Beyond the Kalman Filter: Particle filters for tracking applications

N. J. Gordon

Tracking and Sensor Fusion Group
Intelligence, Surveillance and Reconnaissance Division
Defence Science and Technology Organisation
PO Box 1500, Edinburgh, SA 5111, AUSTRALIA.
Neil.Gordon@dsto.defence.gov.au

Contents

What is tracking?

- Use models of the real world to
  - estimate the past and present
  - predict the future
- Achieved by extracting underlying information from sequence of noisy/uncertain observations
- Perform inference on-line
- Evaluate evolving sequence of probability distributions

Recursive filter

System model

$$x_t = f_t(x_{t-1}, \epsilon_t) \quad \leftrightarrow \quad p\,(x_t \,|\, x_{t-1})$$

Measurement model

$$y_t = h_t(x_t, \nu_t) \quad \leftrightarrow \quad p\,(y_t \,|\, x_t)$$

Information available

$$y_{1:t} = (y_1, \ldots, y_t)$$

$$p(x_0)$$

Want

$$p\,(x_{0:t+i} \,|\, y_{1:t})$$

and especially

$$p\,(x_t \,|\, y_{1:t})$$

DEFENCE SCIENCE & TECHNOLOGY

Recursive filter

Prediction

$$p\left(x_t \mid y_{1:t-1}\right) = \int p\left(x_t \mid x_{t-1}\right) p\left(x_{t-1} \mid y_{1:t-1}\right) dx_{t-1}$$

Update

$$p\left(x_t \mid y_{1:t}\right) = \frac{p\left(y_t \mid x_t\right) p\left(x_t \mid y_{1:t-1}\right)}{p\left(y_t \mid y_{1:t-1}\right)}$$

$$p\left(y_t \mid y_{1:t-1}\right) = \int p\left(y_t \mid x_t\right) p\left(x_t \mid y_{1:t-1}\right) dx_t$$

Alternatively ...

$$p\left(x_{0:t} \mid y_{1:t}\right) = p\left(x_{0:t-1} \mid y_{1:t-1}\right) \frac{p\left(y_t \mid x_t\right) p\left(x_t \mid x_{t-1}\right)}{p\left(y_t \mid y_{1:t-1}\right)}$$

$$p\left(x_t \mid y_{1:t}\right) = \int p\left(x_{0:t} \mid y_{1:t}\right) dx_{0:t-1}$$

Tracking : what are the problems?

- On-line processing
- Target manoeuvres
- Missing measurements
- Spurious measurements
- Multiple objects and/or sensors
- Finite sensor resolution
- Prior constraints
- Signature information

Nonlinear/non-Gaussian models

Analytic approximations

- EKF and variants : linearisation, Gaussian approx, unimodal
- Score function moment approximation : Masreliez (75), West (81), Fahrmeir (92), Pericchi, Smith (92)
- Series based approximation to score functions : Wu, Cheng (92)

Numerical approximations

- Discrete grid: Pole, West (88)
- Piecewise pdf: Kitagawa (87), Kramer, Sorenson (88)
- Series expansion: Sorenson, Stubberud (68)
- Gaussian mixtures: Sorenson, Alspach (72), West (92)
- Unscented filter: Julier, Uhlman (95)

Monte Carlo Approximations

- Sequential Importance Sampling (SIS): Handschin & Mayne, *Automatica*, 1969.
- Improved SIS: Zaritskii *et al.*, *Automation and Remote Control*, 1975.
- Rao-Blackwellisation: Akashi & Kumamoto, *Automatica*, 1977...

$\Rightarrow$ Too computationally demanding 20-30 years ago

Sequential Monte Carlo (SMC)

- SMC methods lead to estimate of the complete probability distribution
- Approximation centred on the pdf rather than compromising the state space model
- Known as Particle filters, SIR filters, bootstrap filters, Monte Carlo filters, Condensation etc

Why random samples?

- nonlinear/non-Gaussian
- whole pdf
- moments/quantiles
- HPD interval
- re-parameterisation
- constraints
- association hypotheses independent over time
- multiple models trivial
- scalable - how big is ∞
- parallelisable

Comparison

Kalman filter

- analytic solution
- restrictive assumptions
- deduce state from measurement
- KF “optimal”
- EKF “sub-optimal”

Particle filter

- sequential MC solution
- based on simulation
- no modelling restrictions
- predicts measurements from states
- optimal (with ∞ computational resources)

Book Advert

Sequential Monte Carlo methods in practice

Editors: Doucet, de Freitas, Gordon

Springer-Verlag (2001)

- Theoretical Foundations
- Efficiency Measures
- Applications :
  - Target tracking, missile guidance, image tracking, terrain referenced navigation, exchange rate prediction, portfolio allocation, in-situ ellipsometry, pollution monitoring, communications and audio engineering.

Useful Information

Books

- "Sequential Monte Carlo methods in practice", Doucet, de Freitas, Gordon, Springer, 2001.
- "Monte Carlo strategies in scientific computing", Liu, Springer, 2001.
- "Beyond the Kalman filter : Tracking applications of particle filters", Ristic, Arulampalam, Gordon, Artech House, 2003?

Papers

- "On sequential Monte Carlo sampling methods for Bayesian filtering", Statistics in Computing, Vol 10, No 3, pgs 197-208, 2000.
- IEEE Trans. Signal Processing special issue, February 2002.

Web site

- www.cs.ubc.ca/ nando/smc/index.html (includes software)

DEFENCE SCIENCE & TECHNOLOGY

Particle Filter

- Represent uncertainty over $x_{1:t}$ using diversity of weighted particles

$$\left\{x^i_{1:t}, w^i_t\right\}_{i=1}^N$$

- Ideally:

$$x^i_{1:t} \sim p(x_{1:t} \mid y_{1:t}) = \frac{p(x_{1:t}, y_{1:t})}{p(y_{1:t})}$$

where

$$p(y_{1:t}) = \int p(x_{1:t}, y_{1:t}) dx_{1:t}$$

- What if we can't sample $p(x_{1:t} \mid y_{1:t})$?

