then we have $E[I_i] = E[I_j] + E[I_k]$. It could also be possible depending on the definition that the maximum intensity is conserved. Another option is that there is no conservation of intensity and all the parameters are free. In any case, once it is determined how to introduce the dependency of the parameters, if any, the likelihood is computed the same way in Chapter 3.

2.5 False Alarm Model

The model used for false alarms will be broken up into separate parts (state, location, attributes) in a manner similar to that for targets. However, since false alarms only exist for an infinitesimal length of time, they can only be observed once. Hence, these models are much less complicated than those for actual targets.

2.5.1 False Alarm State

We will assume that the distribution of false alarms is a Spatial Poison Process which is independent over time. Hence, the number of false alarms at time t, $N_f(t)$, is distributed as a Poisson random variable,

$$N_f(t) \sim \text{Poisson}(\lambda_f A) \tag{2.12}$$

which is *iid* over time. Here A is the area of the field of vision or image window, so that λ_f is the expected number of false alarms in a unit area.

If we are using adaptive thresholding, then we will have a higher concentration of false alarms in the gating regions where we are using a smaller threshold. Hence, we would need to allow for a higher rate parameter, λ_f^*, for the gating regions. Suppose G denotes the set in $\Re^2$ which is the collection of all gating regions. We would then have a heterogeneous spatial Poisson process with parameter

$$\lambda_{x,y} = \lambda_f I_{G^c}(x,y) + \lambda_f^* I_G(x,y) \tag{2.13}$$

It may also be beneficial to use the following form of adaptive thresholding. Let the baseline amount of thresholding over the field of vision be τ_0. We can subtract an amount $\tau_{x,y}$ to this baseline threshold which is proportional to the conditional density of a new observation given the previous observations in a track. In this way, we threshold at a much lower level near where we expect the new observations to be. For this type of thresholding, we could assume a heterogeneous spatial Poisson process with a parameter $\lambda_f(x,y)$ that is inversely proportional to $\tau_0 - \tau_{x,y}$ so that

$$\lambda_f(x,y) = \frac{c}{(\tau_0 - \tau_{x,y})} \tag{2.14}$$

would be a continuous function of x, y-space. If we denote the field of vision as the set $F \in \Re^2$ then for either of the two models for $\lambda_f(x, y)$ we have

$$N_f(t) \sim \text{Poisson}\left(\int_F \lambda_f(x, y)\right). \tag{2.15}$$

There is no need for a missing state model for false alarms. The conceptual reason for this is that there is nothing there to be missing in the first place. The existence of a false alarm implies that it is detectable. If for some reason we were to assume a probability of detecting a false alarm P_d, on top of a probability of its existence, then the number of observable false alarms $\tilde{N}_f(t)$ would still have a Poisson distribution with rate parameter $\tilde{\lambda}_f = P_d \lambda_f$. Hence, we would just have to absorb the probability of detecting a present false alarm into our rate parameter anyway. Hence, the false alarm state model is assumed to describe how many false observations will be detected at a given time.

2.5.2 False Alarm Location

For the homogeneous spatial Poisson process, the location of a given false alarm observation X will be assumed to have a uniform distribution that is independent of other false alarms in space and time,

$$X \sim \text{Unif}(L_X, U_X). \tag{2.16}$$

where L_X and U_X are the range limits in the field of vision or image window. He distribution for Y would be defined similarly.

For the heterogeneous Poisson models, the distribution will be more concentrated where $\lambda_f(x, y)$ is higher. This means we have to consider the distribution of X and Y together.

The density of X, Y for a heterogeneous Passion process is actually proportional to the rate $\lambda_f(x, y)$.

2.5.3 False Alarm Attributes

Whatever attributes that are modeled for targets, must also be modeled for false alarms. This is so that a likelihood computed assuming a group of sequential observations originated from a target would be comparable to the likelihood assuming the observations were false alarms. We propose that a false alarm attribute have independent and identical distribution for each false alarm occurrence.

For example, if we were to use the size or R_1 and R_2 in our model, each target would have its own size parameters, but all false alarms would have the same parameters. We would let the minor radius for all false alarms have the distribution,

$$R_1 \sim \mathrm{logN}(\mu_{R_{1,f}}, \sigma^2_{R_{1,f}}) \tag{2.17}$$

and similarly for R_2. We will model the angle of orientation of the major axis for all false alarms as

$$Q_2 \sim \mathrm{VonMises}(\alpha_f, \beta_f) \tag{2.18}$$

and lastly for intensity we will use

$$I \sim \mathrm{logN}(\mu_{I_f}, \sigma^2_{I_f}) \tag{2.19}$$

Chapter 3

Model Likelihood

In this Chapter we will present the likelihood of the model described in Chapter 2. We will use the notation $[X]$ to denote the probability density function of the random variable X, $[X](x)$ to denote $[X]$ evaluated at x and $[X \mid Y]$ to denote the conditional density of X given Y. We wish to write out the density, or likelihood, for the following collection of random variables that correspond to targets,

$$(\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}) \tag{3.1}$$

The bold $\mathcal{W}$, $\mathcal{X}$,and $\mathcal{Y}$ denote the collection of those variables for all targets at all times. These variables will be more formally defined in the following sections. For ease of presentation, we are restricting the focus to location information right now. We will incorporate the attribute contribution to the likelihood later.

In addition, we wish to write out a density for the following collection of random variables that correspond to false alarms.

$$(\boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f) \tag{3.2}$$

where $\boldsymbol{N}_f = (N_f(t_1), \ldots, N_f(t_n))$ and the bold $\mathcal{X}_f$ and $\mathcal{Y}_f$ denote the collection of the locations for all false alarms at all times. The overall model likelihood function is then given by

$$[(\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}), (\boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f)] = [\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}][\boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f] \tag{3.3}$$

since the false alarms are assumed to be completely independent of the targets unless we are using adaptive thresholding in which case the false alarm density will depend on the target density. In the following sections we will write out the target density, $[\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}]$, and and the false alarm density, $[\boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f]$.

3.1 Target Density

Using the properties of conditional probability, we can write the target density as

$$[\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}] = [\mathcal{U}, \mathcal{V}] \cdot [\mathcal{W} \mid \mathcal{U}, \mathcal{V}] \cdot [\mathcal{X} \mid \mathcal{U}, \mathcal{V}, \mathcal{W}] \cdot [\mathcal{Y} \mid \mathcal{U}, \mathcal{V}, \mathcal{W}] \tag{3.4}$$

since $\mathcal{X}$ and $\mathcal{Y}$, are independent given $(\mathcal{U}, \mathcal{V}, \mathcal{W})$. We will call the conditional densities in (3.4), in order from left to right, the target state density, missing state density, and target location densities respectively. We will describe each of these in the following sections.

3.1.1 Target State Density

Since the State Model has independent increments, the Target State density can be written as

$$[\mathcal{U}, \mathcal{V}] = [N_0] \prod_{j=1}^{n} [U_{b,j}, U_{d,j}, U_{s,j}, U_{m,j} \mid N(t_j)][V_{b,j}, V_{d,j}, V_{s,j}, V_{m,j} \mid N(t_j), U_{b,j}, U_{d,j}, U_{s,j}, U_{m,j}] \tag{3.5}$$

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

$$[N_0](k) = \frac{\lambda_0^k e^{-\lambda_0}}{k!}. \tag{3.6}$$

To write out the exact density for $(U_{b,j}, U_{d,j}, U_{s,j}, U_{m,j} \mid N(t_j))$, is difficult since 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 $(U_{b,i}, U_{d,i}, U_{s,i}, U_{m,i})$ 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 occurance of events stays constant during the interval $[t_j, t_{j+1})$. If we let $\bar{N}_j = (N(t_j) + N(t_{j+1}))/2$, which is the average number of targets alive during the interval, then we can assume that the rate of each of the events during the interval is

$$\bar{\lambda}_{b,j} = \lambda_b \tag{3.7}$$

$$\bar{\lambda}_{d,j} = \lambda_d \bar{N}_j \tag{3.8}$$

$$\bar{\lambda}_{s,j} = \lambda_s \bar{N}_j \tag{3.9}$$

$$\bar{\lambda}_{m,j} = \lambda_m \bar{N}_j \tag{3.10}$$

for birth, death, splitting, and merging respectively. With this assumption, the variables $(U_{b,i}, U_{d,i}, U_{s,i}, U_{m,i})$ 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,

$$[U_{b,j} \mid N(t_j)]\,(u) \approx (\lambda_b \Delta t_j)^u e^{-\lambda_b \Delta t_j}/u! \tag{3.11}$$

$$[U_{d,j} \mid N(t_j)]\,(u) \approx (\bar{\lambda}_{d,j} \Delta t_j)^u e^{-\bar{\lambda}_{d,j} \Delta t_j}/u! \tag{3.12}$$

$$[U_{s,j} \mid N(t_j)]\,(u) \approx (\bar{\lambda}_{s,j} \Delta t_j)^u e^{-\bar{\lambda}_{s,j} \Delta t_j}/u! \tag{3.13}$$

$$[U_{m,j} \mid N(t_j)]\,(u) \approx (\bar{\lambda}_{m,j} \Delta t_j)^u e^{-\bar{\lambda}_{m,j} \Delta t_j}/u! \tag{3.14}$$

For the variables $(V_{b,i}, V_{d,i}, V_{s,i}, V_{m,i})$ if we make the same assumption that $N(t) = \bar{N}_j$ is constant during the interval $[t_j, t_{j+1})$, then we have

$$\begin{aligned}
[V_{b,j} \mid N(t_j), U_{b,j}]\,(v) &\approx 1 && (3.15)\\
[V_{d,j} \mid N(t_j), U_{d,j}]\,(v) &\approx (1/\bar{N}_j)^{U_{d,j}} && (3.16)\\
[V_{s,j} \mid N(t_j), U_{s,j}]\,(v) &\approx (1/\bar{N}_j)^{U_{s,j}} && (3.17)\\
[V_{m,j} \mid N(t_j), U_{m,j}]\,(v) &\approx \left(1/\binom{\bar{N}_j}{2}\right)^{U_{m,j}} && (3.18)
\end{aligned}$$

and we can write (3.5) as

$$[\mathcal{U},\mathcal{V}] \approx [N_0]\prod_{j=1}^{n}[U_{b,j}][V_{b,j} \mid U_{b,j}]\cdot[U_{d,j}][V_{d,j} \mid U_{d,j}]\cdot[U_{s,j}][V_{s,j} \mid U_{s,j}]\cdot[U_{m,j}][V_{m,j} \mid U_{m,j}]. \tag{3.19}$$

3.1.2 Missing State Density

Recall that $W_i(t)$ represents the observability (0 or 1) of the i^{th} target at time t. Let $\boldsymbol{W}_i = (W_i(t_1), \ldots, W_i(t_n))$ and $\mathcal{W} = (\boldsymbol{W}_1, \ldots, \boldsymbol{W}_m)$ where m is the number of targets that existed before time t_n. The time of initiation of the i^{th} target is denoted by ξ_i. Also let the time of termination of the i^{th} target be given by ζ_i. For convenience if the i^{th} target is still alive at time t_n, we will let $\zeta_i = t_n$.

The state variables $\mathcal{U}$ and $\mathcal{V}$ do not specify the exact values of ξ_i and ζ_i. They do however specify which interval, between observations they are in. This is all we need for $\mathcal{W}$ since its dependence on $\mathcal{U}$ and $\mathcal{V}$ is only on whether or not a target exists at the observed time points. For the following, if it is known that ξ_i is in the interval (t_j, t_{j+1}), we will set

$\xi_i = t_j + \Delta t_j/2$.

The white noise model for $\mathcal{W}$ of Section 2.2 assumes probability P_d of observing the i^{th} target if it exists at a given time, independent of other times. If the target does not exist at time t then $W_i(t) = 0$. Under this model, the conditional density of $\mathcal{W}$ given the state variables in (3.4) can be written out using indicator functions to separate the cases when the i^{th} target exists and does not. This density is then given by

$$[\mathcal{W} \mid \mathcal{U}, \mathcal{V}](w) = \prod_{i=1}^{m}\prod_{j=1}^{n}\{I_{[t_1,\xi_i)\cup(\zeta_i,t_n]}(t_j)(1-w_{ij}) + I_{[\xi_i,\zeta_i]}(t_j)\left((1-w_{ij})(1-P_d)+w_{ij}P_d\right)\} \tag{3.20}$$

where w_{ij} is representing an observed value of $W_i(t_j)$.

If instead we assume the Markov Chain model of Section 2.2, then we will make the following assumptions. For a time t before a target is initiated $W_i(t) = 0$. When target i is initiated, $W_i(\xi_i) = 1$. During the lifespan of a path, it switches back and forth between missing and observable with rates ν_0 and ν_1. As soon as path i is terminated by the State Model, $W_i(t)$ becomes 0 again for $t > \zeta_i$.

Since for $i \neq j$, W_i is independent of W_j given the state variables and W_i is Markov, we have

$$[\mathcal{W} \mid \mathcal{U}, \mathcal{V}](w) = \prod_{i=1}^{m}\prod_{j=1}^{n}[W_i(t_j) \mid \mathcal{U}, \mathcal{V}, W_i(t_{j-1}) = w_{i,j-1}](w_{i,j}) \tag{3.21}$$

Now let $P_{j,k}(\Delta t)$ be the probability of W_i going from state j to state k in a time Δt , assuming that the path exists during this time as determined by the State Model. Simple calculations can show these transition probabilities to be

$$P_{j,k}(\Delta t) = \frac{\nu_{1-k}}{\nu_0+\nu_1} + \frac{\nu_j}{\nu_0+\nu_1}e^{-(\nu_0+\nu_1)\Delta t} \tag{3.22}$$

Let $\Delta t_j = t_j - t_{j-1}$. Also for notational convenience, adopt the convention that $t_0 < t_1$ even though we do not observe the process at t_0. We can then write the conditional density of $W_i(t_j)$ on the right side of (3.21) by using indicators to break it apart into the times when the target exists and does not and also into the case of the first observation of an existing target and a consecutive one.

$$[W_i(t_j) \mid \mathcal{U}, \mathcal{V}, W_i(t_{j-1}) = w_{i,j-1}](w_{ij}) =$$
$$I_{[t_1,\xi_i)\cup(\zeta_i,t_n]}(t_j)(1-w_{ij}) + I_{[t_0,\xi_i)}(t_{j-1})I_{[\xi_i,\zeta_i]}(t_j)P_{1,w_{ij}}(t_j-\xi_i) +$$
$$I_{[\xi_i,\zeta_i]}(t_{j-1})I_{[\xi_i,\zeta_i]}(t_j)P_{w_{i,j-1},w_{ij}}(\Delta t_j) \tag{3.23}$$

Equation (3.23) along with (3.21) describe the the conditional density of W given the state variables in (3.4) for the Markov Chain model.

3.1.3 Target Location Density

Since $X_i(t)$ is normally distributed for all t, the observed location of all targets at all time points, has a multivariate normal distribution. Let the times at which the i^{th} target is observable be denoted by $\boldsymbol{t}_i = (t_{i,1}, \ldots, t_{i,n_i})$. Also let $\boldsymbol{X}_i = (X_i(t_{i,1}), \ldots, X_i(t_{i,n_i}))$ and lastly let $\mathcal{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m)$ be the collection of all observed locations of all targets. Then,

$$\mathcal{X} \sim N(\mu_X, \Sigma_X) \tag{3.24}$$

where we will define μ_X, and Σ_X below.

Recall from Section 2.3 that this mean and covariance will depend on the time of initiation, ξ, of the targets. We will adopt the convention of the previous section here and set $\xi_i = t_j + \Delta t_j/2$ if ξ_i is known to be in the interval (t_j, t_{j+1}). 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 intervals are sufficiently small though, this approximation will be quite close to the true density.

Also recall from section 2.3 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 from Section 2.3.2 for the i^{th} merging event, $i = 1, \ldots, N_d$. Then let

$$D_i = X_{d_{i,1}}(\xi_{d_{i,3}}) - X_{d_{i,2}}(\xi_{d_{i,3}}) + \psi_i \tag{3.25}$$

be the difference, D, for the i^{th} merging event. The random variable D_i is also normally distributed. Finally, let $\boldsymbol{D} = (D_1, \ldots, D_{N_d})$. Then we also can write,

$$\boldsymbol{D} \sim N(\mu_D, \Sigma_D)$$

For the collection of both $\mathcal{X}$ and $\boldsymbol{D}$ we have

$$\begin{pmatrix} \mathcal{X} \\ \boldsymbol{D} \end{pmatrix} \sim N(\mu, \Sigma) \tag{3.26}$$

where

$$\mu = \begin{pmatrix} \mu_X \\ \mu_D \end{pmatrix} \tag{3.27}$$

and

$$\Sigma = \left(\begin{array}{c|c} \Sigma_X & \Sigma_{X,D} \\ \hline \Sigma'_{X,D} & \Sigma_D \end{array} \right) \tag{3.28}$$

The mean vectors and covariance matrices will be described in the following. Let $\mu_i(t) = EX_i(t)$ and $\mu_{D_i} = ED_i$. These functions are given for the IBM model in Appendix A. Then let $\boldsymbol{\mu}_i = (\mu_i(t_{i,1}), \ldots, \mu_i(t_{i,n_i}))$ and we can now write the mean vectors in (3.27) as $\mu_X = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_m)$ and $\mu_D = (\mu_{D_1}, \ldots, \mu_{D_{N_d}})$.

Define the matrices $\Sigma_{i,j}$ to be the covariances between all of the observations of target path i with all of the observations of path j. Specifically the $(k,l)^{th}$ element of this matrix can be written

$$\Sigma_{i,j}(k,l) = \mathrm{Cov}(X_i(t_{i,k}), X_j(t_{j,l})) \quad k = 1, \ldots, n_i; \quad l = 1, \ldots, n_j \tag{3.29}$$

Also define the matrices $\Sigma_{i,D}$ and Σ_D, by their $(k,l)^{th}$ element as follows

$$\Sigma_{i,D}(k,l) = \mathrm{Cov}(X_i(t_{i,k}), D_l) \quad k = 1, \ldots, n_i; \quad l = 1, \ldots, N_d \tag{3.30}$$

$$\Sigma_D(k,l) = \mathrm{Cov}(D_k, D_l) \quad k = 1, \ldots, N_d; \quad l = 1, \ldots, N_d \tag{3.31}$$

The covariance functions in (3.29), (3.30), and (3.31) for the IBM model are given in Appendix A.

Now we can write the covariance matrix for $\mathcal{X}$ as

$$\Sigma_X = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} & \cdots & \Sigma_{1,m} \\ \Sigma_{2,1} & \Sigma_{2,2} & \cdots & \Sigma_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{m,1} & \Sigma_{m,2} & \cdots & \Sigma_{m,m} \end{pmatrix} \tag{3.32}$$

and that for $(\mathcal{X}, \boldsymbol{D})$ as

$$\Sigma_{X,D} = \begin{pmatrix} \Sigma_{1,D} \\ \vdots \\ \Sigma_{m,D} \end{pmatrix} \tag{3.33}$$

This completes the description of the distribution of $(\mathcal{X}; \boldsymbol{D})$ given in (3.26).

Now we are ready to write out the conditional distribution of $\mathcal{X}$ given $\boldsymbol{D} = 0$, which we will just call the distribution of $\mathcal{X}$ from this point forward. Multivariate normal theory tells us that

$$\mathcal{X} \mid D = 0 \sim N(\mu^*, \Sigma^*) \tag{3.34}$$

where

$$\mu^* = \mu_X - \Sigma_{X,D}\Sigma_D^{-1}\mu_D \tag{3.35}$$

and

$$\Sigma^* = \Sigma_X - \Sigma_{X,D}\Sigma_D^{-1}\Sigma_{X,D}' \tag{3.36}$$

The density of X is then just the multivariate normal density with parameters, μ^* and Σ^*. This will require taking the inverse of Σ^*, but this can be done somewhat efficiently since Σ^* is a relatively sparse matrix. Unless path i is a relative of path j, in the sense that one is a by-product of splits or mergers of the other, they will have 0 covariance.

Because of the conditioning on $\boldsymbol{D}$, this model cannot be conveniently put into state space form. Hence, the Kalman filter cannot be used to update the conditional distribution of a new observation given the previous observations. However, it is possible that the Innovations Algorithm could be used to make the likelihood calculation of this model more efficient.

3.2 False Alarm Density

In a manner very similar to target density we can write the density of of the false alarm variables from (3.2) as,

$$[\boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f] = [\boldsymbol{N}_f] \cdot [\mathcal{X}_f \mid \boldsymbol{N}_f] \cdot [\mathcal{Y} \mid \boldsymbol{N}_f] \tag{3.37}$$

We will call these conditional densities, false alarm state density, and false alarm location densities respectively.

3.2.1 False Alarm State Density

The density for false alarm state or $\boldsymbol{N}_f = (N_f(t_1), \ldots, N_f(t_n))$ is straight forward to write out since it is assumed in Section 2.5.1 to be *iid* Poisson distributed random variables with rate $\lambda_f A$. Hence the density of $\boldsymbol{N}_f$ is,

$$[\boldsymbol{N}_f](\boldsymbol{k}) = \prod_{j=1}^{n} \frac{(\lambda_f A)_j^k e^{-\lambda_f A}}{k_j!} \tag{3.38}$$

For the case of adaptive thresholding, we simply replace the rate parameter $\lambda_f A$ at each time with the integral over the field of vision, F, of the rate function $\lambda_f(x, y, t)$. Note that the rate function must be a function of time as well since the thresholding values will change for each time step. For adaptive thresholding, the density for $\boldsymbol{N}_f$ is conditional on the target model variables since the gating regions depend on the state variables and location. We are however suppressing this dependency in the notation. This density is then,

$$[\boldsymbol{N}_f](\boldsymbol{k}) = \prod_{j=1}^{n} \frac{1}{k_j!} \left(\int_F \lambda_f(x,y) \right)^{k_j} e^{-\int_F \lambda_f(x,y)} \tag{3.39}$$

where we have suppressed the dependency on the target model for notational convenience.

3.2.2 False Alarm Location Density

The location density for false alarms is also simple compared to that for targets since the false alarms locations are assumed to be independent over time and space. Let the x component of the i^{th} false alarm at time t be denoted as $X_{f,i}(t)$ for $i = 1, \ldots, N_f(t)$. Also let

$$\mathcal{X}_f = \{X_{f,i}(t_j) : i = 1, \ldots, N_f(t_j),\ \ j = 1, \ldots, n\}$$

be the collection of x locations of all false alarms at all times. We will also use the same notation for Y_f.

For the uniform model of (2.16) the density function for a particular $X_{f,i}(t)$ would be

$$f(x) = \frac{1}{U_X - L_X} \tag{3.40}$$

and the density for $\mathcal{X}_f$ would then be

$$[\mathcal{X}_f \mid \boldsymbol{N}_f](x) = \prod_{j=1}^{n} \prod_{i=1}^{N_f(t_j)} f(x_{ij}) \tag{3.41}$$

where x_{ij} is a dummy variable for the value of $X_{f,i}(t_j)$ The density for $\mathcal{Y}_f$ would be the same except for the obvious difference in parameters and independent of $\mathcal{X}_f$.

For the case of adaptive thresholding the density for $\mathcal{X}_f, \mathcal{Y}_f$ is proportional to the rate function $\lambda_f(x, y)$. Hence the density for $(X_{f,i}(t), Y_{f,i}(t))$ should be

$$f(x, y, t) = c\lambda_f(x, y, t) \tag{3.42}$$

where

$$c = \frac{1}{\int_F \lambda_f(x, y, t)} \tag{3.43}$$

This density obviously depends on the target variables as well but we have again suppressed this dependence to make notation less cumbersome. The density for $(\mathcal{X}_f, \mathcal{Y}_f)$ would then be

$$[\mathcal{X}_f, \mathcal{Y}_f \mid \boldsymbol{N}_f]](x, y) = \prod_{j=1}^{n} \prod_{i=1}^{N_f(t_j)} f(x_{ij}, y_{ij}, t_j) \tag{3.44}$$

3.3 Attributes

The attribute variables are assumed *iid* over time given the Missing State variable $\mathcal{W}$. Hence the densities are fairly straight forward to write out. With attributes, we now have the following collection of random variables for targets,

$$(\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}) \tag{3.45}$$

where $\mathcal{A}$ denotes the collection of all attribute variables. In the following we will assume $\mathcal{A} = (\mathcal{R}_{(1)}, \mathcal{R}_{(2)}, \mathcal{Q}_{(2)}, \mathcal{I})$, which are the smallest radius, largest radius, angle of orientation and intensity for targets respectively. These variables will be more formally defined in the following sections.

We also have the following collection of random variables that correspond to false alarms,

$$(\boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f) \tag{3.46}$$

where $\mathcal{A}_f = (\mathcal{R}_{(1),f}, \mathcal{R}_{(2),f}, \mathcal{Q}_{(2),f}, \mathcal{I}_f)$ which are the same variables as above for false alarms.

The target likelihood function is then given by

$$[\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}] = [\mathcal{U}, \mathcal{V}] \cdot [\mathcal{W} \mid \mathcal{U}, \mathcal{V}] \cdot [\mathcal{X} \mid \mathcal{U}, \mathcal{V}, \mathcal{W}] \cdot [\mathcal{Y} \mid \mathcal{U}, \mathcal{V}, \mathcal{W}] \cdot [\mathcal{A} \mid \mathcal{W}] \tag{3.47}$$

So we can just multiply $[\mathcal{A} \mid \mathcal{W}]$ to the target density without attributes given in (3.4). Technically $\mathcal{A}$ should really be conditioned on $\mathcal{U}$, $\mathcal{V}$, $\mathcal{X}$, and $\mathcal{Y}$ as well, but the way we modeled attributes in the previous chapter, the density of $\mathcal{A}$ would still depend only on W so we dropped the other variables in the notation. Similarly, the false alarm likelihood is given by

$$[\boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f] = [\boldsymbol{N}_f] \cdot [\mathcal{X}_f \mid \boldsymbol{N}_f] \cdot [\mathcal{Y} \mid \boldsymbol{N}_f] \cdot [\mathcal{A} \mid \boldsymbol{N}_f] \tag{3.48}$$

so we can just multiply $[\mathcal{A} \mid \boldsymbol{N}_f]$ to the false alarm density with out attributes in (3.37). And of course the overall density is

$$[(\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}), (\boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f)] = [\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}] \cdot [\boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f]. \quad (3.49)$$

We can of course incorporate any of these attribute variables separately or add other attributes in a similar manner. For this collection above though, we have

$$[\mathcal{A} \mid \mathcal{W}] = [\mathcal{R}_{(1)}, \mathcal{R}_{(2)} \mid \mathcal{W}][\mathcal{Q}_{(2)} \mid \mathcal{W}][\mathcal{I} \mid \mathcal{W}] \quad (3.50)$$

and

$$[\mathcal{A} \mid \boldsymbol{N}_f] = [\mathcal{R}_{(1),f}, \mathcal{R}_{(2),f} \mid \boldsymbol{N}_f][\mathcal{Q}_{(2),f} \mid \boldsymbol{N}_f][\mathcal{I}_f \mid \boldsymbol{N}_f] \quad (3.51)$$

We will describe these densities in the following sections.

3.3.1 Radius Density

For target size, we will use the following notation. Let $R_{1,i}(t)$ and $R_{2,i}(t)$ be the length of minor and major axes of the best fitting ellipse to target i at time t. We observe the min and max of these which are $R_{(1),i}(t)$ and $R_{(2),i}(t)$ respectively. Also let

$$\mathcal{R}_{(1)} = \{(R_{(1)}(t_j) : 1 \leq i \leq m, \ \ 1 \leq j \leq n\}$$

and similarly for $\mathcal{R}_{(2)}$.

Recall that $R_{1,i}(t)$ and $R_{2,i}(t)$ are assumed to be distributed as independent log-normals for all t. The density for $(R_{(1),i}(t), R_{(2),i}(t))$ does not depend on time so we will write it as

$[R_{(1),i}, R_{(2),i}]$. This density is similar to that for order statistics and is given by

$$[R_{(1),i}, R_{(2),i}](r, s) = \{[R_{1,i}](r)[R_{2,i}](s) + [R_{1,i}](s)[R_{2,i}](r)\}\, I_{\{r \leq s\}} \tag{3.52}$$

where $[R_{1,i}]$ and $[R_{2,i}]$ are log-normal densities with parameters $(\mu_{R_{1,i}}, \sigma^2_{R_{1,i}})$ and $(\mu_{R_{2,i}}, \sigma^2_{R_{2,i}})$ respectively as described in Section 2.4.1.

Since the radii of path i at time t_j, are independent of the radii at other times or of other targets, the density for $(\mathcal{R}_{(1)}, \mathcal{R}_{(2)})$ is

$$[\mathcal{R}_{(1)}, \mathcal{R}_{(2)} \mid W](r, s) = \prod_{i=1}^{m} \prod_{\{j: W_{i,j}=1\}} [R_{(1),i}, R_{(2),i}](r_{i,j}, s_{i,j}) \tag{3.53}$$

where $r_{i,j}$ and $s_{i,j}$ are the arguments for the values of $R_{(1),i}(t_j)$ and $R_{(2),i}(t_j)$ respectively and as usual m is the number of targets observed before t_n.

For false alarms, we will use similar notation. Let $(R_{1,f,i}(t)$ and $R_{2,f,i}(t))$ be the length of minor and major axes of the best fitting ellipse to the i^{th} false alarm at time t. We observe the min and max of these which are $R_{(1),f,i}(t)$ and $R_{(2),f,i}(t)$ respectively. Also let

$$\mathcal{R}_{(1),f} = \{(R_{(1),i}(t_j) : 1 \leq j \leq n, \;\; 1 \leq i \leq N_f(t_j), \}$$

and similarly for $\mathcal{R}_{(2),f}$.

The density for false alarms is very similar to that above, but all false alarms at all times are assumed to have the same distribution so

$$[R_{(1),f,i}(t), R_{(2),f,i}(t)] = [R_{(1),f,i'}(t), R_{(2),f,i'}(t)] = [R_{(1),f}, R_{(2),f}]$$

where

$$[R_{(1),f}, R_{(2),f}](r,s) = \{[R_{1,f}](r)[R_{2,f}](s) + [R_{1,f}](s)[R_{2,f}](r)\} I_{\{r \le s\}} \tag{3.54}$$

and $[R_{1,i}]$ and $[R_{2,i}]$ are log-normal densities with parameters $(\mu_{R_{1,f}}, \sigma^2_{R_{1,f}})$ and $(\mu_{R_{2,f}}, \sigma^2_{R_{2,f}})$. So the density of $(\mathcal{R}_{(1),f}, \mathcal{R}_{(2),f})$ is

$$[\mathcal{R}_{(1),f}, \mathcal{R}_{(2),f} \mid \boldsymbol{N}_f](r,s) = \prod_{j=1}^{n} \prod_{i=1}^{N_f(t_j)} [R_{(1),f}, R_{(2),f}](r_{i,j}, s_{i,j}) \tag{3.55}$$

where $r_{i,j}$ and $s_{i,j}$ are the arguments for the values of $R_{(1),f,i}(t_j)$ and $R_{(2),f,i}(t_j)$ respectively.

3.3.2 Angle of Orientation Density

For target orientation or angle, we will use the following notation. Let $Q_{2,i}(t))$ be the angle of orientation of the axis corresponding to R_2 of the best fitting ellipse to target i at time t. We observe $Q_{(2),i}(t))$ which is the angle that corresponds to $R_{(2),i}(t)$. Also let

$$\mathcal{Q}_{(2)} = \{Q_{(2)}(t_j) : 1 \le i \le m, \ \ 1 \le j \le n\}.$$

Consider for now a given target's orientation at a fixed time $Q_{(2),i}(t)$. We will drop the subscript i and argument t for now and write this as $Q_{(2)}$ to make notation less cumbersome. When $R_{(2)} = R_2$, $Q_{(2)} = Q_2$. However, when $R_{(2)} = R_1$, $Q_{(2)} = \lfloor Q_2 + \pi/2 \rfloor_p i$ where $\lfloor x \rfloor_y$ is x mod y.

