Tracking Promyelocytic Leukemia Nuclear Bodies for Disease Detection: a Proof of Concept via Filtering

J. Hailes, M.A. Kouritzin, R. Marsh, J. McCrosky
MITACS PINTS Centre, Dept. of Mathematics, University of Alberta, Edmonton, Canada
jarett.hailes@randomknowledge.net, mkouritz@math.ualberta.ca, marshr@macewan.ca, mccrosky@math.ualberta.ca

Abstract

Nonlinear filtering has proven to be a valuable tool in many applications ranging from search and rescue to performing arts. Herein we discuss an application of filtering to medical imaging. A sample problem in detecting, tracking, and characterizing the movement of promyelocytic leukemia nuclear bodies within a cell nucleus is used to demonstrate the utility of nonlinear filtering in this area. Simulation results are presented and discussed.

1. Introduction

Promyelocytic leukemia (PML) nuclear bodies are subnuclear domains in the eukaryotic cell nucleus [4]. Over 15 major components have been identified within PML nuclear bodies, most notably the promyelocytic leukemia protein itself [5]. PML bodies have been implicated in both oncogenesis [6] and viral infection [7], and they have been found to be lacking in acute promyelocitic leukemia (cancerous) cells. While there are multiple theories about the primary function(s) of PML bodies, there is currently no definitive model for either their purpose or their movement within the nucleus of a cell [4]. Herein, we utilize nonlinear filtering (1) to determine the validity of stochastic process motion models for PML body movement and (2) to identify the number and behaviour of PML bodies within a specified region of the nucleus of a cell. It is believed that this behaviour may eventually be useful in detecting the onset of disease.

2. Nonlinear filtering

Filtering theory is an active research field with a wide range of applications in areas as diverse as signal processing, communication and information networks, and search and rescue. The objects forming the signal or target we wish to track are described and modeled by stochastic dynamical systems which cannot be observed directly or completely. We can only obtain some partial, distorted, and corrupted observations of their current states. The purpose of filtering is to find probabilistic knowledge of the past path, current state, or future changes of the signals in real time based on the back observations. Formally, we wish to estimate the probability of the state of our signal X at some time t given a series of discrete observations Y:

$$P(X_t \in A \mid \sigma\{Y_{t_i}, i = 1,...,k\}) \quad \textbf{(1)}$$

For nonlinear filtering problems, it is usually impossible to find exact finite-dimensional recursive solutions like the Kalman filter. Instead, to obtain a computationally feasible solution, some kind of approximation is required. Particle filters are particularly useful since they are applicable to general Markov models, are easy to implement, and are scalable based on the availability of computer resources. Particle filters use independent copies of the signal called particles to construct an approximation that converges asymptotically to the exact filtering solution as the number of particles used increases. Specifically, such particle filters approximate equation **(1)** using a weighted average of particle positions, according to

$$P(X_t \in A \mid \sigma\{Y_j, j = 1,...,k\}) \approx \frac{\sum_{i=1}^{N} W_t^i 1_{X_t^i \in A}}{\sum_{i=1}^{N} W_t^i} \quad \textbf{(2)}$$

where X_t^i is the ith particle at time t and W_t^i is its assigned weight based upon the observations.

The classical particle filter method (Monte-Carlo method) is composed of a two stage process. First, particles are evolved between observations according to the signal state equation. Second, weights for each particle are recalculated once a new observation arrives, based on the new information as well as the previous weight. Unfortunately, the resulting particle scheme is not effective; because there is no attempt to adapt the particle locations to the signal and most of the particles

tend to become unrepresentative of the real signal. In the past decade, a resampling step has been added to the classical scheme by adapting the particles to observations. However, resampling can add unnecessary randomness to the filtering system, thereby degrading performance. A highly versatile and effective solution is the Selectively Resampling Particle (SERP) filter. The SERP filter controls the extent of resampling by ensuring the range of weights for all particles remains within a specified range $\rho > 1$ [2]. This parameter is selected based on the particular characteristics of the problem, and may be time variant. There are several computational benefits associated with the SERP filter, since the algorithm maintains the number of particles in the system at a constant number and utilizes efficient data structures to maintain the information concerning particle weights and locations. The SERP filter has been shown [1,3] to be extremely powerful in tracking multi-target signals, such as a collection of PML bodies.

