SUPPLEMENTARY DOCUMENT TO ACCOMPANY THE ARTICLE "TRACKING OF MULTIPLE MERGING AND SPLITTING TARGETS WITH APPLICATION TO CONVECTIVE SYSTEMS"

Curtis B. Storlie[a], Thomas C. M. Lee[b,c], Jan Hannig[c], and Douglas Nychka[d]

[a]Department of Mathematics and Statistics, University of New Mexico

[b]Department of Statistics, The Chinese University of Hong Kong

[c]Department of Statistics, Colorado State University

[d]Geophysical Statistics Project, National Center for Atmospheric Research

This document provides many of the technical details of the model proposed in the main article in addition to some more in-depth discussion.

S.1 Calculation of Model Likelihood

In this Section we present the likelihood of the model described in Section 3. 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,

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

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 first restrict the focus to location information. 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

$$(\boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f), \tag{S.1}$$

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

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

as the false alarms are assumed to be completely independent of the targets. 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, $[\boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f]$.

S.1.1 Target Density

Since $\mathcal{X}$ and $\mathcal{Y}$ are independent given $(\mathcal{U}, \mathcal{V}, \mathcal{W})$, we can write the target density as

$$[\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}] \;=\; [\mathcal{U}, \mathcal{V}] \cdot [\mathcal{W} \mid \mathcal{U}, \mathcal{V}] \cdot [\mathcal{X} \mid \mathcal{U}, \mathcal{V}, \mathcal{W}] \cdot [\mathcal{Y} \mid \mathcal{U}, \mathcal{V}, \mathcal{W}]. \tag{S.2}$$

We will call the conditional densities in (S.2), in order from left to right, the target event density, observability density, and target location densities respectively. We will describe each of these in the following sections.

S.1.1.1 Target Event Density

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

$$[\mathcal{U}, \mathcal{V}] = [N_0] \prod_{j=1}^{n} [U_{b,j}, U_{d,j}, U_{s,j}, U_{m,j} \mid N(t_j)][V_{b,j}, V_{d,j}, V_{s,j}, V_{m,j} \mid N(t_j), U_{b,j}, U_{d,j}, U_{s,j}, U_{m,j}], \tag{S.3}$$

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 assumed to be Poisson distributed with parameter λ_0. Therefore

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

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 occurrence 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 $\bar{\lambda}_{b,j} = \lambda_b$, $\bar{\lambda}_{d,j} = \lambda_d \bar{N}_j$, $\bar{\lambda}_{s,j} = \lambda_s \bar{N}_j$ and $\bar{\lambda}_{m,j} = \lambda_m \left(\bar{N}_j - 1\right)$ for birth, death, splitting, and merging events 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,

$$\begin{aligned}
[U_{b,j} \mid N(t_j)]\,(u) &\approx (\lambda_b \Delta t_j)^u e^{-\lambda_b \Delta t_j}/u! \\
[U_{d,j} \mid N(t_j)]\,(u) &\approx (\bar{\lambda}_{d,j} \Delta t_j)^u e^{-\bar{\lambda}_{d,j} \Delta t_j}/u! \\
[U_{s,j} \mid N(t_j)]\,(u) &\approx (\bar{\lambda}_{s,j} \Delta t_j)^u e^{-\bar{\lambda}_{s,j} \Delta t_j}/u! \\
[U_{m,j} \mid N(t_j)]\,(u) &\approx (\bar{\lambda}_{m,j} \Delta t_j)^u e^{-\bar{\lambda}_{m,j} \Delta t_j}/u!
\end{aligned} \tag{S.4}$$

Under the same assumption that $N(t) = \bar{N}_j$ is constant during the interval $[t_j, t_{j+1})$, we have

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

and we can write (S.3) as

$$[\mathcal{U},\mathcal{V}] \approx [N_0]\prod_{j=1}^{n}[U_{b,j}][V_{b,j} \mid U_{b,j}]\cdot[U_{d,j}][V_{d,j} \mid U_{d,j}]\cdot[U_{s,j}][V_{s,j} \mid U_{s,j}]\cdot[U_{m,j}][V_{m,j} \mid U_{m,j}]. \qquad \text{(S.5)}$$

S.1.1.2 Observability Density

Recall that $W_i(t)$ represents the observability (0 or 1) of the i^{th} target at time t, $i = 1,\ldots,M$ where M is the number of targets that existed before time t_n. Let $\mathcal{W} = \{W_i(t_j) : i = 1,\ldots,M,\ j = 1,\ldots,n\}$ 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 events 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 completely specifies $\mathcal{W}$ since its dependence on $\mathcal{U}$ and $\mathcal{V}$ is only on whether or not a target exists at the observed time points. In the sequel, 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 3.2 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 event variables in (S.2) can be written out using indicator functions to separate the cases when the i^{th} target exists and when it does not. This density is then given by

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

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

S.1.1.3 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 during the time sequence $t_1,\ldots,t_n$. Then

$\mathcal{X} \sim \mathcal{N}(\mu_X, \Sigma_X)$, where we will define μ_X, and Σ_X below.

Recall from Section 3.3 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 calculate the exact density, we would need to integrate out on the joint distribution of $\mathcal{X}$ and ξ, given that the ξ_i's are in their respective intervals. Most likely this can only be achieved via numerical approximations.

Also recall from Section 3.3 that we need to condition $\mathcal{X}$ on the random variables $(D_1, \ldots, D_{N_m})$ and evaluate this density when they are zeros. Let $\mathcal{D} = (D_1, \ldots, D_{N_m})'$, and we write

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

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

$$\begin{pmatrix} \mathcal{X} \\ \mathcal{D} \end{pmatrix} \sim \mathcal{N}(\mu, \Sigma) \tag{S.6}$$

where

$$\mu = \begin{pmatrix} \mu_X \\ \mu_D \end{pmatrix} \tag{S.7}$$

and

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

The mean vectors and covariance matrices will be described in the following. Let $\mu_i(t) = E\{X_i(t)\}$ and $\mu_{D_i} = E(D_i)$. These functions are given for the IBM model in Section S.2. 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 (S.7) as $\mu_X = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_m)$ and $\mu_D = (\mu_{D_1}, \ldots, \mu_{D_{N_m}})$.