Hence, the distribution of $Q_{(2)}$ given $(R_{(1)}, R_{(2)})$ is a mixture distribution that takes the value of Q_2 with probability γ and $\lfloor Q_2+\pi/2\rfloor_\pi$ with probability $1-\gamma$, where

$$\begin{aligned}\gamma &= P(R_1 < R_2 \mid R_{(1)}, R_{(2)}) \\ &= \frac{[R_1](R_{(1)})[R_2](R_{(2)})}{[R_1](R_{(1)})[R_2](R_{(2)}) + [R_1](R_{(2)})[R_2](R_{(1)})}.\end{aligned} \tag{3.56}$$

Thus the conditional density of $Q_{(2),i}$ is

$$[Q_{(2),i} \mid R_{(1)}, R_{(2)}](q) = \gamma[Q_{2,i}](q) + (1-\gamma)[Q_{2,i}](\lfloor q+\pi/2\rfloor_\pi) \tag{3.57}$$

where $[Q_{2,i}]$ is the VonMises density on $[0,\pi)$ given by

$$[Q_{2,i}](q) = \frac{e^{\beta_i cos(q-\alpha_i)}}{\pi\Psi_0(\beta_i)} I_{[0,\pi)}(q). \tag{3.58}$$

Here $\Psi_0(x)$ is a modified Bessel function of the first kind of order 0. As with the radii, $Q_{(2),i}(t)$ is independent over time and of other targets so the conditional density of $\mathcal{Q}_{(2)}$ is

$$[\mathcal{Q}_{(2)} \mid \mathcal{W}, \mathcal{R}_{(1)}, \mathcal{R}_{(2)}](q) = \prod_{i=1}^{m} \prod_{\{j:W_{i,j}=1\}} [Q_{(2)} \mid R_{(1),i}(t_j), R_{(2),i}(t_j)](q_{i,j}) \tag{3.59}$$

where $q_{i,j}$ are the arguments for the values of $Q_{(2),i}(t_j)$.

Again the situation for false alarms is very similar. We will let $Q_{(2),f,i}(t)$ be the angle of orientation corresponding to $R_{(2),f,i}(t)$ and

$$\mathcal{Q}_{(2),f} = \{Q_{(2),i}(t_j) : 1 \le j \le n,\ \ 1 \le i \le N_f(t_j)\}.$$

Let $[Q_{2,f}]$ be the same density as in (3.57) only with parameters α_f and β_f in place of α_i

and β_i. False alarms are *iid* so

$$[\mathcal{Q}_{(2),f}(t) \mid \mathcal{W}, \mathcal{R}_{(1)}, \mathcal{R}_{(2)}](q) \prod_{j=1}^{n} \prod_{i=1}^{N_f(t_j)} [Q_{(2),f} \mid R_{(1),f,i}(t_j), R_{(2),f,i}(t_j)](q_{i,j}). \tag{3.60}$$

3.3.3 Intensity Density

Intensity does not have any of the dependency like that built into size and orientation, so the density is quite a bit simpler. For target intensity, we will use familiar notation. Let $I_i(t)$ be the intensity of target i at time t. Also let

$$\mathcal{I} = \{I_i(t_j) : 1 \leq i \leq m,\ \ 1 \leq j \leq n\}.$$

For any target the density of $I_i(t)$ does not depend on time so we will write it as $[I_i]$. Recall from Section 2.4.2 that $[I_i]$ is assumed to be a log-normal density with parameters $(\mu_{I_i}, \sigma^2_{I_i})$. The density of $\mathcal{I}$ is then

$$[\mathcal{I} \mid \mathcal{W}](\iota) \prod_{i=1}^{m} \prod_{\{j:W_{i,j}=1\}} [I_i](\iota_{i,j}) \tag{3.61}$$

where as usual $\iota_{i,j}$ are the arguments for the values of $I_i(t_j)$.

For false alarm intensity, we again assume the same density $[I_f]$ for all false alarms which is log-normal with parameters $(\mu_{I_f}, \sigma^2_{I_f})$. The density of $\mathcal{I}_f$ is then

$$[\mathcal{I} \mid \mathcal{W}](\iota) \prod_{j=1}^{n} \prod_{i=1}^{N_f(t_j)} [I_f](\iota_{i,j}). \tag{3.62}$$

Chapter 4

Parameter Estimation

In this chapter we describe how we chose to estimate the parameters of the model given in Chapter 2. In most cases these estimates are the maximum likelihood estimators (MLE's). In some cases however, the MLE is not in closed form and would be too computationally expensive to compute. In these cases we will make use of other reasonable choices for estimates

4.1 Parameters of the State Model

The State Model parameters are λ_0, λ_b, λ_d, λ_s, and λ_m. There is also the false alarm state parameter, λ_f. For the State Model parameters can calculate approximate MLE's based on the approximate likelihood given in Section 3.1.1. The MLE for λ_0 is obviously

$$\hat{\lambda}_0 = N_0 \tag{4.1}$$

where recall N_0 is the initial number of targets.

Now consider the estimation of the death rate, λ_d. From the approximation in (3.12) we can consider the $U_{d,j}$ for $j = 1, \ldots, n$ as independent Poisson observations with parameter $\bar{N}_j \lambda_{d,j} \Delta t_j$. Recall that $\bar{N}_j$ is the average number of targets alive in the interval $[t_j, t_{j+1})$. Denote the collection of $\bar{N}_j$'s by $\bar{\boldsymbol{N}}$. Then the contribution to the likelihood in (3.19) from $\boldsymbol{U}_d$ is

$$\left[\boldsymbol{U}_d \mid \bar{\boldsymbol{N}}\right](\boldsymbol{u}) = \prod_{j=1}^{n} (\bar{N}_j \lambda_{d,j} \Delta t_j)^{u_j} e^{-\bar{N}_j \lambda_{d,j} \Delta t_j} / u_j! \tag{4.2}$$

So the derivative of the log likelihood is

$$\frac{d}{d\lambda_d} \log \left[\boldsymbol{U}_d \mid \bar{\boldsymbol{N}}\right](\boldsymbol{u}) = \sum_{j=1}^{n} \frac{u_j}{\lambda_d} - \bar{N}_j \Delta t_j. \tag{4.3}$$

Setting (4.3) equal to zero gives

$$\hat{\lambda}_d = \frac{\sum_{j=1}^{n} U_{d,j}}{\sum_{j=1}^{n} \bar{N}_j \Delta t_j} \tag{4.4}$$

In similar fashion the approximate MLE's of λ_b, λ_s, and λ_m can be shown to be

$$\hat{\lambda}_b = \frac{\sum_{j=1}^{n} U_{b,j}}{\sum_{j=1}^{n} \Delta t_j} \tag{4.5}$$

$$\hat{\lambda}_d = \frac{\sum_{j=1}^{n} U_{s,j}}{\sum_{j=1}^{n} \bar{N}_j \Delta t_j} \tag{4.6}$$

$$\hat{\lambda}_d = \frac{\sum_{j=1}^{n} U_{m,j}}{\sum_{j=1}^{n} (\bar{N}_j - 1) \Delta t_j} \tag{4.7}$$

Lastly, consider estimation of the false alarm rate, λ_f. The number of false alarms at each time, $N_f(t_j)$ is Poisson with parameter $\lambda_f A$, where A is the area of the field of vision.

Hence the MLE for λ_f is

$$\hat{\lambda}_f = \frac{\sum_{j=1}^{n} N_f(t_j)}{nA} \tag{4.8}$$

4.2 Parameters of the Missing State Model

If we assume the simple *iid* model for missing observations, then the MLE for the missing state model parameter, P_d, is the ratio of the number of times targets were observable to the number of times they existed.

$$\hat{\lambda}_f = \frac{\sum_{i=m}^{n} \sum_{j=1}^{n} W_i(t_j) I_{[\xi_i,\zeta_i]}(t_j)}{\sum_{i=m}^{n} \sum_{j=1}^{n} I_{[\xi_i,\zeta_i]}(t_j)} \tag{4.9}$$

To estimate the parameters ν_0 and ν_1 of the Markov Chain Model, we will settle for approximate MLE's. We will assume that transitions from missing ($W = 0$) to observable ($W = 1$) or observable to missing, happen at the observed time points, t_j's. Let $\tau_{i,k}$ be the time of the k^{th} transition from $0 \to 1$ or $1 \to 0$ for for the i^{th} target, W_i. We are assuming then that we know the exact value of the $\tau_{i,k}$'s. Let $n_{t,i}$ be the number of transitions made by W_i during the time interval we observe the process, $[0, t_n]$. For notational convenience let $\tau_{i,0} = \xi_i$. Remember it is also assumed that $W_i(\xi_i) = 1$. The approximate likelihood is then

$$\begin{aligned} L(\nu_0, \nu_1) \quad = \quad & \prod_{i=1}^{m} \Bigg\{ \prod_{\substack{k=1 \\ k \text{ odd}}}^{n_{t,i}} \nu_1 e^{-\nu_1(\tau_{i,k}-\tau_{i,k-1})} \prod_{\substack{k=1 \\ k \text{ even}}}^{n_{t,i}} \nu_0 e^{-\nu_0(\tau_{i,k}-\tau_{i,k-1})} \\ & \left(e^{-\nu_0(\zeta_i-\tau_{i,n_{t,i}})} I_{\{n_{t,i} \text{ odd}\}} + e^{-\nu_1(\zeta_i-\tau_{i,n_{t,i}})} I_{\{n_{t,i} \text{ even}\}} \right) \Bigg\} \end{aligned} \tag{4.10}$$

The derivative of the log likelihood with respect to ν_1 is then

$$\frac{\partial}{\partial \nu_1} \log L(\nu_0, \nu_1) = \sum_{i=1}^{m} \frac{\lceil n_{t,i}/2 \rceil}{\nu_1} - \sum_{\substack{k=1 \\ k \text{ odd}}}^{n_{t,i}} (\tau_{i,k} - \tau_{i,k-1}) - (\zeta_i - \tau_{i,n_{t,i}}) I_{\{n_{t,i} \text{ even}\}}. \tag{4.11}$$

Setting the above equation equal to zero gives the approximate MLE for ν_1 to be

$$\hat{\nu}_1 = \frac{\sum_{i=1}^{m} \lceil n_{t,i}/2 \rceil}{\sum_{i=1}^{m} \sum_{\substack{k=1 \\ k \text{ odd}}}^{n_{t,i}} (\tau_{i,k} - \tau_{i,k-1}) + (\zeta_i - \tau_{i,n_{t,i}}) I_{\{n_{t,i} \text{ even}\}}}. \tag{4.12}$$

In a similar manner, the approximate MLE for ν_0 is

$$\hat{\nu}_0 = \frac{\sum_{i=1}^{m} \lfloor n_{t,i}/2 \rfloor}{\sum_{i=1}^{m} \sum_{\substack{k=1 \\ k \text{ even}}}^{n_{t,i}} (\tau_{i,k} - \tau_{i,k-1}) + (\zeta_i - \tau_{i,n_{t,i}}) I_{\{n_{t,i} \text{ odd}\}}}. \tag{4.13}$$

4.3 Location Parameters

The derivative of the location density is difficult to work with analytically because of the matrix algebra involved. Exact MLE's would require a time consuming iterative method. We therefore decided to use alternatives to the MLE's for the location parameter estimates. We will present these estimates for the x-coordinate parameters. The estimates for the y-coordinate parameters will be the same with the obvious changes.

4.3.1 White Noise Variance

We will first consider estimation of the white noise error variance $\sigma^2_{X,e}$. For the IBM model, the observed location for a path is

$$X_i^*(t_j) = X_i(\xi_i) + tX_i'(\xi_i) + \sigma_i^2 \int_0^{t_j - \xi_i} B_i(s)ds + \varepsilon_{i,j} \tag{4.14}$$

So if we make a derivative approximation, we have

$$\frac{D_{i,j}}{\Delta t_j} = \frac{X_i^*(t_{j+1}) - X_i^*(t_j)}{\Delta t_j} \approx X_i'(\xi_i) + \sigma_i^2 B_i(t_j - \xi_i) + \frac{1}{\Delta t_j}(\varepsilon_{i,j+1} - \varepsilon_{i,j}). \tag{4.15}$$

If we then take the consecutive differences of the $D_j/\Delta t_j$ we have

$$\begin{aligned} D_{i,j}^2 &= \frac{D_{i,j+1}}{\Delta t_{j+1}} - \frac{D_{i,j}}{\Delta t_j} \\ &\approx \sigma_i^2(B_i(t_{j+1} - \xi_i) - B_i(t_j - \xi_i)) \\ &\quad + \frac{1}{\Delta t_j \Delta t_{j+1}} \left(\Delta t_j \varepsilon_{i,j+2} - (\Delta t_{j+1} + \Delta t_j)\varepsilon_{i,j+1} + \Delta t_{j+1}\varepsilon_{i,j} \right). \end{aligned} \tag{4.16}$$

The covariance of consecutive D_j^2's is given by

$$\operatorname{Cov}(D_j^2, D_{j+1}^2) \approx K_j \sigma_{X,e}^2 \tag{4.17}$$

where

$$K_j = -\frac{\Delta t_j(\Delta t_{j+1} + \Delta t_{j+2}) + \Delta t_{j+2}(\Delta t_j + \Delta t_{j+1})}{\Delta t_j \Delta t_{j+1}^2 \Delta t_{j+2}}. \tag{4.18}$$

Hence a method of moments estimate for the measurement error variance is

$$\hat{\sigma}_{X,e}^2 = \frac{1}{N} \sum_{i=1}^{m} \sum_{j \in O_i} \frac{D_{i,j}^2 D_{i,j+1}^2}{K_j} \tag{4.19}$$

where O_i is the set of indices, j, that we have four consecutive times $t_j, t_{j+1}, t_{j+2}, t_{j+3}$ where the i^{th} target is observable,

$$O_i = \{j : W_i(t_j) = W_i(t_{j+1}) = W_i(t_{j+2}) = W_i(t_{j+3}) = 1\}$$

and $N = \sum_i n(O_i)$, is the total number of terms in the sum in (4.19).

4.3.2 IBM Variance Scalar

For the estimate of the variance scalar, σ_i^2 for the i^{th} target, we will make use of the estimate for $\sigma_{X,e}^2$ and use a local linear regression to estimate $X_i(t_j)$'s given the observations $X_i^*(t_j) = X_i(t_j) + \varepsilon_{i,j}$. Once we have an estimate for the $X_i(t_j)$'s, we can form an estimate σ_i^2.

The criterion for selection of the bandwidth, h will based on the following rule presented on pages 100-101 of [54]. Dropping the subscript i, we have n observations $X^*(t_j)$ and we wish to estimate $X(t_j)$. Denote this estimate as $\hat{m}(t_j, h)$. Then as described in [54], the mean square error is

$$E\left[\sum_{j=1}^{n}(X^*(t_j) - \hat{m}(t_j,h))^2\right] = E\left[\sum_{j=1}^{n}(X(t_j) - \hat{m}(t_j,h))^2\right] + \sigma_{X_e}^2(n - 2\mathrm{tr}(S)) \quad (4.20)$$

so

$$\frac{1}{n}\sum_{j=1}^{n}(X^*(t_j) - \hat{m}(t_j,h))^2 \approx \frac{1}{n}\sum_{j=1}^{n}(X(t_j) - \hat{m}(t_j,h))^2 + \frac{\hat{\sigma}_{X_e}^2}{n}(n - 2\mathrm{tr}(S)). \quad (4.21)$$

Since it is our goal to minimize the bias which is the first term on the right side of (4.20), we will use the bandwidth, h, that minimizes the quantity

$$R(h) = \frac{1}{n}\sum_{j=1}^{n}(X^*(t_j) - \hat{m}(t_j,h))^2 - \frac{\hat{\sigma}_{X_e}^2}{n}(n - 2\mathrm{tr}(S)). \quad (4.22)$$

We will only use this approach to estimate $X_i(t_j)$ if there are more than K observations for the i^{th} path. We usually set $K = 6$ in practice.

Now we turn to the problem of estimating σ_i^2. We can do this in the following way. From the above discussion, we now have an estimate, $\hat{X}_i(t_j)$, for $X_i(t_j)$ and

$$\hat{X}_i(t_j) \approx X_i(0) + tX_i'(0) + \sigma_i^2 G_i(t_j) \tag{4.23}$$

where $G_i(t)$ is an IBM, $G_i(t) = \int_0^t B_i(s)ds$. The consecutive difference quotient is

$$\frac{\hat{D}_{i,j}}{\Delta t_j} = \frac{\hat{X}_i(t_{j+1}) - \hat{X}_i(t_j)}{\Delta t_j} \approx X_i'(0) + \frac{\sigma_i^2}{\Delta t_j}(G_i(t_{j+1}) - G_i(t_j)) \tag{4.24}$$

And taking consecutive differences of the $\hat{D}_{i,j}/\Delta t_j$'s gives

$$\hat{D}_{i,j}^2 = \frac{\hat{D}_{i,j+1}}{\Delta t_{j+1}} - \frac{\hat{D}_{i,j}}{\Delta t_j} \approx \frac{\sigma_i^2}{\Delta t_j \Delta t_{j+1}} \left(\Delta t_j G_i(t_{j+2}) - (\Delta t_j + \Delta t_{j+1}) G_i(t_{j+1}) + \Delta t_{j+1} G_i(t_j)\right). \tag{4.25}$$

The variance of $\hat{D}_{i,j}^2$ is then

$$\mathrm{Var}(\hat{D}_{i,j}^2) \approx C_j \sigma_i^2 \tag{4.26}$$

where C_j is given by

$$\begin{aligned} C_j \;=\; & \frac{1}{(\Delta t_j \Delta t_{j+1})^2} \Bigg[\Delta t_j^2 \frac{t_{j+2}^3}{3} + (\Delta t_j + \Delta t_{j+1})^2 \frac{t_{j+1}^3}{3} + \Delta t_{j+1}^2 \frac{t_j^3}{3} \\ & -\Delta t_j (\Delta t_j + \Delta t_{j+1}) \left(t_{j+1}^2 t_{j+2} - \frac{t_{j+1}^3}{3} \right) - \Delta t_j \Delta t_{j+1} \left(t_j^2 t_{j+2} - \frac{t_j^3}{3} \right) \\ & -\Delta t_{j+1} (\Delta t_j + \Delta t_{j+1}) \left(t_j^2 t_{j+1} - \frac{t_j^3}{3} \right) \Bigg]. \end{aligned} \tag{4.27}$$

Hence a method of moments estimate for σ_i^2 is

$$\hat{\sigma}_i^2 = \frac{1}{N} \sum_{j \in O_i} \frac{(D_{i,j}^2)^2}{C_j} \tag{4.28}$$

where here O_i is the set of indices, j, that we have three consecutive times t_j, t_{j+1}, t_{j+2} where the i^{th} target is observable,

$$O_i = \{j : W_i(t_j) = W_i(t_{j+1}) = W_i(t_{j+2}) = 1\}$$

and $N = \sum_i n(O_i)$, is the total number of terms in the sum in (4.28).

Again, we only estimate σ_i^2 in this way if we have greater than K observations for the i^{th} path. If the i^{th} path has less than K observations, then we let σ_i^2 equal the weighted average of the σ_i^2 estimates of the other paths.

4.3.3 Initial Conditions

To estimate the initial conditions parameters, μ_{X_0}, σ_{X_0}, $\mu_{X_0'}$, and $\sigma_{X_0'}$, we will also take advantage of the local regression fits $\hat{X}_i(t)$. We can use the local regression to estimate $X_i(\xi_i)$. Let $t_{i,j}$ be the j^{th} time at which the i^{th} path is observed for $j = 1, \ldots, n_i$. We can then estimate $X_i'(\xi_i)$ as

$$\hat{X}_i'(\xi_i) = \frac{\hat{X}_i(t_{i,1}) - \hat{X}_i(\xi_i)}{t_{i,1} - \xi_i}. \tag{4.29}$$

If the i^{th} path has fewer than K observations, then we can simply let $\hat{X}_i(\xi_i) = X_i(t_{i,1})$ and $\hat{X}_i'(\xi_i) = (X_i(t_{i,2}) - X_i(t_{i,1}))/(t_{i,2} - t_{i,1})$.

Let $B = \{i$: target i is a an initial target or a birth$\}$, and let $n(B)$ be the number elements in B. We can construct estimates for the initial conditions parameters as

$$\hat{\mu}_{X_0} = \frac{1}{n(B)} \sum_{i \in B} \hat{X}_i(\xi_i) \tag{4.30}$$

$$\hat{\sigma}^2_{X_0} = \frac{1}{n(B)} \sum_{i \in B} \left(\hat{X}_i(\xi_i) - \hat{\mu}_{X_0}\right)^2 \tag{4.31}$$

$$\hat{\mu}_{X'_0} = \frac{1}{n(B)} \sum_{i \in B} \hat{X}'_i(\xi_i) \tag{4.32}$$

$$\hat{\sigma}^2_{X'_0} = \frac{1}{n(B)} \sum_{i \in B} \left(\hat{X}'_i(\xi_i) - \hat{\mu}_{X'_0}\right)^2 \tag{4.33}$$

4.3.4 Splitting and Merging

Here we will construct estimates for the parameters involved in the initial conditions of a split or merger, σ_{X_s}, $\sigma_{X'_s}$, σ_{X_m}, $\sigma_{X'_m}$, and σ_{X_d}. In order to do this we need estimates for $X_i(\zeta_i)$ and $X'_i(\zeta_i)$. We can also use the local regression to estimate $X_i(\zeta_i)$ and in a similar manner we can estimate $X'_i(\xi_i)$ as

$$\hat{X}'_i(\zeta_i) = \frac{\hat{X}_i(\zeta_i) - \hat{X}_i(t_{i,n})}{\zeta_i - t_{i,n}}. \tag{4.34}$$

Adopt the convention of Section 2.3 and denote the indexes of the parents of target i (if it has any) as $p_{i,1}$ and $p_{i,2}$. Recall that σ_{X_s} is the variance of

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

and $\sigma_{X'_s}$ is the variance of

$$\psi'_{s,i} = X'_{p_{i,1}}(\xi_i) - X'_i(\xi_i)$$

for any path i that is the result of a split. If we let $S = \{i : \text{target } i \text{ is the result of a split}\}$

and $n(S)$ be the number elements in S, then we can construct estimates for these parameters as

$$\hat{\sigma}^2_{X_s} = \frac{1}{n(S)} \sum_{i \in S} \left(\hat{X}_{p_{i,1}}(\zeta_{p_{i,1}}) - \hat{X}_i(\xi_i) \right)^2 \quad (4.35)$$

$$\hat{\sigma}^2_{X'_s} = \frac{1}{n(S)} \sum_{i \in S} \left(\hat{X}'_{p_{i,1}}(\zeta_{p_{i,1}}) - \hat{X}'_i(\xi_i) \right)^2 \quad (4.36)$$

Similarly, σ_{X_m} is the variance of

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

and $\sigma_{X'_m}$ is the variance of

$$\psi'_{m,i} = \frac{1}{2} X'_{p_{i,1}}(\xi_i) + \frac{1}{2} X'_{p_{i,2}}(\xi_i) - X'_i(\xi_i)$$

for any path i that is the result of a merger. So let $M = \{i : \text{target } i \text{ is the result of a merger}\}$ and we can construct estimates of these parameters as

$$\hat{\sigma}^2_{X_m} = \frac{1}{n(M)} \sum_{i \in M} \left(\frac{1}{2} \hat{X}_{p_{i,1}}(\zeta_{p_{i,1}}) + \frac{1}{2} \hat{X}_{p_{i,2}}(\zeta_{p_{i,2}}) - \hat{X}_i(\xi_i) \right)^2 \quad (4.37)$$

$$\hat{\sigma}^2_{X'_m} = \frac{1}{n(M)} \sum_{i \in M} \left(\frac{1}{2} \hat{X}'_{p_{i,1}}(\zeta_{p_{i,1}}) + \frac{1}{2} \hat{X}'_{p_{i,2}}(\zeta_{p_{i,2}}) - \hat{X}'_i(\xi_i) \right)^2 \quad (4.38)$$

Lastly, σ_{X_d} is the variance of

$$\psi_{d,i} = X_{p_{i,1}}(\xi_i) - X_{p_{i,2}}(\xi_i)$$

for any path i that is the result of a merger. So its estimate is given by

$$\hat{\sigma}^2_{X_m} = \frac{1}{n(M)} \sum_{i \in M} \left(\hat{X}_{p_{i,1}}(\zeta_{p_{i,1}}) - \hat{X}_{p_{i,2}}(\zeta_{p_{i,2}}) \right)^2 \tag{4.39}$$

4.4 Size Parameters

Estimation of the size parameters $\mu_{R_{1,i}}$, $\sigma_{R_{1,i}}$, $\mu_{R_{2,i}}$, and $\sigma_{R_{2,i}}$ is complicated by the restriction that mean size must be conserved. Let the size of a target i be defined to be $S_i(t) = R_{1,i}(t)R_{2,i}(t)$ as in Section 2.4.1. So the constraints are that

$$ES_i + ES_{i+1} = ES_{p_{i,1}} \tag{4.40}$$

if targets i and $i+1$ are the result of a split and

$$ES_i = ES_{p_{i,1}} + ES_{p_{i,2}} \tag{4.41}$$

if target i is the result of a merger.

A brief overview of the plan here is to first estimate the mean size for each target, ES_i, under the constraints above. Then estimate the scale parameter, $\sigma^2_{S_i}$, for S_i. We will use these to get an estimate for the shape parameter, μ_{S_i}, of S_i. Lastly, we can then estimate the parameters $\mu_{R_{1,i}}$, $\sigma_{R_{1,i}}$, $\mu_{R_{2,i}}$, and $\sigma_{R_{2,i}}$ by maximum likelihood under the constraints that $\mu_{R_{1,i}} + \mu_{R_{2,i}} = \mu_{S_i}$ and $\sigma^2_{R_{1,i}} + \sigma^2_{R_{2,i}} = \sigma^2_{S_i}$. This procedure will ensure that that the mean size is conserved by these parameter estimates.

Again let $t_{i,j}$ be the j^{th} time at which the i^{th} path is observed for $j = 1, \ldots, n_i$. Notice that for size, we do not have the ambiguity problem that can occur with the radii. For example $S = R_1 R_2 = R_{(1)} R_{(2)}$, so estimating the actual parameters of the size, S_i, is not complicated by only observing the order statistics of the radii. To first estimate the ES_i, we used a weighted least squares approach. The weights are to be inversely proportional to the sample variance of the observations for S_i. Let $\text{var}(S_i)$ denote the sample variance of the $S_i(t_{i,j})$ observations for $j = 1, \ldots, n_i$. Then we wish to find the values of ES_i that minimize

$$\sum_{i=1}^{m} \sum_{j=1}^{n_i} \frac{1}{\text{var}(S_i)} (S_i(t_{i,j}) - ES_i)^2 \tag{4.42}$$

subject to the constraints in (4.40) and (4.41). This is carried out using the Lagrangian Multiplier method. Denote the resulting minimizers of expression (4.42) as $\widehat{ES}_i$.

We will then estimate the scale parameter for S_i, $\sigma^2_{S_i} = \sigma^2_{R_{1,i}} + \sigma^2_{R_{2,i}}$ by the unconstrained MLE. This is just the sample variance of the $\log(S_i(t_{i,j}))$ observations for $j = 1, \ldots, n_i$. Denote this estimate as $\hat{\sigma}^2_{S_i}$. Notice that since S_i is log-normal

$$ES_i = e^{\mu_{S_i} + \frac{1}{2}\sigma^2_{S_i}}$$

where $\mu_{S_i} = \mu_{R_{1,i}} + \mu_{R_{2,i}}$ is the shape parameter of S_i. So once the estimates $\widehat{ES}_i$ and $\hat{\sigma}^2_{S_i}$ are obtained, we can let

$$\hat{\mu}_{S_i} = \log(\widehat{ES}_i) - \frac{1}{2}\hat{\sigma}^2_{S_i}. \tag{4.43}$$

Finally, we can estimate the parameters $\mu_{R_{1,i}}$, $\sigma_{R_{1,i}}$, $\mu_{R_{2,i}}$, and $\sigma_{R_{2,i}}$ by maximum likelihood under the constraints that $\hat{\mu}_{R_{1,i}} + \hat{\mu}_{R_{2,i}} = \hat{\mu}_{S_i}$ and $\hat{\sigma}^2_{R_{1,i}} + \hat{\sigma}^2_{R_{2,i}} = \hat{\sigma}^2_{S_i}$. If we set $\mu_{R_{2,i}} = \mu_{S_i} - \mu_{R_{1,i}}$ and $\sigma^2_{R_{2,i}} = \sigma^2_{S_i} - \sigma^2_{R_{1,i}}$, this is equivalent to the estimation of $\mu_{R_{1,i}}$

and $\sigma_{R_{1,i}}$ with $\mu_{R_{1,i}}$ unconstrained and $\sigma_{R_{1,i}}$ confined to the interval $(0, \hat{\sigma}^2_{S_i})$. Recall from equation (3.52) that this likelihood is a product of sums, and we will therefore need an iterative method to maximize it. Thus this estimation is carried out using a Newton Raphson algorithm. Notice, however that this is only a two dimensional maximization and we can use the unconstrained MLE's assuming $R_1 = R_{(1)}$ for the parameters as starting points. The optimization can therefore be carried out quite quickly. This is the reason we chose to first reduce the problem to a two dimensional estimation for each target instead of applying a Newton Raphson approach to the entire problem to begin with.

4.5 Orientation Parameters

For the estimation of the angle of orientation parameters, α_i and β_i, we again use maximum likelihood. Recall from (3.56) and (3.57) that the likelihood for the $Q_i(t_j)$ depends on the $R_{(1),i}(t_j)$, $R_{(2),i}(t_j)$ and their corresponding parameters $\mu_{R_{1,i}}$, $\sigma_{R_{1,i}}$, $\mu_{R_{2,i}}$, and $\sigma_{R_{2,i}}$. So we can substitute the parameter estimates $\hat{\mu}_{R_{1,i}}$, $\hat{\sigma}_{R_{1,i}}$, $\hat{\mu}_{R_{2,i}}$, and $\hat{\sigma}_{R_{2,i}}$ from Section 4.4 into the density for $\mathcal{Q}$ given in (3.59). We then again use Newton Raphson to find the values of α_i and β_i that maximize the likelihood given in (3.59).

Chapter 5

The Tracking Estimate

The general frame work of the tracking problem is the following. We collect data at the following times $t_1, \ldots, t_n$. At each time there are m_j observations which correspond to either targets or false alarms. Let $Z_i(t_j)$ be the i^{th} observation at time t_j. Each $Z_i(t_j)$ is a vector of the location values and any attribute values for either a target or false alarm.

Let $\boldsymbol{Z}_j$ be the collection of observations at time t_j, $\boldsymbol{Z}_j = (z_1(t_j), \ldots, z_{m_j}(t_j))$, and let

$$\mathcal{Z} = \{Z_i(t_j) : 1 \leq j \leq n, \;\; 1 \leq i \leq m_j \;\}$$

be the collection of observations at all times.

For each observation, $Z_i(t_j)$, we need to decide whether it originated from a target or a false alarm. Also, if from a target, we need to know which target to assign it 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 $p_i(t_j)$ which is the index of the target that observation $Z_i(t_j)$ originated from. We can assume that the index for a false alarm is

0. Let

$$\mathcal{P} = \{(p_i(t_j) : j = 1, \dots, n, \;\; i = 1, \dots, m_j\}.$$

So for a given $\mathcal{Z}$, $\mathcal{P}$ will specify the tracks of each target, but we must also specify the events (births, deaths, splits, and mergers) that took place with $\mathcal{U}$ and $\mathcal{V}$. The variables $\mathcal{U}$ and $\mathcal{V}$ together with $\mathcal{P}$ will denote a solution to the tracking problem. Let our estimate of the tracking solution $(\mathcal{U}, \mathcal{V}, \mathcal{P})$ be denoted

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

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

To achieve our tracking estimate, we will compute the conditional density of $(\mathcal{U}, \mathcal{V}, \mathcal{P})$ given $\mathcal{Z}$,

$$[\mathcal{U}, \mathcal{V}, \mathcal{P} \mid \mathcal{Z} = z](u, v, p). \tag{5.1}$$

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

$$(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}}) = \arg\max_{u,v,p} [\mathcal{U}, \mathcal{V}, \mathcal{P} \mid \mathcal{Z} = z](u, v, p) \tag{5.2}$$

Even more, we know that $[\mathcal{U}, \mathcal{V}, \mathcal{P} \mid \mathcal{Z} = z](\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$ is the probability that $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$ is the correct solution given the data $\mathcal{Z}$.