The basic SERP algorithm is as follows:

1. Initialize all particles $\left\{X_0^{\,i}\right\}_{i=1}^{N}$ with a given distribution over the problem domain D.
2. For k = 1, 2, …
   a) Evolve all particles over time interval $\tau = t_k - t_{k-1}$. Call the particles just prior to observation Y_{t_k} $\left\{X_{t_k^-}^{\,i}\right\}_{i=1}^{N}$.
   b) Calculate the weights $\tilde{W}_k^{\,i}$ for $\left\{X_{t_k^-}^{\,i}\right\}_{i=1}^{N}$ using only the kth observation and then update each particle's weight according to $W_{t_k}^{\,i} = \tilde{W}_k^{\,i} W_{t_{k-1}}^{\,i}$.
   c) Resample until $W_{t_k}^{\,i} < \rho\, W_{t_k}^{\,j};\ \forall\ i,\ j$.

As demonstrated in [3] the SERP filter can be used to select between different signal models and arrive at a probability indicating the relative likelihood of each model being correct. This is extremely useful when identifying valid motion models for processes when little empirical or theoretical background is available.

3. Problem description

Using the SERP filter, we wish to determine the number and current locations of PML bodies located within a region of a cell nucleus. The domain of this problem is the nucleus of a cell, which is roughly an ellipsoid with an approximate diameter of 1.7 μm. For simplification purposes, the nucleus is modeled here as a 3-dimensional spherical region $D \subset \Re^3$. Included in the domain are dense regions of chromatin, which possibly influence the motion of the PML bodies. The chromatin are modeled as cylindrical structures aligned along the z-axis of the domain with a radius of 0.1 domain units and a height that extends from the bottom to the top of the sphere. Between 3 and 5 chromatin are placed within the domain at pre-specified points. Another possible region within the cell that may affect the movement of the PML bodies are 'splicing factors'. These factors are modeled as polygons of arbitrary size. All other materials found within the cell nucleus are deemed to be inconsequential to the evolution of the signal.

3.1. Signal model

Within the domain there will be zero or more PML targets. In our initial work, each target is assumed to move independently in order to simplify the modeling of the signal. A single PML target $\overline{X}_t \in D$ is modeled as a spherical object 0.05 domain units in diameter. Each target has a three dimensional state (x_t, y_t, z_t) representing its position within the domain.

It has been noted that PML bodies seem to avoid passing through chromatin as well as the nuclear wall, but otherwise move in a random fashion. In addition, these movements possibly become less energetic when close to the wall of the cell nucleus. Therefore, it is reasonable to model each target as a random process governed by the following Ito stochastic differential equations (SDEs):

$$
\begin{aligned}
dx_t^i &= \sigma dW_t^{\,x}\left(1-\sqrt{(x_t^i)^2+(y_t^i)^2+(z_t^i)^2}\right)^5 \\
&\quad -\beta\, x_t^i\sqrt{(x_t^i)^2+(y_t^i)^2+(z_t^i)^2} \\
&\quad -\sum_{j=1}^{m}\frac{\delta(\overline{C}_x^{\,j}-x_t^i)\left(2-\sqrt{(x_t^i-\overline{C}_x^{\,j})^2+(y_t^i-\overline{C}_y^{\,j})^2}\right)^4}{\sqrt{(x_t^i-\overline{C}_x^{\,j})^2+(y_t^i-\overline{C}_y^{\,j})^2}} \\
dy_t^i &= \sigma dW_t^{\,y}\left(1-\sqrt{(x_t^i)^2+(y_t^i)^2+(z_t^i)^2}\right)^5 \\
&\quad -\beta\, y_t^i\sqrt{(x_t^i)^2+(y_t^i)^2+(z_t^i)^2} \\
&\quad -\sum_{j=1}^{m}\frac{\delta(\overline{C}_y^{\,j}-y_t^i)\left(2-\sqrt{(x_t^i-\overline{C}_x^{\,j})^2+(y_t^i-\overline{C}_y^{\,j})^2}\right)^4}{\sqrt{(x_t^i-\overline{C}_x^{\,j})^2+(y_t^i-\overline{C}_y^{\,j})^2}} \\
dz_t^i &= \sigma dW_t^{\,z}\left(1-\sqrt{(x_t^i)^2+(y_t^i)^2+(z_t^i)^2}\right)^5 \\
&\quad -\beta\, z_t^i\sqrt{(x_t^i)^2+(y_t^i)^2+(z_t^i)^2}
\end{aligned}
\qquad \textbf{(4)}
$$

where W_t^x, W_t^y, W_t^z are independent standard Brownian motions, $\overline{C}_x^j, \overline{C}_y^j$ are the spatial coordinates of the point of the j^{th} Chromatin positioned within the domain, and $\beta, \gamma, \delta,$ and σ are constant parameters. A signal is defined as a collection of *n* targets $X_t = \{\overline{X}_t^1, ..., \overline{X}_t^n\}$.