Particle Filter - Importance Sampling

- Sample from a convenient proposal distribution $q(x_{1:t} \mid y_{1:t})$
- Use importance sampling to modify weights

$$\begin{aligned}\int p(x_{1:t} \mid y_{1:t}) f(x_{1:t}) dx_{1:t} &= \int \frac{p(x_{1:t} \mid y_{1:t})}{q(x_{1:t} \mid y_{1:t})} q(x_{1:t} \mid y_{1:t}) f(x_{1:t}) dx_{1:t} \\ &\approx \sum_{i=1}^{N} w_t^i f(x_{1:t}^i)\end{aligned}$$

where

$$\begin{aligned} x_{1:t}^i &\sim q(x_{1:t} \mid y_{1:t}) \\ w_t^i &= \frac{p(x_{1:t} \mid y_{1:t})}{q(x_{1:t} \mid y_{1:t})}\end{aligned}$$

Particle Filter - Importance Sampling

- Pick a convenient proposal
- Define the un-normalised weight:

$$\tilde{w}_t^i = \frac{p(x_{1:t}, y_{1:t})}{q(x_{1:t} \mid y_{1:t})}$$

- Can then calculate approximation to $p(y_{1:t})$

$$p(y_{1:t}) \approx \sum_{i=1}^{N} \tilde{w}_t^i$$

- Normalised weight is