We will now consider how the to calculate the density in (5.1). Since we have written out the density for

$$(\mathcal{U},\mathcal{V},\mathcal{W},\mathcal{X},\mathcal{Y},\mathcal{A},\boldsymbol{N}_f,\mathcal{X}_f,\mathcal{Y}_f,\mathcal{A}_f)$$

in (3.49), we will consider how we may use this to calculate the density in (5.1)

Notice that there is 1 to 1 mapping

$$g:(\mathcal{P},\mathcal{Z})\rightarrow(\mathcal{W},\mathcal{X},\mathcal{Y},\mathcal{A},\boldsymbol{N}_f,\mathcal{X}_f,\mathcal{Y}_f,\mathcal{A}_f,\mathcal{Z})$$

So for a given $\mathcal{Z}$, the information contained in $\mathcal{P}$ and $(\mathcal{W},\mathcal{X},\mathcal{Y},\mathcal{A},\boldsymbol{N}_f,\mathcal{X}_f,\mathcal{Y}_f,\mathcal{A}_f)$ is the same. Let

$$g^*:(\mathcal{P},\mathcal{Z})\rightarrow(\mathcal{W},\mathcal{X},\mathcal{Y},\mathcal{A},\boldsymbol{N}_f,\mathcal{X}_f,\mathcal{Y}_f,\mathcal{A}_f)$$

be the function g without the last variable in its output. Then we can write

$$\begin{aligned}
[\mathcal{U},\mathcal{V},\mathcal{P}\mid\mathcal{Z}](u,v,p\mid z) &= P\{\mathcal{U}=u,\mathcal{V}=v,\mathcal{P}=p\mid\mathcal{Z}=z\}\\
&= P\{\mathcal{U}=u,\mathcal{V}=v,(\mathcal{W},\mathcal{X},\mathcal{Y},\mathcal{A},\boldsymbol{N}_f,\mathcal{X}_f,\mathcal{Y}_f,\mathcal{A}_f)=g^*(p,z)\mid\mathcal{Z}=z\}\\
&= [\mathcal{U},\mathcal{V},(\mathcal{W},\mathcal{X},\mathcal{Y},\mathcal{A},\boldsymbol{N}_f,\mathcal{X}_f,\mathcal{Y}_f,\mathcal{A}_f)\mid\mathcal{Z}](u,v,g^*(p,z)\mid z) \qquad (5.3)
\end{aligned}$$

It is assumed that the distribution of $\mathcal{Z}$ given $(\mathcal{U},\mathcal{V},\mathcal{W},\mathcal{X},\mathcal{Y},\mathcal{A},\boldsymbol{N}_f,\mathcal{X}_f,\mathcal{Y}_f,\mathcal{A}_f)$ is point uniform on the possible permutations of the values of $(\mathcal{X},\mathcal{Y},\mathcal{A},\mathcal{X}_f,\mathcal{Y}_f,\mathcal{A}_f)$ within each time t_j, so

$$[\mathcal{Z}\mid\mathcal{U},\mathcal{V},\mathcal{W},\mathcal{X},\mathcal{Y},\mathcal{A},\boldsymbol{N}_f,\mathcal{X}_f,\mathcal{Y}_f,\mathcal{A}_f](z\mid u,v,w,x,y,a,n_f,x_f,y_f,a_f)=\frac{1}{\prod_{j=1}^{n}m_j!}I_B(z) \qquad (5.4)$$

where

$$B = \{z : g^*(p, z) = (w, x, y, a, n_f, x_f, y_f, a_f) \text{ for some } p\}$$

So we can write out the likelihood of $(\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}, \boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f, \mathcal{Z})$ by multiplying the likelihood given in (3.49) by that in (5.4). To then obtain the density in (5.3), note that

$$[\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}, \boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f, \mid \mathcal{Z}] \propto [\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}, \boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f, \mathcal{Z}] \tag{5.5}$$

and also realize that for a given z, there are a countable number of arguments, $\alpha_i = (u_i, v_i, w_i, x_i, y_i, a_i, n_{f,i}, x_{f,i}, y_{f,i}, a_{f,i})$, that will make

$$[\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}, \boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f, \mid \mathcal{Z}](\alpha_i \mid z) > 0 \tag{5.6}$$

since the values of $\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f$ must be a permutation of the values in $\mathcal{Z}$ at each time. This means that we must have

$$\begin{aligned}
&[\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}, \boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f, \mid \mathcal{Z}](\alpha_i \mid z) \\
&\qquad = \frac{[\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}, \boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f, \mathcal{Z}](\alpha_i, z)}{\sum_{j=1}^{\infty} [\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}, \boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f, \mathcal{Z}](\alpha_j, z)} \\
&\qquad = \frac{[\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}, \boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f](\alpha_i)}{\sum_{j=1}^{\infty} [\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}, \boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f](\alpha_j)}
\end{aligned}$$

where the second equality comes from the fact that the contribution of $\mathcal{Z}$ to the density is a constant by (5.4). Now by equation (5.3) we have

$$[\mathcal{U}, \mathcal{V}, \mathcal{P} \mid \mathcal{Z}](u, v, p \mid z) = \frac{[\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}, \boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f](u, v, g^*(p, z))}{\sum_{j=1}^{\infty} [\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}, \boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f](u_j, v_j, g^*(p_j, z)))} \tag{5.7}$$

where $\{(u_j, v_j, p_j) : j = 1, 2, \ldots\}$ is an enumeration of the possible tracking solutions. Equation (5.7) gives us the distribution of the possible tracking solutions given the data $\mathcal{Z}$, which 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 is captured by that set is $100(1-\alpha)\%$. We could also calculate the probability that a given observation is a target or a false alarm or the probability that two targets merged, etc.

The difficulty of course is that we cannot calculate all possible tracking solutions in practice. Suppose we make a reasonable restriction to the set of solutions such as every target must be observed at least once. Then the number of solutions is finite, but still too large in practice to calculate a likelihood for all of them. However, notice that the solution (u, v, p) that maximizes (5.7) is the same one that maximizes the numerator. Hence we can use the Multiple Hypothesis Tracking (MHT) Algorithm described in Chapter 1 to search for the solution with the highest model likelihood. The variant of the MHT that we used will be described in more detail next in Section 5.2, but first we will discuss what we can do with the results of the MHT.

The MHT will give us an approximation to the solution that maximizes the numerator of (5.7). Hence this takes care of the point estimate $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$, but we would still like to make the probability statements about the solution and this requires having an estimate of the distribution. This can still be done as well though, since the MHT approximately gives us the M best solutions (those with highest likelihood) as well. Label these M solutions (u_i, v_i, p_i) for $i = 1, \ldots, M$ and let $\mathcal{M}$ denote the set of these M solutions. If we assume that the correct solution is in $\mathcal{M}$, then we can calculate the conditional density of $(\mathcal{U}, \mathcal{V}, \mathcal{P})$ given $\mathcal{Z} = z$ and the event $(\mathcal{U}, \mathcal{V}, \mathcal{P}) \in \mathcal{M}$. This would be given by

$$[\mathcal{U},\mathcal{V},\mathcal{P} \mid \mathcal{Z},(\mathcal{U},\mathcal{V},\mathcal{P}) \in \mathcal{M}](u,v,p \mid z)$$

$$= \frac{[\mathcal{U},\mathcal{V},\mathcal{W},\mathcal{X},\mathcal{Y},\mathcal{A},\boldsymbol{N}_f,\mathcal{X}_f,\mathcal{Y}_f,\mathcal{A}_f](u,v,g^*(p,z))}{\sum_{j=1}^{M}[\mathcal{U},\mathcal{V},\mathcal{W},\mathcal{X},\mathcal{Y},\mathcal{A},\boldsymbol{N}_f,\mathcal{X}_f,\mathcal{Y}_f,\mathcal{A}_f](u_j,v_j,g^*(p_j,z)))}. \tag{5.8}$$

We can then use the distribution given in (5.8) to give the probability that our estimate $(\hat{\mathcal{U}},\hat{\mathcal{V}},\hat{\mathcal{P}})$ is correct or the probability that a given observation is a false alarm. In addition, when we estimate target position and velocity, we could use the distribution of tracking solutions to get mean and variance estimates for these as well. Note that this type of estimation for target position and velocity is very similar in flavor to the JPDA approach described in Section 1.3.3.

5.2 Optimization Algorithm

As stated before, the optimization algorithm is a variant of the MHT algorithm outlined in Section 1.3.2. Note that in our framework, a possible solution, $(\hat{\mathcal{U}},\hat{\mathcal{V}},\hat{\mathcal{P}})$, is a hypothesis in the terminology of Section 1.3.2. The algorithm used here will proceed as follows.

At each time, t_j, when we receive a new set of observations, $\boldsymbol{Z}_j$, we will assume that each observation, $Z_i(t_j)$ is either:

1. an observation from an existing target track,
2. the first observation from a target resulting from birth,
3. the first observation from a target resulting from split,
4. the first observation from a target resulting from merger, or
5. a false alarm.

Tracks that did not receive a new observation to continue the track must either

1. go missing (stay missing), or

2. terminate.

At time t_1 each observation is treated as an initial target or a false alarm. Now assume that we have a set of likely solutions (hypotheses) for the observations through time t_{j-1}. We would then take the new observations, $\boldsymbol{Z}_j$, at time t_j and form updated solutions based on all possible combinations of the above possibilities. We then hold on to several of the solutions with the highest likelihood to use to form solutions at the next time step, t_{j+1}. The actual number of solutions to make it through to the next time will vary. Let $\max\{L_j\}$ be the log-likelihood of the best solution at time t_j. We would then hold on to all solutions that had log-likelihood greater than $\max\{L_j\} - l_d$. For speed concerns, we also set a limit, M_s, for the maximum number of solutions that make it through to the next time. The control parameters l_d and M_S varied depending on the complexity of the problem, but were usually set to $l_d = 10$ and $M_s = 200$ for the problems of Chapters 7 and 8.

Of course it is very inefficient to examine all possible combinations at each time, so we will form gates for each of the tracks. A gate is a prediction region for a new observation at time t_j given the previous observations assumed to be part of the track. We used a probability of $p_g = 0.999$ for the gating or prediction region. We then limit the possible observations for inclusion into a track to only those that fall in the gating region for that track.

We can also do a similar form of gating for observations that we are considering to be the first observations of new tracks resulting from the split of an existing track. We can just add

σ_{X_s} and σ_{Y_s} to the x and y components of the conditional variance for the prediction of a new observation in a track. We can then form the prediction region or gate using the inflated variance. This is equivalent to forming a prediction region for $(X_i(t_j), Y_i(t_j)) + (\psi_{X,s}, \psi_{Y,s})$. We would then only consider pairs of new observations within this region to possibly be a split pair from the existing track.

For possible mergers, we can also form a similar region. We can compute the prediction region for the difference between a pair of existing tracks plus a random ψ_d term, for example $(X_1(t_j) - X_2(t_j), Y_1(t_j) - Y_2(t_j)) + (\psi_{X,d}, \psi_{Y,d})$. If this prediction region for this quantity includes zero, we will consider the two targets for a possible merger. Suppose targets 1 and 2 can be considered for merger, then we must also find a new observation to be the first observation of the track that they merge into. So we must form another prediction region for $1/2(X_1(t_j) + X_2(t_j), Y_1(t_j) + Y_2(t_j)) + (\psi_{X,m}, \psi_{Y,m})$. We would then only consider new observations within this region to be from a new track resulting from the merger of tracks 1 and 2.

The prediction regions described above are calculated via the Kalman Filter without the built in dependency of merging and splitting. That is the first observation is assumed fixed and the others are calculated assuming the model given in (2.3). It is possible to improve these regions if we had an iterative method such as the Kalman Filter to compute the updated conditional distribution of new observations from a track. This would also increase the efficiency of likelihood computation for a possible solution. As mentioned before, this is something to look into.

Suppose we are running the algorithm on a fixed number, n, of time points, and we obtain the set of likely solutions for the last time t_n. It is possible in the MHT that the a solution that would eventually be optimal (have the highest likelihood) has a likelihood that is not very high early on when considering only a subset of all the times. We can only hold on to a limited number of solutions at each time, so it is possible that the optimal solution will be discarded at an earlier time less than t_n and thus never recovered.

In this case however, the MHT will likely produce a solution that is close to the optimal one with just a few things different, i.e. false alarms connected together in a short track, declaring a birth and two deaths instead of a merger, etc. We can improve the set of solutions obtained at the last time, t_n, by a greedy exchange algorithm similar to that described by Sethi [55]. This basically considers making several simple changes to a solution. If a change results in an increased likelihood, then make the change. This process continues until there are no more beneficial changes to be made.

There are a multitude of possible changes that we could consider making, but from initial results of the MHT, it seemed to do a pretty good job of classifying the splitting and merging events correctly, as well as identifying the correct correspondences of observations within tracks. However, where it seemed to struggle the most, was to form short tracks that were made up of only false alarms. This is likely because, it had to discard the correct solution, which labeled these observations as false alarms, before it realized it would have to kill this track. Hence the simple greedy exchange algorithm we use here, will only consider the possibility of changing short tracks ($\leq M_n$ observations) into false alarm observations. For the results in Chapters 7 and 8 we set $M_n = 3$.

So for each of the K_n solutions produced by the MHT, we will go through and consider changing any track with less than 4 observations to a collection of false alarm observations. If one of these changes improve the likelihood, then we will keep it.

5.3 Incorporating Parameter Estimation

Up until this point, we have ignored the idea of estimating model parameters for the purposes of the tracking estimate. If the problem requires estimation of model parameters, then we can do so with those given in Chapter 4. The problem is that we do not know the values of the variables that go into these estimates unless we specify a tracking solution, $(\mathcal{U}, \mathcal{V}, \mathcal{P})$. Hence, we will allow each of the solutions that we consider in the MHT algorithm to have its own parameter estimates. Note that this will make some solutions have an overly optimistic likelihood and bias the distribution given in (5.8). We can limit the amount of this bias by setting reasonable limits for the parameter estimates. This can be done very effectively in most problems, since the researcher is usually quite familiar with the number of targets and false alarms to expect and the velocity of the targets, etc.

Also in the first few time points, some of our estimates in Chapter 4 cannot be computed because of lack of data. In these cases, we need an initial guess for some of the parameter values. We simply used the midpoint (or geometric midpoint for variance parameters) of the parameter limits for an initial guess until enough data was available to estimate these parameters.

Chapter 6

Theoretical Considerations

In this chapter we will derive some results for the tracking estimate of Chapter 5. Intuitively, the estimate should get better if we sample at a finer time step, Δt. The question then is how much better does the estimate get? Here we will show that under certain conditions the tracking estimate converges to the correct solution almost surely.

We will assume the following. For each $k = 1, 2, \ldots$ we look in on the process and collect observations $\mathcal{Z}_k = (\boldsymbol{Z}_{k,1} \ldots \boldsymbol{Z}_{k,n_k})$ at times $0 \leq t_1^k < t_2^k < \cdots < t_{n_k}^k \leq T$. Let $\Delta t_j^k = t_{j+1}^k - t_j^k$. Further let $t_{i,j}^k$, $j = 1, \ldots, n_i^k$ denote the j^{th} time that the i^{th} target is observed, where n_i^k is the number of times the the i^{th} target is observed. So then also $\Delta t_{i,j}^k = t_{i,j+1}^k - t_{i,j}^k$. At times it will be convenient to write $t_{i,j} = t_{i,j}^k$, $\Delta t_{i,j}^k = \Delta t_{i,j}$ 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 $G_i(t)$ in equations (2.3), (2.5), and (2.7). We will replace equation (2.3) with

$$X_i(t) = X_i(\xi_i) + \sigma_i^2 B_i(t - \xi_i) \tag{6.1}$$

and similarly adjust (2.5), and (2.7). 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 given by

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \frac{(X_i(t_{i,j+1}) - X_i(t_{i,j}))^2}{\Delta t_{i,j}} I_{E_{i,j}} \tag{6.2}$$

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

$$\left\{ \frac{(X_i(t_{i,j+1}) - X_i(t_{i,j}))^2 + (Y_i(t_{i,j+1}) - Y_i(t_{i,j}))^2}{\Delta t_{i,j}} \leq K_1 \sqrt{\log \log N} \right\} \tag{6.3}$$

for some constant $K_1 > 0$. The indicator in (6.2) 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. The conditions needed for Theorem 1 are as follows.

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

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

Condition 3. *The random motion component, $G_i(t)$, of the location model in Section 2.3, is a Brownian Motion for all targets.*

Condition 4. *The likelihood is calculated under the assumption that $G_i(t)$ is a Brownian Motion for all targets*

Condition 5. *The parameter estimates are confined to a compact set such that λ_0, λ_b, λ_d,*

λ_s, and λ_m are greater than zero and all the variance components of the location model (except σ_e^2) are greater than 0.

Condition 6. *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 (6.2).*

Condition 7. *There is no iid measurement error, $\sigma_e^2 = 0$.*

Condition 8. *The estimate $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$ given in (5.2) is restricted to those with less than M tracks where $M \geq m(\omega)$, which is the number of tracks 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.

Condition 9. *There are no missing observations, $P_d = 0$.*

Condition 10. *There are no false alarms, $\lambda_f = 0$.*

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

Theorem 1. *Assume Conditions 1-10. Let $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$ be a sequence of correct tracking solutions. Let $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ be given by (5.2) restricted by Condition 8 for each k. Then $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k = (\mathcal{U}, \mathcal{V}, \mathcal{P})_k$ eventually almost surely (a.s.)*

The conclusion in Theorem 1 means 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 $k > K(\omega)$.

Suppose we replace Condition 3 with the following condition.

Condition 3′. *The random motion component, $G_i(t)$, of the location model in Section 2.3, is any continuously differentiable random process that satisfies the result of Lemma 7.*

Lemma 7 is basically saying that there is no chance that two paths will ever be equal at the same time. Then we also have the following theorem.

Theorem 2. *Assume Conditions 1, 2, 3′ and 4-10. Let $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$ be a sequence of correct tracking solutions. Then $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k = (\mathcal{U}, \mathcal{V}, \mathcal{P})_k$ eventually a.s.*

At first glance it may seem somewhat restrictive to assume $\sigma_e^2 = 0$, but there can still be measurement error. As long as it is a continuously differentiable process, we could add it to the $G_i(t)$ component of the model and we have the same result.

We will prove the theorem by computing the ratio of the likelihood of the best possible alternative, $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})$, to that of the correct solution, $(\mathcal{U}, \mathcal{V}, \mathcal{P})$. We will then show that this ratio converges to zero.

These results assume a Brownian Motion model for likelihood computation. It would seem that for most cases the integrated Brownian Motion Model (IBM) would be a better fit to the data. But the Brownian Motion case will give some intuition into what should happen in the IBM case. For example we may conjecture that for a likelihood calculated with an IBM, we would get the same result in Theorem 2 for a twice continuously differentiable process.

Before we can compare likelihoods, we must first consider how an alternative solution can differ from the correct one. 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.

1. It incorrectly breaks a target track into two target tracks.

2. It incorrectly labels a merger as two deaths and a birth.

3. It incorrectly labels a split as a death and two births.

4. It incorrectly connects a death with a birth to make one track.

5. It incorrectly labels two deaths and a birth as a merger.

6. It incorrectly labels a death and two births as a split.

If we were considering false alarms, then there are two more possibilities

7. It incorrectly labels a false alarm as a target.

8. It incorrectly labels a target as a false alarm

The following propositions deal with each of the first six possible differences between solutions listed above. Each one assumes something about two sequences of tracking estimates $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ and $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$.

Recall from Chapter 5 that $\mathcal{Z}$ is the collection of locations and attributes for targets and/or false alarms and there is a mapping $g^*(\mathcal{P}, \mathcal{Z}) = (\mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}, \boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f)$. For the following we will assume that we do not have any attributes however so let

$$g^*(\hat{\mathcal{P}}, \mathcal{Z}) = (\hat{\mathcal{W}}_k, \hat{\mathcal{X}}_k, \hat{\mathcal{Y}}_k) \tag{6.4}$$

$$g^*(\tilde{\mathcal{P}}, \mathcal{Z}) = (\tilde{\mathcal{W}}_k, \tilde{\mathcal{X}}_k, \tilde{\mathcal{Y}}_k) \tag{6.5}$$

where recall that $\mathcal{X}$ and $\mathcal{Y}$ are the collection of x and y coordinates respectively of the locations of the targets and $\mathcal{W}$ is the collection of missing state variables. The attribute and false alarm variables are absent in (6.4) and (6.5) because we are not considering attributes and condition 10 implies there are no false alarms.

Recall that $t_{i,j}^k$, $j = 1, \ldots, n_i^k$ 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 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.

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.

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

Proposition 1a. *Assume Conditions 1-10. 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 target i_1 and the birth of target i_2 during a fixed time interval*

$[t^k_{j'}, t^k_{j'+1})$ *for each* k. *Then,*

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

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 $(\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{P}})_k$; see Figure 6.1. Note that this implies $\hat{m} = \tilde{m} - 1$ where $\hat{m}$ is the number of tracks in $(\hat{\mathcal{U}},\hat{\mathcal{V}},\hat{\mathcal{P}})_k$. From (5.7) of Chapter 5 the ratio of the densities in the proposition

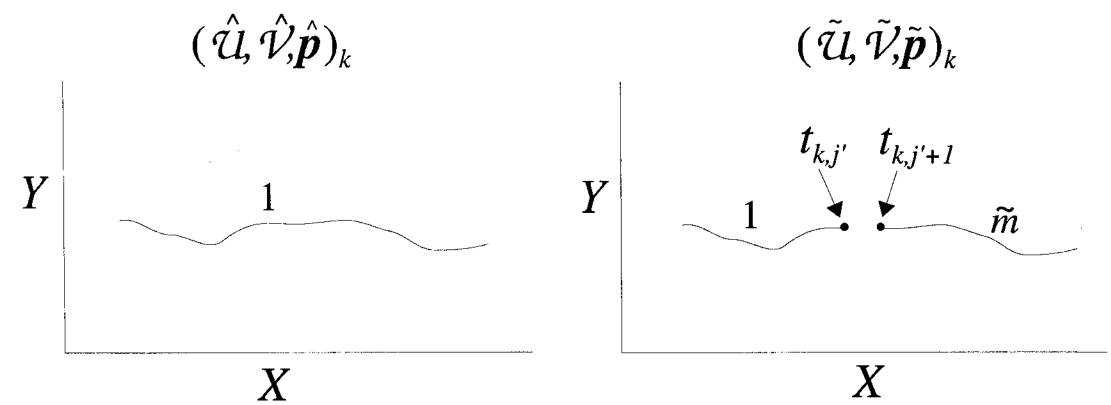

Figure 6.1: $(\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{P}})_k$ incorrectly breaks target track 1 into two tracks

can be written as

$$\frac{[(\mathcal{U},\mathcal{V},\mathcal{P})_k \mid \mathcal{Z}_k]((\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{P}})_k)}{[(\mathcal{U},\mathcal{V},\mathcal{P})_k \mid \mathcal{Z}_k]((\hat{\mathcal{U}},\hat{\mathcal{V}},\hat{\mathcal{P}})_k)} = \frac{[\mathcal{U}_k,\mathcal{V}_k](\tilde{\mathcal{U}}_k,\tilde{\mathcal{V}}_k)\cdot[\mathcal{W}_k \mid \tilde{\mathcal{U}}_k,\tilde{\mathcal{V}}_k](\tilde{\mathcal{W}}_k)}{[\mathcal{U}_k,\mathcal{V}_k](\hat{\mathcal{U}}_k,\hat{\mathcal{V}}_k)\cdot[\mathcal{W}_k \mid \hat{\mathcal{U}}_k,\hat{\mathcal{V}}_k](\hat{\mathcal{W}}_k)}\cdot \frac{[\mathcal{X}_k \mid (\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{W}})_k](\tilde{\mathcal{X}}_k)\cdot[\mathcal{Y}_k \mid (\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{W}})_k](\tilde{\mathcal{Y}}_k)}{[\mathcal{X}_k \mid (\hat{\mathcal{U}},\hat{\mathcal{V}},\hat{\mathcal{W}})_k](\hat{\mathcal{X}}_k)\cdot[\mathcal{Y}_k \mid (\hat{\mathcal{U}},\hat{\mathcal{V}},\hat{\mathcal{W}})_k](\hat{\mathcal{Y}}_k)} \tag{6.6}$$

since again there are no attribute or false alarm variables. Also, since we are assuming

$P_d = 1$ by condition 9, we do not really need to consider the ratio

$$\frac{[\tilde{\mathcal{W}}_k \mid \tilde{\mathcal{U}}_k, \tilde{\mathcal{V}}_k](\tilde{\mathcal{W}}_k)}{[\hat{\mathcal{W}}_k \mid \hat{\mathcal{U}}_k, \hat{\mathcal{V}}_k](\hat{\mathcal{W}}_k)}$$

unless one of the solutions has a missing value in which case its density would be identically 0.

Let $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,n_i})$ and let

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

where D_j are the difference variables resulting from merger as given as in (3.25). By convention let $\mathcal{F}_{i,0} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{i-1}, D_1, \ldots, D_{N_m})$. 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

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

for example. So we can write the ratio of the x location densities in (6.6) by breaking it apart into tracks $2, \ldots, \tilde{m} - 1$, which the two solutions have in common, then handle tracks

1 and $\tilde{m}$ separately

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

$$= \frac{\prod_{i=2}^{\tilde{m}-1} \prod_{j=1}^{\tilde{n}_i} [\tilde{X}_{i,j} \mid \tilde{\mathcal{F}}_{i,j-1}](\tilde{X}_{i,j})}{\prod_{i=2}^{\tilde{m}-1} \prod_{j=1}^{\hat{n}_i} [\hat{X}_{i,j} \mid \hat{\mathcal{F}}_{i,j-1}](\hat{X}_{i,j})} \cdot \frac{\prod_{j=1}^{\tilde{n}_1} [\tilde{X}_{1,j} \mid \tilde{\mathcal{F}}_{1,j-1}](\tilde{X}_{1,j})}{\prod_{j=1}^{\tilde{n}_1} [\hat{X}_{1,j} \mid \hat{\mathcal{F}}_{1,j-1}](\hat{X}_{1,j})} \cdot \frac{[\tilde{X}_{\tilde{m},1} \mid \tilde{\mathcal{F}}_{\tilde{m},0}](\tilde{X}_{\tilde{m},1})}{[\hat{X}_{1,\tilde{n}_1+1} \mid \hat{\mathcal{F}}_{1,\tilde{n}_1}](\hat{X}_{1,\tilde{n}_1+1})} \cdot \frac{\prod_{j=1}^{\tilde{n}_{\tilde{m}}} [\tilde{X}_{\tilde{m},j} \mid \tilde{\mathcal{F}}_{\tilde{m},j-1}](\tilde{X}_{\tilde{m},j})}{\prod_{j=\tilde{n}_1+2}^{\hat{n}_1} [\hat{X}_{1,j} \mid \hat{\mathcal{F}}_{1,j-1}](\hat{X}_{1,j})} \quad (6.7)$$

But for $i = 2, \ldots, \tilde{m}-1$, and $j = 1, \ldots, \hat{n}_i$, we have $\hat{X}_{i,j} = \tilde{X}_{i',j}$ for some i'. The number of observations $\hat{n}_i = \tilde{n}_{i'}$ for all of these tracks are the same as well. Also $\hat{X}_{1,j} = \tilde{X}_{1,j}$ for $j = 1, \ldots, \tilde{n}_1$ and $\hat{X}_{1,\tilde{n}_1+j} = \tilde{X}_{\tilde{m},j}$ for $j = 2, \ldots, \tilde{n}_{\tilde{m}}$. So by Lemma 6 the first, second and third terms of (6.7) are no more than $a^{\tilde{m}}(\log k)^{\tilde{m}b}$ for some constants a and b so that

$$\frac{[\tilde{\mathcal{X}}_k \mid (\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{W}})_k](\tilde{\mathcal{X}}_k)}{[\hat{\mathcal{X}}_k \mid (\hat{\mathcal{U}},\hat{\mathcal{V}},\hat{\mathcal{W}})_k](\hat{\mathcal{X}}_k)} \leq \frac{[\tilde{X}_{\tilde{m},1} \mid \tilde{\mathcal{F}}_{\tilde{m},0}](\tilde{X}_{\tilde{m},1})}{[\hat{X}_{1,\tilde{n}_1+1} \mid \hat{\mathcal{F}}_{1,\tilde{n}_1}](\hat{X}_{1,\tilde{n}_1+1})} \cdot a^{\tilde{m}}(\log k)^{\tilde{m}b} \quad \text{a.s.} \qquad (6.8)$$

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

$$\sup_{(\hat{\theta}_k,\tilde{\theta}_k)\in\Theta} \frac{[\tilde{\mathcal{X}}_k \mid (\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{W}})_k](\tilde{\mathcal{X}}_k)}{[\hat{\mathcal{X}}_k \mid (\hat{\mathcal{U}},\hat{\mathcal{V}},\hat{\mathcal{W}})_k](\hat{\mathcal{X}}_k)} \leq \sup_{(\hat{\theta}_k,\tilde{\theta}_k)\in\Theta} \frac{[\tilde{X}_{m,1} \mid \tilde{\mathcal{F}}_{m,0}](\tilde{X}_{m,1})}{[\hat{X}_{1,\tilde{n}_1+1} \mid \hat{\mathcal{F}}_{1,\tilde{n}_1}](\hat{X}_{1,\tilde{n}_1+1})} a^M(\log k)^{Mb} \text{ a.s.} \; (6.9)$$

Notice that the term in the numerator of (6.9) 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 $\sigma^2_{X_0}$. Also, the denominator can be written out by utilizing Lemma 3. So we have,

$$\sup_{(\hat{\theta}_k,\tilde{\theta}_k)\in\Theta} \frac{[\tilde{\mathcal{X}}_k \mid (\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{W}})_k](\mathcal{X}_k)}{[\hat{\mathcal{X}}_k \mid (\hat{\mathcal{U}},\hat{\mathcal{V}},\hat{\mathcal{W}})_k](\mathcal{X}_k)} \leq$$

$$\frac{\phi(0;0,\sigma^2_{X_0}) \cdot a^M (\log k)^{Mb}}{\inf\limits_{(\hat{\theta}_k,\tilde{\theta}_k)\in\Theta} \exp\left\{ -\frac{\left(\hat{X}_1(t^k_{j'+1})-\hat{X}_1(t^k_{j'})+O(k^{-1})\right)^2}{2\hat{\sigma}^2\Delta t^k_{j'}+O(k^{-2})} \right\} / \sqrt{2\pi\hat{\sigma}^2\Delta t^k_{j'} + O(k^{-2})}}. \tag{6.10}$$

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

$$\sup_{(\hat{\theta}_k,\tilde{\theta}_k)\in\Theta} \frac{[\tilde{\mathcal{X}}_k,\tilde{\mathcal{Y}}_k \mid (\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{W}})_k](\tilde{\mathcal{X}}_k,\tilde{\mathcal{Y}}_k)}{[\hat{\mathcal{X}}_k,\hat{\mathcal{Y}}_k \mid (\hat{\mathcal{U}},\hat{\mathcal{V}},\hat{\mathcal{W}})_k](\hat{\mathcal{X}}_k,\hat{\mathcal{Y}}_k)} \leq \frac{\phi(0;0,\sigma^2_{X_0})\phi(0;0,\sigma^2_{Y_0}) \cdot a^M(\log k)^{Mb}}{\inf\limits_{(\hat{\theta}_k,\tilde{\theta}_k)\in\Theta} \mathcal{D}} \tag{6.11}$$

where

$$\mathcal{D} = \frac{\exp\left\{ -\left(\frac{\left(\hat{X}_1(t^k_{j'+1})-\hat{X}_1(t^k_{j'})+O(k^{-1})\right)^2}{2\hat{\sigma}^2\Delta t^k_{j'}+O(k^{-2})} + \frac{\left(\hat{Y}_1(t^k_{j'+1})-\hat{Y}_1(t^k_{j'})+O(k^{-1})\right)^2}{2\hat{\eta}^2\Delta t^k_{j'}+O(k^{-2})} \right) \right\}}{2\pi\hat{\sigma}\hat{\eta}\Delta t^k_{j'} + O(k^{-3/2})}. \tag{6.12}$$

Now

$$\frac{\left(\hat{X}_1(t^k_{j'+1}) - \hat{X}_1(t^k_{j'})\right)^2}{\hat{\sigma}^2} \stackrel{\mathcal{L}}{=} \frac{\sigma^2}{\hat{\sigma}^2}(B(t^k_{j'+1}) - B(t^k_{j'})) \tag{6.13}$$

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, $\hat{\sigma}^2 \to \sigma^2$.