It is also possible that PML targets alter their movement when interacting with 'splicing factors' within the cell. These materials alter the amount of randomness and interacting with both the chromatin and the edge wall. In these circumstances, PML targets will have two sets of parameters: $(\beta_1, \delta_1, \sigma_1)$ are used when a target is outside of a splicing factor, while $(\beta_2, \delta_2, \sigma_2)$ are used within the splicing factor.

3.2. Observation model

Observations on the random system are taken via a digital camera suspended above the nucleus of a cell. The camera takes photos of the system at periodic intervals. In order to distinguish between the various elements within the nucleus, the cell has been injected with dies which facilitate identification. Chromatin appear as blue objects, the PML bodies are red and all other objects, including the splicing factors, appear as various shades of green. The mathematical observation utilized in the filtering process is the intensity of red from each pixel of the digital image; an integer value ranging from 0 to 255. In real life, the digital images taken are 'noisy', in that they are potentially corrupted by external light sources (which can appear red on the image) as well as the fact that the cell is not stationary in the contact solution. The resultant equation is:

$$Y_{t_k}^{i,j} = h^{(i,j)}(X_t) + V_k \qquad \textbf{(5)}$$

where $h^{(i,j)}(X_t)$ = 128 if a target is over pixel (*i,j*) and 0 otherwise, and V_k is a zero-mean Gaussian random variable with a standard deviation of σ_V. Furthermore, an additional three dimensional Brownian motion term is added to the signal's current state at time *t* prior to the sensor function *h* in order to simulate the movement of the cell nucleus.

4. Filtering simulations

The goal of the simulations is to utilize the SERP filter to determine which SDE model most likely matches the true motion of the PML bodies given in a series of observations. Each model specifies two different characteristics: how many targets are present within the domain and the effect of splicing factors on the movement of the targets. The latter binary option is dependent on the values of the second set of parameters, $(\beta_2, \delta_2, \sigma_2)$. In the case that the splicing factors do not affect PML movement, they are identical to $(\beta_1, \delta_1, \sigma_1)$. Otherwise, they will differ.

The following parameters were set experimentally based on initial observations regarding the nature of PML bodies. Outside of the splicing factors each model utilizes the same set of parameters $(\beta_1, \delta_1, \sigma_1)$ = (0.3, 0.001, 0.3). Inside of the splicing factors the second model utilizes the parameters $(\beta_2, \delta_2, \sigma_2)$ = (0.0003, 0.00001, 0.0003) while the first model utilizes $(\beta_2, \delta_2, \sigma_2)$ = (0.3, 0.001, 0.3). A value of σ_V = 120 is used for the observation noise. Observations are received at periodic intervals with a time step of *dt* = 0.01. The number of targets is fixed at 2, limiting the number of models under consideration by the filter to two, one for the case where splicing factors affect PML movement, and one for the case where it does not. The total number of particles used by the model selection SERP filter is 350,000, divided amongst the different model options. Simulations are run for 50 iterations, or a total simulation time of 0.5. Two large triangular splicing factors in between the chromatin are placed within the nucleus.

In order to measure the accuracy of the SERP filter, the probability of a given model being correct is considered for a particular model. The time to determination of the correct model and target count, as well as the number of times the correct determination occur over several runs are considered. Depending on the desired threshold of certainty (i.e. the likelihood of a model being correct), the fidelity and time to localization varies.

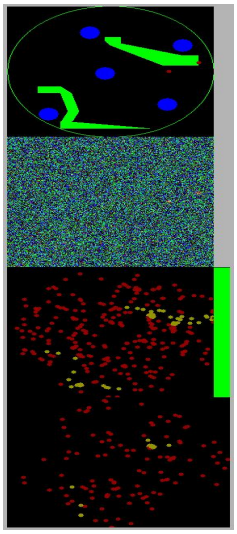

Figure 1. Simulation graphics

5. Results

Two sets of simulations were run; in one simulation the first model is used as the true signal while in the other the second model is employed. Figure 2 depicts the average probability of the signal model being correctly identified by the filter over the simulated time frame. These results demonstrate that the SERP filter is able to determine which model most accurately fits the filter with a high probability within the specified simulation time frame. Table 1 demonstrates the proportion of trial runs that the SERP filter selected the correct model upon completion of the simulation, given a certain probability threshold.