$$w_t^i = \frac{p(x_{1:t} \mid y_{1:t})}{q(x_{1:t} \mid y_{1:t})} = \frac{p(x_{1:t}, y_{1:t})}{q(x_{1:t} \mid y_{1:t})} \frac{1}{p(y_{1:t})} = \frac{\tilde{w}_t^i}{\sum_{i=1}^{N} \tilde{w}_t^i}$$

Particle Filter - SIS

- To perform **Sequential Importance Sampling**, SIS

$$q(x_{1:t} \mid y_{1:t}) \triangleq \underbrace{q(x_{1:t-1} \mid y_{1:t-1})}_{\text{Keep existing path}} \underbrace{q(x_t \mid x_{t-1}, y_t)}_{\text{Extend path}}$$

- The un-normalised weight then takes the appealing form

$$\begin{aligned}
\tilde{w}_t^i &= \frac{p(x_{1:t}^i, y_t \mid y_{1:t-1})}{q(x_{1:t}^i \mid y_{1:t})} \\
&= \frac{p(x_{1:t-1}^i \mid y_{1:t-1})}{q(x_{1:t-1}^i \mid y_{1:t-1})} \frac{p(x_t^i, y_t \mid x_{1:t-1}^i)}{q(x_t^i \mid x_{t-1}^i, y_t)} \\
&= w_{t-1}^i \underbrace{\frac{p(y_t \mid x_{t-1}^i) p(x_t^i \mid x_{t-1}^i)}{q(x_t^i \mid x_{t-1}^i, y_t)}}_{\text{Incremental weight}}
\end{aligned}$$

Illustration of SIS

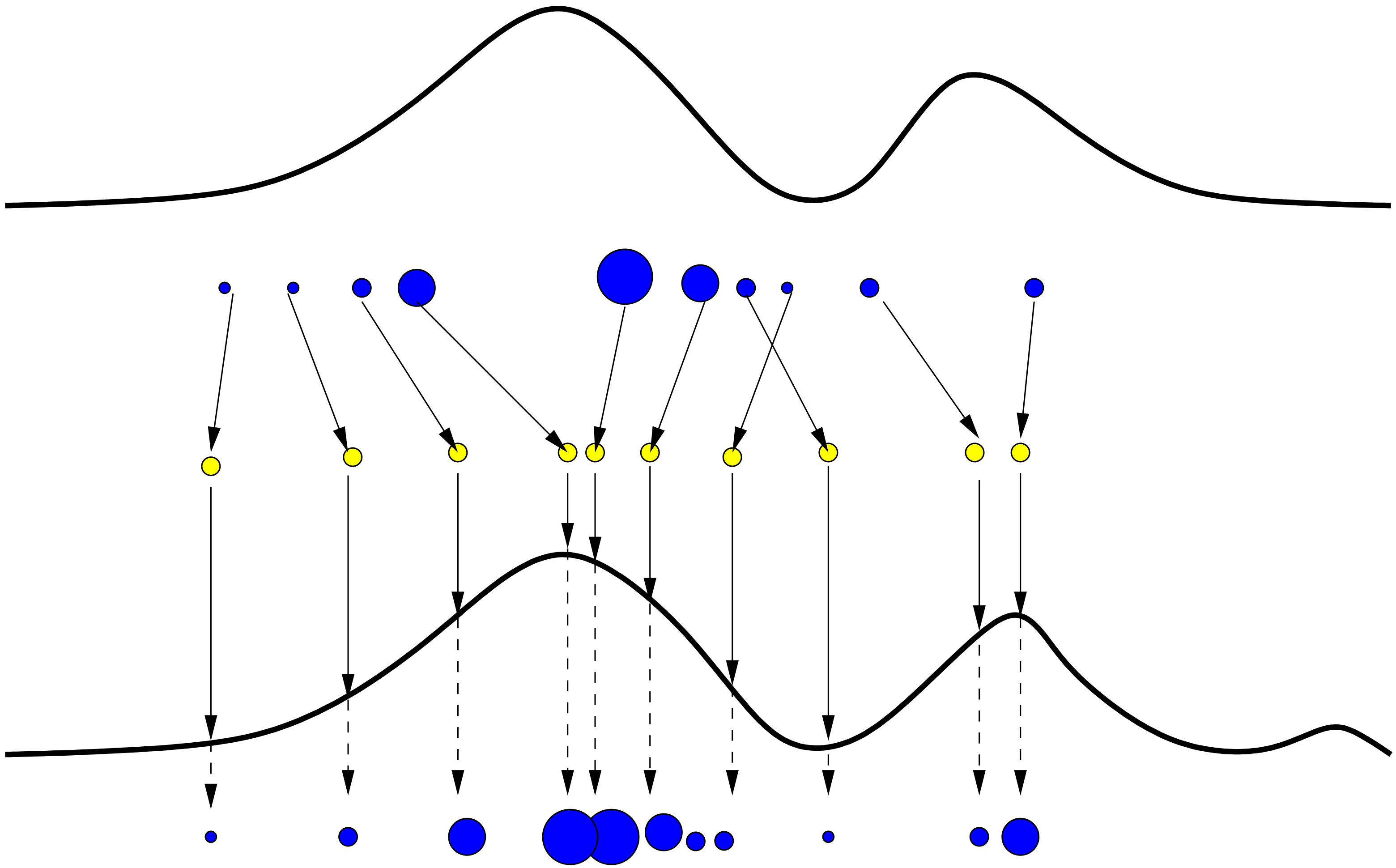
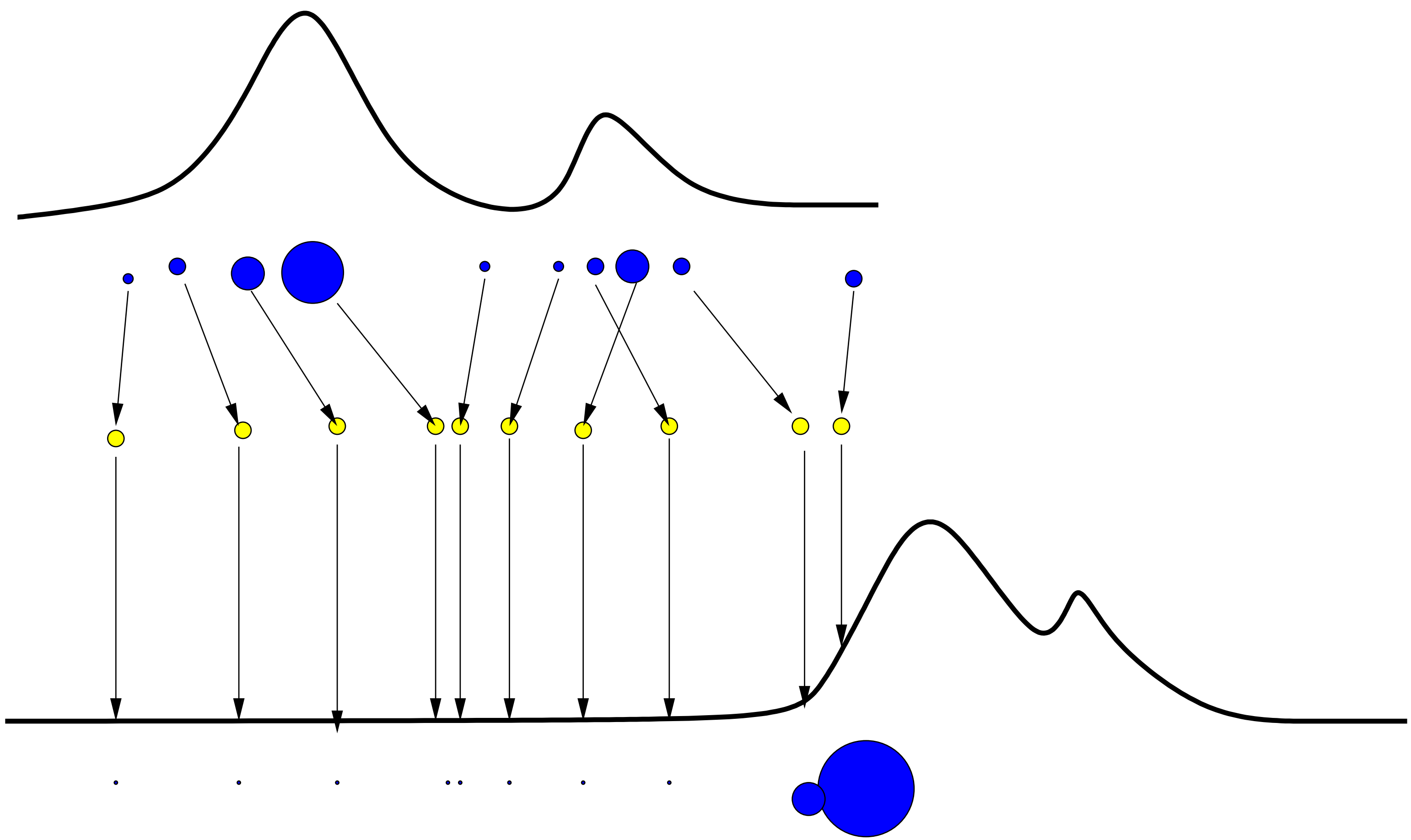