Using this fact along with (6.13) on $\mathcal{D}$, we have

$$\inf_{(\hat{\theta}_k,\tilde{\theta}_k)\in\Theta} \mathcal{D} = \frac{\exp\left\{-\frac{(1+o(1))\left(B(t^k_{j'+1})-B(t^k_{j'})\right)^2+(1+o(1))\left(\tilde{B}(t^k_{j'+1})-\tilde{B}(t^k_{j'})\right)^2}{2\Delta t^k_{j'}+O(k^{-2})}\right\}}{2\pi\sigma_1\eta_1\Delta t^k_{j'}+O(k^{-3/2})} \tag{6.14}$$

$$= \frac{\exp\left\{-\frac{\left(B(t^k_{j'+1})-B(t^k_{j'})\right)^2+\left(\tilde{B}(t^k_{j'+1})-\tilde{B}(t^k_{j'})\right)^2}{2\Delta t^k_{j'}\log\log(1/\Delta t^k_{j'})+\log\log(1/\Delta t^k_{j'})O(k^{-2})}(1+o(1))\log\log(1/\Delta t^k_{j'})\right\}}{C\Delta t^k_{j'}+O(k^{-3/2})} \tag{6.15}$$

$$= \frac{\exp\left\{-\frac{\left(B(t^k_{j'+1})-B(t^k_{j'})\right)^2+\left(\tilde{B}(t^k_{j'+1})-\tilde{B}(t^k_{j'})\right)^2}{2\Delta t^k_{j}\log\log(1/\Delta t^k_{j'})}(1+o(1))\log\log(1/\Delta t^k_{j'})\right\}}{C\Delta t^k_{j'}+O(k^{-3/2})} \tag{6.16}$$

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

$$\inf_{(\hat{\theta}_k,\tilde{\theta}_k)\in\Theta} \mathcal{D} \geq \frac{\exp\left\{-(1+o(1))\log\log(1/\Delta t^k_{j'})\right\}}{C\Delta t^k_{j'}+O(k^{-3/2})}. \tag{6.17}$$

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 (6.17) 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

$$\inf_{(\hat{\theta}_k,\tilde{\theta}_k)\in\Theta} \mathcal{D} = \frac{\log(1/\Delta t^k_{j'})^{-1+o(1)}}{C\Delta t^k_{j'}+O(k^{-3/2})} \tag{6.18}$$

$$\geq C'k(\log k)^{-1} \text{ a.s.} \tag{6.19}$$

for some constant C' where we used Condition 1 to get from (6.18) to (6.19). This says that the location contribution to the density of target 1 at time $t^k_{j'}$ goes to ∞ like $k/(\log(k))^c$. This along with (6.11) gives,

$$\sup_{(\hat{\theta}_k,\tilde{\theta}_k)\in\Theta} \frac{[\tilde{\mathcal{X}}_k, \tilde{\mathcal{Y}}_k \mid (\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{W}})_k](\tilde{\mathcal{X}}_k, \tilde{\mathcal{Y}}_k)}{[\hat{\mathcal{X}}_k, \hat{\mathcal{Y}}_k \mid (\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{W}})_k](\hat{\mathcal{X}}_k, \hat{\mathcal{Y}}_k)} \leq O(k^{-1}(\log(k))^c) \text{ a.s.} \tag{6.20}$$

We also need to consider the state density contribution to the ratio of (6.6). The state densities $[(\mathcal{U},\mathcal{V})_k]((\hat{\mathcal{U}},\hat{\mathcal{V}}_k)$ and $[(\mathcal{U},\mathcal{V})_k]((\tilde{\mathcal{U}},\tilde{\mathcal{V}})_k)$ will be different only in their contribution of the number of events during the interval $[t^k_{j'}, t^k_{j'+1})$. For $(\tilde{\mathcal{U}},\tilde{\mathcal{V}})_k$, there is one more death and one more birth in the interval than for $(\hat{\mathcal{U}},\hat{\mathcal{V}})_k$, so we have

$$\begin{aligned}
\frac{[(\mathcal{U},\mathcal{V})_k]((\tilde{\mathcal{U}},\tilde{\mathcal{V}})_k)}{[(\mathcal{U},\mathcal{V})_k]((\hat{\mathcal{U}},\hat{\mathcal{V}})_k)} &= \frac{P(U_{b,j'}=\hat{U}_{b,j'}+1)P(U_{d,j'}=\hat{U}_{d,j'}+1)P(U_{s,j'}=\hat{U}_{s,j'})P(U_{m,j'}=\hat{U}_{b,j'})P(V_k=\tilde{V}_k)}{P(U_{b,j'}=\hat{U}_{b,j'})P(U_{d,j'}=\hat{U}_{d,j'})P(U_{s,j'}=\hat{U}_{s,j'})P(U_{m,j'}=\hat{U}_{b,j'})P(V_k=\hat{V}_k)} \\
&= \frac{\left(\frac{(\lambda_b\Delta t^k_{j'})^{(\hat{U}_{b,j'}+1)}}{(\hat{U}_{b,j'}+1)!}e^{-\lambda_b\Delta t^k_{j'}}\right)\left(\frac{(\lambda_d\tilde{N}_{j'}\Delta t^k_{j'})^{(\hat{U}_{d,j'}+1)}}{(\hat{U}_{d,j'}+1)!}e^{-\lambda_d\tilde{N}_{j'}\Delta t^k_{j'}}\right)P(V_k=\tilde{V}_k)}{\left(\frac{(\lambda_b\Delta t^k_{j'})^{(\hat{U}_{b,j'})}}{(\hat{U}_{b,j'})!}e^{-\lambda_b\Delta t^k_{j'}}\right)\left(\frac{(\lambda_d\tilde{N}_{j'}\Delta t^k_{j'})^{(\hat{U}_{d,j'})}}{(\hat{U}_{d,j'})!}e^{-\lambda_d\tilde{N}_{j'}\Delta t^k_{j'}}\right)P(V_k=\hat{V}_k)} \\
&= \frac{(\lambda_b\Delta t^k_{j'})(\lambda_d\tilde{N}_{j'}\Delta t^k_{j'})}{(\hat{U}_{b,j'}+1)(\hat{U}_{d,j'}+1)}
\end{aligned} \tag{6.21}$$

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

$$\begin{aligned}
\sup_{(\hat{\theta}_k,\tilde{\theta}_k)\in\Theta} \frac{[(\mathcal{U},\mathcal{V})_k]((\tilde{\mathcal{U}},\tilde{\mathcal{V}})_k)}{[(\mathcal{U},\mathcal{V})_k]((\hat{\mathcal{U}},\hat{\mathcal{V}})_k)} &= O(\Delta t^k_{j'})^2 \\
&= O(k^{-2}) \text{ a.s.}
\end{aligned} \tag{6.22}$$

Putting this together with (6.20) gives us,

$$\sup_{(\hat{\theta}_k,\tilde{\theta}_k)\in\Theta} \frac{[(\mathcal{U},\mathcal{V},\mathcal{P})_k \mid \mathcal{Z}_k]((\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{P}})_k)}{[(\mathcal{U},\mathcal{V},\mathcal{P})_k \mid \mathcal{Z}_k]((\hat{\mathcal{U}},\hat{\mathcal{V}},\hat{\mathcal{P}})_k)} \leq O(k^{-3}(\log(k))^c) \text{ a.s.} \tag{6.23}$$

as we desired.

□

Proposition 1b. *Assume Conditions 1-10. 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* $[t^k_{j'}, t^k_{j'+1})$. *Then,*

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

for some positive constant c *which depends on* ω.

Proof. This proof follows the exact same logic as the previous proof of Proposition 1a. The only difference being that now $t^k_{j'}$ is an arbitrary time. Hence the only change will be to apply Lemma 2 instead of Lemma 1 to (6.14). So for the $\mathcal{D}$ of (6.11) we end up with

$$\mathcal{D} \geq \frac{\exp\left\{-\sup_j \frac{\left(B(t^k_{j+1}) - B(t^k_j)\right)^2 + \left(\tilde{B}(t^k_{j+1}) - \tilde{B}(t^k_j)\right)^2}{2\Delta t^k_j \log(1/\Delta t^k_j)} (1 + o(1)) \log(1/\Delta t^k_{j'})\right\}}{C\Delta t^k_{j'} + O(k^{-3/2})}. \tag{6.24}$$

Now use Lemma 2 which results in

$$\inf_{(\hat{\theta}_k,\tilde{\theta}_k)\in\Theta} \mathcal{D} \geq \frac{\exp\left\{-(1+o(1))\log(1/\Delta t^k_{j'})\right\}}{C\Delta t^k_{j'} + O(k^{-3/2})} \tag{6.25}$$

$$= \frac{\Delta(t^k_{j'})^{1+o(1)}}{C\Delta t^k_{j'} + O(k^{-3/2})} \tag{6.26}$$

$$= O(1) \text{ a.s.} \tag{6.27}$$

This gives us,

$$\sup_{(\hat{\theta}_k,\tilde{\theta}_k)\in\Theta} \frac{[\mathcal{X}_k,\mathcal{Y}_k \mid (\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{W}})_k](\tilde{\mathcal{X}}_k,\tilde{\mathcal{Y}}_k)}{[\mathcal{X}_k,\mathcal{Y}_k \mid (\hat{\mathcal{U}},\hat{\mathcal{V}},\hat{\mathcal{W}})_k](\hat{\mathcal{X}}_k,\hat{\mathcal{Y}}_k)} \leq a^M(\log k)^{Mb} \text{ a.s.} \tag{6.28}$$

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

$$\sup_{(\hat{\theta}_k,\tilde{\theta}_k)\in\Theta} \frac{[(\mathcal{U},\mathcal{V})_k]((\tilde{\mathcal{U}},\tilde{\mathcal{V}})_k)}{[(\mathcal{U},\mathcal{V})_k]((\hat{\mathcal{U}},\hat{\mathcal{V}})_k)} = O(k^{-2}) \text{ a.s.} \tag{6.29}$$

This along with (6.28) gives us,

$$\sup_{(\hat{\theta}_k,\tilde{\theta}_k)\in\Theta} \frac{[(\mathcal{U},\mathcal{V},\mathcal{P})_k \mid \mathcal{Z}_k]((\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{P}})_k)}{[(\mathcal{U},\mathcal{V},\mathcal{P})_k \mid \mathcal{Z}_k]((\hat{\mathcal{U}},\hat{\mathcal{V}},\hat{\mathcal{P}})_k)} \leq O(k^{-2}(\log k)^c) \text{ a.s.} \tag{6.30}$$

□

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 1-10. 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 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*

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

for some positive constant c *which depends on* ω.

Proof. WLOG assume that $i_1 = 1$, $i_2 = 2$ and $i_3 = 3$; see Figure 6.2.

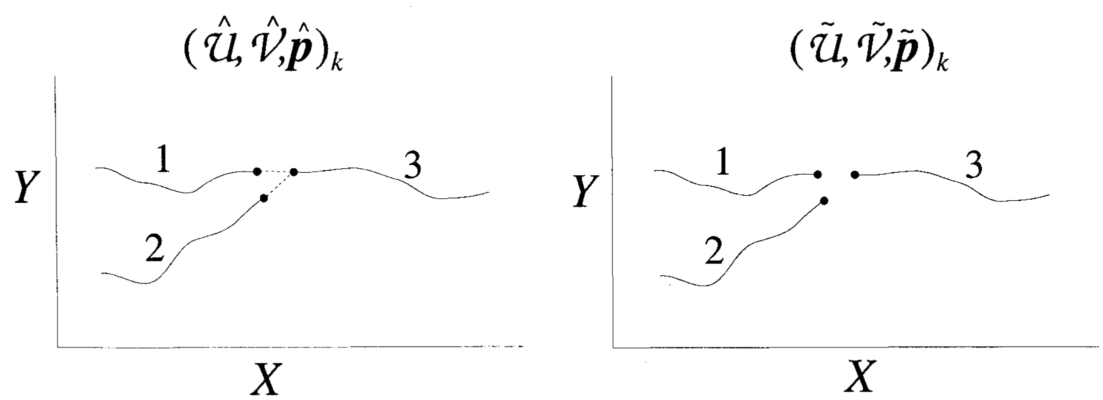

Figure 6.2: $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ incorrectly breaks apart a merger into two deaths and a birth

Here we have that for $i = 1, \ldots, m$, and $j = 1, \ldots, \hat{n}_i$, 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 (6.6) is no more than $a^{\tilde{m}}(\log k)^{\tilde{m}b}$ for some constants a and b,

$$\frac{[\tilde{\mathcal{X}}_k \mid (\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k](\tilde{\mathcal{X}})}{[\hat{\mathcal{X}}_k \mid (\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k](\hat{\mathcal{X}})} \leq a^{\tilde{m}}(\log k)^{\tilde{m}b} \text{ a.s.} \tag{6.31}$$

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

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[\tilde{\mathcal{X}}_k \mid (\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k](\tilde{\mathcal{X}})}{[\hat{\mathcal{X}}_k \mid (\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k](\hat{\mathcal{X}})} \leq a^M (\log k)^{Mb} \quad \text{a.s.} \tag{6.32}$$

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

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[\tilde{\mathcal{X}}_k, \hat{\mathcal{Y}}_k \mid (\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{W}})_k](\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}_k)}{[\hat{\mathcal{X}}_k, \hat{\mathcal{Y}}_k \mid (\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{W}})_k](\hat{\mathcal{X}}, \hat{\mathcal{Y}}_k)} \leq a^{2M} (\log k)^{2Mb} \quad \text{a.s.} \tag{6.33}$$

For the state model contribution, the state 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 (6.22) then we have

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[(\mathcal{U}, \mathcal{V})_k]((\tilde{\mathcal{U}}, \tilde{\mathcal{V}})_k)}{[(\mathcal{U}, \mathcal{V})_k]((\hat{\mathcal{U}}, \hat{\mathcal{V}})_k)} \leq O(k^{-2}) \text{ a.s.} \tag{6.34}$$

This along with (6.33) gives us the desired result.

□

Proposition 3. *Assume Conditions 1-10. 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*

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

for some positive constant c which depends on ω.

Proof. This is symmetric with respect to the difference between solutions in Proposition 2; see Figure 6.3. The proof will therefore be identical.

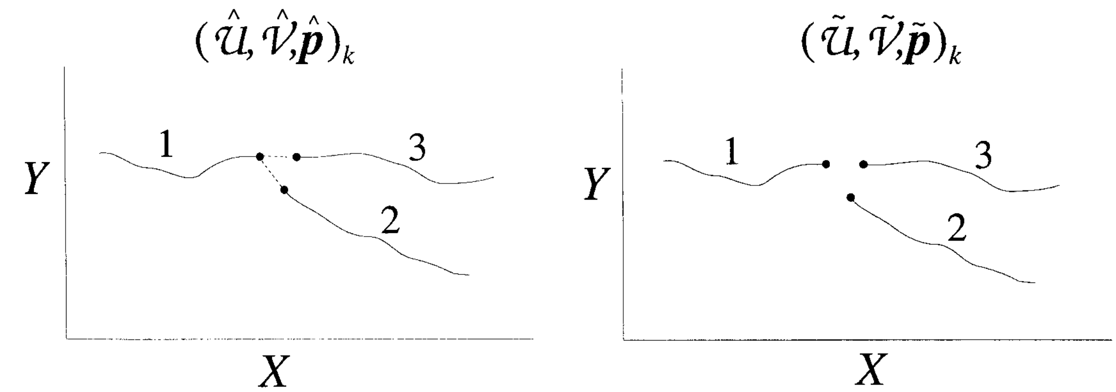

Figure 6.3: $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ incorrectly breaks apart a split into a death and two births

□

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 $[t^k_{j'}, t^k_{j'+1})$. These two events are not necessarily labeled correctly. However target i_1 at time $t^k_{j'}$ is assumed to be distinct from target i_2 at time $t^k_{j'+1}$. 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 6.4. Proposition 4 then says that there can be no differences of this type eventually.

Proposition 4. *Assume Conditions 1-10. 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. 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*

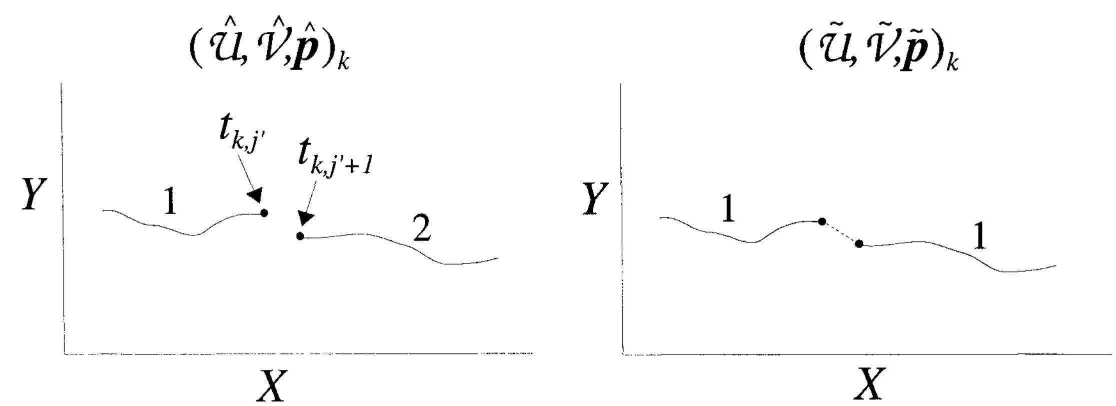

Figure 6.4: $(\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{P}})_k$ incorrectly joins target tracks 1 and 2 into one track

and i_2 which are distinct into one target track. Then Θ is the empty set.

Proof. WLOG assume that $i_1 = 1$ and $i_2 = 2$ as in Figure 6.4. The observations that get incorrectly joined together in a track in the solution $(\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{P}})$ are $(\hat{X}_1(t^k_{j'}),\hat{Y}_1(t^k_{j'}))$ and $(\hat{X}_2(t^k_{j'+1}),\hat{Y}_2(t^k_{j'+1}))$. Now these are observations from distinct targets in the correct solution so by Lemma 7 and the continuity of Brownian paths, we know that

$$
\begin{aligned}
&\inf_j \left\{ \left(\hat{X}_2(t^k_{j'+1}) - \hat{X}_1(t^k_{j'})\right)^2 + \left(\hat{Y}_2(t^k_{j'+1}) - \hat{Y}_1(t^k_{j'})\right)^2 \right\} \\
&\qquad = \inf_j \left\| \left(\hat{X}_1(t^k_{j'}),\hat{Y}_1(t^k_{j'})\right) - \left(\hat{X}_2(t^k_{j'}+1),\hat{Y}_2(t^k_{j'+1})\right) \right\|^2 \\
&\qquad \geq C_1 > 0 \text{ a.s.} \qquad (6.35)
\end{aligned}
$$

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.

□

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 $[t^k_{j'}, t^k_{j'+1})$ 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 1-10. 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,*

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

Proof. WLOG let $i_1 = 1$, $i_2 = 2$, and $i_3 = 3$; see Figure 6.5. Let the unknown times at

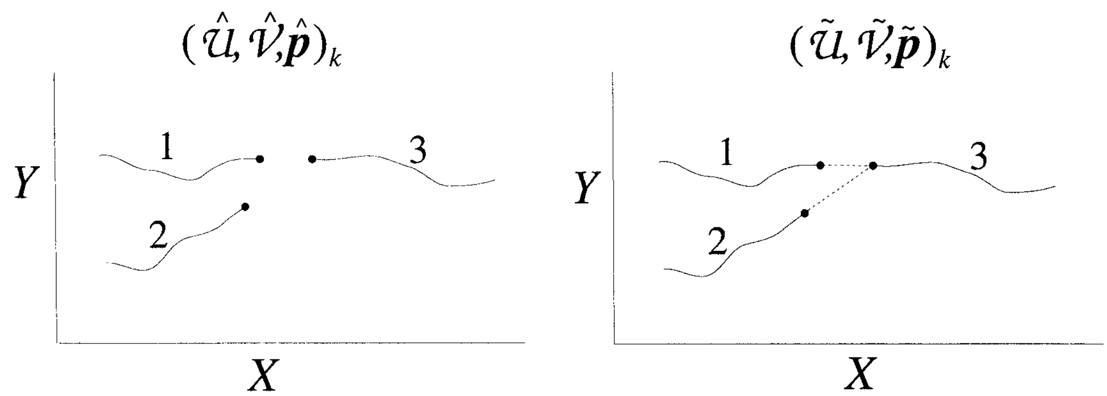

Figure 6.5: $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ incorrectly merges targets 1 and 2 into target 3

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 state 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, $t^k_{j'}$, 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. Since the probability of detection, P_d, is assumed to be 1 by Condition 9, $[\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k, \mathcal{Z}_k]((\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k, \mathcal{Z}_k) = 0$ unless it merges targets 1 and 2 before before τ_1. But both targets still exists in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ at time $t^k_{j'}$ so this would violate the hypothesis. So eventually $[\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k, \mathcal{Z}_k]((\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k, \mathcal{Z}_k) = 0$.

□

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 1-10. 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*

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

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 $i = 1, \ldots, m$ and $j = 1, \ldots, \hat{n}_i$, 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

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[\tilde{\mathcal{X}}_k, \hat{\mathcal{Y}}_k \mid (\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{W}})_k](\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}_k)}{[\hat{\mathcal{X}}_k, \hat{\mathcal{Y}}_k \mid (\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{W}})_k](\hat{\mathcal{X}}, \hat{\mathcal{Y}}_k)} \leq a^{2M} (\log k)^{2Mb} \text{ a.s.} \tag{6.36}$$

For the state 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 $[t^k_{j'}, t^k_{j'+1})$. 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 state 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 $[t^k_{j'}, t^k_{j'+1})$. The state model for $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ has two more events in $[t^k_{j'}, t^k_{j'+1})$ than does $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$. In the same manner we derived (6.21) we can see that

$$\frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k]((\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k)}{[(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k]((\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k)} = \frac{(\lambda_d(\tilde{N}_{j'} - 1)\Delta t^k_{j'})(\tilde{U}_{b,j'} + 1)(\tilde{U}_{d,j'} + 1)(\tilde{U}_{d,j'} + 2)}{(\lambda_b \Delta t^k_{j'})(\lambda_d \tilde{N}_{j'} \Delta t^k_{j'})^2(\hat{U}_{m,j'} + 1)}$$

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

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k]((\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k)}{[(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k]((\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k)} \leq ck^2 \quad \text{a.s.} \tag{6.37}$$

Combining this with (6.36) gives the desired result.

□

Proposition 6a is the counterpart of Proposition 5a for splitting instead of merging; see Figure 6.6.

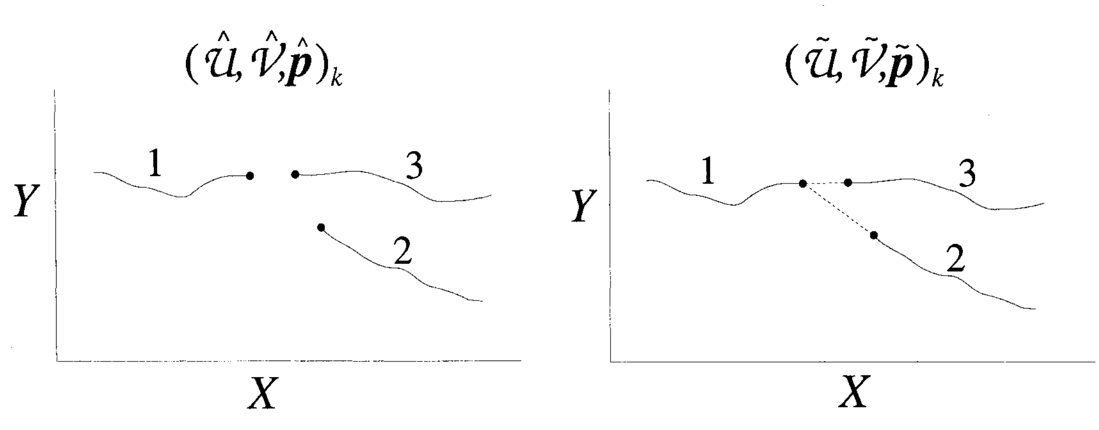

Figure 6.6: $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ incorrectly merges targets 1 and 2 into target 3

Proposition 6a. *Assume Conditions 1-10. 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_1 *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,*

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

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

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 1-10. 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_1 *and* i_3 *that were born in* $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$. *Then*

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

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.

□

Proof of Theorem 1. Let $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$ 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 $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$. 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 6.7 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 6.7. 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 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

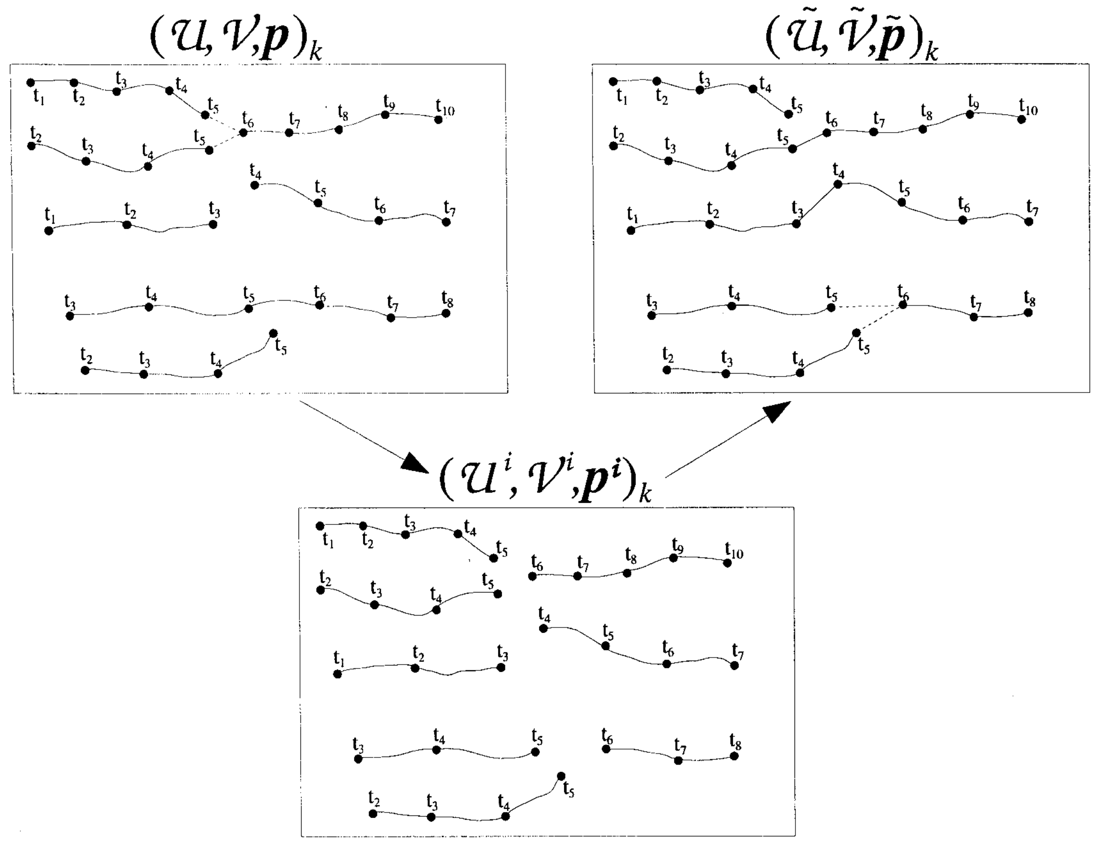

Figure 6.7: Break $(\mathcal{U},\mathcal{V},\mathcal{P})_k$ into track segments and connect them to get $(\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{P}})_k$

sequence as,

$$(\mathcal{U},\mathcal{V},\mathcal{P})_k, (\mathcal{U}^1,\mathcal{V}^1,\mathcal{P}^1)_k, (\mathcal{U}^2,\mathcal{V}^2,\mathcal{P}^2)_k, \ldots, (\mathcal{U}^l,\mathcal{V}^l,\mathcal{P}^l)_k, (\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{P}})_k \tag{6.38}$$

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

$$\begin{aligned}\frac{[(\mathcal{U},\mathcal{V},\mathcal{P})_k \mid \mathcal{Z}]((\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{P}})_k)}{[(\mathcal{U},\mathcal{V},\mathcal{P})_k \mid \mathcal{Z}]((\mathcal{U},\mathcal{V},\mathcal{P})_k)} &= \frac{[(\mathcal{U},\mathcal{V},\mathcal{P})_k \mid \mathcal{Z}]((\mathcal{U}^1,\mathcal{V}^1,\mathcal{P}^1)_k)}{[(\mathcal{U},\mathcal{V},\mathcal{P})_k \mid \mathcal{Z}]((\mathcal{U},\mathcal{V},\mathcal{P})_k)} \cdot \frac{[(\mathcal{U},\mathcal{V},\mathcal{P})_k \mid \mathcal{Z}]((\mathcal{U}^2,\mathcal{V}^2,\mathcal{P}^2)_k)}{[(\mathcal{U},\mathcal{V},\mathcal{P})_k \mid \mathcal{Z}]((\mathcal{U}^1,\mathcal{V}^1,\mathcal{P}^1)_k)} \\ &\quad \cdots \frac{[(\mathcal{U},\mathcal{V},\mathcal{P})_k \mid \mathcal{Z}]((\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{P}})_k)}{[(\mathcal{U},\mathcal{V},\mathcal{P})_k \mid \mathcal{Z}]((\mathcal{U}^l,\mathcal{V}^l,\mathcal{P}^l)_k)}\end{aligned} \tag{6.39}$$

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

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

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 $\Theta' = \Theta \backslash \{(\mathcal{U}, \mathcal{V}, \mathcal{P})_k\}$. We need to show that the supremum over $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k \in \Theta'$ of the ratio in (6.39) converges to 0 as $k \to \infty$. By Propositions 1a, 1b, 2, 3, 5a, and 6a, any of the ratios in (6.39) that have differences 1a, 1b, 2, 3, 5a, and 6a, are $\le O(k^{-2}(\log k)^c)$. However, if any of the terms in (6.39) 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 $[t^k_{j'}, t^k_{j'+1})$, 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 $[t^k_{j'}, t^k_{j'+1})$ 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 $[t^k_{j'}, t^k_{j'+1})$, 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 $[t^k_{j'}, t^k_{j'+1})$ 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 (6.39) 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 $[t^k_{j'}, t^k_{j'+1})$ 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 $[t^k_{j'}, t^k_{j'+1})$ 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 $O(k^{-2}(\log k)^c)$. Hence the overall contribution of difference 5b to the ratio of (6.39) for cases 2-5 is no more than $O(k^{-2}(\log k)^{2c})$.