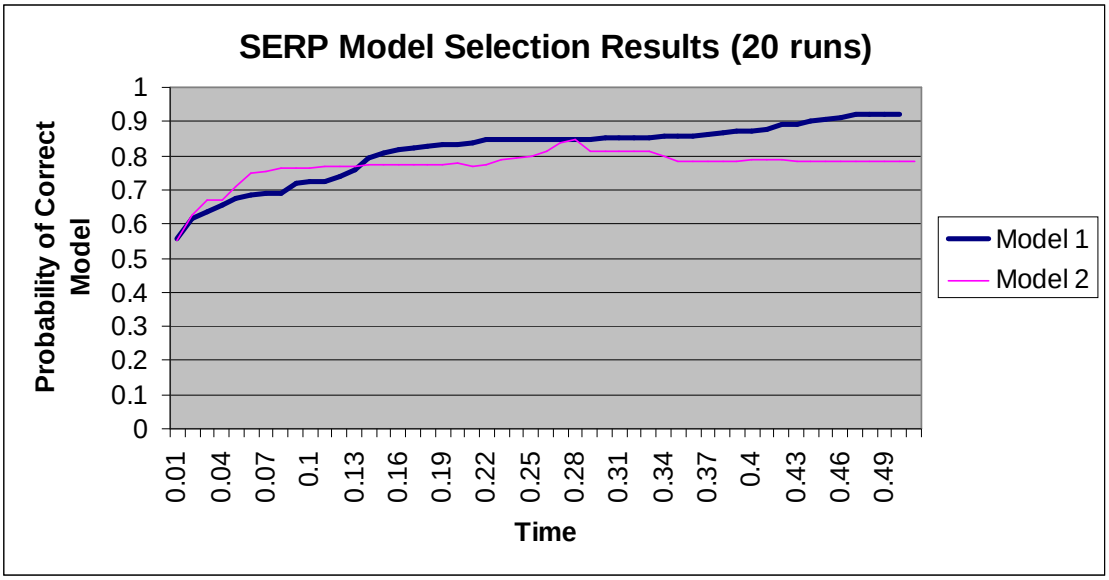

Figure 2. Model likelihoods

Table 1. Number of correct selections

Probability Threshold	Percentage of correct selections	
	Model 1	Model 2
65%	95%	80%
75%	95%	75%
85%	90%	70%
95%	70%	70%

6. Conclusion

Through nonlinear filtering, we are able to determine the number of PML bodies, their current location within

a cell nucleus, and discriminate between various motion models on simulated data. These methods will undoubtedly help in the characterisation of PML bodies in the future. Further research is required in order to expand upon this proof of concept, with suggested advances in the modeling of PML bodies as well as testing these models against real data. An interesting extension would involve modeling the PML bodies are multiple interacting targets; a much more challenging problem but one that can be handled using the SERP algorithm. In addition, these methods can be further refined in order to utilize data from actual samples in order to properly characterize the motion of PML bodies.

7. Acknowledgments

The authors gratefully acknowledge the support of The National Science and Engineering Research Council of Canada (NSERC), the Mathematics of Information Technology and Complex Systems (MITACS) network, and the Cross Cancer Institute of Edmonton.

8. References

[1] D. Ballantyne, J. Hailes, M.A. Kouritzin, H. Long, and J. Wiersma, "A hybrid weighted interacting particle filter for multi-target tracking", in *Signal Processing, Sensor Fusion, and Target Recognition XII*, I. Kadar, ed., *Proceedings of SPIE*, 5096, pp. 244-255, 2003.

[2] D. Ballantyne, S. Kim, M.A. Kouritzin, "A Weighted Interacting Particle-based Nonlinear Filter", in *Signal Processing, Sensor Fusion, and Target Recognition XI*, I. Kadar, ed., *Proceedings of SPIE*, 4729, pp. 236-247, 2002.

[3] D. Blount, J. McCrosky, and M. Kouritzin, "Artificial learning approaches for multi-target tracking", to appear in *Automatic Target Recognition XIV*, Sadjadi, ed., *Proceedings of SPIE*, 2004.

[4] C. Eskiw, D. Bazett-Jones, "The promyelocytic leukemia nuclear body: sites of activity?", in *Biochem Cell Biology*, vol 80, no. 3, pp. 301-310, 2002.

[5] M. Hodges, C. Tissot, K. Howe, D. Grimwade, P. Freemont, "Structure, Organization, and Dynamics of Promyelocytic Leukemia Protein Nuclear Bodies", in *The American Journal of Human Genetics*, vol 63, no. 2, pp. 297-304, 1998.

[6] M. Koken, G. Linares-Cruz, F. Quignon, A. Viron, M. Chelbi-Alix, J. Sobczak-Thepot, L. Juhlin, L. Degos, F. Calvo, H. de The H. "The PML growth-suppressor has an altered expression in human oncogenesis", Oncogene, 10(7). 1315-24, 1996.

[7] J. Ahn and G. Hayward. "Disruption of PML-associated nuclear bodies by IE1 correlates with efficient early stages of viral gene expression and DNA replication in human cytomegalovirus infection. Virology, 274(1), 39-55, 2000.