Illustration of SIS - data conflict



SIS

Problem : Whatever the importance function, degeneracy is observed (Kong, Liu and Wong 1994).

- **Introduce a selection scheme to discard/multiply particles $x^i_{0:t}$ with respectively high/low importance weights**
- **Resampling maps the weighted random measure $(x^i_{0:t}, w_t)$ onto the equally weighted random measure $(x^i_{0:t}, N^{-1})$**
- **Scheme generates N_i children such that $\sum_{i=1}^{N} N_i = N$ and satisfies $E(N_i) = Nw^i_k$**

Illustration of SIR

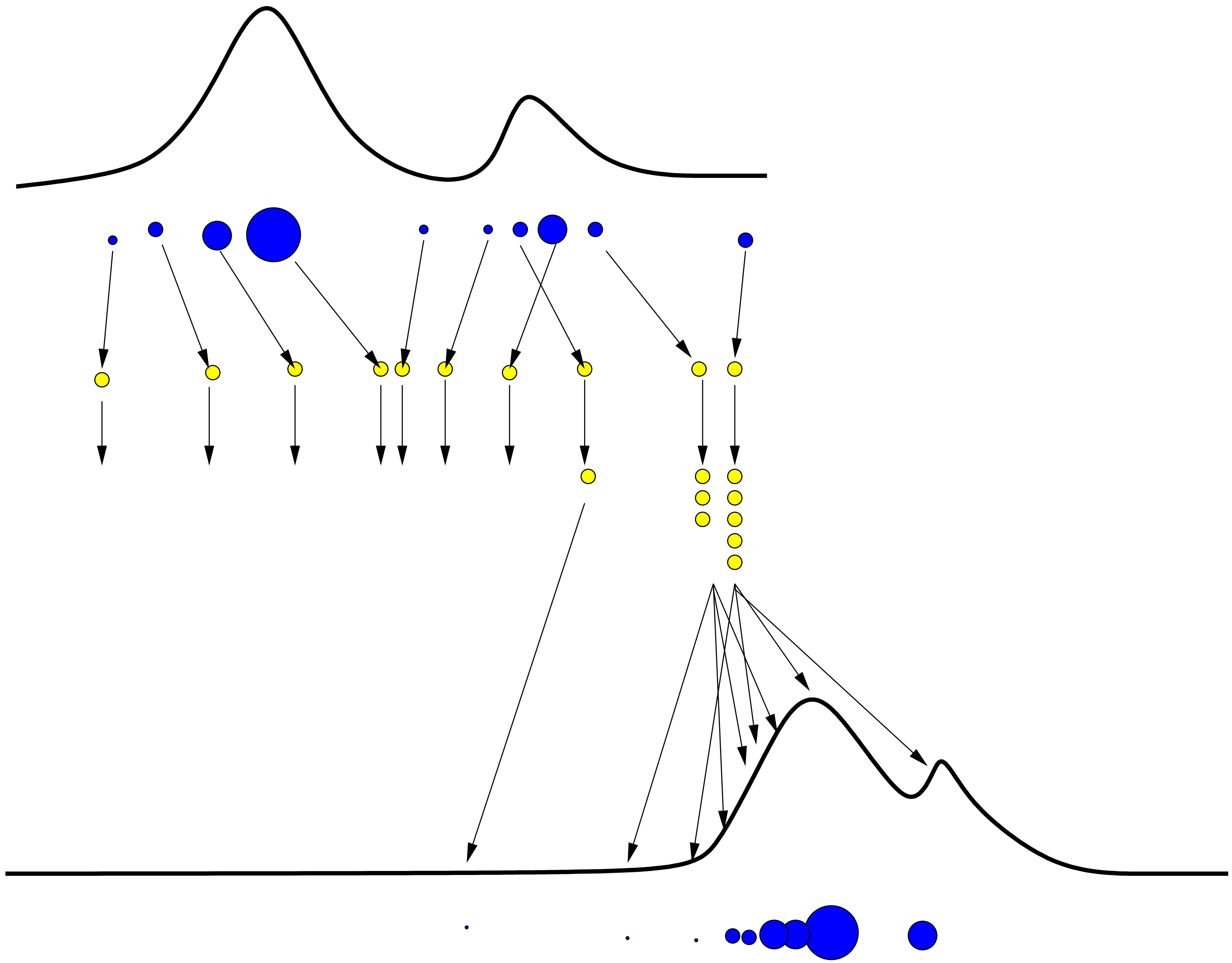
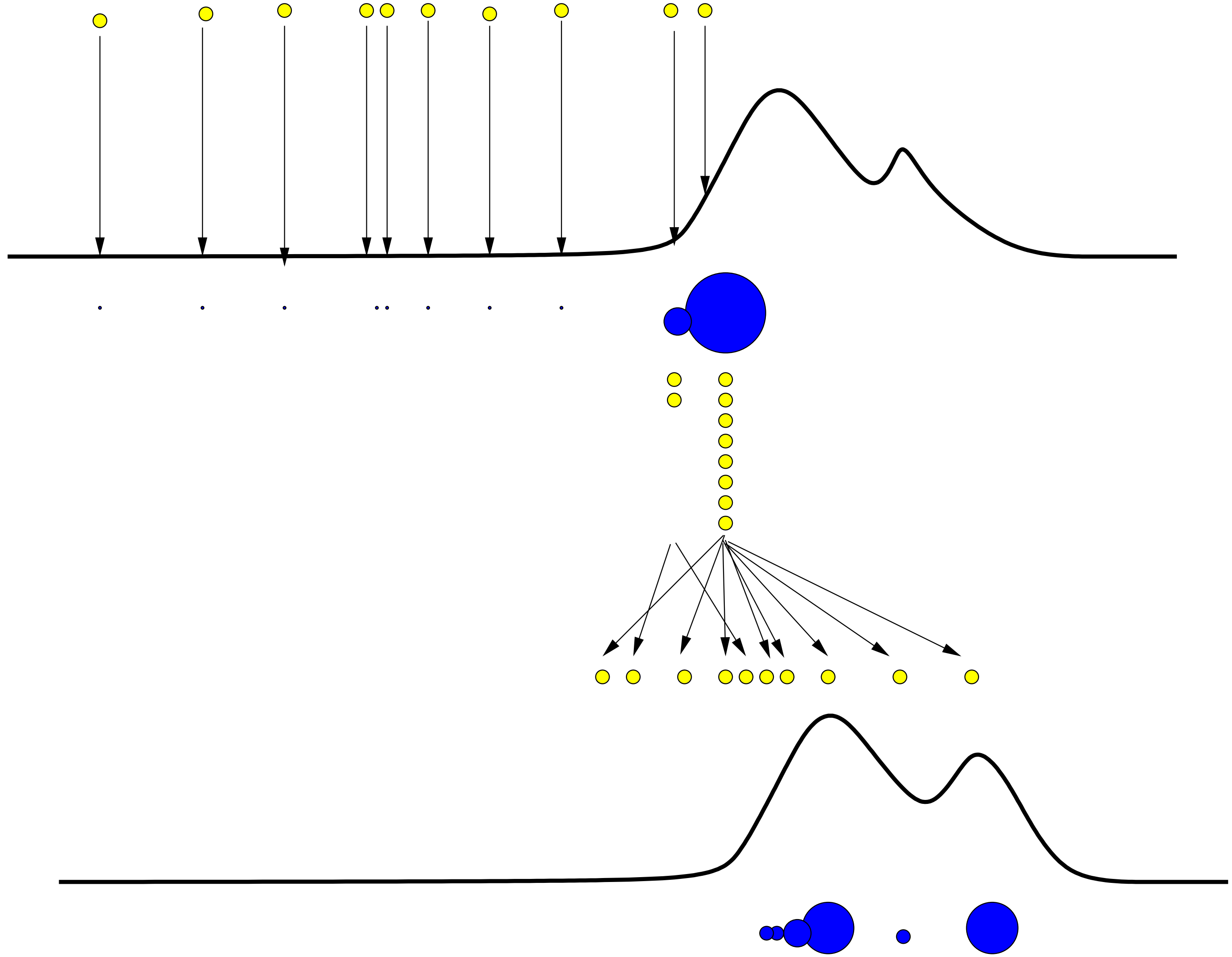


Illustration of SIR



Ingredients for Particle filter

- Importance sampling function
  - Prior $p(x_t \mid x_{t-1}^{(i)})$
  - Optimal $p(x_t \mid x_{t-1}^{(i)}, y_t)$
  - UKF, linearised EKF, ...
- Redistribution scheme
  - Multinomial
  - Deterministic
  - Residual
  - Stratified
- Careful initialisation procedure (for efficiency)

Improvements to SIR

To alleviate degeneracy problems many other methods have been proposed

- Local linearisation (Doucet, 1998; Pitt & Shephard, 1999) using the EKF to estimate the importance distribution or UKF (Doucet et al, 1999)
- Rejection methods (Müller, 1991; Hürzeler & Künsch, 1998; Doucet, 1998; Pitt & Shephard, 1999)
- Auxiliary particle filters (Pitt & Shephard, 1999)
- Kernel smoothing (Gordon, 1993; Liu & West, 2000; Musso et al, 2000)
- MCMC methods (Müller, 1992; Gordon & Whitby, 1995; Berzuini et al, 1997; Gilks & Berzuini, 1999; Andrieu et al, 1999)
- Bridging densities : (Clapp & Godsill, 1999)

Auxiliary SIR - ASIR

- Introduced by Pitt and Shephard 1999.
- Use importance sampling function $q(x_t, i \mid y_{1:t})$
- Auxiliary variable i refers to index of particle at time $t-1$
- Importance distribution chosen to satisfy

$$q(x_t, i \mid y_{1:t}) \propto p(y_t \mid \mu_t^i) p(x_t \mid x_{t-1}^i) w_{t-1}^i$$

  - μ_t^i is some characterisation of x_t given x_{t-1}^i
  - eg, $\mu_t^i = \mathbb{E}(x_t \mid x_{t-1}^i)$ or $\mu_t^i \sim p(x_t \mid x_{t-1}^i)$
- This gives

$$w_t^j \propto w_{t-1}^{i^j} \frac{p(y_t \mid x_t^j) p(x_t^j \mid x_{t-1}^{i^j})}{q(x_t^j, i^j \mid y_{1:t})} = \frac{p(y_t \mid x_t^j)}{p(y_t \mid \mu_t^{i^j})}$$