Therefore if there is one difference 5b applied in any interval then the overall contribution of that difference to the ratio in (6.39) 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

$[t^k_{j'}, t^k_{j'+1})$ 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 (6.39) is no more than $O(k^{-1}(\log k)^{2c})$

We now consider N^* differences 5b and/or 6b applied in a single interval $[t^k_{j'}, t^k_{j'+1})$, 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 $[t^k_{j'}, t^k_{j'+1})$. 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 $[t^k_{j'}, t^k_{j'+1})$. 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 $[t^k_{j'}, t^k_{j'+1})$ previously in our sequence of solutions. So the contribution of the N^* differences 5b and/or 6b to the ratio of (6.39) 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

$$N_5 = \left\lceil \frac{N^*}{2} \right\rceil \text{ and } N_6 = \left\lfloor \frac{N^*}{2} \right\rfloor$$

or vise 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

$$(R_{1b})^{(5/4)N^*}(R_{5b})^{N^*} = O(k^{-(1/2)N^*}(\log k)^{(9/4)N^* c}).$$

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 (6.39) is no more than $O(k^{-1}(\log k)^c)$ for some constant c.

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

$$\sup_{(\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{P}})_k \in \Theta} \frac{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}]((\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k)}{[(\mathcal{U}, \mathcal{V}, \mathcal{P})_k \mid \mathcal{Z}]((\mathcal{U}, \mathcal{V}, \mathcal{P})_k)} \leq O(k^{-1}(\log k)^c) \tag{6.40}$$

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

$$\sup_{(\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{P}})_k\in\Theta} \frac{[(\mathcal{U},\mathcal{V},\mathcal{P})_k \mid \mathcal{Z}]((\tilde{\mathcal{U}},\tilde{\mathcal{V}},\tilde{\mathcal{P}})_k)}{[(\mathcal{U},\mathcal{V},\mathcal{P})_k \mid \mathcal{Z}]((\mathcal{U},\mathcal{V},\mathcal{P})_k)} \leq O(k^{-2}(\log k)^c) \tag{6.41}$$

Hence in general we have (6.40). This means for a given ω there exists a $K(\omega)$ s.t. for all $k > K(\omega)$,

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

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

□

Proof of Theorem 2. Whenever we used the Law of the Iterated Logarithm or Levy's Modulus of Continuity in the Lemmas or Propositions, we can substitute the assumption that $X_i(t)$ has a continuous derivative and hence the derivative is bounded on the interval $[0, T]$. This implies a faster convergence of differences between consecutive observations than we had for Brownian Motion. Hence the rates in the propositions will only improve and the proof proceeds the same as in Theorem 1.

□

Chapter 7

Simulated Data Results

In this chapter, we present some results of the tracking algorithm on simulated data. For all of these simulations, the data, $\mathcal{Z}$, is assumed to come from the model given in Chapter 2. The random motion component, $G_i(t)$ is an integrated Brownian Motion for all targets. The parameters used to simulate the different cases will be given below. All of the simulations use common location parameters. These values were meant to make the target tracks produced from the model behave like the storm tracks of the Chapter 8. So in all of the realizations we set, $\mu_{X_0} = -113$, $\sigma^2_{X_0} = 100$, $\mu_{X'_0} = 1.5$, $\sigma^2_{X'_0} = .1$, $\sigma^2_i = 0.1$ for all i, $\sigma^2_{X_s} = .5$, $\sigma^2_{X'_s} = .01$, $\sigma^2_{X_m} = .125$, $\sigma^2_{X'_m} = .01$, $\sigma^2_{X_d} = 1$, $\sigma^2_{X_e} = 0$, $\mu_{Y_0} = 37.5$, $\sigma^2_{Y_0} = 100$, $\mu_{Y'_0} = 0$, $\sigma^2_{Y'_0} = 2$, $\eta^2_i = .1$ for all i, $\sigma^2_{Y_s} = .5$, $\sigma^2_{Y'_s} = .5$, $\sigma^2_{Y_m} = .125$, $\sigma^2_{Y'_m} = .01$, $\sigma^2_{Y_d} = 1$, and $\sigma^2_{Y_e} = 0$, where $\mu_{X_0}, \sigma^2_{X_0}, \ldots, \sigma^2_{X_e}$ are defined in Section 2.3. The parameters $\mu_{Y_0}, \sigma^2_{Y_0}, \ldots, \sigma^2_{Y_e}$ are the counterparts for the y-coordinate. Also σ^2_i and η^2_i are the variance scalers multiplied to $G_i(t)$ for $X_i(t)$ and $Y_i(t)$ respectively.

We also set the probability of detection to be $P_d = 0.95$ in all simulations. The parameters λ_0, λ_f, λ_b, λ_d, and λ_s, and λ_m are different for each simulation and will be described for each case.

For the parameter estimation, we restricted the parameter values to the followings sets $\lambda_0 \in [0, 25]$, $\lambda_f \in [0, 0.02]$, $\lambda_b \in [0.001, .25]$, $\lambda_d \in [0.001, .15]$, $\lambda_s \in [0.001, .15]$, $\lambda_m \in [0.001, .15]$, $P_d \in [0.5, 1.0]$, $\mu_{X_0} \in [-120, -85]$, $\sigma^2_{X_0} \in [500, 1000]$, $\mu_{X'_0} \in [0, 5]$, $\sigma^2_{X'_0} \in [0.001, 5.0]$, $\sigma^2_i \in [0.001, 10.0]$, $\sigma^2_{X_s} \in [0.001, 1.5]$, $\sigma^2_{X'_s} \in [0.0, 1.0]$, $\sigma^2_{X_m} \in [0.001, 0.5]$, $\sigma^2_{X'_m} \in [0.0, 1.0]$, $\sigma^2_{X_d} \in [0.001, 5.0]$, $\sigma^2_{X_e} \in [0.0, 1.0]$, $\mu_{Y_0} \in [25, 50]$, $\sigma^2_{Y_0} \in [500, 1000]$, $\mu_{Y'_0} \in [-5, 5]$, $\sigma^2_{Y'_0} \in [0.5, 10.0]$, $\eta^2_i \in [.001, 10.0]$, $\sigma^2_{Y_s} \in [0.001, 1.5]$, $\sigma^2_{Y'_s} \in [0.0, 1.0]$, $\sigma^2_{Y_m} \in [0.001, 0.5]$, $\sigma^2_{Y'_m} \in [0.0, 1.0]$, $\sigma^2_{Y_d} \in [0.001, 5.0]$, and $\sigma^2_{Y_e} \in [0.0, 1.0]$. Notice that it seems awkward to have the lower parameter estimate limits for $\sigma^2_{X_0}$ and $\sigma^2_{Y_0}$ be higher than the actual values are set in the simulation. This was done because for this problem with $\Delta t = 1$ and clutter, the algorithm seemed to perform better when there was a little extra penalty for a target birth.

7.1 Simulations without Clutter

This set of simulations assumes that there are no false alarms or that $\lambda_f = 0$. This was essentially done to get a baseline idea of how well the algorithm works to identify birth, death, splitting and merging. There are six cases that we considered here: (i) birth only, (ii) death only, (iii) splitting only, (iv) merging only, (v) completely random model realizations, and (vi) completely random model realizations with size. For each case we generated $N = 100$ realizations. These simulations take place on the time interval $[0, 9]$ with $\Delta t_j = 1$ for all j so that $\boldsymbol{t} = (0, 1, \ldots, 9)$.

In these simulations we have the following hypotheses we wish to investigate.

1. The percentage of births, deaths, splits, and mergers labeled correctly in each of the

first four simulations will be roughly equal to the rates of correctly labeled events in the full model realizations of simulation (v).

2. Since birth is symmetric to death in reverse time, we would expect that the rate of correctly labeled births would be similar to that of correctly labeled deaths.

3. Since also splitting is symmetric to merging in reverse time, we would expect that the rate of correctly labeling these two events would be similar.

4. The results with additional size information in simulation (vi) should be an improvement over those in simulation (v).

7.1.1 Birth Only

In these simulations, we set $\lambda_0 = 2.0$, $\lambda_b = 0.20$ so that we would have an average of approximately 2 births in a time interval $[0, 9]$. We then set $\lambda_d = \lambda_s = \lambda_m = 0$ so we could isolate the tracking algorithm's ability to identify birth events. We also restricted the realizations to have at least one birth event. An example of a realization from this model can be seen in Figure 7.1. This plot has all the observations for each time on one plot. Observations from time n are labeled 'n'. The correct solution and the top four alternative solutions are given along with their estimated probability given by (5.8).

7.1.2 Death Only

In these simulations, we set $\lambda_0 = 4.0$, $\lambda_d = 0.10$. This makes for an average of about 2.5 deaths in the time interval and we restricted the set of realizations to have at least one death. We then set $\lambda_b = \lambda_s = \lambda_m = 0$. An example of a realization from this model can be seen in Figure 7.2. Again, the the correct solution and the top four alternative solutions are

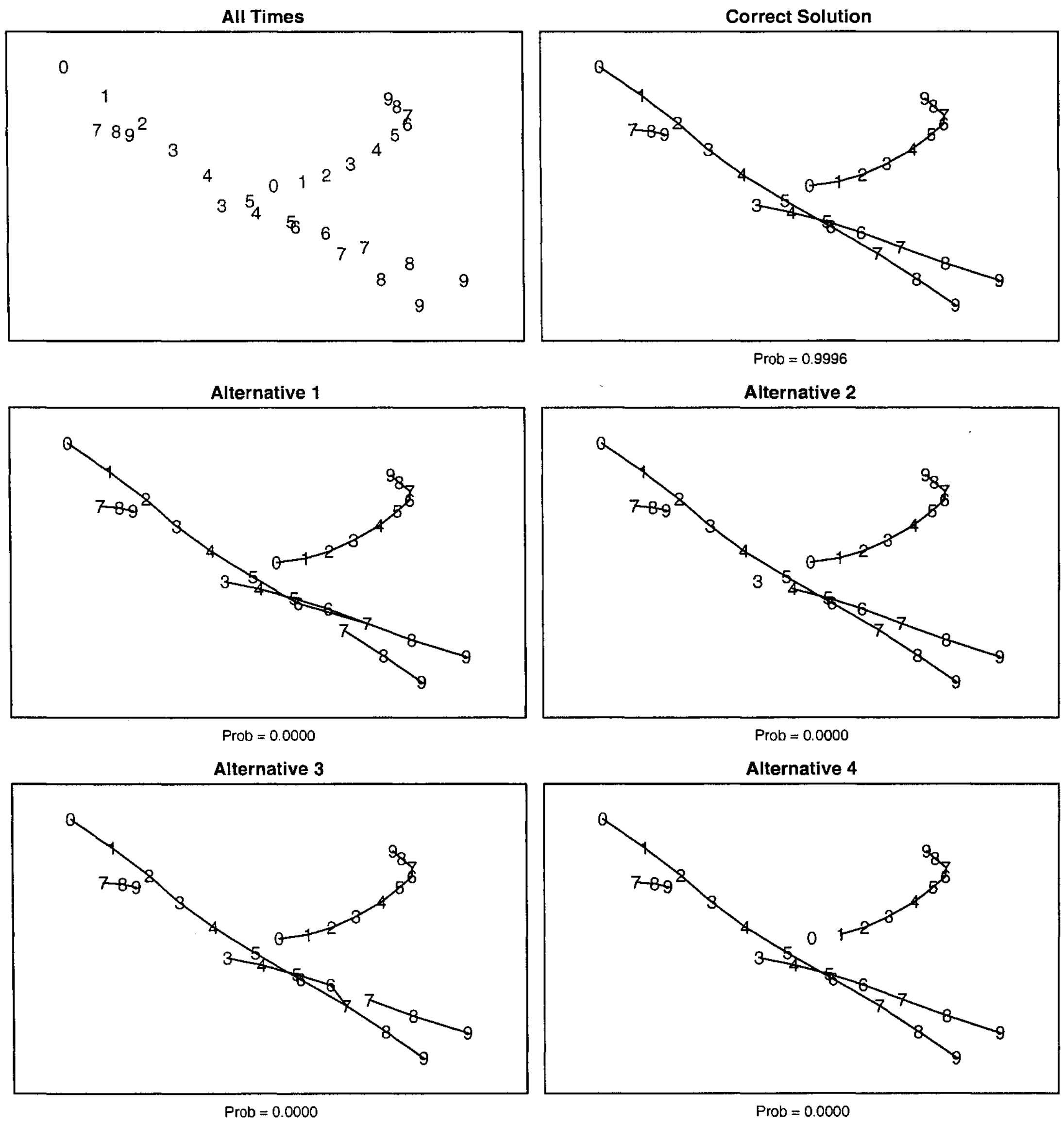

Figure 7.1: Birth Only Realization with Possible Solutions

given. Notice that the correct solution and the top alternative are indistinguishable from these plots. What actually happened was that the target that reaches the top of the plot at time 8, actually dies before time 9. The first alternative labeled it as missing at time 9 however, so it did not kill it. This is a common problem for deaths that occur near the end of the time window.

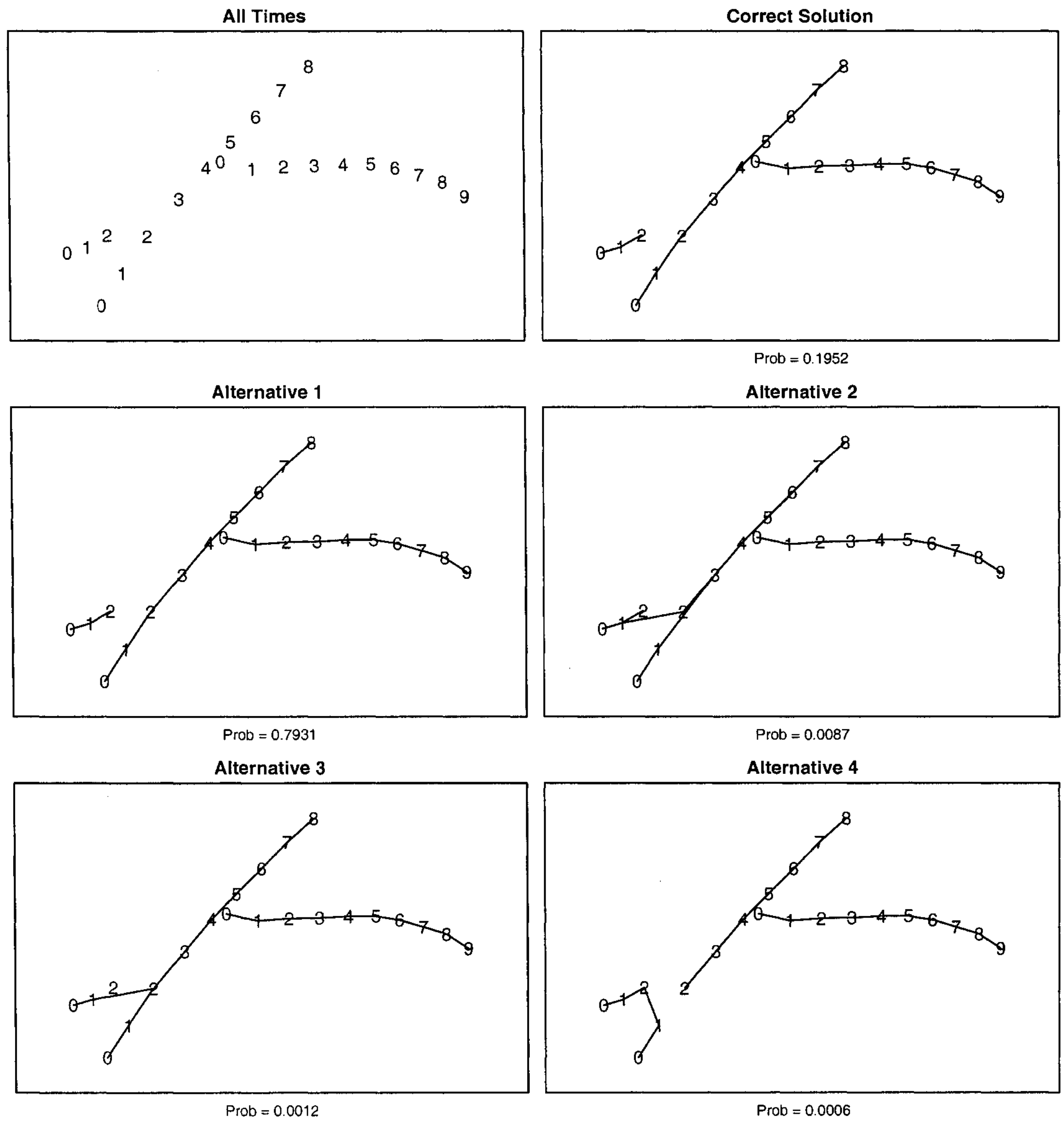

Figure 7.2: Death Only Realization with Possible Solutions

7.1.3 Splitting Only

In these simulations, we forced there to be exactly one target that split into two targets at a random uniformly distributed time in the interval $(1.0, 8.0)$. An example of one of these realizations is given in Figure 7.3

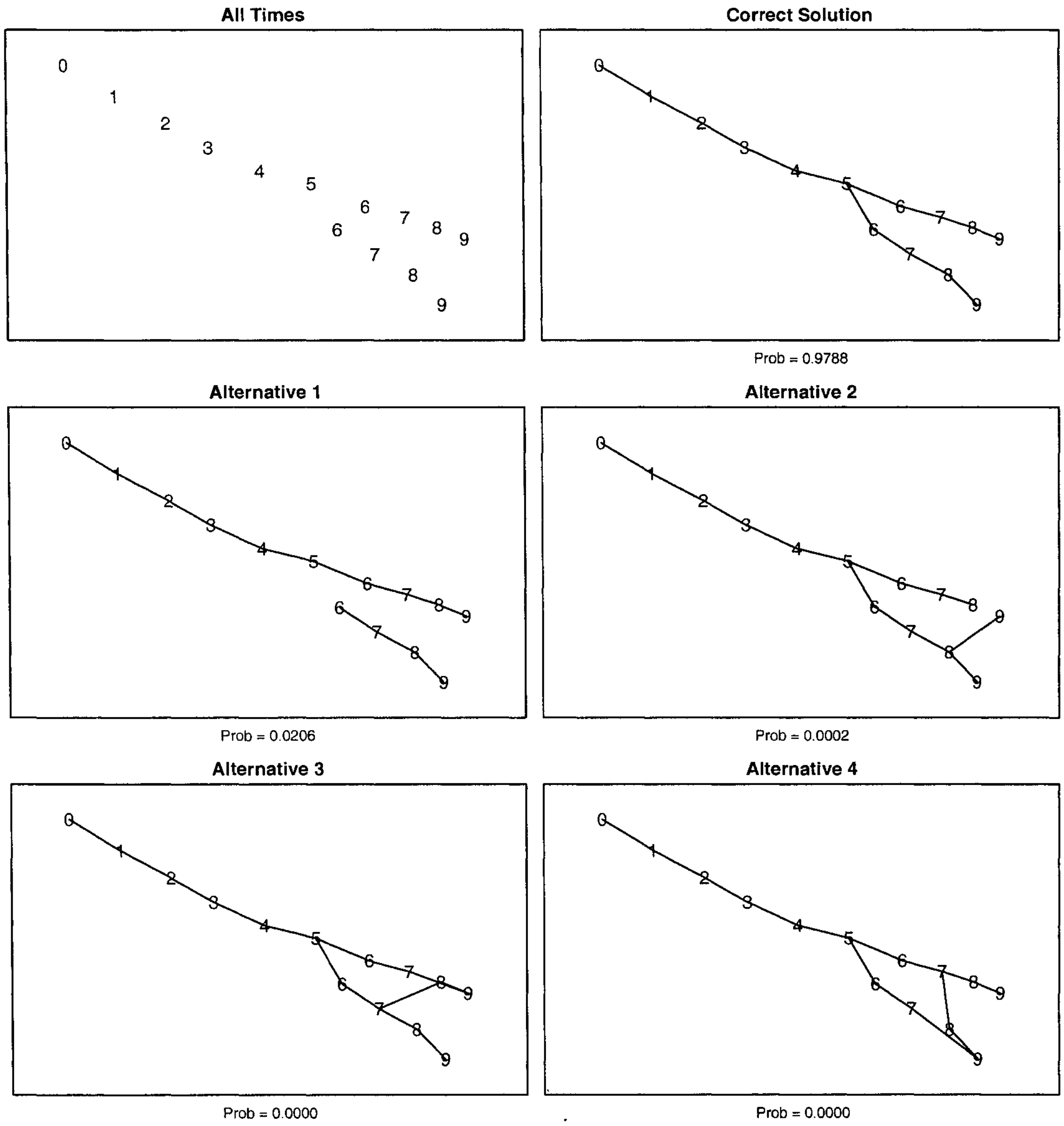

Figure 7.3: Splitting Only Realization with Possible Solutions

7.1.4 Merging Only

In the same manner as for the splitting simulations, here we forced exactly one merger by two targets at a uniform time in the interval (1.0, 8.0). An example of one of these realizations is given in Figure 7.4

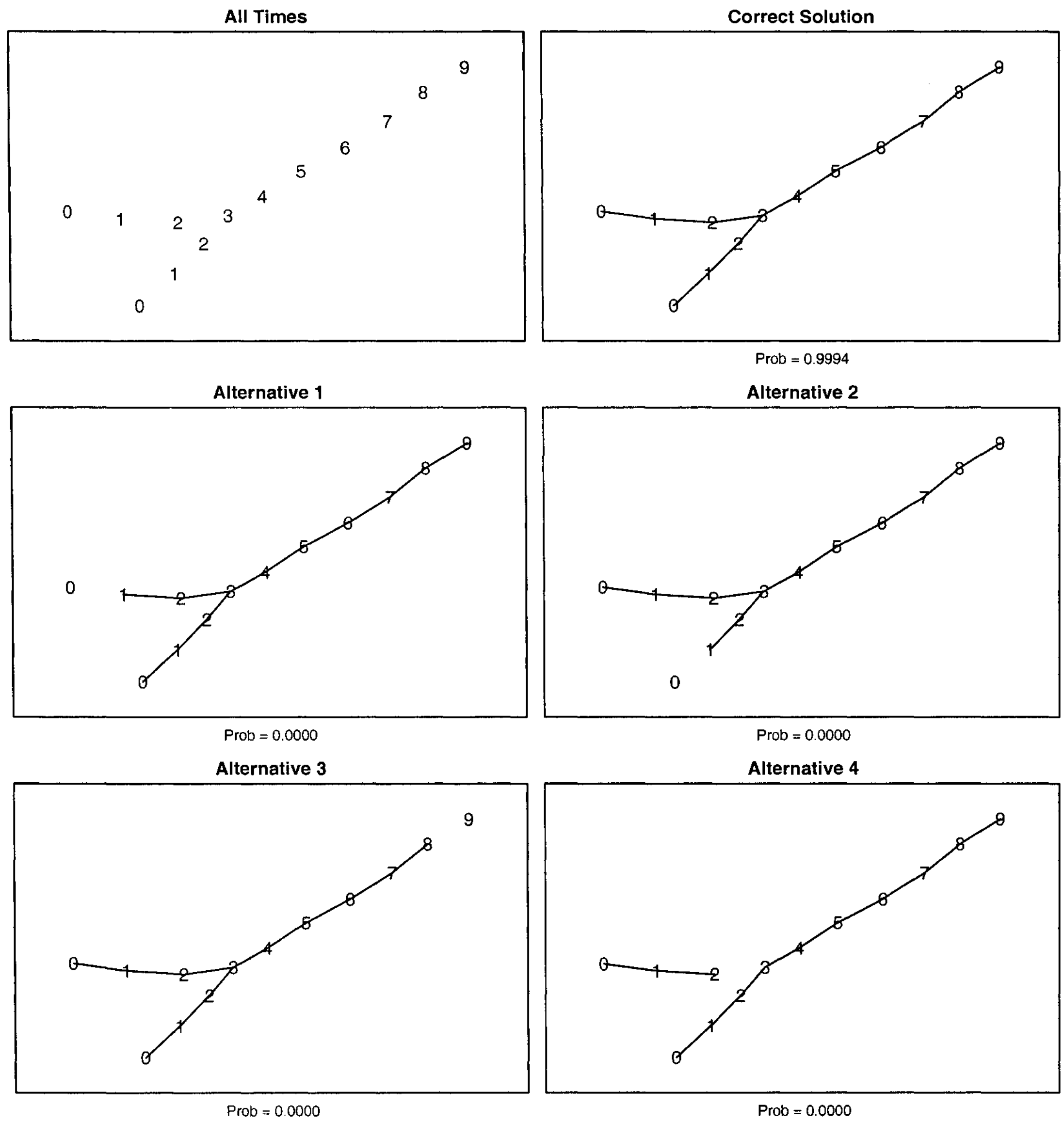

Figure 7.4: Merging Only Realization with Possible Solutions

7.1.5 Completely Random Model Realizations

These are completely unrestricted realizations from the model with state parameters set as $\lambda_0 = 4$, $\lambda_b = 0.1$, $\lambda_d = .02$, $\lambda_s = 0.06$, and $\lambda_m = .08$. The death rate was lower than the rate of splitting and merging since for the storm problem there seemed to be more splitting and merging than death. An example of one of these realizations is given in Figure 7.5

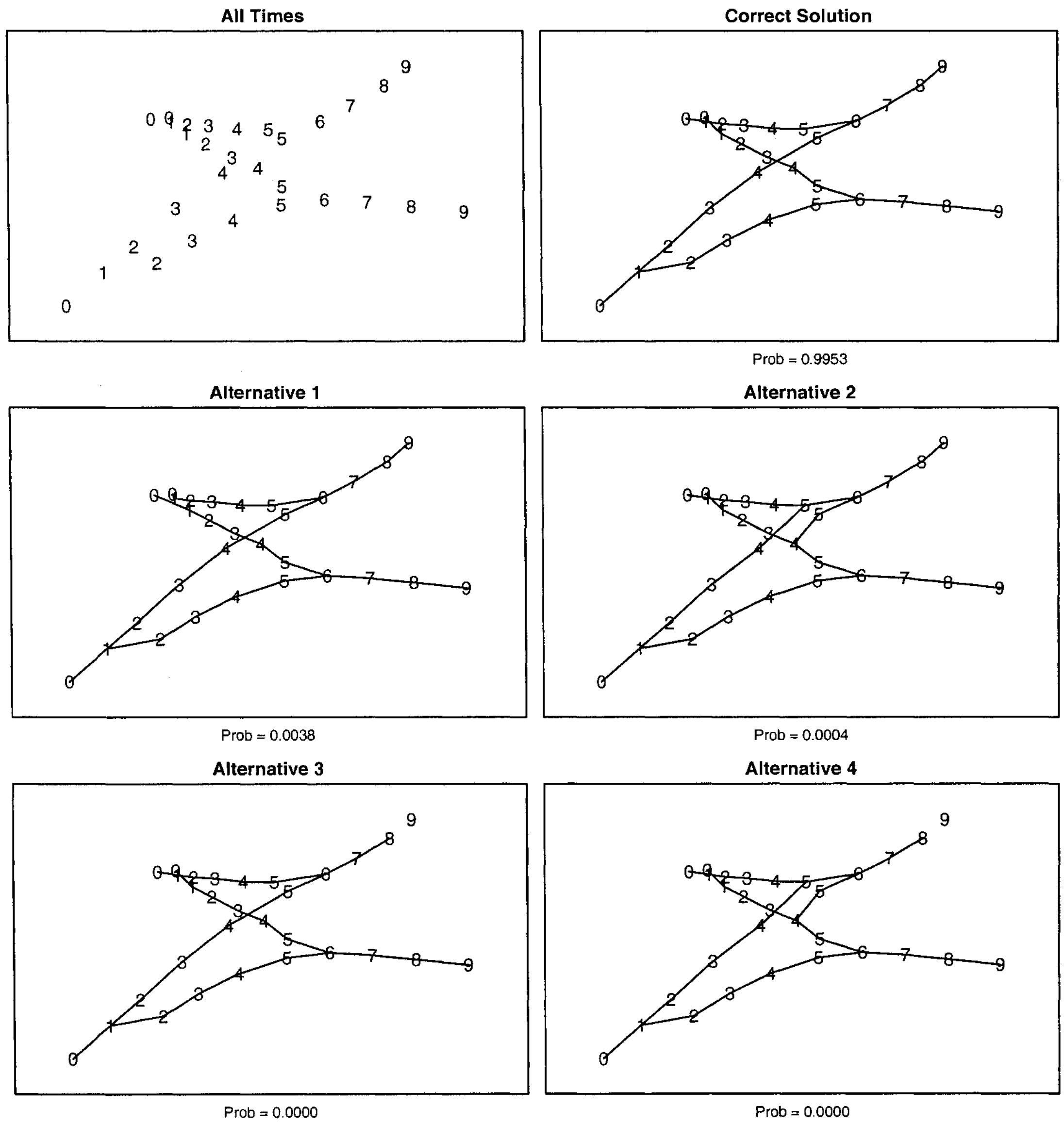

Figure 7.5: Random Model Realization with Possible Solutions

7.1.6 Completely Random Model Realizations with Size

These are the same realizations from the model in Section 7.1.5, but now with size information to be used in the tracking algorithm. You can see that there is now size information in the plots in Figure 7.6. This plot is displaying circles with the appropriate size, not ellipses,

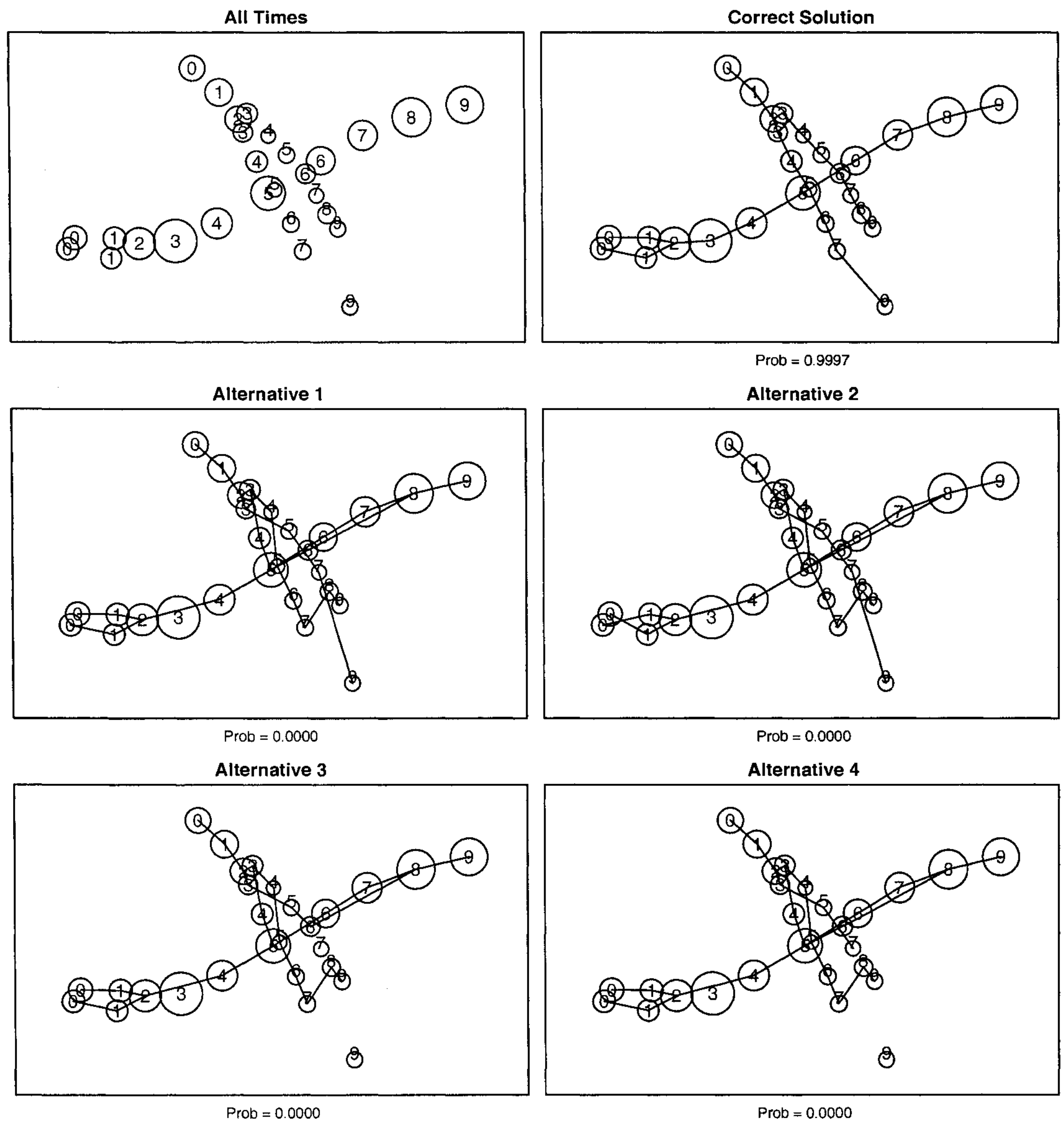

Figure 7.6: Random Model Realization with Size

so we cannot see the actual radii, but they are what is being used in the algorithm.

