# The Unscented Particle Filter

**Rudolph van der Merwe**
(OGI)

**Nando de Freitas**
(UC Berkeley)

**Arnaud Doucet**
(Cambridge University)

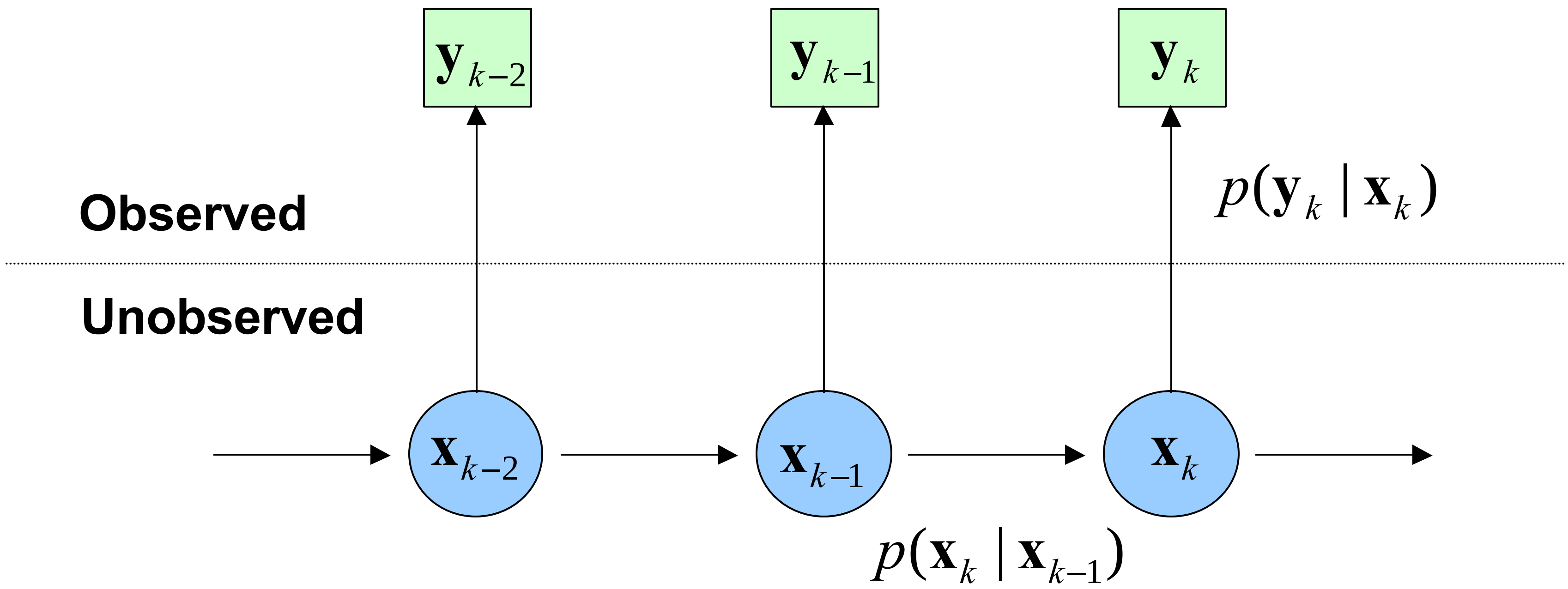
**Eric Wan**
(OGI)

# Outline

- **Optimal Estimation & Filtering**
- **Optimal Recursive Bayesian Solution**
- **Practical Solutions**
  - Gaussian approximations (EKF, UKF)
  - Sequential Monte Carlo methods (Particle Filters)
- **The Unscented Particle Filter**
  - The Unscented Transformation and UKF
  - Applications of UT/UKF to Particle Filters
- **Experimental Results**
- **Conclusions**

# Filtering

- **General problem statement**




- Filtering is the problem of sequentially estimating the states (parameters or hidden variables) of a system as a set of observations become available on-line.

# Filtering

- **Solution of sequential estimation problem given by**
  - Posterior density : $$p(\mathbf{X}_k \mid \mathbf{Y}_k)$$

$$\mathbf{X}_k = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k\}$$
$$\mathbf{Y}_k = \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_k\}$$

  - By recursively computing a marginal of the posterior, the *filtering density*,

$$p(\mathbf{x}_k \mid \mathbf{Y}_k)$$

one need not keep track of the complete history of the states.

# Filtering

- **Given the filtering density, a number of estimates of the system state can be calculated:**
  - Mean (optimal MMSE estimate of state)

$$\hat{\mathbf{x}}_k = E\left[x_k \mid \mathbf{Y}_k\right] = \int \mathbf{x}_k \, p(\mathbf{x}_k \mid \mathbf{Y}_k) d\mathbf{x}_k$$

  - Mode
  - Median
  - Confidence intervals
  - Kurtosis, etc.