ASIR

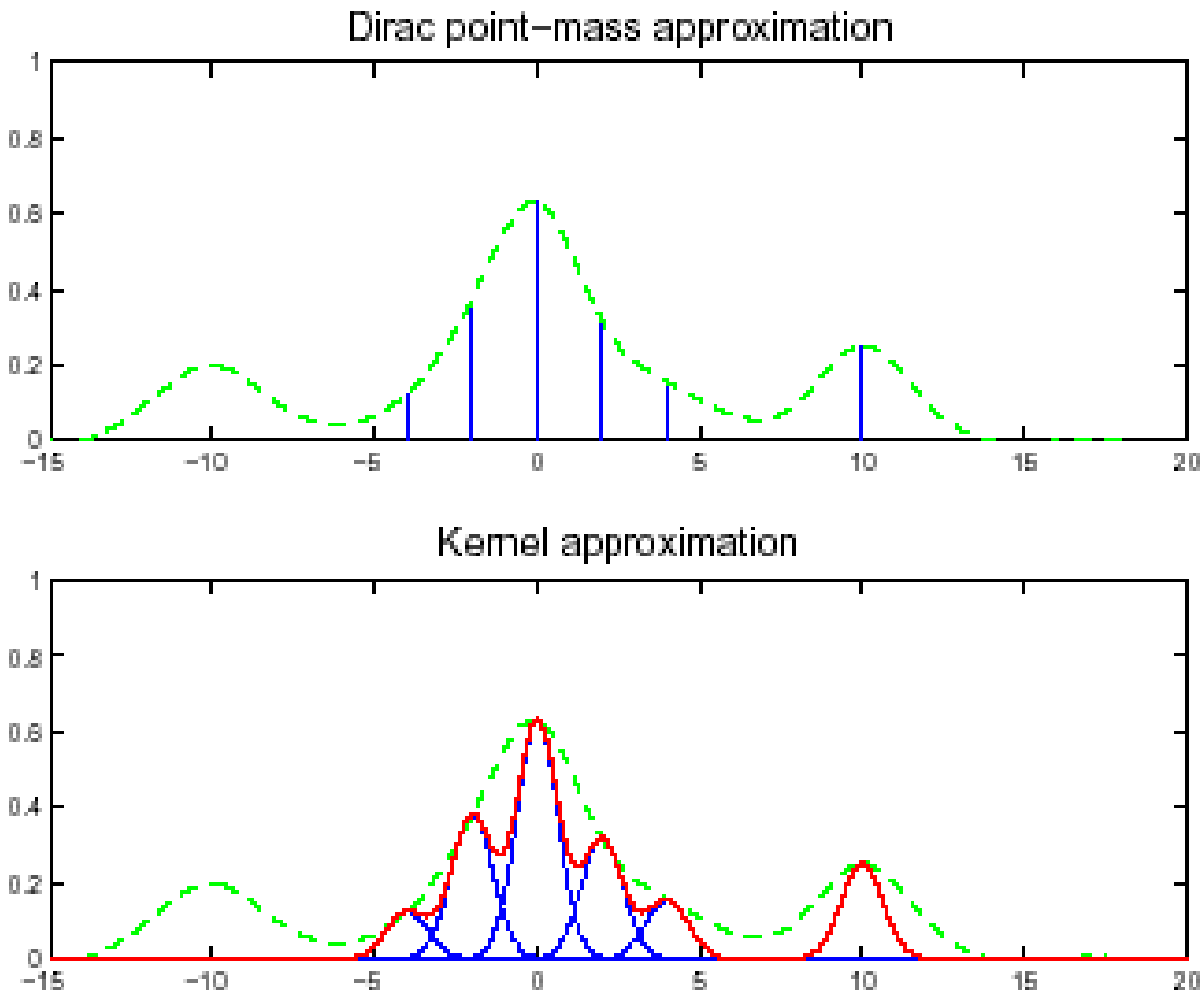
- Naturally uses points at $t-1$ which are "close" to measurement y_t
- If process noise is small then ASIR less sensitive to outliers than SIR
  - This is because single point μ_t^i characterises $p(x_t \mid x_{t-1})$ well
- But if process noise is large then ASIR can degrade performance
  - Since a single point μ_t^i does not characterise $p(x_t \mid x_{t-1})$

Regularised PF - RPF

- Resampling introduced to reduce degeneracy
- But, also reduces diversity
- RPF proposed as a solution
- Uses continuous Kernel based approximation

$$\hat{p}(x_{1:t} \mid y_{1:t}) \approx \sum_{i=1}^{N} w_t^i K_h \left(x_t - x_t^i \right)$$

RPF


Dirac point−mass approximation
Kernel approximation
0
0.2
0.4
1
−15
−10
−5
0
5
10
15
20

RPF

- Kernel K(.) and bandwidth h chosen to minimise MISE

$$MISE(\hat{p}) = \mathbb{E}\left[\int \{\hat{p}(x_t \mid y_{1:t}) - p(x_t \mid y_{1:t})\}^2 \, dx_t\right]$$

- For equally weighted samples, optimal choice is Epanechnikov kernel

$$K_{opt} = \begin{cases} \frac{n_x+2}{2c_{n_x}}(1 - \parallel x \parallel^2) & \textbf{if} \parallel x \parallel < 1 \\ 0 & \textbf{otherwise} \end{cases}$$

- Optimal bandwidth can be obtained as a function of underlying pdf
- Assume this is Gaussian with unit covariance matrix

$$h_{opt} = \left[8N(n_x+4)(2\sqrt{\pi})^{n_x} c_{n_x}^{-1}\right]^{1/(n_x+4)}$$

DEFENCE SCIENCE & TECHNOLOGY

MCMC Moves

- RPF moves are blind
- Instead, introduce Metropolis style acceptance step
  - Resampled x_k^R and created $x_{1:k}^R = \left\{ x_k^R, x_{1:k-1}^R \right\}$
  - Resampled x_k^R and then sampled from a proposal distribution $x_k^P \sim q(. \mid x_k^R)$ and created $x_{1:k}^P = \left\{ x_k^P, x_{1:k-1}^R \right\}$
- Assume $q(. \mid .)$ symmetric

$$x_{1:k} = \begin{cases} x_{1:k}^P & \text{with probability } \alpha \\ x_{1:k}^R & \text{otherwise} \end{cases}$$

$$\begin{aligned} \alpha &= \min\left(1, \frac{p(x_{1:k}^P \mid y_{1:k}) q(x_k^R \mid x_k^P))}{p(x_{1:k}^R \mid y_{1:k}) q(x_k^P \mid x_k^R)}\right) \\ &= \min\left(1, \frac{p(y_k \mid x_k^P) p(x_k^P \mid x_{k-1}^R)}{p(y_k \mid x_k^R) p(x_k^R \mid x_{k-1}^R)}\right) \end{aligned}$$

Tracking dim targets

- Detection and tracking of low SNR targets - better not to threshold the sensor data !
- The concept: track-before-detect
- Conventional TBD approaches:
  - Hough transform (Carlson et al, 1994)
  - dynamic programming (Barniv, 1990; Arnold et al, 1993)
  - maximum likelihood (Tonissen, 1994)
- The performance improved by 3-5 dB in comparison to the MHT (thresholded data).