Sizes parameters were randomly generated for each realization. For each $(\mu_{R_{1,i}}, \mu_{R_{2,i}})$ pair we would generate two random variables $z_1 \sim \mathcal{N}(0.6, .01)$, $z_2 \sim \mathcal{N}(0.6, .01)$ and set $\mu_{R_{1,i}} = z_1 \wedge z_2$, $\mu_{R_{2,i}} = z_1 \vee z_2$. We then set the log-normal scale parameters, $\sigma^2_{R_{1,i}} = 0.001$

	Birth	Death	Split	Merge	CR	CR w/Size
% Best Est Correct	88.0	78.0	94.0	89.0	91.0	95.0
% Births Correct	94.6	NA	NA	NA	97.6	100.0
% Deaths Correct	NA	79.2	NA	NA	90.2	94.9
% Splits Correct	NA	NA	96.0	NA	97.4	97.3
% Mergers Correct	NA	NA	NA	97.0	95.1	98.7
% Falling in 95% CS	92.0	89.0	95.0	99.0	95.0	97.0
(5%)	0.000	0.000	0.156	0.222	0.062	0.596
Prob of True (25%)	0.986	0.711	0.948	0.963	0.920	0.996
(50%)	0.996	0.996	0.987	0.994	0.994	0.996
(5%)	0.787	0.684	0.665	0.671	0.668	0.966
Prob of Best Est (25%)	0.996	0.992	0.954	0.963	0.948	0.996
(50%)	0.996	0.996	0.988	0.994	0.994	0.996
(5%)	0.000	0.000	0.000	0.000	0.000	0.000
Prob of 2nd Est (25%)	0.000	0.000	0.001	0.000	0.000	0.000
(50%)	0.000	0.000	0.007	0.002	0.001	0.000
(5%)	0.852	0.720	1.000	0.882	0.954	0.996
Track Purity (25%)	1.000	1.000	1.000	1.000	1.000	1.000
(50%)	1.000	1.000	1.000	1.000	1.000	1.000

Table 7.1: Results of 100 Realizations Without Clutter

and $\sigma^2_{R_{2,i}} = 0.001$ for all i. In the parameter estimation, parameter limits for size were set for $\mu_{S,i} = \mu_{R_{1,i}} + \mu_{R_{2,i}}$ and $\sigma^2_{S,i} = \sigma^2_{R_{1,i}} + \sigma^2_{R_{2,i}}$. The parameter limits for $\mu_{S,i}$ were set to be the min and max of the observed values of the log sizes, $\mu_{S,i} \in [\min\{\log(S_{i,j})\}, \max\{\log(S_{i,j})\}]$ and $\sigma^2_{S,i} \in [0.001, 1.0]$. Also recall that $\mu_{S,i}$ is also restricted by merging and splitting so that the mean size of the parent(s) adds to the mean size of the child(ren). The radius parameters were otherwise free in the maximum likelihood estimation.

7.1.7 Results

The results of each of the simulations in Section 7.1 are given as the columns of Table 7.1. In the following we describe each of the summary statistics that make up the rows of Table 7.1.

% Best Est Correct This is the percentage of times that $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$ was equal to the correct solution $(\mathcal{U}, \mathcal{V}, \mathcal{P})$.

% Births Correct Percentage of all birth events in the simulation that were labeled correctly by the estimate, $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$.

% Deaths Correct Percentage of all death events in the simulation that were labeled correctly by the estimate.

% Splits Correct Percentage of all splitting events in the simulation that were labeled correctly by the estimate.

% Mergers Correct Percentage of all merging events in the simulation that were labeled correctly by the estimate.

% Falling in 95% CI We form a 95% confidence set of solutions for each realization. This is the percentage of times that the 95% confidence set contained the correct solution.

Prob of True This is the estimated probability of the correct solution, $(\mathcal{U}, \mathcal{V}, \mathcal{P})$, given by (5.8). These three rows are respective quantiles from the 100 realizations for these probabilities

Prob of Best Est This is the estimated probability that the estimate $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$ in (5.2) is correct given the data, again presented by the quantiles.

Prob of 2nd Est This is the estimated probability that $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})$, which is the second best estimate was correct given the data presented by the quantiles.

Track Purity These three rows are quantiles for the overall track purity for each realization. The overall track purity is defined below.

Consider a given track in the correct solution, $(\mathcal{U}, \mathcal{V}, \mathcal{P})$. Track purity for this track is defined to be the proportion of observations in that track that are labeled as part of the same track in the estimate $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$. The overall track purity is then the proportion of observations in all tracks in $(\mathcal{U}, \mathcal{V}, \mathcal{P})$ that are labeled together in tracks in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$. This is the same as the weighted average (by number of observations in the track) of individual track purities. For example if $(\mathcal{U}, \mathcal{V}, \mathcal{P})$ had two tracks; track 1 with 5 observations and track 2 with 10 observations. And the estimate, $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$, has three tracks; track 1, track 2, and track 3. Where track 1 in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$ is identical to track 1 in $(\mathcal{U}, \mathcal{V}, \mathcal{P})$. Track 2 in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$ is the first 7 observations of track 2 in $(\mathcal{U}, \mathcal{V}, \mathcal{P})$ and track 3 in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$ is the last 3 observations of track 2 in $(\mathcal{U}, \mathcal{V}, \mathcal{P})$. Then the track purity for track 1 is 1.0. The track purity for track 2 is 0.7 and the overall track purity is $[5(1.0) + 10(0.7)]/15 = 0.8$.

Refer back to the four hypotheses we posed earlier. The first hypothesis is that the rates of events labeled correctly in the first four simulations will equal the rates of events labeled correctly in the completely random (CR) model case which is the fifth simulation. We can see from Table 7.1 that the percentages of estimates that equaled the correct solution for the first four simulations is roughly equal to the percentage for the fifth simulation (the CR model realizations). Also, the percentage of births labeled correctly in the birth only simulation (94.6%) is roughly equal to that in the CR model (97.6%). The percentage of deaths labeled correctly in the death only case (79.2%) is slightly lower than that for the CR model (90.2%) which is somewhat surprising. If anything we would expect this to be the other way around, so this difference could be due to random chance. The death rate is higher in the death only simulations, which would make for shorter tracks though. So this may have something to do with this as well. The percentage of splits correct in the splitting only case is also comparable to the CR model, (96.0%) versus (97.4%). Lastly the percentage of mergers correct in the merging only case is also similar to that for the CR model, (97.0%) to (95.1%). Hence it appears that we have some strong evidence that the first hypothesis is correct.

For the second hypothesis, the percentage of births correct in the birth only case (94.6%) is a bit higher than the percentage of deaths correct in the death only case (79.2%). The standard deviation in these estimates is about 4% so it is likely that these rates are not the same. One explanation for this may be the problem mentioned with respect to Figure 7.2. This problem is for deaths that occur near the end of the time window. These will sometimes be labeled as missing observations for a time or two at the end instead of a death, since the likelihood doesn't find it beneficial to kill the target until it has more missing observations. Another explanation was hinted at in the previous paragraph, that a high death rate leads to short tracks. The algorithm, might then misclassify these short tracks as clutter.

For the third hypothesis, the percentage of splits correct in the splitting only case (96.0%) is very similar to the percentage of mergers correct in the merging only case (97.0%). This is fairly strong evidence that third hypothesis is correct.

Our last hypothesis states that the size information will improve the results. This is somewhat evident from the number of correct estimates increasing from (91.0%) in the CR model without size to (95.0%) in the CR model with size which is not a large increase, but there was not too much room for improvement. In any case, the standard error of these rates is about 3% so this is not conclusive evidence that size information helps. However, if we look at the probability of the correct solution given the data we see that the fifth percentile for the without size case is 0.062 while that for the with size case is 0.596. Hence for the extreme cases, it appears that the size information is quite helpful.

It is also worth noting that the 95% confidence sets for these simulations had fairly accurate coverage at 92.0%, 89.0%, 95.0%, 99.0%, 95.0%, and 97.0% for the six simulations respectively. The track purity values were also very high for each case.

7.2 Simulations with Clutter

This set of simulations allows false alarms to appear at each time with rate $\lambda_f = 0.007$. The limits of the field of vision are $(-120, -85) \times (25, 50)$, so this corresponds to an area, $A = 1125$. Hence, $\lambda_f A = 7.9$, so we can expect about 8 false alarms at each time. We repeated the same six simulations from Section 7.1 again with $N = 100$ realizations each so as to compare results. We also wish to investigate the same hypotheses 1 - 4, posed in the previous section now with the presence of clutter.

An example of a realization from the CR model now with clutter is given in Figures 7.7 and 7.8. It is informative here to examine Figure 7.7 which gives the plot of the observations at each time step. This way we can get a feel for how many false alarms there really are at each time. Figure 7.8 makes it seem like there is more clutter than there actually is. In any case, it is still very difficult to discern any sort of pattern from the plots at each time. We would be better off using the "All Times" plot in Figure 7.8, but it is still hard to see the true solution here as well.

7.2.1 Results

The results of each of the simulations in Section 7.2 are given as the columns of Table 7.2. Here there are a few additional summary statistics that result from the addition of false alarms. These are described below.

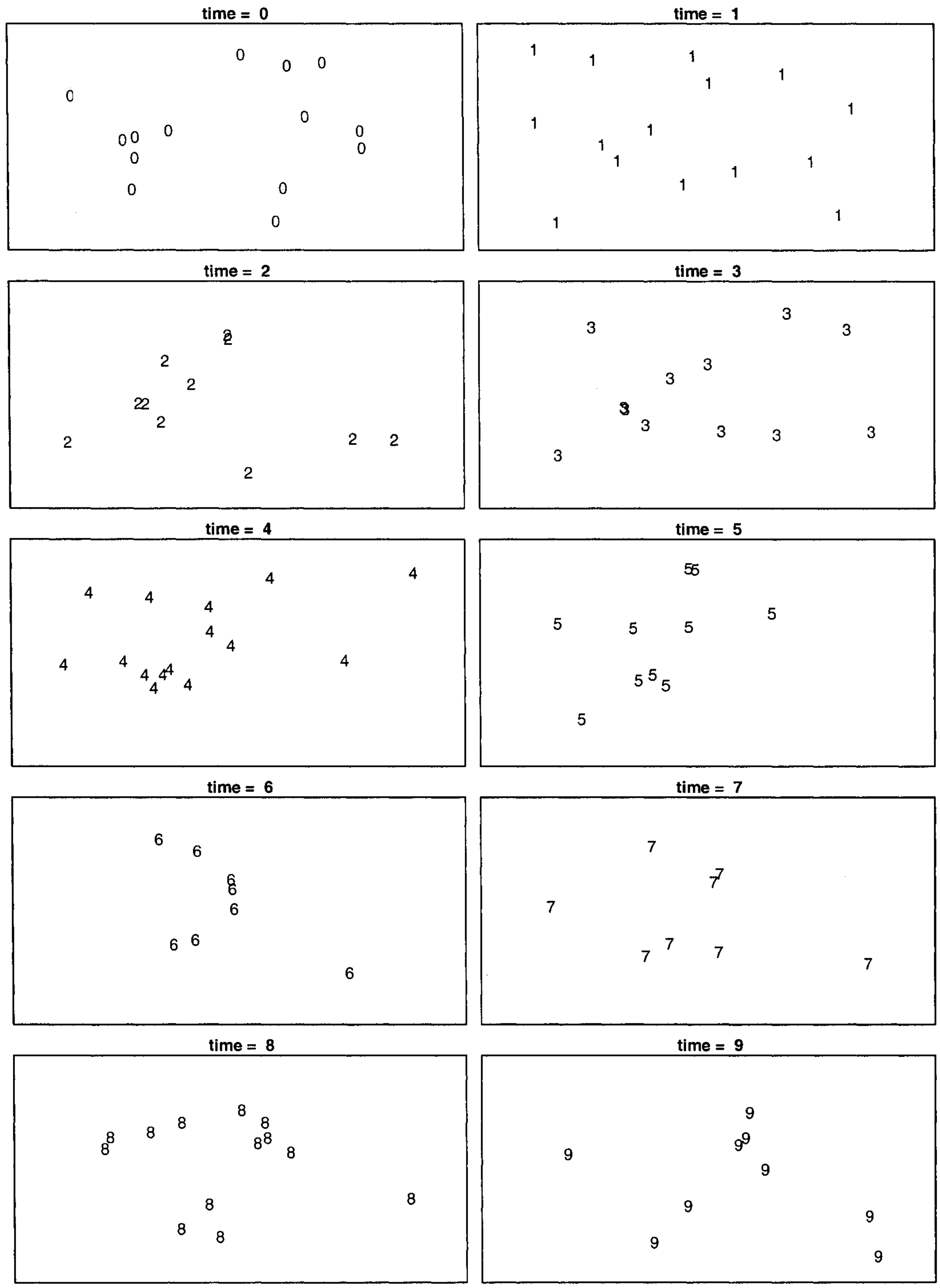

Figure 7.7: Observations at Each Time of a CR Model Realization with Clutter

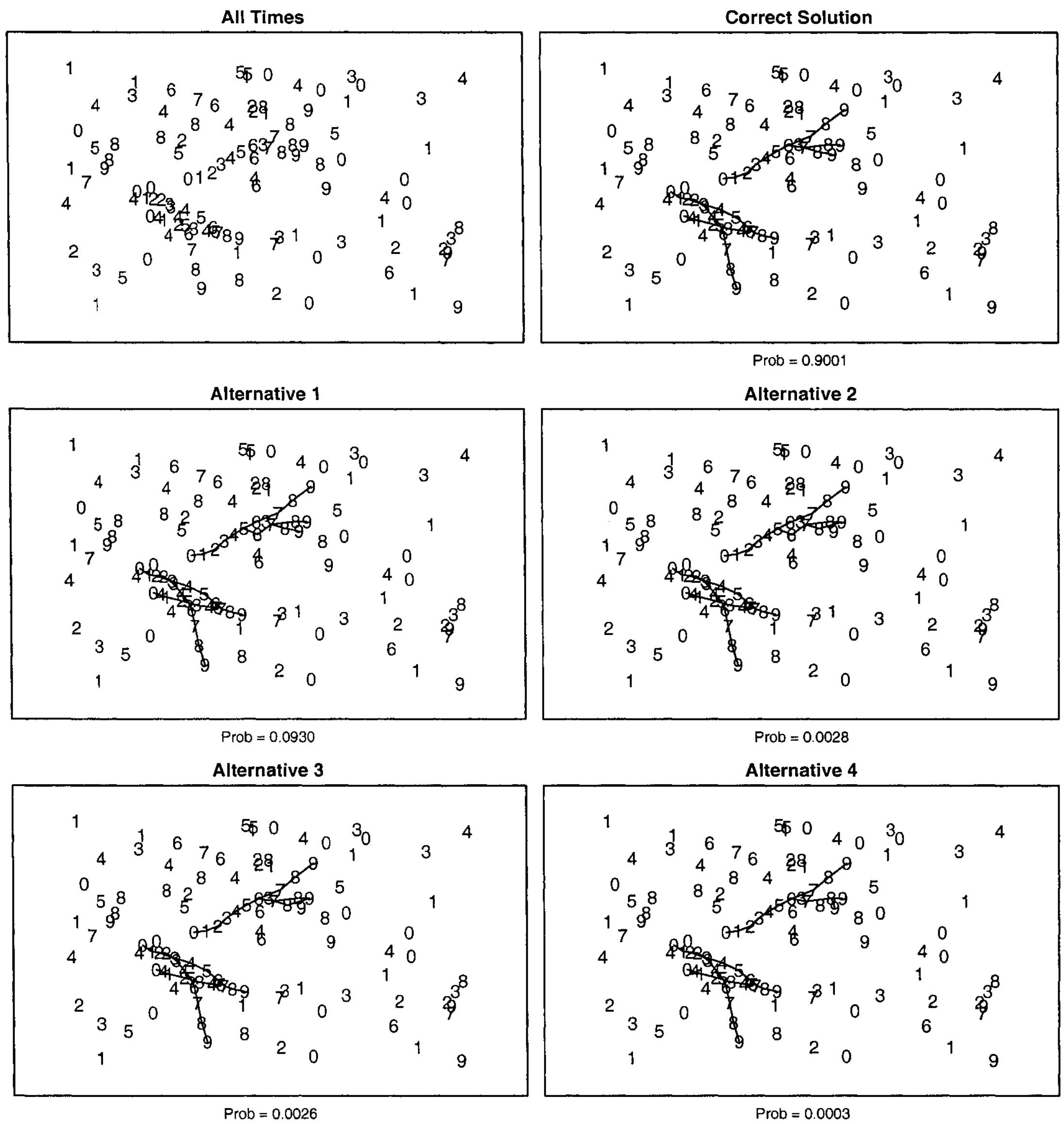

Figure 7.8: Birth Only Realization with Clutter

% Targets Correct This is the percentage of all targets in the simulation that were labeled correctly as targets by the estimate $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$.

% FAs Correct Percentage of all false alarms in the simulation that were labeled correctly by the estimate.

	Birth	Death	Split	Merge	CR	CR w/Size
% Best Est Correct	72.0	60.0	61.0	82.0	67.0	92.0
% Births Correct	94.8	NA	NA	NA	83.1	100.0
% Deaths Correct	NA	79.2	NA	NA	70.2	95.7
% Splits Correct	NA	NA	93.0	NA	87.0	100.0
% Mergers Correct	NA	NA	NA	97.0	90.7	98.7
% Targets Correct	99.3	99.0	98.9	99.9	99.0	99.8
% FAs Correct	97.4	95.2	97.4	99.5	99.2	99.6
% Falling in 95% CS	94.0	79.0	86.0	98.0	81.0	96.0
(5%)	0.015	0.000	0.001	0.059	0.000	0.068
Prob of True (25%)	0.292	0.090	0.074	0.598	0.207	0.996
(50%)	0.850	0.685	0.746	0.947	0.996	0.996
(5%)	0.315	0.209	0.294	0.363	0.265	0.790
Prob of Best Est (25%)	0.614	0.508	0.583	0.695	0.582	0.996
(50%)	0.906	0.877	0.891	0.947	0.996	0.996
(5%)	0.000	0.000	0.000	0.000	0.000	0.000
Prob of 2nd Est (25%)	0.000	0.000	0.000	0.000	0.000	0.000
(50%)	0.051	0.076	0.046	0.031	0.000	0.000
(5%)	0.881	0.610	0.831	0.938	0.856	1.000
Track Purity (25%)	1.000	1.000	1.000	1.000	0.956	1.000
(50%)	1.000	1.000	1.000	1.000	1.000	1.000
(5%)	0.903	1.000	0.997	1.000	0.869	1.000
Prob of Target (25%)	1.000	1.000	1.000	1.000	1.000	1.000
(50%)	1.000	1.000	1.000	1.000	1.000	1.000
(5%)	0.922	0.705	0.897	0.922	1.000	1.000
Prob of FA (25%)	1.000	1.000	1.000	1.000	1.000	1.000
(50%)	1.000	1.000	1.000	1.000	1.000	1.000

Table 7.2: Results of 100 Realizations With Clutter

Prob of Target This is the probability given the data that a given target at the last time step should be labeled a target. The three rows are the quantiles of these probabilities over all of the targets in the last time step in all of the realizations.

Prob of FA This is the same as "Prob of Target" only for false alarms.

We will now reconsider the four hypotheses we posed in Section 7.1 but this time with the presence of clutter.

Recall that the first hypothesis states that the first four simpler simulations will translate their error rates to the more complicated CR model case. From Table 7.2 we can see that average of the percentages of estimates that were equal to the correct solution for the first four simulations, $(72.0\% + 60.0\% + 61.0\% + 82.0\%)/4 = 68.8\%$, is roughly equal to the percentage of correct estimates for the CR model realizations (67.0%). Also, the percentage of births labeled correctly in the birth only simulation (94.8%) is somewhat higher but comparable to that in the CR model (83.1%). The percentage of deaths labeled correctly in the death only case (79.2%) is also comparable to that for the CR model (70.2%). The percentage of splits correct in the splitting only case is close to the CR model, (93.0%) versus (87.0%). Lastly the percentage of mergers correct in the merging only case is also similar to that for the CR model, (97.0%) to (90.7%). Hence, it once again seems that we have some evidence that the first hypothesis is correct even with the addition clutter.

For the second hypothesis, the percentage of births correct in the birth only case (94.8%) is again a bit higher than the percentage of deaths correct in the death only case (79.2%). This is likely due to the same problems mentioned for the case without clutter. That is deaths near the end of the time window and high death rate resulting in shorter tracks. In fact, the fifth percentile for track purity in the death only case is only 0.610 here which leads us to believe there are a few instances where the algorithm decided to label a short track as clutter instead of paying the price for a birth and a death.

For the third hypothesis, the percentage of splits correct in the splitting only case (93.0%) is again quite similar to the percentage of mergers correct in the merging only case (97.0%). So again there is some good evidence that third hypothesis is correct.

Recall that the last hypothesis says that the size information will improve the results. This was not abundantly clear in the simulations without clutter. However, in the presence of clutter, the size information adds quite a bit of discernment power. The percentage of correct estimates jumps from 67.0% for the CR model without size to 92.0% for the CR model with size. Also if we look at the probability given the data that the correct solution is correct we see these are substantially higher when we include size. Lastly, the coverage of the 95% confidence sets is significantly improved from 81.0%, to 96.0% when we use size in the algorithm.

The coverage of the 95% confidence sets for these simulations 94.0%, 79.0%, 86.0%, 98.0%, 81.0%, and 96.0% for the six simulations dropped off some from the simulations without clutter. One explanation for this, referring back to (5.8), is that these sets assume that that the correct answer is in the collection of solutions we obtained from the MHT algorithm. If it is not always in this collection, which it is not, then of course our distribution given of solutions given in (5.8) will not be correct. Also, since we estimate parameters for each of the possible solutions, this also introduces some bias. Overall though, these confidence sets and probabilities give a very good general idea how confident we are in the estimated solution(s).

Notice that although the estimate is not always the correct solution for these simulations, the track purity values are always high. Only 5% of track purities for any of the cases was below 0.88 with the exception of the death only simulation which had 5% below 0.610. The percentages of Targets correct and false alarms correct were also uniformly high. These were usually around 99% for most cases and never lower than 95.2% for any of the simulations.

	CR, Δt=1.0	CR, Δt=0.5	CR, Δt=0.1
% Best Est Correct	67.0	79.0	99.0
% Births Correct	83.1	88.2	100.0
% Deaths Correct	70.2	91.1	98.2
% Splits Correct	87.0	95.7	100.0
% Mergers Correct	90.7	97.7	100.0
% Targets Correct	99.0	99.7	100.0
% FAs Correct	99.2	99.9	100.0
% Falling in 95% CS	81.0	90.0	100.0
(5%)	0.000	0.006	0.834
Prob of True (25%)	0.207	0.622	0.994
(50%)	0.996	0.970	0.996
(5%)	0.265	0.499	0.836
Prob of Best Est (25%)	0.582	0.803	0.994
(50%)	0.996	0.974	0.996
(5%)	0.000	0.000	0.000
Prob of 2nd Est (25%)	0.000	0.000	0.000
(50%)	0.000	0.014	0.000
(5%)	0.856	0.940	1.000
Track Purity (25%)	0.956	1.000	1.000
(50%)	1.000	1.000	1.000
(5%)	0.869	0.999	1.000
Prob of Target (25%)	1.000	1.000	1.000
(50%)	1.000	1.000	1.000
(5%)	1.000	1.000	1.000
Prob of FA (25%)	1.000	1.000	1.000
(50%)	1.000	1.000	1.000

Table 7.3: Results of 100 Realizations With Clutter

7.3 Decreasing Time Increments

This set of simulations uses a model identical to that of the CR model realizations with clutter of Section 7.2. Here however, we use three different time increments, $\Delta t = 1.0$, $\Delta t = 0.5$, and $\Delta t = 0.1$. The conjecture here is that even with false alarms, missing observations, and a likelihood based on an integrated Brownian Motion model, there is still a convergence of the estimate to the correct solution as in Chapter 6. The results of these simulations can be found in Table 7.3.

From Table 7.3 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 67.0% for the $\Delta t = 1.0$ case, to 79.0% for the $\Delta t = 0.5$ case, to 99.0% for the $\Delta t = 0.1$ case. Also for the probability of the correct solution given the data, 25% of the $\Delta t = 1.0$ probabilities are less than 0.207, but only 5% of the $\Delta t = 0.1$ probabilities are less than 0.834. It is very likely that there is a convergence similar to that in Chapter 6 happening for this situation as well.

Chapter 8

Application to Rainfall Data

In this chapter we consider tracking the storms that evolve during the morning of July 14, 1996 in the images of Figure 1.1. Here we will make use of all the attribute data that we have available as well as location. These would include size (radii), orientation, and intensity. For the purposes of this problem, we are only interested in tracking the mesoscale convective systems, which we will conveniently define to be any storm with major axis longer than 100 km. This corresponds to roughly 1^o latitude or longitude in Figure 1.1. Before, we can track the storms though, we must first explain how to identify them from the images and measure their location and attributes. This is done next.

8.1 Detection Algorithm

The problem of target or object identification in images has been studied quite thoroughly. It is not the goal of this thesis to make a contribution in this area, hence a detailed description of these techniques will not be given. We will however describe the details of the particular identification technique we chose to use on the storm tracking problem. For a good summary of other imaging techniques, see [51].

Recall, the goal of the detection algorithm is to go through each image and record the location of each target, in this case storm, that it finds. In our case, we will record the size and orientation of the storms as well.

An image consists of intensity values $I_{i,j}$ for each of the pixels. We start by thresholding the intensities at a value α. At this point, all pixels with intensities $I_{i,j} < \alpha$ get set to zero. We then consider all of the pixels with $I_{i,j} > \alpha$ and we wish to group these pixels together to make up the targets.

Simply stated, all pixels with $I_{i,j} > \alpha$ that are connected to each other are part of the same target. We just need to define what it means for two pixels to be connected. There are two definitions that are allowed here. Two pixels are 4-connected if the share one of their 4 sides with each other. Two pixels are 8-connected if the share a common side or corner. We have found that the 4-connected definition works well for the storms problem, but certainly the best one to use is problem dependent.

We now have a collection of targets, defined by their corresponding cluster of pixels. To specify location, size and orientation of the targets, we will fit an ellipse to each target (cluster of pixels). This will be accomplished by estimating a bivariate Gaussian distribution for each target and using the 99% contour of the density.

The mean and covariance of the Gaussian distribution used to fit an ellipse to a given target are given by the following. Suppose $x_{i,j}$, $y_{i,j}$, are the coordinates of the center of pixel i, j. The moments, $\hat{\mu}_x$, $\hat{\mu}_y$, $\hat{\sigma}_x$, $\hat{\sigma}_y$, and $\hat{\sigma}_{xy}$ for a given target are given by the MLE's under the assumption that the locations of the pixels, $(x_{i,j}, y_{i,j})$, that make up that target, are *iid* observations from a bivariate normal distribution.

The location of the target is then given by $(\hat{\mu}_x, \hat{\mu}_y)$. The length of the radii of the minor and major axes, R_1 and R_2, of the target are given by those of the 99% contour ellipse. Similarly for the the angle of orientation of the major axis Q_2. Refer back to Figure 2.2 for an illustration of this. The intensity I for a storm is defined here to be the average of the 10 highest pixel intensities $I_{i,j}$ which make up the storm.

For this application, the pixel intensities ranged anywhere from 0.00 to 150.00 mm/hour of rainfall which roughly equates to 0.0 to 6.0 inches of rain per hour. Most pixels that made up storms had intensities between 1.00 and 10.00 mm/hour. We used a threshold of $\alpha = 0.10$ with the 4-connected definition. In addition, we are only considering mesoscale systems here, which are storms with $R_2 > 1°$. All other storms are discarded, so this could be considered a second stage of thresholding.

8.2 Results

Here we give some results of the application of the tracking algorithm to this problem. Recall that the images in Figure 1.1 are separated by 30 minutes. So this small illustration covers a time span of 2.5 hours. Figure 8.1 shows the images after the processing of the detection algorithm. It is the ellipses in these images that we will actually feed into the tracking algorithm.

Figure 8.2 shows the results of applying the tracking algorithm to the images in Figure 8.1. Here we used the size, orientation, and intensity attributes along with the location in the tracking algorithm. The parameter limits were set to the same as those in the simulations of Chapter 7. The tracks that the storms took given by the tracking algorithm are drawn in red. When there is a merger or a split, the red track line connects the parents

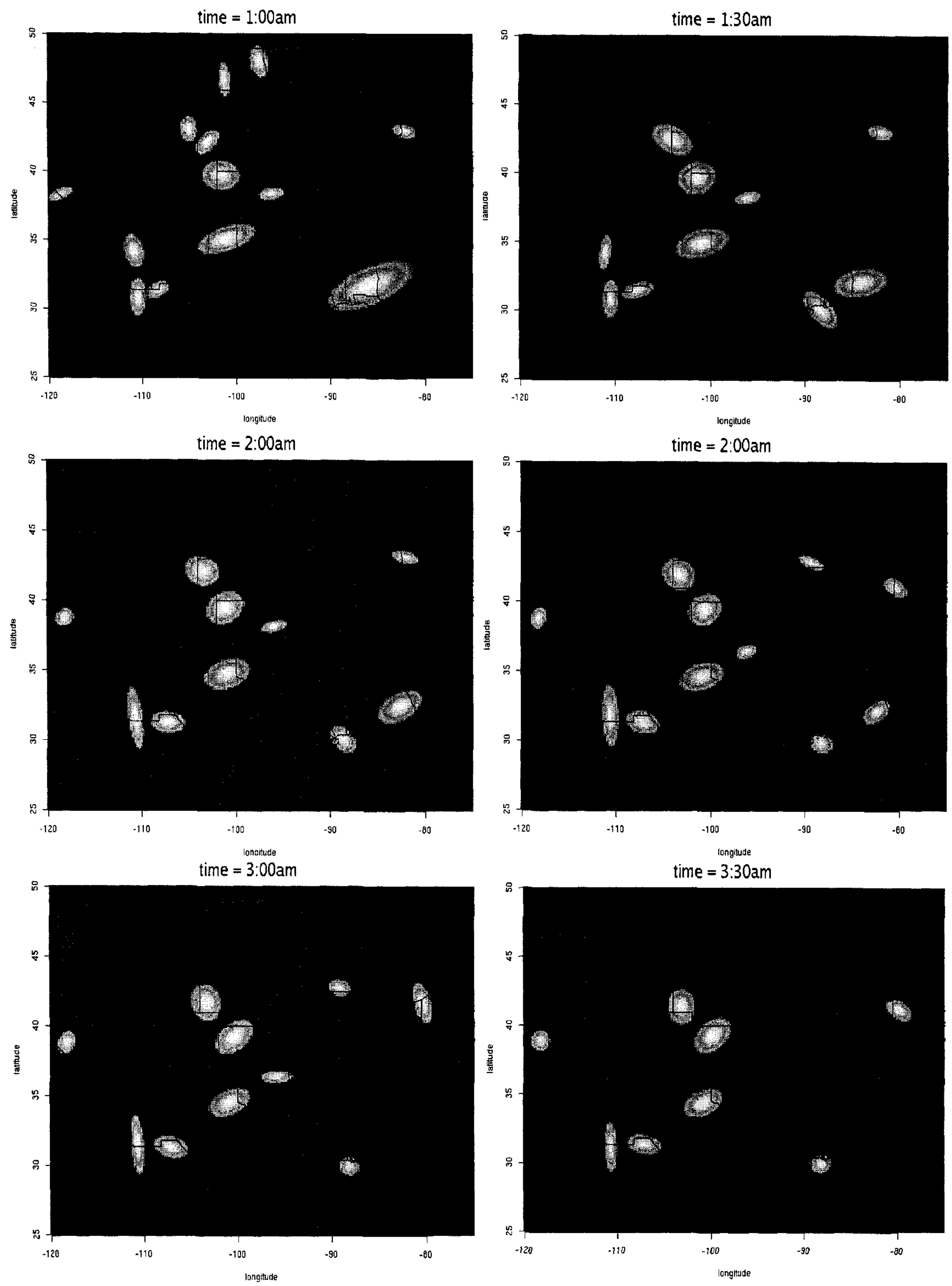

Figure 8.1: Best Fitting Ellipses of Radar Reflectivity Images

to the children. We can see an example of this between 1:00am and 1:30am where the two storms in the south west corner of South Dakota merge together into one storm. Also in this same time interval, the large system over Alabama and the panhandle of Florida splits apart into two smaller systems. The reader is also referred to the following website, http://www.stat.colostate.edu/~storlie/ncar/, to see a video of the raw data, the processed data, and the solution with the track lines. The corresponding links from the above web-page are "Rain Fall Video", "Cleaned Rainfall Images", and "Paths Given by Tracking Algorithm". These videos cover a longer time span as well, from 1:00am to 1:00pm on July 14, 1996.