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 as

$$\Sigma_{i,j}(k,l) = \mathrm{Cov}(X_i(t_{i,k}), X_j(t_{j,l})), \qquad k = 1, \ldots, n_i; \quad l = 1, \ldots, n_j. \tag{S.8}$$

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

$$\Sigma_{i,D}(k,l) = \mathrm{Cov}(X_i(t_{i,k}), D_l) \quad k = 1, \ldots, n_i; \quad l = 1, \ldots, N_m \tag{S.9}$$

$$\Sigma_D(k,l) = \mathrm{Cov}(D_k, D_l) \quad k = 1, \ldots, N_m; \quad l = 1, \ldots, N_m. \tag{S.10}$$

The covariance functions in (S.8), (S.9), and (S.10) for the IBM model are given in Section S.2.

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

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

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

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

This completes the description of the distribution of $(\mathcal{X}, \mathcal{D})$ given in (S.6).

We can then compute the conditional distribution of $\mathcal{X}$ given $\mathcal{D} = 0$, which we will just call the distribution of $\mathcal{X}$ from this point onward. From standard multivariate normal theory we have

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

where

$$\mu^* = \mu_X - \Sigma_{X,D}\Sigma_D^{-1}\mu_D \quad \text{and} \quad \Sigma^* = \Sigma_X - \Sigma_{X,D}\Sigma_D^{-1}\Sigma'_{X,D}.$$

The density of X is then just the multivariate normal density with parameters μ^* and Σ^*. This will require computing the inverse of Σ^*, which can be done quite 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 splitting or merging of the other, they will have 0 covariance. Unfortunately, because of the conditioning on $\mathcal{D}$, this model cannot be posed in state space form. Hence, the corresponding filters cannot be used to update the conditional distribution of a new observation given the previous observations.

S.1.2 False Alarm Density

In a manner similar to target density we can write the density of the false alarm variables from (S.1) as

$$[\boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f] \;=\; [\boldsymbol{N}_f] \cdot [\mathcal{X}_f, \mathcal{Y}_f \mid \boldsymbol{N}_f] \tag{S.11}$$

where $\boldsymbol{N}_f$, $\mathcal{X}_f$, and $\mathcal{Y}_f$ will be precisely defined below. It was assumed in Section 3.5 that false alarms occur at each time frame as *iid* realizations from a Poisson Process with intensity function $\rho(x, y)$. Hence $\boldsymbol{N}_f = (N_f(t_1), \ldots, N_f(t_n))$ are *iid* Poisson distributed random variables with rate $\lambda_f = \int \rho(x, y)dxdy$ The corresponding density of $\boldsymbol{N}_f$ is then

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

Now 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. Similar notation will be used for $\mathcal{Y}_f$. Due to the Poisson process assumption, the density function for a particular $(X_{f,i}(t), Y_{f,i}(t))$ is $f(x, y) = \rho(x, y)/\lambda_f$ and hence the density for $(\mathcal{X}_f, \mathcal{Y}_f)$ is

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

where x_{ij} is a dummy variable for the value of $X_{f,i}(t_j)$ and similarly for y_{ij}.

S.1.3 Attributes

The attribute variables are assumed *iid* over time given the Observability variable $\mathcal{W}$, thus the densities are quite straightforward to calculate. With the presence of attributes, we now have the following collection of random variables for targets

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

where $\mathcal{A}$ denotes the collection of all attribute variables. Below 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 formally defined later.

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

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

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 but for false alarms.

The target likelihood function is then given by

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

So we can just multiply $[\mathcal{A} \mid \mathcal{W}]$ to the target density without attributes given in (S.2). Technically $\mathcal{A}$ should also be conditioned on $\mathcal{U}$, $\mathcal{V}$, $\mathcal{X}$, and $\mathcal{Y}$ as well, but given the way that we modeled attributes in the previous section, the density of $\mathcal{A}$ would still depend only on $\mathcal{W}$, and hence we dropped the other variables in the notation. Similarly, the false alarm likelihood is given by

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

so we can just multiply $[\mathcal{A} \mid \boldsymbol{N}_f]$ to the false alarm density without attributes in (S.11). Therefore the overall density is

$$[(\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}), (\boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f)] \;=\; [\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}] \cdot [\boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f]. \tag{S.12}$$

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

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

and

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

We will describe these densities in the following sections.

S.1.3.1 Radius Density

Let $R_{1,i}(t)$ and $R_{2,i}(t)$ respectively be the length of minor and major axes of the best fitting ellipse to target i at time t. We only observe the min and max of these from the data 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 $\mathcal{R}_{(2)}$, where recall M in the total number of targets that existed before time t_n.

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

$$[R_{(1),i}, R_{(2),i}](r, s) = \{[R_{1,i}](r)[R_{2,i}](s) + [R_{1,i}](s)[R_{2,i}](r)\}\, I_{\{r \leq s\}} \tag{S.13}$$

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 3.4.

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

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

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.

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

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

and $[R_{1,i}]$ and $[R_{2,i}]$ are respectively 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

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

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.

S.1.3.2 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 actually observe $Q_{(2),i}(t)$ which is the angle that corresponds to $R_{(2),i}(t)$. Also let

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

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

$$\begin{aligned}\gamma &= P(R_1 < R_2 \mid R_{(1)}, R_{(2)}) \\ &= \frac{[R_1](R_{(1)})[R_2](R_{(2)})}{[R_1](R_{(1)})[R_2](R_{(2)}) + [R_1](R_{(2)})[R_2](R_{(1)})}.\end{aligned} \tag{S.14}$$

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

$$[Q_{(2),i} \mid R_{(1)}, R_{(2)}](q) = \gamma[Q_{2,i}](q) + (1-\gamma)[Q_{2,i}](\lfloor q + \pi/2 \rfloor_\pi), \tag{S.15}$$

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

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

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

$$[\mathcal{Q}_{(2)} \mid \mathcal{W}, \mathcal{R}_{(1)}, \mathcal{R}_{(2)}](q) = \prod_{i=1}^{M} \prod_{\{j:W_{i,j}=1\}} [Q_{(2)} \mid R_{(1),i}(t_j), R_{(2),i}(t_j)](q_{i,j}) \tag{S.16}$$

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

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

Let $[Q_{2,f}]$ be the same density as in (S.15) only with parameters α_f and β_f in place of α_i and β_i. False alarms are *iid* so

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

S.1.3.3 Intensity Density

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 3.4 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

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

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

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

S.2 Mean and Covariance Calculations

Here we calculate the mean functions $E\{X_i(t)\}$, $E(D_i)$ and the covariance functions $\text{Cov}(X_i(s), X_j(t))$, $\text{Cov}(X_i(s), D_j)$, and $\text{Cov}(D_i, D_j)$. Recall in Section S.1.1.3 that these are calculated before conditioning on any merging events.