Recursive Bayesian TBD

- Drawbacks of conventional TBD approaches: batch processing; prohibit or penalise deviations from the straight line motion; require enormous computational resources.
- A recursive Bayesian TBD (Salmond, 2001), implemented as a particle filter
  - no need to store/process multiple scans
  - target motion - stochastic dynamic equation
  - valid for non-gaussian and structured background noise
  - the effect of point spread function, finite resolution, unknown and fluctuating SNR are accommodated
  - target presence and absence explicitly modelled
- Run Demo

Mathematical formulation

- Target state vector

$$\mathbf{x}_k = [x_k \quad \dot{x}_k \quad y_k \quad \dot{y}_k \quad I_k]^T.$$

- State dynamics:

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k, \mathbf{v}_k),$$

- Target existence - two state Markov chain, $E_k \in \{0, 1\}$ with TPM

$$\Pi = \begin{bmatrix} 1 - P_b & P_b \\ P_d & 1 - P_d \end{bmatrix}.$$

Mathematical formulation (Cont'd)

- **Sensor model: 2D map, image of a region $\Delta_x \times \Delta_y$.**
- **At each resolution cell (i,j) measured intensity:**

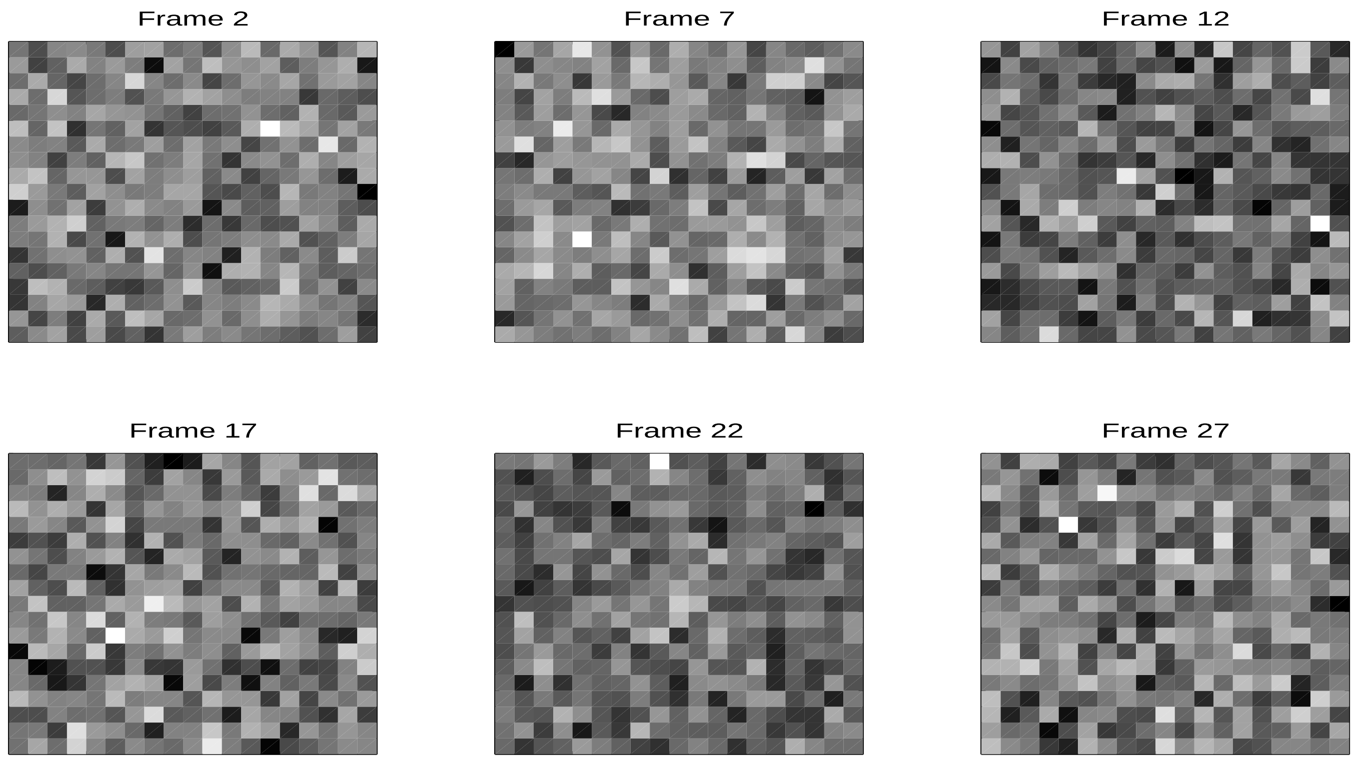
$$z_k^{(i,j)} = \begin{cases} h_k^{(i,j)}(\mathbf{x}_k) + w_k^{(i,j)} & \textbf{if target present} \\ \\ w_k^{(i,j)} & \textbf{if target absent} \end{cases}$$

where

$$h_k^{(i,j)}(\mathbf{x}_k) \approx \frac{\Delta_x \Delta_y I_k}{2\pi\Sigma^2} \exp\left\{ -\frac{(i\Delta_x - x_k)^2 + (j\Delta_y - y_k)^2}{2\Sigma^2} \right\}$$

Example

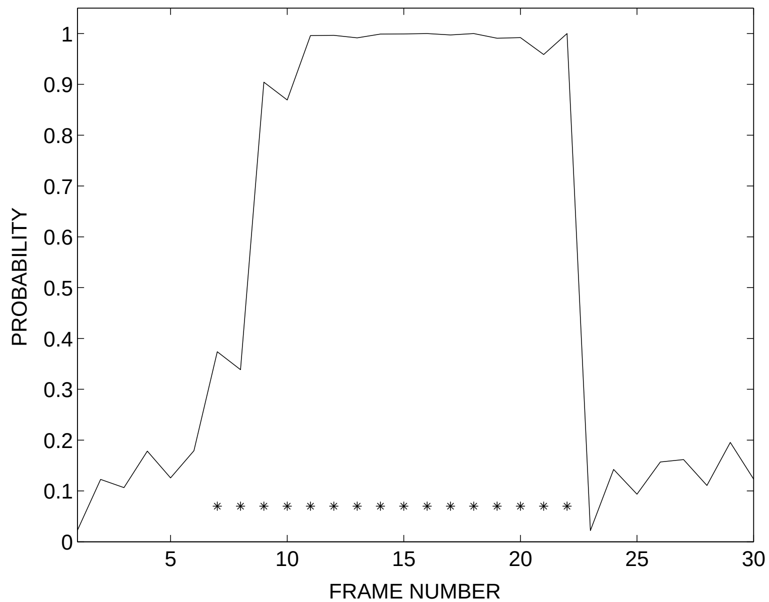
- 30 frames and target present in 7 to 22
- SNR = 6.7 dB (unknown)
- 20×20 cells