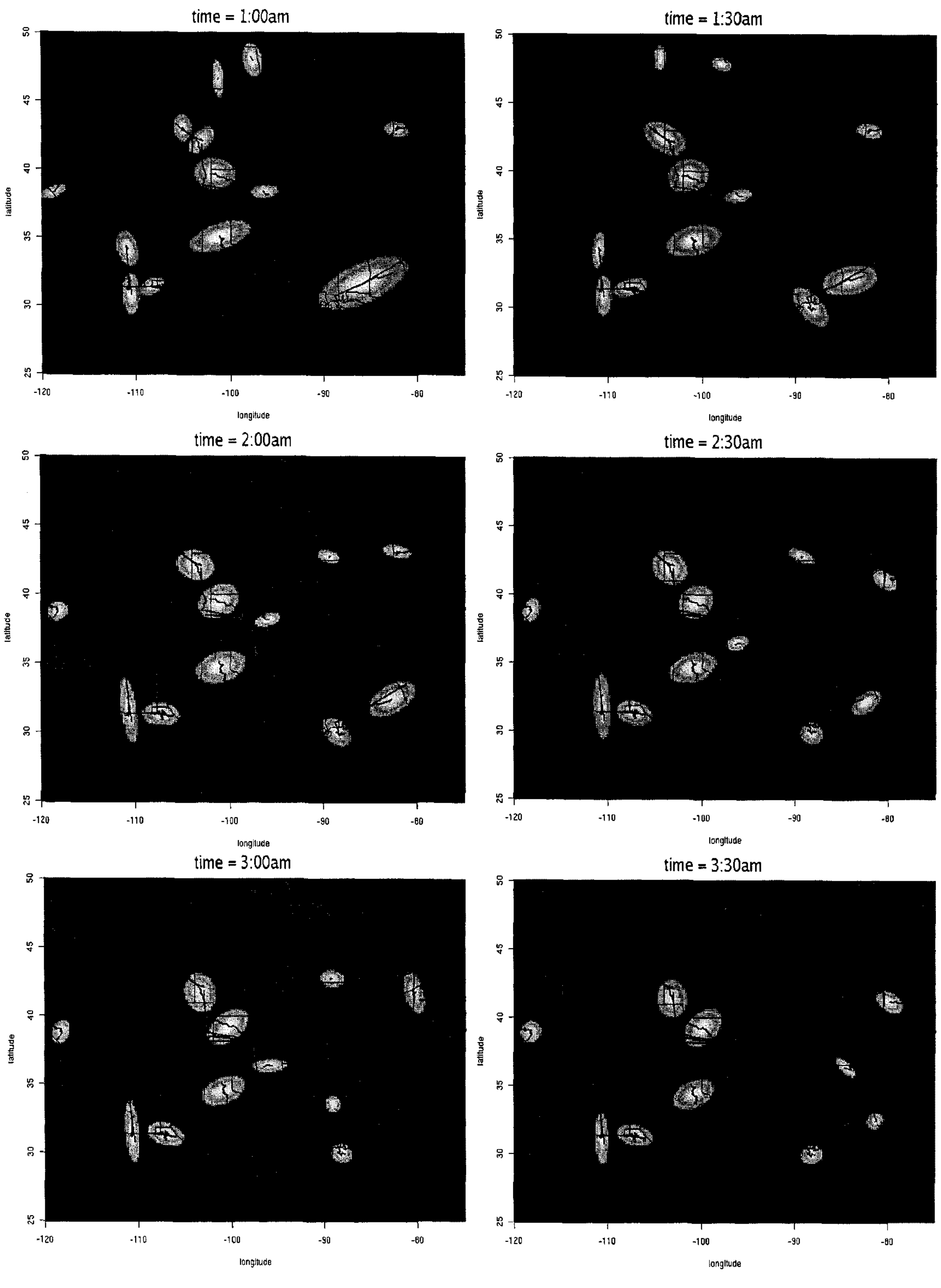

Figure 8.2: Best Fitting Ellipses with Tracks

Chapter 9

Conclusions & Further Work

In this thesis we have presented a novel approach to multiple target tracking which allows for birth, death, splitting and merging of targets. The model we proposed incorporated the events of birth, death, splitting, and merging of targets into the likelihood, as well as missed detections and false alarms. Track estimation was then accomplished purely by considering the distribution of the relevant variables given the data. There was no need for hypothesis testing to initiate and terminate tracks as an intermediate step. We also addressed the problem of how to quantify the confidence in the results of a tracking estimate.

In addition, we gave some theoretical justification for this approach to multiple target tracking in terms of sufficient conditions under which the estimate will converge to the correct solution. In doing so we realized the importance of using a robust estimate of the variance component for the random process $G_i(t)$. The results obtained by the simulated data also gave strong evidence that this estimation procedure works very well even in the presence of clutter and missing observations from targets.

Finally the utility of the method was demonstrated on radar reflectivity data collected over the United States. The algorithm successfully recovered the storm tracks along with the merging and splitting events.

In the future we would like to address many other concerns related to multiple target tracking estimation which were not investigated here. These are summarized below.

- Theoretical Results
  - Allow for missing observations and clutter in the Theorems of Chapter 6.
  - Examine the properties of assuming an integrated Brownian Motion model.
- Develop a procedure for sample size (or Δt) determination for tracking problems.
- Improvement to the Optimization Algorithm
  - Develop a way to approximate the conditional distribution of a new observation by updating, similar to the Kalman Filter, for this model.
  - Possibly incorporate a track orientated approach or a method similar to Cox's implementation to efficiently find the N best solutions.
- Application to Storms
  - Improve the identification procedure to more intelligently identify storm objects.
  - Recover the tracks for several months worth of real data and compare the results to those of regional climate models.
- Apply the method to the 2 dimensional turbulence simulations of Section 1.1.2.

Appendix A

Mean and Covariance Calculations

Here we calculate the mean functions $\mathrm{E}X_i(t)$, $\mathrm{E}D_i$ and the covariance functions $\mathrm{Cov}(X_i(s), X_j(t))$, $\mathrm{Cov}(X_i(s), D_j)$, and $\mathrm{Cov}(D_i, D_j)$. Recall in Section 3.1.3 that these are calculated before conditioning on any merging events.

As in Section 4.3, let

$$
\begin{aligned}
B &= \{i : \text{target } i \text{ is an initial target or a birth}\} \\
S &= \{i : \text{target } i \text{ is the result of a split}\} \\
M &= \{i : \text{target } i \text{ is the result of a merger}\}
\end{aligned}
$$

Also let $n(B)$, $n(S)$, and $n(M)$ be the number elements in these sets respectively. The location equations for a target resulting from birth, split, and merger from (2.3), (2.7), and

(2.5) are given here again for convenient reference as

$$X_i(t) = \begin{cases} X_i(\xi_i) + X_i'(\xi_i)(t-\xi_i) + \sigma_i G_i(t-\xi_i) & \text{for } i \in B \\ X_{p_{i,1}}(\xi_i) + \psi_{s,i} + \left[X'_{p_{i,1}}(\xi_i) + \psi'_{s,i}\right](t-\xi_i) + \sigma_i G_i(t-\xi_i) & \text{for } i \in S \\ \frac{1}{2}\left(X_{p_{i,1}}(\xi_i) + X_{p_{i,2}}(\xi_i)\right) + \psi_{m,i} + & \text{for } i \in M \\ \left[\frac{1}{2}\left(X'_{p_{i,1}}(\xi_i) + X'_{p_{i,2}}(\xi_i)\right) + \psi'_{m,i}\right](t-\xi_i) + \sigma_i G_i(t-\xi_i) & \end{cases}$$

where we are assuming that $G_i(t)$ is an IBM. Also recall that we actually observe $X_i^*(t_j) = X_i(t_j) + \varepsilon_j$ for each time point t_j. We give the target velocities for the three cases as well,

$$X_i'(t) = \begin{cases} X_i'(\xi_i) + \sigma_i B_i(t-\xi_i) & \text{for } i \in B \\ X'_{p_{i,1}}(\xi_i) + \psi'_{s,i} + \sigma_i B_i(t-\xi_i) & \text{for } i \in S \\ \frac{1}{2}\left(X'_{p_{i,1}}(\xi_i) + X'_{p_{i,2}}(\xi_i)\right) + \psi'_{m,i} + \sigma_i B_i(t-\xi_i) & \text{for } i \in M. \end{cases}$$

Lastly, the variable D_i for $i = 1, \ldots, n(M)$ from (3.25) where $n(M)$ is the number of mergers is given as

$$D_i = X_{d_{i,1}}(\xi_{d_{i,3}}) - X_{d_{i,2}}(\xi_{d_{i,3}}) + \psi_{d,i} \tag{A.1}$$

We will use the following notation to denote the means and covariance of path locations and velocities

$$\mu_i(t) = \mathrm{E}X_i(t) \tag{A.2}$$

$$\mu_i'(t) = \mathrm{E}X_i'(t) \tag{A.3}$$

$$\gamma_{i,j}^*(s,t) = \mathrm{Cov}(X_i^*(s), X_i^*(t)) \tag{A.4}$$

$$\gamma_{i,j}(s,t) = \mathrm{Cov}(X_i(s), X_j(t)) \tag{A.5}$$

$$\gamma_{i,j}'(s,t) = \mathrm{Cov}(X_i(s), X_j'(t)) \tag{A.6}$$

$$\gamma_{i,j}''(s,t) = \mathrm{Cov}(X_i'(s), X_j'(t)) \tag{A.7}$$

$$\gamma_i(s,t) = \mathrm{Cov}(X_i(s), X_i(t)) \tag{A.8}$$

$$\gamma_i'(s,t) = \mathrm{Cov}(X_i(s), X_i'(t)) \tag{A.9}$$

$$\gamma_i''(s,t) = \mathrm{Cov}(X_i'(s), X_i'(t)) \tag{A.10}$$

Note that for the purposes of likelihood calculation, we are only interested in the functions given in (A.2) and (A.4) above. However, the expressions for these two functions will depend on the others, so in the following, we will need to write out expressions for all of these functions.

A.1 Mean Functions

We can write out the mean functions for location for the three cases of birth, split, and merger as a recursive formula,

$$\mu_i(t) = \begin{cases} \mu_{X_0} + (t-\xi_i)\mu_{X_0'} & \text{if } i \in B \\ \mu_{p_{i,1}}(\xi_i) + (t-\xi_i)\mu'_{p_{i,1}}(\xi_i) & \text{if } i \in S \\ \frac{1}{2}\mu_{p_{i,1}}(\xi_i) + \frac{1}{2}\mu_{p_{i,2}}(\xi_i) + \frac{t-\xi_i}{2}\left(\mu'_{p_{i,1}}(\xi_i) + \mu'_{p_{i,2}}(\xi_i)\right) & \text{if } i \in M. \end{cases} \tag{A.11}$$

Eventually this recursion will lead back to a parent target which is an initial target or a birth, at which point the recursion will terminate. We can also write out the mean velocities for the three cases in a similar manner,

$$\mu_i(t) = \begin{cases} \mu_{X_0'} & \text{if } i \in B \\ \mu'_{p_{i,1}}(\xi_i) & \text{if } i \in S \\ \frac{1}{2}\left(\mu'_{p_{i,1}}(\xi_i) + \mu'_{p_{i,2}}(\xi_i)\right) & \text{if } i \in M. \end{cases} \tag{A.12}$$

Of course the mean of D_i can be written as

$$\mathrm{E}D_i = \mu_{d_{i,1}}(\xi_{d_{i,3}}) - \mu_{d_{i,2}}(\xi_{d_{i,3}}). \tag{A.13}$$

A.2 Covariance Functions

Now we will consider the calculation of the covariance functions. First realize that

$$\gamma^*_{i,j}(s,t) = \gamma_{i,j}(s,t) + \sigma^2_{X_e} I_{\{i=j\}} I_{\{s=t\}}. \tag{A.14}$$

Also for $\mathrm{Cov}(X_i(s), D_j)$, and $\mathrm{Cov}(D_i, D_j)$ we have

$$\begin{aligned}
\mathrm{Cov}(X_i(s), D_j) &= \gamma_{i,d_{j,1}}(s, \xi_{d_{j,3}}) - \gamma_{i,d_{j,2}}(s, \xi_{d_{j,3}}) && \text{(A.15)}\\
\mathrm{Cov}(D_i, D_j) &= \gamma_{d_{i,1},d_{j,1}}(\xi_{d_{1,3}}, \xi_{d_{j,3}}) - \gamma_{d_{i,1},d_{j,2}}(\xi_{d_{1,3}}, \xi_{d_{j,3}}) - \gamma_{d_{i,2},d_{j,1}}(\xi_{d_{1,3}}, \xi_{d_{j,3}}) + \\
&\quad \gamma_{d_{i,2},d_{j,2}}(\xi_{d_{1,3}}, \xi_{d_{j,3}}) && \text{(A.16)}
\end{aligned}$$

so we need just to write out an expression for $\gamma_{i,j}(s,t)$. This will require the following definition. Let two paths i and j be connected if one is a byproduct of a splits and/or mergers of the other. That is they are relatives in the sense that there is a way to trace back the genealogy from one to the other in the family tree. Define the indicator $\delta_{i,j}$ to be

$$\delta_{i,j} = \begin{cases} 1 & \text{if path } i \text{ is connected to path } j \\ 0 & \text{otherwise} \end{cases} \tag{A.17}$$

It is clear that $\gamma_{i,j}(s,t) = 0$, whenever $\delta_{i,j} = 0$, since paths are independent unless they are connected. Consider now calculating the covariance function $\gamma_{i,j}(s,t)$ when $\delta_{i,j} = 1$, $i < j$ and $j \in S$.

$$\begin{aligned}
\gamma_{i,j}(s,t) &= \mathrm{Cov}\Big(X_i(s)\ ,\ X_{p_{j,1}}(\xi_j) + \psi_{s,j} + \left[X'_{p_{j,1}}(\xi_j) + \psi'_{s,j}\right](t - \xi_j) + \sigma_j G_j(t - \xi_j)\Big) \\
&= \gamma_{i,p_{j,1}}(s, \xi_j) + (t - \xi_j)\gamma'_{i,p_{j,1}}(s, \xi_j). && \text{(A.18)}
\end{aligned}$$

Using this same idea, we can calculate the case for $\delta_{i,j} = 1$, $i < j$ and $j \in M$ as well. If $i < j$ and $j \in B$ then necessarily $\delta_{i,j} = 0$. This is true because if $i < j$ and $j \in B$, then be the way we have organized the labels, $\xi_i \leq \xi_j$, and if target j is a birth, there is no way they can be connected. Since we always break down the larger index into the contribution from its parents, we will eventually get to the covariance of a parent(s) that is a birth or initial target and the recursion will terminate.

Hence we have

$$\gamma_{i,j}(s,t) = \begin{cases} \gamma_i(s,t) & \text{if } i = j \\ \gamma_{i,p_{j,1}}(s,\xi_i) + (t-\xi_i)\gamma'_{i,p_{j,1}}(s,\xi_i) & \text{if } \delta_{i,j}=1,\ i<j,\ j\in S \\ \frac{1}{2}\left(\gamma_{i,p_{j,1}}(s,\xi_i) + \gamma_{i,p_{j,2}}(s,\xi_i)\right) + & \\ \frac{t-\xi_i}{2}\left(\gamma'_{i,p_{j,1}}(s,\xi_i) + \gamma'_{i,p_{j,2}}(s,\xi_i)\right) & \text{if } \delta_{i,j}=1,\ i<j,\ j\in M \\ \gamma_{j,i}(t,s) & \text{if } \delta_{i,j}=1,\ i>j \\ 0 & \text{otherwise.} \end{cases} \tag{A.19}$$

We can also calculate $\gamma'_{i,j}(s,t)$ in the same manner as in (A.18). Although, now we cannot use the symmetry of the function if $i > j$. So consider calculating $\gamma'_{i,j}(s,t)$ for the case when $\delta_{i,j} = 1$, $i > j$ and $i \in S$. We still want to break down the larger index into its parents, so we can write this as

$$\begin{aligned} \gamma'_{i,j}(s,t) &= \mathrm{Cov}\Big(X_{p_{i,1}}(\xi_i) + \psi_{s,i} + \left[X'_{p_{i,1}}(\xi_i) + \psi'_{s,i}\right](s-\xi_i) + \sigma_i G_i(s-\xi_i)\ ,\ X'_j(t)\Big) \\ &= \gamma'_{p_{i,1},j}(\xi_i,t) + (s-\xi_i)\gamma''_{p_{i,1},j}(\xi_i,t). \end{aligned} \tag{A.20}$$

The other cases are similar and $\gamma'_{i,j}(s,t)$ can be written as

$$\gamma'_{i,j}(s,t) = \begin{cases} \gamma'_i(s,t) & \text{if } i = j \\ \gamma'_{i,p_{j,1}}(s,\xi_i) & \text{if } \delta_{i,j} = 1,\ i < j,\ j \in S \\ \frac{1}{2}\left(\gamma'_{i,p_{j,1}}(s,\xi_i) + \gamma'_{i,p_{j,2}}(s,\xi_i)\right) & \text{if } \delta_{i,j} = 1,\ i < j,\ j \in M \\ \gamma'_{p_{i,1},j}(\xi_i,t) + (s-\xi_i)\gamma''_{p_{i,1},j}(\xi_i,t) & \text{if } \delta_{i,j} = 1,\ i > j,\ j \in S \\ \frac{1}{2}\left(\gamma'_{p_{i,1},j}(\xi_i,t) + \gamma'_{p_{i,2},j}(\xi_i,t)\right) + & \\ \frac{s-\xi_i}{2}\left(\gamma''_{p_{i,1},j}(\xi_i,t) + \gamma''_{p_{i,2},j}(\xi_i,t)\right) & \text{if } \delta_{i,j} = 1,\ i > j,\ j \in M \\ 0 & \text{otherwise.} \end{cases} \tag{A.21}$$

In the same way we as in (A.19), we can calculate $\gamma''_{i,j}(s,t)$, since we again have symmetry in the function.

$$\gamma''_{i,j}(s,t) = \begin{cases} \gamma''_i(s,t) & \text{if } i = j \\ \gamma''_{i,p_{j,1}}(s,\xi_i) & \text{if } \delta_{i,j} = 1,\ i < j,\ j \in S \\ \frac{1}{2}\left(\gamma''_{i,p_{j,1}}(s,\xi_i) + \gamma''_{i,p_{j,2}}(s,\xi_i)\right) & \text{if } \delta_{i,j} = 1,\ i < j,\ j \in M \\ \gamma''_{j,i}(t,s) & \text{if } \delta_{i,j} = 1,\ i > j \\ 0 & \text{otherwise.} \end{cases} \tag{A.22}$$

Moving on now to the function $\gamma_i(s,t)$, we can use the same technique in (A.18) but now breaking down both arguments to the covariance since they are the same path. For example, if target i is a birth or an initial target, then we have

$$\begin{aligned}\gamma_i(s,t) &= \mathrm{Cov}\big(X_i(\xi_i) + X_i'(\xi_i)(s-\xi_i) + \sigma_i G_i(s-\xi_i)\ , \\ &\qquad X_i(\xi_i) + X_i'(\xi_i)(t-\xi_i) + \sigma_i G_i(t-\xi_i)\big) \\ &= \sigma_{X_0}^2 + (s-\xi_i)(t-\xi_i)\sigma_{X_0'}^2 + \sigma_i^2 \mathrm{Cov}(G_i(s-\xi), G_i(t-\xi)) \qquad \text{(A.23)}\end{aligned}$$

where for an IBM

$$\mathrm{Cov}(G_i(s), G_i(t)) = \frac{(s\wedge t)^2(s \vee t)}{2} - \frac{(s\wedge t)^3}{6} \qquad \text{(A.24)}$$

If target i is a split then we have

$$\begin{aligned}\gamma_i(s,t) &= \mathrm{Cov}\Big(X_{p_{i,1}}(\xi_i) + \psi_{s,i} + \Big[X'_{p_{i,1}}(\xi_i) + \psi'_{s,i}\Big](s-\xi_i) + \sigma_i G_i(s-\xi_i)\ , \\ &\qquad X_{p_{i,1}}(\xi_i) + \psi_{s,i} + \Big[X'_{p_{i,1}}(\xi_i) + \psi'_{s,i}\Big](t-\xi_i) + \sigma_i G_i(t-\xi_i)\Big) \\ &= \gamma_{p_{i,1}}(\xi_i,\xi_i) + \sigma_{X_s}^2 + (t+s-2\xi_i)\gamma'_{p_{i,1}}(\xi_i,\xi_i) + \\ &\qquad (s-\xi_i)(t-\xi_i)\left(\gamma''_{p_{i,1}}(\xi_i,\xi_i) + \sigma_{X_s'}^2\right) + \sigma_i^2\mathrm{Cov}(G_i(s-\xi), G_i(t-\xi)) \qquad \text{(A.25)}\end{aligned}$$

and the calculation is very similar for a merger. The general form for $\gamma_i(s,t)$ is then given by

$$
\gamma_i(s,t) = \begin{cases}
\sigma^2_{X_0} + (s-\xi_i)(t-\xi_i)\sigma^2_{X_0'} + \sigma_i^2 \mathrm{Cov}(G_i(s-\xi), G_i(t-\xi)) & \text{if } i \in B \\
\gamma_{p_{i,1}}(\xi_i,\xi_i) + \sigma^2_{X_s} + (t+s-2\xi_i)\gamma'_{p_{i,1}}(\xi_i,\xi_i) + \\
(s-\xi_i)(t-\xi_i)\left(\gamma''_{p_{i,1}}(\xi_i,\xi_i) + \sigma^2_{X_s'}\right) + \sigma_i^2 \mathrm{Cov}(G_i(s-\xi), G_i'(t-\xi)) & \text{if } i \in S \\
\frac{1}{4}\left(\gamma_{p_{i,1}}(\xi_i,\xi_i) + \gamma_{p_{i,2}}(\xi_i,\xi_i) + 2\gamma_{p_{i,1},p_{i,2}}(\xi_i,\xi_i)\right) + \sigma^2_{X_m} + \\
\frac{s+t-2\xi_i}{4}\left(\gamma'_{p_{i,1}}(\xi_i,\xi_i) + \gamma'_{p_{i,2}}(\xi_i,\xi_i) + \gamma'_{p_{i,1},p_{i,2}}(\xi_i,\xi_i) + \gamma'_{p_{i,2},p_{i,1}}(\xi_i,\xi_i)\right) + \\
\frac{(s-\xi_i)(t-\xi_i)}{4}\left(\gamma''_{p_{i,1}}(\xi_i,\xi_i) + \gamma''_{p_{i,2}}(\xi_i,\xi_i) + 2\gamma''_{p_{i,1},p_{i,2}}(\xi_i,\xi_i)\right) + \\
(s-\xi_i)(t-\xi_i)\sigma^2_{X_m'} + \sigma_i^2 \mathrm{Cov}(G_i(s-\xi), G_i(t-\xi)) & \text{if } i \in M.
\end{cases} \tag{A.26}
$$

Making use of the same strategy as in (A.23) and (A.25) we can calculate $\gamma_i'(s,t)$ to be

$$
\gamma_i(s,t) = \begin{cases}
(s-\xi_i)\sigma^2_{X_0'} + \sigma_i^2 \mathrm{Cov}(G_i(s-\xi), G_i'(t-\xi)) & \text{if } i \in B \\
\gamma'_{p_{i,1}}(\xi_i,\xi_i) + (s-\xi_i)\left(\gamma''_{p_{i,1}}(\xi_i,\xi_i) + \sigma^2_{X_s'}\right) + \sigma_i^2 \mathrm{Cov}(G_i(s-\xi), G_i'(t-\xi)) & \text{if } i \in S \\
\frac{1}{4}\left(\gamma'_{p_{i,1}}(\xi_i,\xi_i) + \gamma'_{p_{i,2}}(\xi_i,\xi_i) + \gamma'_{p_{i,1},p_{i,2}}(\xi_i,\xi_i) + \gamma'_{p_{i,2},p_{i,1}}(\xi_i,\xi_i)\right) + \\
\frac{s-\xi_i}{4}\left(\gamma''_{p_{i,1}}(\xi_i,\xi_i) + \gamma''_{p_{i,2}}(\xi_i,\xi_i) + 2\gamma''_{p_{i,1},p_{i,2}}(\xi_i,\xi_i)\right) + (s-\xi_i)\sigma^2_{X_m'} + \\
\sigma_i^2 \mathrm{Cov}(G_i(s-\xi), G_i'(t-\xi)) & \text{if } i \in M.
\end{cases} \tag{A.27}
$$

For the IBM,

$$
\begin{aligned}
\mathrm{Cov}(G_i(s), G_i'(t)) &= \mathrm{E}\left\{\left(\int_0^s B_i(u)du\right) B_i(t)\right\} \\
&= \int_0^s \mathrm{E}\left\{B_i(u)B_i(t)\right\} du \\
&= \int_0^s (u \wedge t)du \\
&= \frac{(s \wedge t)^2}{2} + t(s - s \wedge t)
\end{aligned}
\tag{A.28}
$$

Lastly, we can use the same strategy to calculate $\gamma_i''(s,t)$,

$$
\gamma_i(s,t) = \begin{cases}
\sigma^2_{X_0'} + \sigma_i^2 \mathrm{Cov}(G_i'(s-\xi), G_i'(t-\xi)) & \text{if } i \in B \\
\gamma''_{p_{i,1}}(\xi_i, \xi_i) + \sigma^2_{X_s'} + \sigma_i^2 \mathrm{Cov}(G_i'(s-\xi), G_i'(t-\xi)) & \text{if } i \in S \\
\frac{1}{4}\left(\gamma''_{p_{i,1}}(\xi_i,\xi_i) + \gamma''_{p_{i,2}}(\xi_i,\xi_i) + 2\gamma''_{p_{i,1},p_{i,2}}(\xi_i,\xi_i)\right) + \sigma^2_{X_m'} + & \\
\sigma_i^2 \mathrm{Cov}(G_i'(s-\xi), G_i'(t-\xi)) & \text{if } i \in M
\end{cases}
\tag{A.29}
$$

and this completes the description of the covariances.

Appendix B

Lemmas

Here we present the Lemmas used in the proofs of Chapter 6. All of the results in this chapter assume the Brownian Motion model given in (6.1). We begin with the Law of Iterated Logarithm for 2 dimensional Brownian Motion. The proof of Lemma 1 for Brownian Motion in d dimensions can be found in [35]. Let $\|\cdot\|$ denote the Euclidean norm.

Lemma 1. *(Law of the Iterated Logarithm) For B a Brownian Motion in $\Re^2$,*

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

Lemma 2 is Levy's Modulus of Continuity for 2 dimensional Brownian Motion.

Lemma 2. *(Levy's Modulus of Continuity) For B a Brownian Motion in $\Re^2$,*

$$P\left\{\overline{\lim_{\varepsilon\to 0}}\left(\sup_{\substack{0\le t_1\le t_2\le 1\\ t_2-t_1\le\varepsilon}}\frac{\|B(t_2)-B(t_1)\|}{(2\varepsilon\log(1/\varepsilon))^{1/2}}\right)=1\right\}=1 \tag{B.2}$$

Proof. This follows the same general argument as for the one dimensional case given in [49]. We will only show that

$$P\left\{\overline{\lim_{\varepsilon\to 0}}\left(\sup_{\substack{0\le t_1\le t_2\le 1\\ t_2-t_1\le\varepsilon}}\frac{\|B(t_2)-B(t_1)\|}{(2\varepsilon\log(1/\varepsilon))^{1/2}}\right)\le 1\right\}=1 \tag{B.3}$$

since this is all that is needed in the propositions of Chapter 6. To prove the other direction is actually simpler and can be done by following the logic of the proof given in [49] with changes similar to those that we make here.