Let

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

Also let $n(\mathcal{B})$, $n(\mathcal{S})$ and $n(\mathcal{M})$ be the number elements in these sets respectively. The location equations for a target resulting from birth, splitting and merging events from (3.3), (3.7) and (3.5) are given here again for convenient reference as

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

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,

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

Lastly we recall the expression for the variable $D_i = X_{d_{i,1}}(\xi_{d_{i,3}}) - X_{d_{i,2}}(\xi_{d_{i,3}}) + \psi_{d,i}, i = 1, \ldots, n(\mathcal{M})$.

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

$$\begin{aligned} \mu_i(t) &= E\{X_i(t)\} && \text{(S.17)} \\ \mu_i'(t) &= E\{X_i'(t)\} \\ \gamma^*_{i,j}(s,t) &= \text{Cov}(X_i^*(s), X_i^*(t)) && \text{(S.18)} \\ \gamma_{i,j}(s,t) &= \text{Cov}(X_i(s), X_j(t)) \\ \gamma'_{i,j}(s,t) &= \text{Cov}(X_i(s), X_j'(t)) \\ \gamma''_{i,j}(s,t) &= \text{Cov}(X_i'(s), X_j'(t)) \\ \gamma_i(s,t) &= \text{Cov}(X_i(s), X_i(t)) \\ \gamma_i'(s,t) &= \text{Cov}(X_i(s), X_i'(t)) \\ \gamma_i''(s,t) &= \text{Cov}(X_i'(s), X_i'(t)) \end{aligned}$$

Note that for the purposes of likelihood calculation, we are only interested in the functions given in (S.17) and (S.18) above. However, the expressions for these two functions will depend on the others, so in the following, we will need to derive expressions for all of these functions.

S.2.1 Mean Functions

We can express the mean functions for location for the three cases of birth, splitting and merging events as a recursive formula,

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

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 express the mean velocities for the three cases in a similar manner,

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

Of course the mean of D_i can be written as

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

S.2.2 Covariance Functions

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

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

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

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

so we just need to derive 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 by-product of a splitting and/or merging of the other. Define the indicator $\delta_{i,j}$ to be

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

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 $j \in \mathcal{S}$:

$$\begin{aligned}
\gamma_{i,j}(s,t) &= \text{Cov}\Big(X_i(s)\ ,\ X_{p_{j,1}}(\xi_j) + \psi_{s,j} + \Big[X'_{p_{j,1}}(\xi_j) + \psi'_{s,j}\Big](t - \xi_j) + \sigma_j G_j(t - \xi_j)\Big) \\
&= \gamma_{i,p_{j,1}}(s, \xi_j) + (t - \xi_j)\gamma'_{i,p_{j,1}}(s, \xi_j). \qquad \text{(S.19)}
\end{aligned}$$

Using this same idea, we can calculate the case for $\delta_{i,j} = 1$, $i < j$ and $j \in \mathcal{M}$ as well. If $i < j$ and $j \in \mathcal{B}$ then necessarily $\delta_{i,j} = 0$. This is true because if $i < j$ and $j \in \mathcal{B}$, then because of the way we have organized the indices, $\xi_i \leq \xi_j$. Hence if $j \in \mathcal{B}$ then there is now way that path j or any of its children could have split or merged to create path i since it existed already before path j. Furthermore, path j resulted from a birth so there is also no way that it could be created from path i or any of its children. Since we always decompose the larger index into the contribution from its parents, we will eventually converge to the covariance of a parent(s) that is a birth or initial target and the recursion will terminate.

Hence we have

$$\gamma_{i,j}(s,t) = \begin{cases} \gamma_i(s,t) & \text{if } i=j \\ \gamma_{i,p_{j,1}}(s,\xi_i) + (t-\xi_i)\gamma'_{i,p_{j,1}}(s,\xi_i) & \text{if } \delta_{i,j}=1,\ i<j,\ j\in\mathcal{S} \\ \frac{1}{2}\left(\gamma_{i,p_{j,1}}(s,\xi_i) + \gamma_{i,p_{j,2}}(s,\xi_i)\right) + \frac{t-\xi_i}{2}\left(\gamma'_{i,p_{j,1}}(s,\xi_i) + \gamma'_{i,p_{j,2}}(s,\xi_i)\right) & \text{if } \delta_{i,j}=1,\ i<j,\ j\in\mathcal{M} \\ \gamma_{j,i}(t,s) & \text{if } \delta_{i,j}=1,\ i>j \\ 0 & \text{otherwise.} \end{cases} \tag{S.20}$$

We can also calculate $\gamma'_{i,j}(s,t)$ in the same manner as in (S.19). 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 $i\in\mathcal{S}$. We still need to decompose the larger index into its parents, and we write this as

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

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

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

We can calculate $\gamma''_{i,j}(s,t)$ in the same way as in (S.20), since we again have symmetry in the function:

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

Now for the function $\gamma_i(s,t)$. We can use the same technique in (S.19) but decompose 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

$$\begin{aligned} \gamma_i(s,t) &= \mathrm{Cov}\big(X_i(\xi_i)+X'_i(\xi_i)(s-\xi_i)+\sigma_i G_i(s-\xi_i)\ ,\ X_i(\xi_i)+X'_i(\xi_i)(t-\xi_i)+\sigma_i G_i(t-\xi_i)\big) \\ &= \sigma^2_{X_0}+(s-\xi_i)(t-\xi_i)\sigma^2_{X'_0}+\sigma^2_i \mathrm{Cov}(G_i(s-\xi),G_i(t-\xi)), \end{aligned} \tag{S.21}$$

where for an IBM

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

If target i is a split then we have

$$\begin{aligned} \gamma_i(s,t) &= \mathrm{Cov}\Big(X_{p_{i,1}}(\xi_i)+\psi_{s,i}+\left[X'_{p_{i,1}}(\xi_i)+\psi'_{s,i}\right](s-\xi_i)+\sigma_i G_i(s-\xi_i)\ , \\ &\qquad X_{p_{i,1}}(\xi_i)+\psi_{s,i}+\left[X'_{p_{i,1}}(\xi_i)+\psi'_{s,i}\right](t-\xi_i)+\sigma_i G_i(t-\xi_i)\Big) \\ &= \gamma_{p_{i,1}}(\xi_i,\xi_i)+\sigma^2_{X_s}+(t+s-2\xi_i)\gamma'_{p_{i,1}}(\xi_i,\xi_i)+(s-\xi_i)(t-\xi_i)\left(\gamma''_{p_{i,1}}(\xi_i,\xi_i)+\sigma^2_{X'_s}\right)+ \\ &\quad \sigma^2_i \mathrm{Cov}(G_i(s-\xi),G_i(t-\xi)) \end{aligned} \tag{S.22}$$

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

$$\gamma_i(s,t) = \begin{cases} \sigma^2_{X_0} + (s-\xi_i)(t-\xi_i)\sigma^2_{X'_0} + \sigma_i^2 \mathrm{Cov}(G_i(s-\xi), G_i(t-\xi)) & \text{if } i \in \mathcal{B} \\ \gamma_{p_{i,1}}(\xi_i,\xi_i) + \sigma^2_{X_s} + (t+s-2\xi_i)\gamma'_{p_{i,1}}(\xi_i,\xi_i) + \\ (s-\xi_i)(t-\xi_i)\left(\gamma''_{p_{i,1}}(\xi_i,\xi_i) + \sigma^2_{X'_s}\right) + \sigma_i^2 \mathrm{Cov}(G_i(s-\xi), G'_i(t-\xi)) & \text{if } i \in \mathcal{S} \\ \frac{1}{4}\left(\gamma_{p_{i,1}}(\xi_i,\xi_i) + \gamma_{p_{i,2}}(\xi_i,\xi_i) + 2\gamma_{p_{i,1},p_{i,2}}(\xi_i,\xi_i)\right) + \sigma^2_{X_m} + \\ \frac{s+t-2\xi_i}{4}\left(\gamma'_{p_{i,1}}(\xi_i,\xi_i) + \gamma'_{p_{i,2}}(\xi_i,\xi_i) + \gamma'_{p_{i,1},p_{i,2}}(\xi_i,\xi_i) + \gamma'_{p_{i,2},p_{i,1}}(\xi_i,\xi_i)\right) + \\ \frac{(s-\xi_i)(t-\xi_i)}{4}\left(\gamma''_{p_{i,1}}(\xi_i,\xi_i) + \gamma''_{p_{i,2}}(\xi_i,\xi_i) + 2\gamma''_{p_{i,1},p_{i,2}}(\xi_i,\xi_i)\right) + \\ (s-\xi_i)(t-\xi_i)\sigma^2_{X'_m} + \sigma_i^2 \mathrm{Cov}(G_i(s-\xi), G_i(t-\xi)) & \text{if } i \in \mathcal{M}. \end{cases}$$

Using the same strategy as in (S.21) and (S.22) we can calculate $\gamma'_i(s,t)$ as

$$\gamma_i(s,t) = \begin{cases} (s-\xi_i)\sigma^2_{X'_0} + \sigma_i^2 \mathrm{Cov}(G_i(s-\xi), G'_i(t-\xi)) & \text{if } i \in \mathcal{B} \\ \gamma'_{p_{i,1}}(\xi_i,\xi_i) + (s-\xi_i)\left(\gamma''_{p_{i,1}}(\xi_i,\xi_i) + \sigma^2_{X'_s}\right) + \sigma_i^2 \mathrm{Cov}(G_i(s-\xi), G'_i(t-\xi)) & \text{if } i \in \mathcal{S} \\ \frac{1}{4}\left(\gamma'_{p_{i,1}}(\xi_i,\xi_i) + \gamma'_{p_{i,2}}(\xi_i,\xi_i) + \gamma'_{p_{i,1},p_{i,2}}(\xi_i,\xi_i) + \gamma'_{p_{i,2},p_{i,1}}(\xi_i,\xi_i)\right) + \\ \frac{s-\xi_i}{4}\left(\gamma''_{p_{i,1}}(\xi_i,\xi_i) + \gamma''_{p_{i,2}}(\xi_i,\xi_i) + 2\gamma''_{p_{i,1},p_{i,2}}(\xi_i,\xi_i)\right) + (s-\xi_i)\sigma^2_{X'_m} + \\ \sigma_i^2 \mathrm{Cov}(G_i(s-\xi), G'_i(t-\xi)) & \text{if } i \in \mathcal{M}. \end{cases}$$

For the IBM,

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

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

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

and this completes the description of the covariances.

S.3 Further Details for the Tracking Estimate

In this section we give the details behind the calculation of the conditional density

$$[\mathcal{U}, \mathcal{V}, \mathcal{P} \mid \mathcal{Z} = z](u, v, p) \tag{S.23}$$

used to achieve our tracking estimate in (4.9).

In (S.12) of Section S.1 we have written out the density for

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

which is the collection of all of the variables in the model of Section 3. Here the script letter denotes the collection of those variables for all targets at all times. For example, $\mathcal{X} = \{(X_i(t_j) : j = 1, \ldots, n; i = 1, \ldots, m\}$ is the collection of all x-coordinate values for each target at all times it was observed; see Section S.1.1.3.