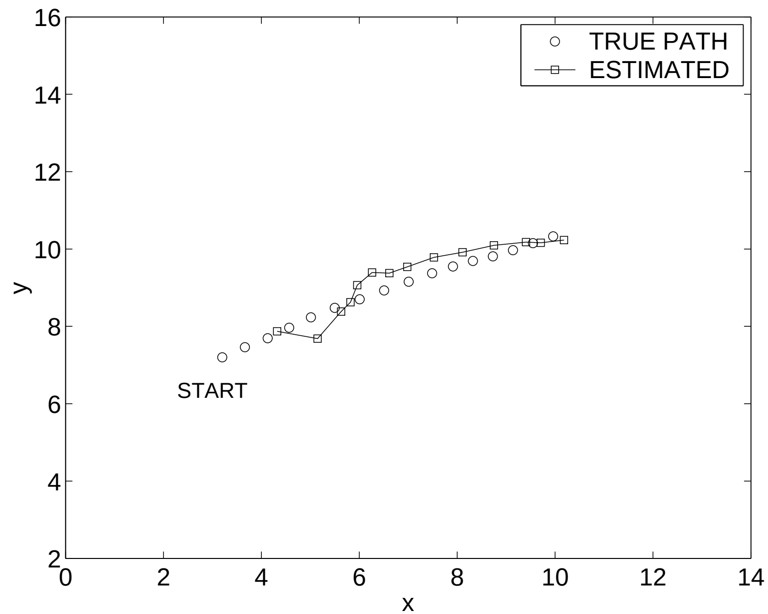


Particle filter output (6 states)

Figures suppressed to reduce file size.

Particle filter output (6 states)





PF based TBD - Performance

- Derived CRLBs (as a function of SNR)
- Compared PF-TBD to CRLBs
- Detection and Tracking reliable at 5 dB (or higher)

Sonobuoy and Submarine

- Noisy bearing measurements from drifting sensors
- Uncertainty in sensor locations
- Sensor loss
- High proportion of spurious bearing measurements
- Run demo

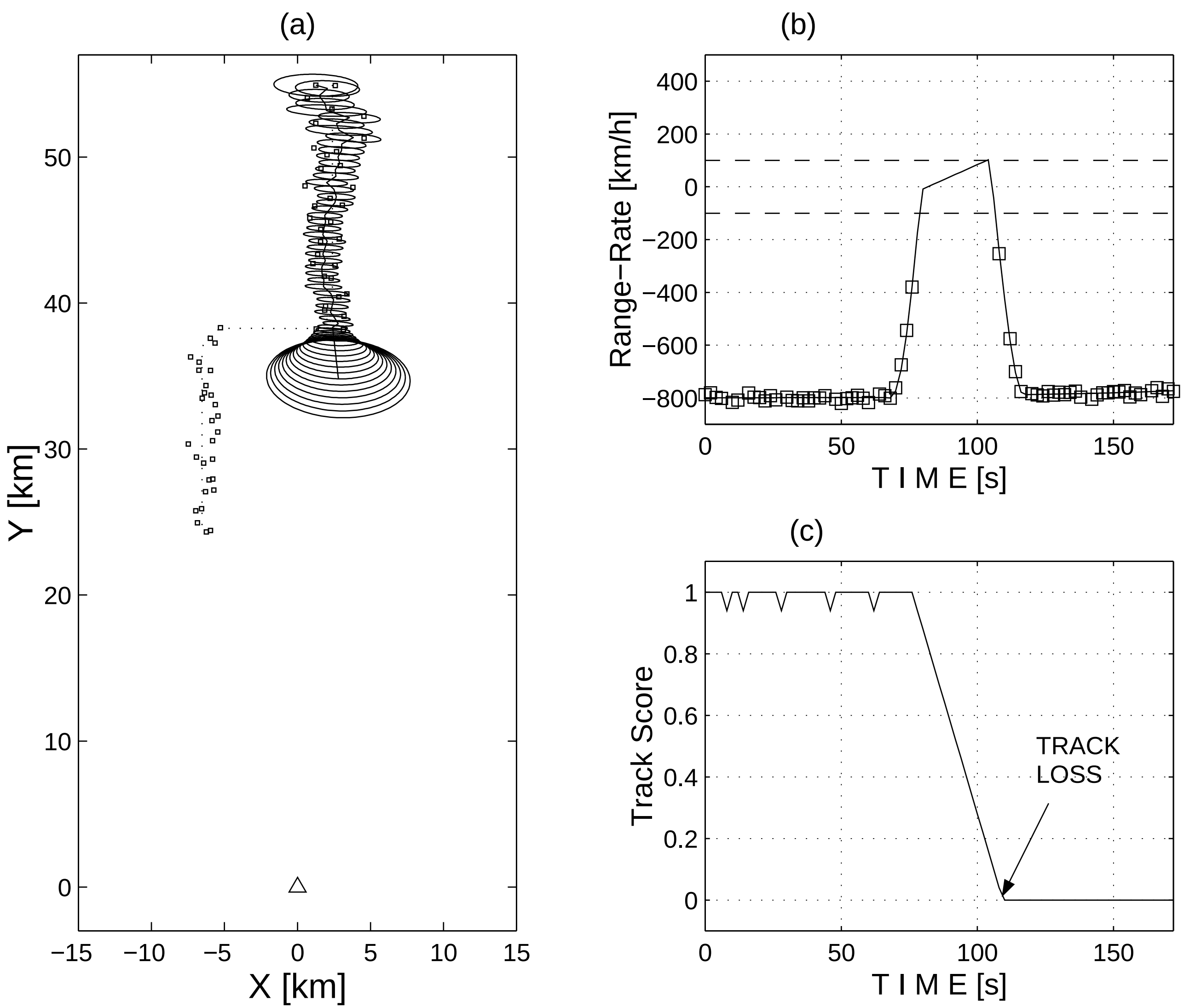
Blind Doppler

- Blind Doppler zone to to filter out ground clutter + ground moving targets
- A simple EP measure against any CW or pulse Doppler radar
- Causes track loss
- Aided by on-board RWR or ESM