Pick $\delta\in(0,1)$ and $\varepsilon>0$ such that $(1+\varepsilon)^2(1-\delta)>1+\delta$. Let K be the set of pairs (i,j) of integers such that $0\le i<j<2^n$ and $0<j-i<2^{n\delta}$ and for each pair set $k=j-i$. Also let $h(t)=(2t\log(1/t))^{1/2}$. Set

$$L=P\left\{\max_{(i,j)\in K}\frac{\|B(j2^{-n})-B(i2^{-n})\|}{h(k2^{-n})}\ge 1+\varepsilon\right\}. \tag{B.4}$$

Now

$$\frac{\|B(j2^{-n})-B(i2^{-n})\|}{h(k2^{-n})}\overset{\mathcal{L}}{=}(2\log(k^{-1}2^n))^{-1/2}X^{1/2} \tag{B.5}$$

where $X\sim\chi^2_2$. The distribution of $Y=cX^{1/2}$ is given by

$$[Y](y)=\frac{2}{c^2\Gamma(1)}ye^{-y^2/2c^2}. \tag{B.6}$$

Hence,

$$\begin{aligned}
L &\leq \sum_{(i,j)\in K} \int_{1+\varepsilon}^{\infty} \frac{2}{\Gamma(1)}(2\log(k^{-1}2^n))ye^{-(2\log(k^{-1}2^n))y^2/2}dy \\
&= C\sum_{(i,j)\in K}(\log(k^{-1}2^n))\int_{(1+\varepsilon)(2\log(k^{-1}2^n))^{1/2}}^{\infty} ye^{-y^2/2}dy \\
&= C\sum_{(i,j)\in K}(\log(k^{-1}2^n))\exp\left(-(1+\varepsilon)^2(\log(k^{-1}2^n))\right) \\
&= C\sum_{(i,j)\in K}(\log(k^{-1}2^n))(k^{-1}2^n)^{-(1+\varepsilon)^2}
\end{aligned}$$

where C is a constant. Since k^{-1} is always less than $2^{-n\delta}$, we further have

$$L \leq C2^{-n(1-\delta)(1+\varepsilon)^2}\sum_{(i,j)\in K}(\log(k^{-1}2^n))$$

Moreover, there are at most $2^{n(1+\delta)}$ points in K and for each of them k is greater than 1, since $i < j$, so $\log(k^{-1}2^n) \leq \log(2^n)$. And finally we have

$$\begin{aligned}
L &\leq C2^{-n(1-\delta)(1+\varepsilon)^2}2^{n(1+\delta)}(\log(2^n)) \\
&= C'n2^{-n((1-\delta)(1+\varepsilon)^2-(1+\delta))}
\end{aligned}$$

By the choice of ε and δ this is the general term of a convergent series; by the first Borel-Cantelli Lemma, for almost every ω there exists an integer $N(\omega)$ such that for $n > N(\omega)$,

$$\|B(j2^{-n}) - B(i2^{-n})\| \leq (1+\varepsilon)h(k2^{-n})$$

where $(i,j) \in K$ and $k = j - i$. Moreover, the integer $N(\omega)$ may be chosen so that for

$n > N(\omega)$,

$$\sum_{m>n} h(2^{-m}) \leq C' h(2^{-n}) \sum_{p \geq 1} \left(p 2^{-p}\right)^{1/2} < \delta h(2^{-(n+1)(1-\delta)}). \tag{B.7}$$

Let ω be a path for which these properties hold. Pick $0 \leq t_1 < t_2 \leq 1$ such that $t_2 - t_1 < 2^{-N(\omega)(1-\delta)}$. Next pick an integer $n > N(\omega)$ such that

$$2^{-n} < 2^{-(n+1)(1-\delta)} \leq t_2 - t_1 < 2^{-n(1-\delta)}.$$

We may find integers i, j, p_r, q_s, such that

$$t_1 = i 2^{-n} - 2^{-p_1} - 2^{-p_2} - \ldots, \qquad t_2 = j 2^{-n} + 2^{-q_1} + 2^{-q_2} + \ldots$$

with $n < p_1 < p_2 < \ldots$, $n < q_1 < q_2 < \ldots$ and $0 < j - i \leq (t_2 - t_1) 2^{-n} < 2^{n\delta}$. Since B is continuous,

$$\begin{aligned}
\|B(t_2,\omega) - B(t_1,\omega)\| &\leq \left\|B(i2^{-n},\omega) - B(t_1,\omega)\right\| + \left\|B(j2^{-n},\omega) - B(i2^{-n},\omega)\right\| + \\
&\quad \left\|B(t_2,\omega) - B(j2^{-n},\omega)\right\| \\
&\leq 2(1+\varepsilon) \sum_{p>n} h(2^{-p}) + (1+\varepsilon) h\left((j-i)2^{-n}\right) \\
&\leq 2(1+\varepsilon)\delta h\left(2^{-(n+1)(1-\delta)}\right) + (1+\varepsilon) h\left((j-i)2^{-n}\right).
\end{aligned}$$

Since h is increasing in a neighborhood of 0, for $t_2 - t_1$ sufficiently small, we have

$$\|B(t_2,\omega) - B(t_1,\omega)\| \leq \left(2(1+\varepsilon)\delta + (1+\varepsilon)\right) h(t_2 - t_1). \tag{B.8}$$

But ε and δ can be chosen arbitrarily close to zero which gives

$$\|B(t_2,\omega) - B(t_1,\omega)\| \leq h(t_2 - t_1). \tag{B.9}$$

and completes the proof.

$\square$

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 $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,n_i})$ and let

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

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

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

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

Lemma 3. *Under Conditions 1-10 of Theorem 1, the distribution of $X_{i,j}$ given $\mathcal{F}_{i,j-1}$ is*

Gaussian with mean and variance given by

$$\begin{aligned} \mu &= X_i(t_{i,j-1}) + O(k^{-1}) \\ \sigma^2 &= \Delta t_{i,j}\sigma^2 + O(k^{-2}) \end{aligned} \tag{B.10}$$

for $j = 2, \ldots, n_i$

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

$$[X_{i,j} \mid \mathcal{F}_{i,j-1}](x, f_{i,j}) = \int_{-\infty}^{\infty} [X_{i,j}, X_i(\zeta) \mid \mathcal{F}_{i,j-1}](x, x_\zeta, f_{i,j}) dx_\zeta \tag{B.11}$$

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

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

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 $X_{i,j-1} = x_{j-1}$ and $X_i(\zeta) = x_\zeta$ can be shown to be Gaussian with mean and variance given by

$$\begin{aligned} \mu &= \frac{\zeta - t_j}{\zeta - t_{j-1}} x_{j-1} + \frac{\Delta t_j}{\zeta - t_{j-1}} x_\zeta \\ \tau^2 &= \Delta t_j \sigma^2 - \frac{(\Delta t_j)^2}{\zeta - t_{j-1}} \sigma^2. \end{aligned} \tag{B.12}$$

Also the conditional distribution of $X_i(\zeta)$ given $\mathcal{F}_{i,j} = f_{i,j}$ is Gaussian with mean and variance denoted by μ_{x_ζ} and $\sigma^2_{x_\zeta}$. Notice that we never observe $X_i(\zeta)$ so this conditional distribution has a variance $\sigma^2_{x_\zeta} > 0$. We can then write the integral in (B.11) as

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

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.

□

The following Lemma is a version of the main Theorem proven by Shepp in [56]. 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 $X(t)$ to be equivalent to the measure imposed by a Brownian Motion, μ_B. Let $m(t) = \mathrm{E}X(t)$ and $\gamma(s,t) = \mathrm{E}(X(s) - \mu(s))(X(t) - \mu(t))$

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

$$(\partial/\partial s)\gamma(s, s^+) - (\partial/\partial s)\gamma(s, s^-) \equiv 1, \quad \textit{for } 0 < s < \infty \tag{B.13}$$

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

$$m(t) = \int_0^t k(u)du. \tag{B.14}$$

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 $\{D_j : 1 \le j \le N_m\}$ is equivalent to μ_B

Lemma 5. *Assume that B is a Brownian Motion. Consider the measure, μ^* imposed by B given $(D_1 = 0, \dots, D_{N_m} = 0)$ where D_j is the difference between parents at the time of merger plus an error defined in (3.25). Then $\mu^* \sim \mu_B$.*

Proof. We just need to show that the conditions (B.13) and (B.14) hold. But the second condition is trivially satisfied since $m(t) = 0$. So we just need to show that (B.13) holds.

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

$$\Sigma = \left(\begin{array}{c|c} \Sigma_1 & \Sigma_{12} \\ \hline \Sigma'_{12} & \Sigma_2 \end{array}\right) \tag{B.15}$$

Notice that

$$\Sigma_1 = \begin{pmatrix} s & s \\ s & t \end{pmatrix} \tag{B.16}$$

and

$$\Sigma_{12} = \begin{pmatrix} f_1(s) & f_2(s) & \cdots & f_{N_m}(s) \\ f_1(t) & f_2(t) & \cdots & f_{N_m}(t) \end{pmatrix} \tag{B.17}$$

for some functions $f_1, \ldots, f_{N_m}$. The matrix Σ_2 is symmetric so let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{N_m}$ and $\lambda_1, \ldots, \lambda_{N_m}$ be its eigenvectors and eigenvalues respectively. We can write the covariance matrix of $(B(s), B(t))'$ conditional on $(D_1, \ldots, D_{N_m})'$ as

$$\Sigma^* = \Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'. \tag{B.18}$$

The off diagonal element of the matrix $\Sigma_{12}\Sigma_2^{-1}\Sigma_{12}'$ is given by

$$\begin{aligned} (f_1(s), f_2(s), \ldots, f_{N_m}(s))\,\Sigma_2^{-1} \begin{pmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_{N_m}(t) \end{pmatrix} &= \sum_{i=1}^{N_m} \frac{1}{\lambda_i} (f_1(s), f_2(s), \ldots, f_{N_m}(s))\,\boldsymbol{v}\boldsymbol{v}' \begin{pmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_{N_m}(t) \end{pmatrix} \\ &= \sum_{i=1}^{N_m}\sum_{j=1}^{N_m} \frac{1}{\lambda_i} v_{ij}^2 f_i(s) f_i(t). \end{aligned} \tag{B.19}$$

This gives $\gamma(s,t) = \mathrm{Cov}(B(s), B(t) \mid (D_1, \ldots, D_{N_m})$ in general as

$$\gamma(s,t) = s \wedge t - \sum_{i=1}^{N_m}\sum_{j=1}^{N_m} \frac{1}{\lambda_i} v_{ij}^2 f_i(s) f_i(t). \tag{B.20}$$

Hence we have

$$\frac{\partial}{\partial s}\gamma(s, s^+) = 1 - \sum_{i=1}^{N_m}\sum_{j=1}^{N_m} \frac{1}{\lambda_i} v_{ij}^2 f_i'(s) f_i(s) \tag{B.21}$$

$$\frac{\partial}{\partial s}\gamma(s, s^-) = -\sum_{i=1}^{N_m}\sum_{j=1}^{N_m} \frac{1}{\lambda_i} v_{ij}^2 f_i(s) f_i'(s) \tag{B.22}$$

which satisfies (B.13) and completes the proof.

□

The next two Lemmas are needed to compare the location densities of two different sequences of solutions $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$ and $(\tilde{\mathcal{U}}, \tilde{\mathcal{V}}, \tilde{\mathcal{P}})_k$. 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 1-10 of Theorem 1. 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 one of the differences 1a, 1b, 2, 3, 5a, 5b or 6b from the propositions. Consider a track segment $(x_1, \ldots, x_{n'_k})$ such that*

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

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

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{[\tilde{X}_{i_1}(t_{j'}) \mid \tilde{\mathcal{F}}_{i_1}(t_{j'-1})](x_1)}{[\hat{X}_{i_2}(t_{j'}) \mid \hat{\mathcal{F}}_{i_2}(t_{j'-1})](x_1)} = O(1) \quad \textit{as } k \to \infty \textit{ a.s.} \tag{B.23}$$

and

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} \frac{\prod_{j=2}^{n'_k} [\tilde{X}_{i_1}(t_{j'+j-1}) \mid \tilde{\mathcal{F}}_{i_1}(t_{j'+j-2})](x_j)}{\prod_{j=2}^{n'_k} [\hat{X}_{i_2}(t_{j'+j-1}) \mid \hat{\mathcal{F}}_{i_1}(t_{j'+j-2})](x_j)} \leq a(\log(k))^b \quad \textit{as } k \to \infty \textit{ a.s.} \tag{B.24}$$

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

$$\sup_{(\hat{\theta}_k,\tilde{\theta}_k)\in\Theta} \frac{\prod_{j=1}^{n'_k}[\tilde{X}_{i_1}(t_{j'+j-1}) \mid \tilde{\mathcal{F}}_{i_1}(t_{j'+j-2})](x_j)}{\prod_{j=1}^{n'_k}[\hat{X}_{i_2}(t_{j'+j-1}) \mid \hat{\mathcal{F}}_{i_1}(t_{j'+j-2})](x_j)} \leq a(\log(k))^b \quad \text{as } k \to \infty \text{ a.s.} \tag{B.25}$$

for some constants a and b which depend on ω.

Proof. For (B.23) consider the case where target i_1 is a birth. Then

$$[\tilde{X}_{i_1}(t_{j'}) \mid \tilde{\mathcal{F}}_{i_1}(t_{j'-1})](x_1) \leq \phi(0;0,\sigma^2_{X_0}) \tag{B.26}$$

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

$$\left[\tilde{X}_{i_1}(t_{j'}) \mid \tilde{\mathcal{F}}_{i_1}(t_{j'-1})\right](x_1) \leq \phi(0;0,\sigma^2_{X_s}) \text{ and} \tag{B.27}$$

$$\left[\tilde{X}_{i_1}(t_{j'}) \mid \tilde{\mathcal{F}}_{i_1}(t_{j'-1})\right](x_1) \leq \phi(0;0,\sigma^2_{X_m}) \tag{B.28}$$

respectively. Let C_1 be the maximum of the three quantities in (B.26), (B.27) and (B.28).

Now for any Brownian path on a finite interval

$$P\left(\sup_{0\leq t\leq T} |B(t)| < \infty\right) = 1.$$

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 is a diffuse distribution. 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 (B.24) and (B.25). We only need to show (B.24) since they are equivalent if in (B.25) we relabel x_1 as x_2 and so on. By Lemma 3 can write the ratio in (B.24) as

$$\frac{\prod_{j=2}^{n_k'}[\tilde{X}_{i_1}(t_{j'+j-1}) \mid \tilde{\mathcal{F}}_{i_1}(t_{j'+j-2})](x_j)}{\prod_{j=2}^{n_k'}[\hat{X}_{i_2}(t_{j'+j-1}) \mid \hat{\mathcal{F}}_{i_1}(t_{j'+j-2})](x_j)} = \left(\prod_{j=2}^{n_k} A_j\right) \cdot B \tag{B.29}$$

where

$$A_j = \sqrt{\frac{2\pi\Delta t_{j'}^k\hat{\sigma}^2 + O(k^{-2})}{2\pi\Delta t_{j'}^k\tilde{\sigma}^2 + O(k^{-2})}}, \tag{B.30}$$

$$B = \exp\left\{\sum_{j=2}^{n_k}\left(-\frac{(x_j - x_{j-1} + O(k^{-1}))^2}{2\Delta t_{j'}^k\tilde{\sigma}^2 + O(k^{-2})} + \frac{(x_j - x_{j-1} + O(k^{-1}))^2}{2\Delta t_{j'}^k\hat{\sigma}^2 + O(k^{-2})}\right)\right\} \tag{B.31}$$

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

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, $\tilde{\sigma}^2$ and $\hat{\sigma}^2$ can be different only by one term in the sum. That is for

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{\hat{N}}\sum_{i=1}^{\hat{m}}\sum_{j=1}^{\hat{n}_i-1}\frac{\hat{X}_i(t_{i,j+1})-\hat{X}_i(t_{i,j})^2}{\Delta t_{i,j}}I_{\hat{E}_{i,j}} \\ &= \frac{1}{\hat{N}}\sum_{i=1}^{\hat{m}}\sum_{j=1}^{\hat{n}_i-1}\hat{D}_{i,j}^2 \end{aligned} \tag{B.32}$$

and

$$\begin{aligned}\tilde{\sigma}^2 &= \frac{1}{\tilde{N}}\sum_{i=1}^{\tilde{m}}\sum_{j=1}^{\tilde{n}_i-1}\frac{\tilde{X}_i(t_{i,j+1})-\tilde{X}_i(t_{i,j})^2}{\Delta t_{i,j}}I_{\tilde{E}_{i,j}} \\ &= \frac{1}{\tilde{N}}\sum_{i=1}^{\tilde{m}}\sum_{j=1}^{\tilde{n}_i-1}\tilde{D}_{i,j}^2, \end{aligned} \tag{B.33}$$

where recall $E_{i,j}$ is given by

$$\left\{\frac{(X_i(t_{i,j+1})-X_i(t_{i,j}))^2+(Y_i(t_{i,j+1})-Y_i(t_{i,j}))^2}{\Delta t_{i,j}} \le \sqrt{\log\log N}\right\}. \tag{B.34}$$

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

$$\begin{aligned}\hat{\sigma}^2-\tilde{\sigma}^2 &= \left[(\hat{N}-1)\sum_{i=1}^{\hat{m}}\sum_{j=1}^{\hat{n}_i}\hat{D}_{i,j}^2-\hat{N}\sum_{i=1}^{\tilde{m}}\sum_{j=1}^{\tilde{n}_i}\tilde{D}_{i,j}^2\right]/(\hat{N}(\hat{N}-1)) \\ &= -\left(\sum_{i=1}^{\tilde{m}}\sum_{j=1}^{\tilde{n}_i}\tilde{D}_{i,j}^2\right)/(\hat{N}(\hat{N}-1))+\hat{D}_{i',j'}^2/\hat{N} \\ &= -\tilde{\sigma}^2/\hat{N}+\hat{D}_{i',j'}^2/\hat{N} \end{aligned} \tag{B.35}$$

so that

$$-\tilde{\sigma}^2/\hat{N} \leq \hat{\sigma}^2 - \tilde{\sigma}^2 \leq \hat{D}^2_{i',j'}/\hat{N}$$

but $\hat{D}^2_{i',j'} \leq c\sqrt{\log\log \hat{N}}$ by Condition 6 and $\tilde{\sigma}^2 \leq K$ for some constant $K < \infty$ by Condition 5. 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 considering the quantity A_j in (B.30), since $|\hat{\sigma}^2 - \tilde{\sigma}^2| \leq (c \log\log \hat{N})/\hat{N}$ 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 $n_k \to \infty$ as $k \to \infty$, then

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

so that

$$\prod_{j=2}^{n_k} A_j \leq (\log \hat{N})^{c_2/2}$$

eventually which means that

$$\prod_{j=2}^{n_k} A_j \leq c_3 (\log k)^{c_4}. \tag{B.36}$$

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

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

So we can write B as

$$\begin{aligned} B = \exp\Bigg\{\sum_{j=2}^{n_k}\Bigg(&-\frac{(x_j - x_{j-1})^2 + (x_j - x_{j-1})O(k^{-1}) + O(k^{-2})}{2\Delta t_j^k \hat{\sigma}^2}(1 + O(k^{-1})) \\ &+\frac{(x_j - x_{j-1})^2 + (x_j - x_{j-1})O(k^{-1}) + O(k^{-2})}{2\Delta t_j^k \hat{\sigma}^2}\left(1 + O\left(\sqrt{\log\log \hat{N}}/\hat{N}\right)\right)\Bigg)\Bigg\} \\ = \exp\Bigg\{\sum_{j=2}^{n_k}&\left(\frac{(x_j - x_{j-1})^2 O(k^{-1}) + (x_j - x_{j-1})O(k^{-1}) + (x_j - x_{j-1})^2 O\left(\sqrt{\log\log \hat{N}}/\hat{N}\right) + O(k^{-2})}{2\Delta t_j^k \hat{\sigma}^2}\right)\Bigg\}. \end{aligned} \tag{B.37}$$

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

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

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 $k \to \infty$. Thus

$$\sup_{(\hat{\theta}_k, \tilde{\theta}_k) \in \Theta} B \leq \exp\left\{C_1' + C_2' \log\log k\right\} = a(\log k)^b \tag{B.38}$$

for some constants a and b. Combining this with (B.36) gives the desired result.

$\square$

Lemma 7 states that the probability that any of the 2 dimensional Brownian Motion paths will intersect at a any time t in a finite interval is zero. This Lemma is an immediate consequence of Proposition 1.4.1 on page 353 of [36].

Lemma 7. *Consider any two (x, y) paths (X_1, Y_1) and (X_2, Y_2) from the location model in Section 2.3 where $G_i(t)$ is a Brownian Motion Then,*

$$P\left\{\inf_{0 \leq t \leq 1} \|(X_1(t), Y_1(t))) - (X_2(t), Y_2(t))\| > 0\right\} = 1 \tag{B.39}$$

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

Lemma 8. *The estimate $\hat{\sigma}^2$ given by (6.2) is strongly consistent as $k \to \infty$ when applied to the any sequence tracking solutions, $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})_k$, 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 $Z_j \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. There are a finite number of tracks in $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$ which means $N \geq c_1 k$ for some positive constant c_1, so $N \to \infty$ as $k \to \infty$. Hence it is sufficient to show that $\frac{1}{N} S_N^* \to \sigma^2$ a.s.

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

$$\sum_{n=1}^{\infty} P\left(\left| \frac{S_{u_n}^* - \mathrm{E} S_{u_n}^*}{u_n} \right| > \varepsilon \right) < \infty \tag{B.40}$$

But

$$\begin{aligned} Var(S_n^*) &= \sum_{k=1}^{n} \mathrm{Var}\left((Z_j^*)^2 \right) \leq \sum_{k=1}^{n} \mathrm{E}\left((Z_j^*)^4 \right) \\ &= n\mathrm{E}\left(Z_j^4 I_{E_{i,j}} \right) \leq n\mathrm{E}\left(Z_j^4 \right) = 2n\sigma^2. \end{aligned} \tag{B.41}$$

So it follows by Chebyshev's Inequality that the sum in (B.40) is no more than

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

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

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

Hence by the first Borel-Cantelli Lemma, we have

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

for all $\varepsilon > 0$. Taking a union over positive rational ε gives us that $(S^*_{u_n} - \mathrm{E}S^*_{u_n})/u_n \to 0$ a.s. But

$$\mathrm{E}S^*_n/n = \mathrm{E}(Z^*_j)^2 \to \mathrm{E}Z_j^2 = \sigma^2,$$

so that $S^*_{u_n}/u_n \to \sigma^2$ 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 $u_{n+1}/u_n \to \alpha$ so

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

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

$$\lim_{k\to\infty} \hat{\sigma}_k^2 = \lim_{N\to\infty} \frac{S^*_N}{N} = \sigma^2 \quad \text{a.s.}$$

□

Corollary 1. *The estimate* $\hat{\sigma}^2$ *given by (6.2) is strongly consistent as* $k \to \infty$ *when applied to the correct sequence of tracking solutions,* $(\mathcal{U}, \mathcal{V}, \mathcal{P})_k$.

Proof. Immediate consequence of Lemma 8.

Bibliography

[1] J. Arnold, S. Shaw, and H. Pasternack. Efficient target tracking using dynamic programming. *IEEE Transactions on Aerospace and Electronic Systems*, 29:44–56, 1993.

[2] Y. Bar-Shalom, X.R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation.* New York: John Wiley & Sons, Inc., 2001.

[3] Y. Bar-Shalom and E. Tse. Tracking in a cluttered environment with probabilistic data association. *Automatica*, 2:451–460, 1975.

[4] S. Blackman. Multiple hypothesis tracking for multiple target tracking. *IEEE A&E Systems Magazine*, 19:5–18, 2004.

[5] S. Blackman, R. Dempster, and T. Broida. Multiple hypothesis track confirmation for infrared surveillance systems. *IEEE Transactions on Aerospace and Electronic Systems*, 29:810–824, 1993.

[6] S. Blackman and R. Popoli. *Design and analysis of modern tracking systems.* Boston: Artech House, 1999.

[7] E.A. Bloem and H. Blom. Joint probabilistic data association methods avoiding track coalesence. *Proceedings of the 34th IEEE Conference on Decision and Control*, 3:2752–2757, 1995.

[8] H.A. Blom. An efficient filter for abrubtly changing systems. *Proceedings of the IEEE Conference on Decision and Control*, pages 656–658, 1984.

[9] H.A. Blom and Y. Bar-Shalom. The interacting multiple model algorithm for systems with markovian switching coefficients. *IEEE Transactions on Automatic Control*, 33:780–783, 1988.

[10] A. Bracco, J. C. McWilliams, G. Murante, A. Provenzale, and J. B. Weiss. Revisiting freely decaying two-dimensional turbulence at millennial resolution. *Physics of Fluids*, 12:2931–2941, 2000.

[11] L. Carvalho and C. Jones. A satellite method to identify structural properties of mesoscale convective systems based on the maximum spatial correlation tracking technique (MASCOTTE). *Journal of Applied Meteorology*, 40:1683–1701, 2001.

[12] K.C. Chang. Multitarget tracking with adaptive detection thresholds. *IEEE Transactions on Aerospace and Electronic Systems*, 32:401–406, 1996.

[13] K.C. Chang and Y. Bar-Shalom. Joint probabilistic data association for multitarget tracking with possibly unresolved measurements and maneuvers. *IEEE Transactions on Automatic Control*, 29:585–594, 1984.

[14] H. Chen, X. Li, and Y. Bar-Shalom. On joint track initiation and parameter estimation under measurement origin uncertainty. *IEEE Transactions on Aerospace and Electronic Systems*, 40:675–694, 2004.

[15] D. Cheng, R.E. Mercer, J.L. Barron, and P. Joe. Tracking severe weather storms in dopplar radar images. *International Journal of Imaging Systems and Technology*, 9:201–213, 1998.

[16] S. Cong and L. Hong. Computational complexity analysis for multiple hypothesis tracking. *Mathematical and computer modelling*, 29:1–16, 1999.

[17] I.J. Cox. A review of statistical data association techniques for motion correspondence. *International Journal of Computer Vision*, 10:53–66, 1993.

[18] I.J. Cox. An efficient implementation of reid's multiple hypothesis tracking algorithm and its evaluation for the pupose of visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:138–150, 1996.

[19] F.E. Daum. The importance of resolution in multiple target tracking. *Signal and Data Processing of Small Targets 1994, SPIE Proceedings*, 2235:329–338, 1994.

[20] C.A. Davis, K.W. Manning, R.E. Carbone, S.B. Trier, and J.D. Tuttle. Coherence of warm-season continental rainfll in numerical weather prediction models. *Monthly Weather Review*, 131:2667–2679, 2003.

[21] S. Deb. A generalized s-d assignment algorithm for multisensor-multitarget state estimation. *IEEE Transactions on Aerospace and Electronic Systems*, 33:523–538, 1997.

[22] M. DeFeo, A. Graziano, R. Miglioli, and A. Farina. IMMJPDA versus MHT and kalman filter with NN correlation: A performance comparison. *IEEE Proceedings of Radar, Sonar, and Navigation*, 144:49–56, 1997.

[23] R. Deriche and O. Faugeras. Tracking line segments. *Image and Computer Vision*, 8:261–270, 1990.

[24] O.E. Drummond. Multiple sensor tracking with multiple frame, probabilistic data association. *Signal and Data Processing of Small Targets 1995, SPIE Proceedings*, 2561:322–336, 1995.

[25] H. Feidaas and C. Constantinos. Monitoring mesoscale convective cloud systems associated with heavy storms using meteostat imagery. *American Meteorological Society*, 40:491–512, 2001.

[26] R.J. Fitzgerald. Track biases and coalescence with probabilistic data association. *IEEE Transactions on Aerospace and Electronic Systems*, 21:822–825, 1985.

[27] T.E. Fortmann, Y. Bar-Shalom, and M. Scheffe. Sonar tracking of multiple targets using joint probabilistic data association. *IEEE Journal of Oceanic Engineering*, 8:173–184, 1983.

[28] L.M. Fuentes and S.A. Velastin. Tracking people for automatic surveillance applications. *Lecture Notes in Computer Science*, 2652:238–245, 2003.

[29] S.B. Gelfand and et al. Adaptive detection threshold optimization for tracking in clutter. *IEEE Transactions on Aerospace and Electronic Systems*, 32:514–522, 1996.

[30] H. Gish and R. Mucci. Target state estimation in a multi-target environment. *IEEE Transactions on Aerospace and Electronic Systems*, 23:60–73, 1987.

[31] B. Hammarberg, C. Forster, and E. Torebjork. Parameter estimation of human nerve c-fibers using matched filtering and multiple hypothesis tracking. *IEEE Transactions on Biomedical Engineering*, 49:329–336, 2002.

[32] K.I. Hodges. A general method for tracking analysis and its application to meteorological data. *Monthly Weather Review*, 122:2573–2586, 1994.

[33] K.I. Hodges. Adaptive constraints for feature tracking. *Monthly Weather Review*, 127:1362–1373, 1999.

[34] P. Jensfelt and S. Kristensen. Active global localization for a mobile robot using multiple hypothesis tracking. *IEEE Transactions on Robotics and Atomation*, 17:748–760, 2001.

[35] O. Kallenberg. *Foundations of Modern Probability*. New York: Springer-Verlag, 2nd edition, 2002.

[36] D. Khoshnevisan. *Multiparameter Processes; An Introduction to Random Fields.* New York: Springer-Verlag, 2002.

[37] W. Koch. On bayesian mht for well seperated targets in a densely cluttered environment. *Proceedings of the IEEE International Radar Conference*, pages 323–328, 1995.

[38] W. Koch and G. van Keuk. Multiple hypothesis track maintenance with possibly unresolved measurements. *IEEE Transactions on Aerospace and Electronic Systems*, 33:883–892, 1997.

[39] J.J. Leonard, H.F. Durantwhyte, and I.J. Cox. Dynamic map building for an autonomous mobile robot. *International Journal of Robotics Research*, 11:286–298, 1992.

[40] H. Leung, H. Zhijian, and M. Blanchette. Evaluation of multiple radar target trackers in stressful environments. *IEEE Transactions on Aerospace and Electrical Systems*, 24:663–673, 1979.

[41] Jing Z. Li J. Practical system for tracking multiple maneuvering targets. *Optical Engineering*, 42:2439–2451, 2003.

[42] Li X.R. Li N. Target percievability and its applications. *IEEE Transactions on Signal Processing*, 49:2588–2604, 2001.

[43] S. Mori, C. Chong, E. Tse, and R. Wishner. Tracking and classifying multiple targets without a priori identification. *IEEE Transactions on Automatic Control*, 31:401–409, 1986.

[44] K.G. Murty. *Linear and Combinitorial Programming.* New York: John Wiley & Sons, Inc., 1976.

[45] Claudia Pasquero, Antonello Provenazale, and Jeffrey B. Weiss. Vortex statistics from eulerian and lagrangian time series. *Physical Review Letters*, 89:284501, 2002.

[46] A. Pece. Generative-model-based tracking by cluster analysis of image differences. *Robotics and Autonomous Systems*, 39:181–194, 2002.

[47] A. B. Poore and A. J. Robertson. A new lagrangian relaxation algorithm for a class of multidimensional assignment problems. *Computational Optimization and Applications*, 8:129–150, 1997.

[48] D.B. Reid. An algorithm for tracking multiple targets. *IEEE Transactions on Automatic Control*, 24:843–854, 1979.

[49] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion.* New York: Springer-Verlag, 1999.

[50] J.A. Roecker. Multiple scan joint probabilistic data association. *IEE Transactions on Aerospace and Electronic Systems*, 31:1204–1210, 1995.

[51] A. Rosenfeld and A. Kak. *Digital Picture Processing.* Academic Press, 2nd edition, 1982.

[52] V. Salari and I.K. Sethi. Feature point correspondence in the presence of occlusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:87–91, 1990.

[53] D.J. Salmond. Mixture reduction algorithms for target tracking in clutter. *Signal and Data Processing of Small Targets 1990, SPIE Proceedings*, 1305:434–445, 1990.

[54] M. Schimek. *Smoothing and Regression: Approaches, Computation, and Application.* New York: John Wiley & Sons, Inc., 2000.

[55] I.K. Sethi and R. Jain. Finding trajectories of feature points in a monocular image sequence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:56–72, 1987.

[56] L.A. Shepp. Radon nikodym derivatives of gaussian measures. *The Annals of Mathematical Statistics*, 37:321–354, 1966.

[57] C.B. Storlie, C. Davis, T. Hoar, T. Lee, D. Nychka, J. Weiss, and B. Whitcher. Identifying and tracking turbulence structures. *Proceedings of the 38th Asilomar Conference on Signals, Systems, and Computers*, (To Appear).

[58] S.M. Tonissen and R.J. Evans. Perfomance of dynamic programming techniques for track-before-detect. *IEEE Transactions on Aerospace and Electronic Systems*, 32:1440–1451, 1996.

[59] G.V. Trunk and J.D. Wilson. Track initiation of occasionally unresolved radar targets. *IEEE Transactions on Aerospace and Electronic Systems*, 17:122–130, 1996.

[60] G. van Keuk. Multihypothesis tracking with electronically scanned radar. *IEEE Transactions on Aerospace and Electronic Systems*, 31:916–927, 1995.

[61] J.B. Weiss and J.C. McWilliams. Temporal scaling behavior of decaying two-dimensional turbulence. *Physics of Fluids A: Fluid Dynamics*, 5:608–621, 1993.

[62] P. Willett, R. Niu, and Y. Bar-Shalom. Integration of bayes detection with target tracking. *IEEE Transactions on Signal Processing*, 49:17–29, 2001.

[63] Li X.R. and Y. Bar-Shalom. Detection threshold selection for tracking performance optimization. *IEEE Transactions on Aerospace and Electronic Systems*, 30:742–749, 1994.

[64] M. Yeasin, E. Polat, and R. Sharma. A multiobject tracking framework for interactive multimedia applications. *IEEE Transactions on Multimedia*, 6:398–405, 2004.