# State Space Formulation of System

- **General discrete-time nonlinear, non-Gaussian dynamic system**

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

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{n}_k)$$

state
process noise
known input
noisy observation
measurement noise

- Assumptions :

  1) States follow a first order Markov process

$$p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{x}_{k-2}, \cdots, \mathbf{x}_0) = p(\mathbf{x}_k \mid \mathbf{x}_{k-1})$$

  2) Observations independent given the states

$$p(\mathbf{y}_k \mid \mathbf{x}_k, A) = p(\mathbf{y}_k \mid \mathbf{x}_k)$$

# Recursive Bayesian Estimation

- **Given this state space model, how do we recursively estimate the filtering density ?**

$$
\begin{aligned}
p(\mathbf{x}_k \mid \mathbf{Y}_k) &= \frac{p(\mathbf{Y}_k \mid \mathbf{x}_k)\, p(\mathbf{x}_k)}{p(\mathbf{Y}_k)} \\
&= \frac{p(\mathbf{y}_k, \mathbf{Y}_{k-1} \mid \mathbf{x}_k)\, p(\mathbf{x}_k)}{p(\mathbf{y}_k, \mathbf{Y}_{k-1})} \\
&= \frac{p(\mathbf{y}_k \mid \mathbf{Y}_{k-1}, \mathbf{x}_k)\, p(\mathbf{Y}_{k-1} \mid \mathbf{x}_k)\, p(\mathbf{x}_k)}{p(\mathbf{y}_k \mid \mathbf{Y}_{k-1})\, p(\mathbf{Y}_{k-1})} \\
&= \frac{p(\mathbf{y}_k \mid \mathbf{Y}_{k-1}, \mathbf{x}_k)\, p(\mathbf{x}_k \mid \mathbf{Y}_{k-1})\, p(\mathbf{Y}_{k-1})\, p(\mathbf{x}_k)}{p(\mathbf{y}_k \mid \mathbf{Y}_{k-1})\, p(\mathbf{Y}_{k-1})\, p(\mathbf{x}_k)} \\
&= \frac{p(\mathbf{y}_k \mid \mathbf{x}_k)\, p(\mathbf{x}_k \mid \mathbf{Y}_{k-1})}{p(\mathbf{y}_k \mid \mathbf{Y}_{k-1})}
\end{aligned}
$$

# Recursive Bayesian Estimation

likelihood

prior

$$p(\mathbf{x}_k|\mathbf{Y}_k) = \frac{p(\mathbf{y}_k|\mathbf{x}_k)\,p(\mathbf{x}_k|\mathbf{Y}_{k-1})}{p(\mathbf{y}_k|\mathbf{Y}_{k-1})}$$

posterior

evidence

- **Prior** : $p(\mathbf{x}_k|\mathbf{Y}_{k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})\,p(\mathbf{x}_{k-1}|\mathbf{Y}_{k-1})\,d\mathbf{x}_{k-1}$

  transition density given by *process model*

  (Propagation of past state into future before new observation is made.)

- **Likelihood** : defined in terms of *observation model*

- **Evidence** : $p(\mathbf{y}_k|\mathbf{Y}_{k-1}) = \int p(\mathbf{y}_k|\mathbf{x}_k)\,p(\mathbf{x}_k|\mathbf{Y}_{k-1})\,d\mathbf{x}_k$

# Practical Solutions

- **Gaussian Approximations**
- **Perfect Monte Carlo Simulation**
- **Sequential Monte Carlo Methods : “Particle Filters”**
  - Bayesian Importance Sampling
  - Sampling-importance resampling (SIR)

# Gaussian Approximations

- **Most common approach.**
- **Assume all RV statistics are *Gaussian.***
- **Optimal recursive MMSE estimate is then given by**

$$\hat{\mathbf{x}}_k = (\text{prediction of } \mathbf{x}_k) + k_k[(\text{observation of } \mathbf{y}_k) - (\text{prediction of } \mathbf{y}_k)]$$

- **Different implementations :**
  - *Extended Kalman Filter* (EKF) : optimal quantities approximated via first order Taylor series expansion (linearization) of process and measurement models.
  - *Unscented Kalman Filter* (UKF) : optimal quantities calculated using the Unscented Transformation (accurate to second order for any nonlinearity). Drastic improvement over EKF [Wan, van der Merwe, Nelson 2000].
- **Problem : Gaussian approximation breaks down for most nonlinear real-world applications (multi-modal distributions, non-Gaussian noise sources, etc.)**

# Perfect Monte Carlo Simulation

- **Allow for a complete representation of the posterior distribution.**
- **Map intractable integrals of optimal Bayesian solution to tractable discrete sums of weighted samples drawn from the posterior distribution.**

$$\hat{p}(\mathbf{x}_k \mid \mathbf{Y}_k) = \frac{1}{N}\sum_{i=1}^{N} d\left(\mathbf{x}_k - \mathbf{x}_k^{(i)}\right)$$

$$\mathbf{x}_k^{(i)} \xleftarrow{I.I.D.} p(\mathbf{x}_k \mid \mathbf{Y}_k)$$

- **So, any estimate of the form**

$$E\left[f(\mathbf{x}_k)\right] = \int f(\mathbf{x}_k) p(\mathbf{x}_k \mid \mathbf{Y}_k) d\mathbf{x}_k$$