Recall that $\mathcal{W}$ is the observability variable and $\mathcal{Y}$ is the y-coordinate. We also let $\mathcal{A}$ denote the collection of all attribute variables we wish to include. For example we might have $\mathcal{A} = (\mathcal{R}_{(1)}, \mathcal{R}_{(2)}, \mathcal{Q}_{(2)}, \mathcal{I})$, which are the smallest radius, largest radius, angle of orientation and intensity for targets respectively. Recall that the variable $N_f(t)$ is the number of false alarms at time t so that $\boldsymbol{N}_f = (N_f(t_1), \ldots, N_f(t_n))$ contains the number of false alarms at each time point. The remaining variables $\mathcal{X}_f$, $\mathcal{Y}_f$, and $\mathcal{A}_f$ are the collection of x and y coordinates for false alarms and attributes for false alarms respectively.

We will use the density for $(\mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}, \boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f)$ to calculate the density in (S.23). Notice that there is a one–to–one mapping

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

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

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

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

$$\begin{aligned} [\mathcal{U}, \mathcal{V}, \mathcal{P} \mid \mathcal{Z}](u, v, p \mid z) &= P\{\mathcal{U} = u, \mathcal{V} = v, \mathcal{P} = p \mid \mathcal{Z} = z\} \\ &= P\{\mathcal{U} = u, \mathcal{V} = v, (\mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}, \boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f) = g^*(p, z) \mid \mathcal{Z} = z\} \\ &= [\mathcal{U}, \mathcal{V}, (\mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}, \boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f) \mid \mathcal{Z}](u, v, g^*(p, z) \mid z). \qquad \text{(S.24)} \end{aligned}$$

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

$$[\mathcal{Z} \mid \mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{A}, \boldsymbol{N}_f, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f](z \mid u, v, w, x, y, a, n_f, x_f, y_f, a_f) = \frac{1}{\prod_{j=1}^{n} m_j!} I_B(z), \qquad \text{(S.25)}$$

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 calculate 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 (S.12) by that in (S.25). To then obtain the density in (S.24), note that for a given value of $\mathcal{Z} = z$,

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

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

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

There is actually a finite number of values of $(\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{X}_f, \mathcal{Y}_f, \mathcal{A}_f)$ since they must be a permutation of the values in $\mathcal{Z}$ at each time. But there could be as many as a countable number combinations of births, deaths, splitting and merging events that could be represented by $\mathcal{U}$ and $\mathcal{V}$. This means that we must have

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

where the second equality comes from the fact that the contribution of $\mathcal{Z}$ to the density is a constant by (S.25). Now by equation (S.24) we have

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

where $\{(u_j, v_j, p_j) : j = 1, 2, \dots\}$ is an enumeration of the possible tracking solutions.

As discussed in Section 4 we also wish to calculate the conditional density of $(\mathcal{U}, \mathcal{V}, \mathcal{P})$ given $\mathcal{Z} = z$ and the event $(\mathcal{U}, \mathcal{V}, \mathcal{P}) \in \mathcal{K}$, where $\mathcal{K} = \{(u_i, v_i, p_i) : i = 1, \dots, K\}$. This is given by

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

S.4 Full Description of the Modified MHT Algorithm

As mentioned in Section 4.3, when we receive a new set of observations, $\boldsymbol{Z}_j = (Z_1(t_j), \dots, Z_{m_j}(t_j))$, at time t_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.

Existing tracks that do not receive a new observation to continue the track at time t_j must either

1. go missing (stay missing), or
2. terminate.

At time t_1 we consider all combinations of each observation treated as an initial target observation or a false alarm. Now assume that we have a set of solutions (hypotheses) for the observations through time t_{j-1}. We then take the new observations, $\boldsymbol{Z}_j$, at time t_j and form updated solutions based on all possible combinations of the possibilities listed above. We then hold on to only a subset of these new solutions (those 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 likelihood of the best solution at time t_j. At each time, t_j, we hold on to all solutions that have likelihood greater than $c \max\{L_j\}$ where $c < 1$ is a user-defined parameter. In the interest of speed, we also set a limit, K_s, for the maximum number of solutions that make it through to the next time. The control parameters c and K_s will vary depending on the complexity of the problem. In the problems of Sections 5 and 6 these were set to $c = e^{-10}$ and $K_s = 200$.

Of course it is very inefficient to examine all possible combinations at each time, so we form gates for each of the tracks. A gate is a prediction region for a new observation from a track at time t_j given the previous observations assumed to be part of the track. In Sections 5 and 6 we used a confidence level of $p_g = 0.9999$ for the gate or prediction region. We then limit the possible observations for inclusion into a track to only those that fall into the gate 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 form a prediction region for $(X_i(t_j), Y_i(t_j)) + (\psi_{X,s}, \psi_{Y,s})$. Recall that $\psi_{X,s}$ is the random error term for the amount the child's x-location will be different from the parent's at the time of split and similarly for $\psi_{Y,s}$. This can be accomplished by simply adding $\sigma^2_{X_s}$ and $\sigma^2_{Y_s}$ 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 this inflated variance. We only consider pairs of new observations within this region to possibly be a split pair from the existing track.

For a possible merging event, 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 $(X_1(t_j) - X_2(t_j), Y_1(t_j) - Y_2(t_j)) + (\psi_{X,d}, \psi_{Y,d})$. Recall $\psi_{X,d}$ is the random distance between the parents at the time of a merging event. If the prediction region for this quantity includes zero, we will consider possibility that these targets are the parents in a merging event.

Now suppose targets 1 and 2 can be considered as parents for a possible merging event. We must also find an observation to possibly 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 merging of tracks 1 and 2. Note that these prediction regions assume that the merging event takes place at t_j when it really would have taken place at some time in the interval (t_{j-1}, t_j), but this seems to be adequate provided the time points are not too spread out.

The prediction regions described above can be calculated efficiently via the Kalman Filter by ignoring the dependency resulting from merging and splitting. That is, the first observation is assumed fixed and the others are calculated assuming the model given in (3.3) and independent of other tracks. It is possible to improve these regions by using the innovations algorithm to compute the conditional distribution of new observations from a track taking into account the previous splitting and merging. This would also be more time consuming however, since the actual covariances take longer to compute than those under the assumed independence of tracks.