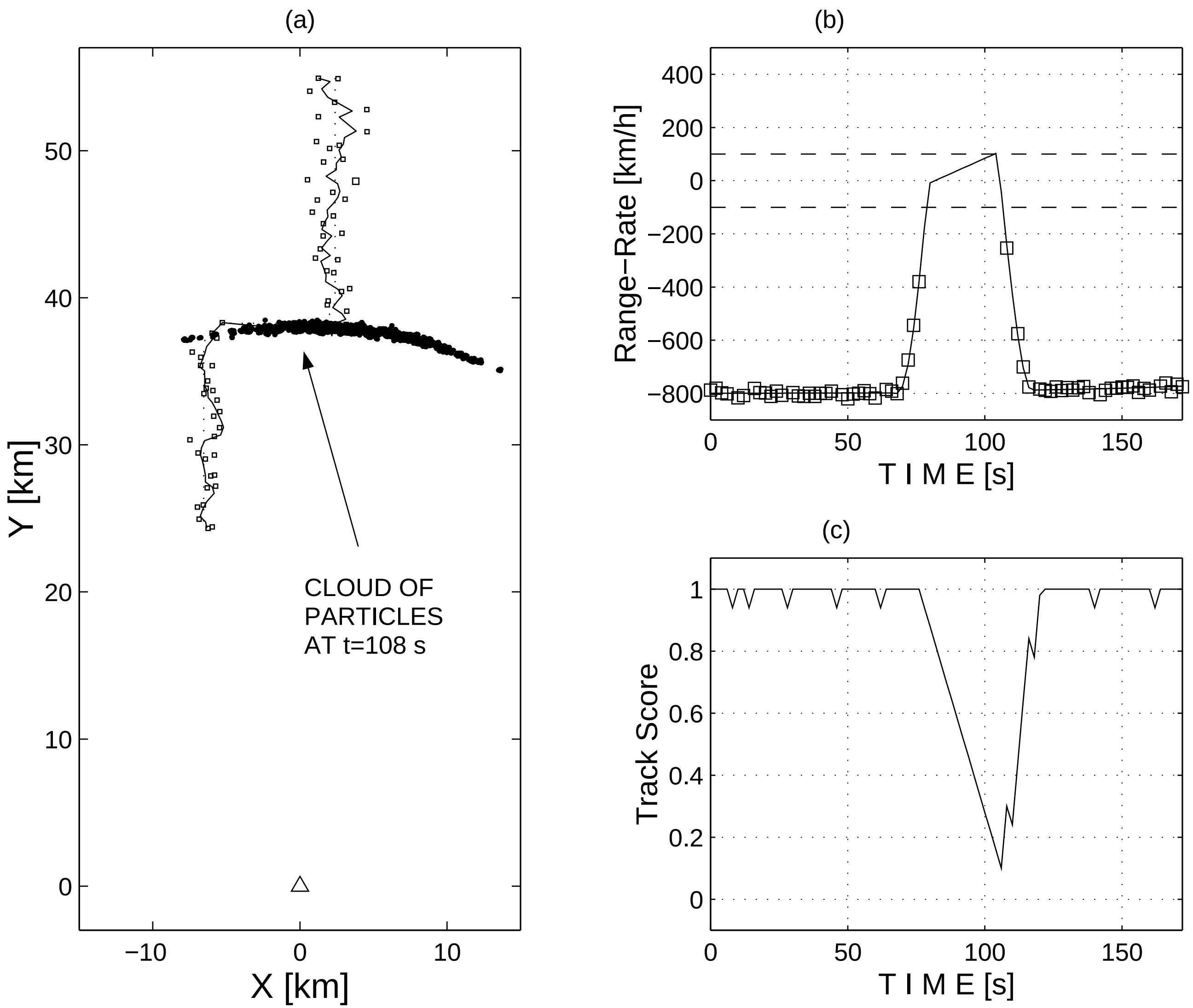
Tracking with hard constraints

- Prior information on sensor constraints
- Posterior pdf is truncated (non-Gaussian)
- Example :
  - 2-D tracking with CV model
  - $(r, \theta, \dot{r})$ measurements
  - $p_d < 1$
  - EKF and Particle Filter with identical gating and track scoring

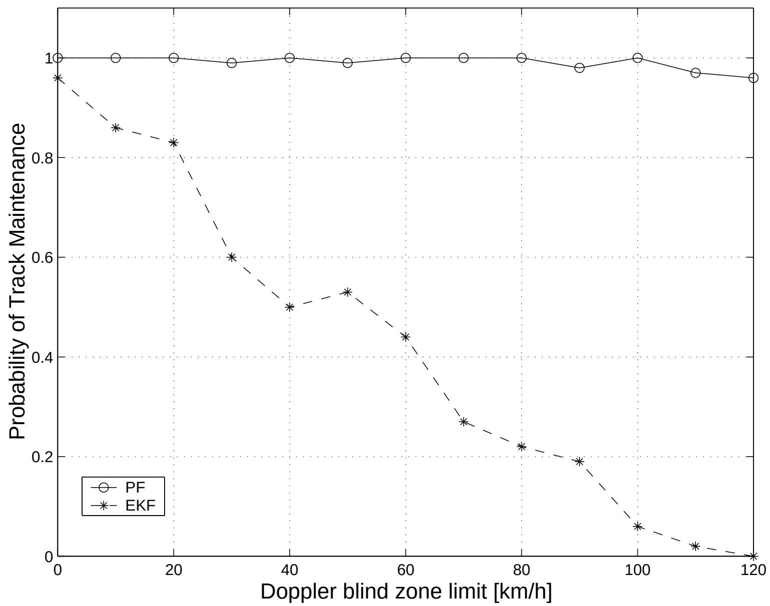
EKF


(a)
Y [km]
X [km]
(b)
Range−Rate [km/h]
T I M E [s]
(c)
Track Score
TRACK
LOSS
T I M E [s]

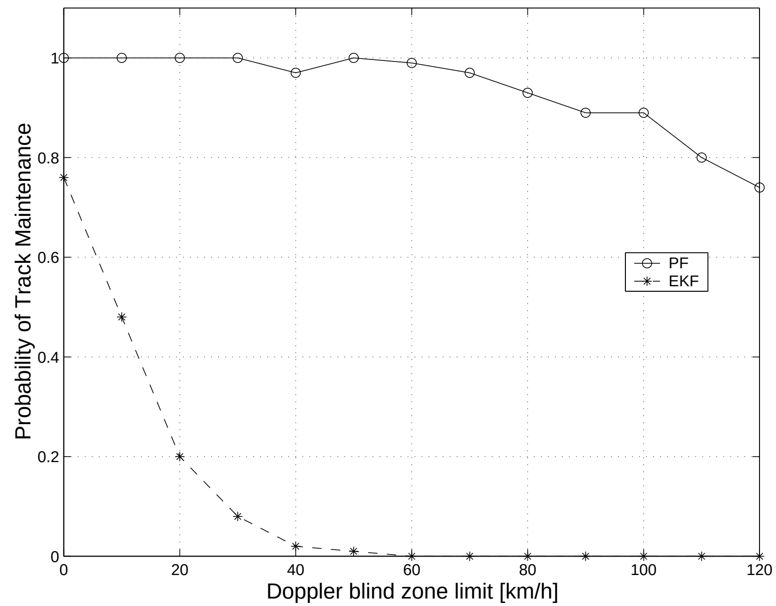
Particle Filter



Track continuity



T=2s



T=4s

Problems hindering SMC

- Convergence results
  - Becoming available (Del Moral, Crisan, Chopin, Lyons, Doucet ...)
- Communication bandwidth
  - Large particle sets impossible to send
- Interoperability
  - Need to integrate with varied tracking algorithms
- Computation
  - Expensive so look to minimise Monte Carlo
- Multi-target problems
  - Rao-Blackwellisation

Final comments

- Sequential Monte Carlo methods
  - “Optimal” filtering for nonlinear/non Gaussian models
  - Flexible/Parallelizable
  - Not a black-box: efficiency depends on careful design.
- If the Kalman filter is appropriate for your application use it
- A Kalman filter is a (one) particle filter