In addition to gating, we can usually separate the entire tracking problem into several smaller tracking problems that are "disjoint" from each other. That is, there are often situations where the area (in space and time) that one group of targets occupies does not intersect the area that another group of targets occupies. These groups of observations can be identified with a simple heuristic approach and the algorithm described above can then be applied to each group separately.

We now discuss a way to improve upon the approximate solution provided by the MHT. 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. Consider the following situation. The solution that would eventually be optimal (have the highest likelihood) at time t_n has a likelihood that is not very high early on in the algorithm when considering only a subset of all the times. We can only hold on to a limited number of possible 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. 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 and Jain (1987). 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 many other changes that we could consider making to improve the performance of the algorithm, but from initial results of the MHT, it seemed to do a very good job of classifying the splitting and merging events correctly according to likelihood, 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 only of false alarms. This is likely because it had to discard the correct solution, before it realized it would have to pay a penalty when it eventually killed this incorrect track after a short time. In any case, the greedy exchange algorithm we use here only considers the possibility of changing short tracks ($\leq k$ observations) into false alarm observations. For the results in Sections 5 and 6 we set $k = 3$. Considering other possible changes in the greedy exchange step would only improve results.

So for each of the solutions produced when the MHT finishes, 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.

S.5 Details on Parameter Estimation

In this section we describe how we estimate the parameters of the model given in Section 3. In most cases these estimates are the maximum likelihood estimates (MLE's). In some cases however, the MLE would be too computationally expensive to compute, and we will use other reasonable choices for estimates.

Also for first few time points of the MHT algorithm, some of the estimates given here cannot be computed because there are two few data points. In these cases, we need an initial guess for some of the parameter values. Here 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.

S.5.1 Parameters of the Event Model

The Event Model parameters are λ_0, λ_b, λ_d, λ_s and λ_m. There is also the false alarm rate parameter, λ_f. For the Event Model parameters one can calculate approximate MLE's based on the approximate likelihood given in Section S.1.1.1. The MLE for λ_0 is obviously

$$\hat{\lambda}_0 = N_0,$$

where recall N_0 is the initial number of targets.

Now consider the estimation of the death rate, λ_d. From the approximation in (S.4) 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 (S.5) from $\boldsymbol{U}_d$ is

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

So the derivative of the log likelihood is

$$\frac{d}{d\lambda_d} \log \left[\boldsymbol{U}_d \mid \bar{\boldsymbol{N}}\right](\boldsymbol{u}) = \sum_{j=1}^{n} \frac{u_j}{\lambda_d} - \bar{N}_j \Delta t_j. \tag{S.26}$$

Setting (S.26) equal to zero gives

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

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

$$\begin{aligned}\hat{\lambda}_b &= \frac{\sum_{j=1}^n U_{b,j}}{\sum_{j=1}^n \Delta t_j} \\ \hat{\lambda}_d &= \frac{\sum_{j=1}^n U_{s,j}}{\sum_{j=1}^n \bar{N}_j \Delta t_j} \\ \hat{\lambda}_d &= \frac{\sum_{j=1}^n U_{m,j}}{\sum_{j=1}^n (\bar{N}_j - 1) \Delta t_j}.\end{aligned}$$

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 λ_f so the MLE for λ_f is

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

S.5.2 Parameters of the Observability Model

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

$$\hat{P}_d = \frac{\sum_{i=1}^M \sum_{j=1}^n W_i(t_j) I_{[\xi_i,\zeta_i]}(t_j)}{\sum_{i=1}^M \sum_{j=1}^n I_{[\xi_i,\zeta_i]}(t_j)}.$$

S.5.3 Location Parameters

The derivative of the location density is difficult to compute analytically because of the matrix algebra involved. Exact MLE's would then 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 notational changes.

S.5.3.1 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

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

So if we make a derivative approximation, we have

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

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

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

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

$$\text{Cov}(D_j^2, D_{j+1}^2) \approx K_j \sigma^2_{X_e},$$

where

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

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

$$\hat{\sigma}^2_{X_e} = \frac{1}{N} \sum_{i=1}^{M} \sum_{j \in O_i} \frac{D_{i,j}^2 D_{i,j+1}^2}{K_j}, \tag{S.27}$$

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 (S.27).

S.5.3.2 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 for σ_i^2.

The criterion for selection of the bandwidth h will based on the following rule presented on pages 100-101 of Schimek (2000). 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 Schimek (2000), the prediction risk is

$$E\left[\sum_{j=1}^{n}(X^*(t_j) - \hat{m}(t_j,h))^2\right] = E\left[\sum_{j=1}^{n}(X(t_j) - \hat{m}(t_j,h))^2\right] + \sigma_{X_e}^2(n - 2\text{tr}(S)) \tag{S.28}$$

so

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

Since it is our goal to minimize the estimation risk which is the first term on the right side of (S.28), we will use the bandwidth, h, that minimizes the quantity

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

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 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

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

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

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

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

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

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

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

where C_j is given by

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

Hence a method of moments estimate for σ_i^2 is

$$\hat{\sigma}_i^2 = \frac{1}{N} \sum_{j \in O_i} \frac{(D_{i,j}^2)^2}{C_j} \tag{S.29}$$

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 (S.29).

Again, we only estimate σ_i^2 in this way if we have greater than $k = 6$ 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.

S.5.3.3 Initial Conditions Parameters

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

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

If the i^{th} path has fewer than $k = 6$ 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 $\mathcal{B} = \{i : \text{target } i \text{ is a an initial target or a birth}\}$, and let $n(\mathcal{B})$ be the number elements in $\mathcal{B}$. We can construct estimates for the initial conditions parameters as

$$\begin{aligned}
\hat{\mu}_{X_0} &= \frac{1}{n(\mathcal{B})} \sum_{i \in \mathcal{B}} \hat{X}_i(\xi_i) \\
\hat{\sigma}^2_{X_0} &= \frac{1}{n(\mathcal{B})} \sum_{i \in \mathcal{B}} \left(\hat{X}_i(\xi_i) - \hat{\mu}_{X_0}\right)^2 \\
\hat{\mu}_{X_0'} &= \frac{1}{n(\mathcal{B})} \sum_{i \in \mathcal{B}} \hat{X}_i'(\xi_i) \\
\hat{\sigma}^2_{X_0'} &= \frac{1}{n(\mathcal{B})} \sum_{i \in \mathcal{B}} \left(\hat{X}_i'(\xi_i) - \hat{\mu}_{X_0'}\right)^2.
\end{aligned}$$

S.5.3.4 Splitting and Merging Parameters

Here we will construct estimates for the parameters involved in the initial conditions of splitting or merging events, 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

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

Adopt the convention of Section 3.3 and denote the indices 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 child of a splitting event. If we let $\mathcal{S} = \{i : \text{target } i \text{ is the child of a splitting event}\}$ and $n(\mathcal{S})$ be the number elements in $\mathcal{S}$, then we can construct estimates for these parameters as

$$\begin{aligned}
\hat{\sigma}^2_{X_s} &= \frac{1}{n(\mathcal{S})}\sum_{i\in\mathcal{S}}\left(\hat{X}_{p_{i,1}}(\zeta_{p_{i,1}}) - \hat{X}_i(\xi_i)\right)^2 \\
\hat{\sigma}^2_{X'_s} &= \frac{1}{n(\mathcal{S})}\sum_{i\in\mathcal{S}}\left(\hat{X}'_{p_{i,1}}(\zeta_{p_{i,1}}) - \hat{X}'_i(\xi_i)\right)^2.
\end{aligned}$$

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 child of a merging event. So let $\mathcal{M} = \{i : \text{target } i \text{ is the child of a merging event}\}$ and we can construct estimates of these parameters as

$$\begin{aligned}
\hat{\sigma}^2_{X_m} &= \frac{1}{n(\mathcal{M})}\sum_{i\in\mathcal{M}}\left(\frac{1}{2}\hat{X}_{p_{i,1}}(\zeta_{p_{i,1}}) + \frac{1}{2}\hat{X}_{p_{i,2}}(\zeta_{p_{i,2}}) - \hat{X}_i(\xi_i)\right)^2 \\
\hat{\sigma}^2_{X'_m} &= \frac{1}{n(\mathcal{M})}\sum_{i\in\mathcal{M}}\left(\frac{1}{2}\hat{X}'_{p_{i,1}}(\zeta_{p_{i,1}}) + \frac{1}{2}\hat{X}'_{p_{i,2}}(\zeta_{p_{i,2}}) - \hat{X}'_i(\xi_i)\right)^2.
\end{aligned}$$

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 child of a merging event. So its estimate is given by

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

S.5.4 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 3.4. So the constraints are that

$$E(S_i) + E(S_{i+1}) = E(S_{p_{i,1}}) \tag{S.30}$$

if targets i and $i+1$ are the children of a splitting event and

$$E(S_i) = E(S_{p_{i,1}}) + E(S_{p_{i,2}}) \tag{S.31}$$

if target i is the child of a merging event.

A brief overview of the plan here is to first estimate the mean size for each target, $E(S_i)$, under the constraints above. Then estimate the scale parameter, $\sigma^2_{S_i}$, for S_i. We will use these to obtain 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 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 $E(S_i)$, 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 $\mathrm{Var}(S_i)$ 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 $E(S_i)$ that minimize

$$\sum_{i=1}^{M} \sum_{j=1}^{n_i} \frac{1}{\mathrm{Var}(S_i)} \{S_i(t_{i,j}) - E(S_i)\}^2 \tag{S.32}$$

subject to the constraints in (S.30) and (S.31). This is carried out using the Lagrangian Multiplier method. Denote the resulting minimizers of expression (S.32) as $\widehat{E}(S_i)$.

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

$$E(S_i) = e^{\mu_{S_i} + \frac{1}{2}\sigma^2_{S_i}},$$

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

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

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 (S.13) 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.

S.5.5 Orientation Parameters

For the estimation of the angle of orientation parameters, α_i and β_i, we again use maximum likelihood. Recall from (S.14) and (S.15) 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 S.5.4 into the density for $\mathcal{Q}$ given in (S.16). We then again use Newton Raphson to find the values of α_i and β_i that maximize the likelihood given in (S.16).

S.6 Detailed Simulation Results

In this section, 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 Section 3. 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 Section 6. 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 3.3. 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 scalars multiplied to $G_i(t)$ in equations (3.3), (3.5), and (3.7) for $X_i(t)$ and $Y_i(t)$ respectively.

All of these simulations allow for false alarms to appear at each time with rate $\lambda_f = 8.0$ so we can expect about 8 false alarms at each time. We also set the probability of detection $P_d = 0.95$. The parameters λ_0, 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 $\lambda_0 \in [0, 25]$, $\lambda_f \in [0, 25]$, $\lambda_b \in [0.001, .25]$, $\lambda_d \in [0.001, .15]$, $\lambda_s \in [0.001, .15]$, $\lambda_m \in [0.001, .15]$, $P_d \in [0.5, 1.0]$, $\mu_{X_0} \in [-120, -85]$, $\sigma^2_{X_0} \in [500, 1000]$, $\mu_{X'_0} \in [0, 5]$, $\sigma^2_{X'_0} \in [0.001, 5.0]$, $\sigma^2_i \in [0.001, 10.0]$,

$\sigma^2_{X_s} \in [0.001, 1.5]$, $\sigma^2_{X'_s} \in [0.0, 1.0]$, $\sigma^2_{X_m} \in [0.001, 0.5]$, $\sigma^2_{X'_m} \in [0.0, 1.0]$, $\sigma^2_{X_d} \in [0.001, 5.0]$, $\sigma^2_{X_e} \in [0.0, 1.0]$, $\mu_{Y_0} \in [25, 50]$, $\sigma^2_{Y_0} \in [500, 1000]$, $\mu_{Y'_0} \in [-5, 5]$, $\sigma^2_{Y'_0} \in [0.5, 10.0]$, $\eta^2_i \in [.001, 10.0]$, $\sigma^2_{Y_s} \in [0.001, 1.5]$, $\sigma^2_{Y'_s} \in [0.0, 1.0]$, $\sigma^2_{Y_m} \in [0.001, 0.5]$, $\sigma^2_{Y'_m} \in [0.0, 1.0]$, $\sigma^2_{Y_d} \in [0.001, 5.0]$, and $\sigma^2_{Y_e} \in [0.0, 1.0]$.

There are six cases that we considered here:

(i) Birth only For this simulation 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 simulation to the set of realizations that have at least one birth event.

(ii) 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 our focus to the set of realizations that had at least one death. We then set $\lambda_b = \lambda_s = \lambda_m = 0$.

(iii) Splitting only In the splitting only simulations, we forced there to be exactly one target that split into two targets at a random uniformly distributed time in the interval $(1.0, 8.0)$.

(iv) Merging only In a similar manner to the splitting only simulations, the merging only simulations, have exactly one merger by two targets at a uniformly distributed time in the interval $(1.0, 8.0)$.

(v) Completely Random These are completely unrestricted realizations from the model with event parameters set as $\lambda_0 = 4$, $\lambda_b = 0.1$, $\lambda_d = .02$, $\lambda_s = 0.06$, and $\lambda_m = .08$.

(vi) Completely Random w/ Size These are the same realizations as in case (v) but now with size information to be used in the tracking algorithm. The radius variables $R_{(1)}$ and $R_{(2)}$ are being used along with location here to compute the likelihood.

For each realization we would generate two random variables $z_1 \sim \mathcal{N}(0.6, .01)$, $z_2 \sim \mathcal{N}(0.8, .01)$ 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, $\sigma^2_{R_{1,i}} = \sigma^2_{R_{2,i}} = 0.025$ 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.

We set the false alarm size parameters to $\mu_{R_{1,f}} = 0.00$, $\mu_{R_{2,f}} = 0.25$, and $\sigma^2_{R_{1,f}} = \sigma^2_{R_{2,f}} = 0.25$. This produces false alarms that are smaller than targets on average, but possibly similar in size to small or medium size targets.

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 $\boldsymbol{t} = (0, 1, \ldots, 9)$. An example of a realization from the completely random (CR) model was given in Figure 5 in the main article. We wish to investigate the same hypotheses 1-4, posed in the previous section now with the presence of clutter (i.e., false alarms).

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 respectively, 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).

S.6.1 Simulation Results for Cases (i)-(vi)

The simulation results of each of the six cases are given as the columns of Table S.1. In the following we describe each of the summary statistics that make up the rows of Table S.1.

		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	-	-	-	83.1	100.0
% Deaths Correct		-	79.2	-	-	70.2	95.7
% Splits Correct		-	-	93.0	-	87.0	100.0
% Mergers Correct		-	-	-	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%)	0.881	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 S.1: Results of 100 Realizations With Clutter

% Best Est Correct This is the percentage of times that $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$ from (4.12) 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% CS 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 posterior probability that the correct solution, $(\mathcal{U}, \mathcal{V}, \mathcal{P})$, is correct calculated using (4.13). 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 (4.9) is correct given the data, again 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 in the paragraph below.

% Targets Correct This is the percentage of all targets at all times 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.

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 present a definition of track purity that is slightly different than that given by Mori, Chang, Chong, and Dunn (1986). In the correct solution, $(\mathcal{U}, \mathcal{V}, \mathcal{P})$, consider a given track i composed of observations produced by target i. Of all the tracks that make up the estimate $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$, find the track i' that contains the most observations in common with track i in $(\mathcal{U}, \mathcal{V}, \mathcal{P})$. The track purity for track i is defined to be the proportion of the observations that make up track i that are also part of track i' in the estimate $(\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{P}})$. The overall track purity is then 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 $[5(1.0) + 10(0.7)]/15 = 0.8$.

Now refer back to the four hypotheses we posed earlier. 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 S.1 we can see that the percentage of births labeled correctly in the birth only simulation (94.8%) is somewhat higher than that in the CR model (83.1%). The percentage of deaths labeled correctly in the death only case (79.2%) is also a bit higher than that for the CR model (70.2%). The percentage of splits correct in the splitting only case is closer to the CR model, (93.0%) versus (87.0%). Lastly the percentage of mergers correct in the merging only case is also slightly higher than in the CR model, (97.0%) to (90.7%). So it appears that these rates for the first four simulations are in general a little bit higher than those for the CR model simulation.

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 again likely due to 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 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 a good indication 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 that 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 simulations 1-6 respectively) dropped off some from the simulations without clutter. One explanation for this, referring back to (4.10), 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, then of course our distribution of solutions given in (4.10) 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 provide us with at least a rough guide as to how confident we should be 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.

S.6.2 Decreasing Time Increments

The set of simulations considered in the section uses a model identical to that of the CR model realizations with clutter of Section S.6.1. 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 there is a convergence of the estimate to the correct solution as the time increment approaches zero.

From Table S.2 we can see that the estimation does improve substantially as Δt becomes 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 appears as though there is a convergence of this estimate to the correct solution.

	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%)	0.856	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 S.2: Results of 100 Realizations With Decreasing Δt

S.7 Detection Algorithm

The problem of target or object identification in images has been studied quite thoroughly. It is not our goal to make a contribution in this area, hence a detailed description of these techniques will not be given. We simply 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 Rosenfeld and Kak (1982).

Recall, the goal of the detection algorithm is to go through each image and record the location of each target (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 $I_{i,j} < \alpha$ will be 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. There are two common definitions of connected pixels. Two pixels are *4-connected* if they share one of their 4 sides with each other. Two pixels are *8-connected* if they 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 fit an ellipse to each target (cluster of pixels). This can 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 for a given target are given by

$$\hat{\mu}_x = \sum \frac{I_{i,j} x_{i,j}}{\sum I_{i,j}}$$

and similarly for $\hat{\mu}_y$, $\hat{\sigma}_x$, $\hat{\sigma}_y$, and $\hat{\sigma}_{xy}$ where the sum is taken over the pixels (i, j), that compose that target.

The location of the target is then given by $(\hat{\mu}_x, \hat{\mu}_y)$. The length of the radii $R_{(1)}$ and $R_{(2)}$ of the target are given by the minor and major axes of the 99% contour ellipse. The angle of orientation $Q_{(2)}$ is also obtained from the ellipse. Refer to Figure 4 for an illustration of this.

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 $R_2 > 1^\circ$. All other storms are discarded, so this could be considered a second stage of thresholding.