**may be approximated by :**

$$E\left[f(\mathbf{x}_k)\right] \approx \frac{1}{N}\sum_{i=1}^{N} f\left(\mathbf{x}_k^{(i)}\right)$$

# Particle Filters

- **Bayesian Importance Sampling**
  - It is often impossible to sample directly from the true posterior density.
  - However, we can rather sample from a known, easy-to-sample, *proposal* distribution,

$$q(\mathbf{x}_k \mid \mathbf{Y}_k)$$

and make use of the following substitution

$$E[f(\mathbf{x}_k)] = \int f(\mathbf{x}_k) \frac{p(\mathbf{x}_k \mid \mathbf{Y}_k)}{q(\mathbf{x}_k \mid \mathbf{Y}_k)} q(\mathbf{x}_k \mid \mathbf{Y}_k)\, d\mathbf{x}_k$$

$$= \int f(\mathbf{x}_k) \frac{p(\mathbf{Y}_k \mid \mathbf{x}_k)\, p(\mathbf{x}_k)}{p(\mathbf{Y}_k)\, q(\mathbf{x}_k \mid \mathbf{Y}_k)} q(\mathbf{x}_k \mid \mathbf{Y}_k)\, d\mathbf{x}_k$$

$$= \int f(\mathbf{x}_k) \frac{w_k(\mathbf{x}_k)}{p(\mathbf{Y}_k)} q(\mathbf{x}_k \mid \mathbf{Y}_k)\, d\mathbf{x}_k$$

$$w_k(\mathbf{x}_k) = \frac{p(\mathbf{Y}_k \mid \mathbf{x}_k)\, p(\mathbf{x}_k)}{q(\mathbf{x}_k \mid \mathbf{Y}_k)}$$

# Particle Filters

$$E\left[f(\mathbf{x}_k)\right] = \frac{1}{p(\mathbf{Y}_k)} \int f(\mathbf{x}_k) w_k(\mathbf{x}_k) q\left(\mathbf{x}_k \mid \mathbf{Y}_k\right) d\mathbf{x}_k$$

$$= \frac{\int f(\mathbf{x}_k) w_k(\mathbf{x}_k) q\left(\mathbf{x}_k \mid \mathbf{Y}_k\right) d\mathbf{x}_k}{\int p(\mathbf{Y}_k \mid \mathbf{x}_k) p(\mathbf{x}_k) \frac{q(\mathbf{x}_k \mid \mathbf{Y}_k)}{q(\mathbf{x}_k \mid \mathbf{Y}_k)} d\mathbf{x}_k}$$

$$= \frac{\int f(\mathbf{x}_k) w_k(\mathbf{x}_k) q\left(\mathbf{x}_k \mid \mathbf{Y}_k\right) d\mathbf{x}_k}{\int w_k(\mathbf{x}_k) q\left(\mathbf{x}_k \mid \mathbf{Y}_k\right) d\mathbf{x}_k}$$

$$= \frac{E_{q(\mathbf{x}_k \mid \mathbf{Y}_k)}\left[w_k(\mathbf{x}_k) f(\mathbf{x}_k)\right]}{E_{q(\mathbf{x}_k \mid \mathbf{Y}_k)}\left[w_k(\mathbf{x}_k)\right]}$$

# Particle Filters

- So, by drawing samples from $q(\mathbf{x}_k \mid \mathbf{Y}_k)$, we can approximate expectations of interest by the following:

$$E\left[f(\mathbf{x}_k)\right] \approx \frac{\frac{1}{N}\sum_{i=1}^{N} w_k(\mathbf{x}_k^{(i)}) f(\mathbf{x}_k^{(i)})}{\frac{1}{N}\sum_{i=1}^{N} w_k(\mathbf{x}_k^{(i)})}$$

$$\approx \sum_{i=1}^{N} \tilde{w}_k(\mathbf{x}_k^{(i)}) f(\mathbf{x}_k^{(i)})$$

- Where the normalized importance weights are given by

$$\tilde{w}_k(\mathbf{x}_k^{(i)}) = \frac{w_k(\mathbf{x}_k^{(i)})}{\sum_{j=1}^{N} w_k(\mathbf{x}_k^{(j)})}$$

# Particle Filters

- Using the state space assumptions (1st order Markov / observational independence given state), the importance weights can be estimated recursively by [proof in De Freitas (2000)]

$$w_k = w_{k-1} \frac{p(\mathbf{y}_k \mid \mathbf{x}_k)\, p(\mathbf{x}_k \mid \mathbf{x}_{k-1})}{q(\mathbf{x}_k \mid \mathbf{X}_{k-1}, \mathbf{Y}_k)}$$

- Problem with SIS is that the variance of the importance weights increase stochastically over time [Kong et al. (1994), Doucet et al. (1999)]

- To solve this, we need to resample the particles
  - keep / multiply particles with high importance weights
  - discard particles with low importance weights

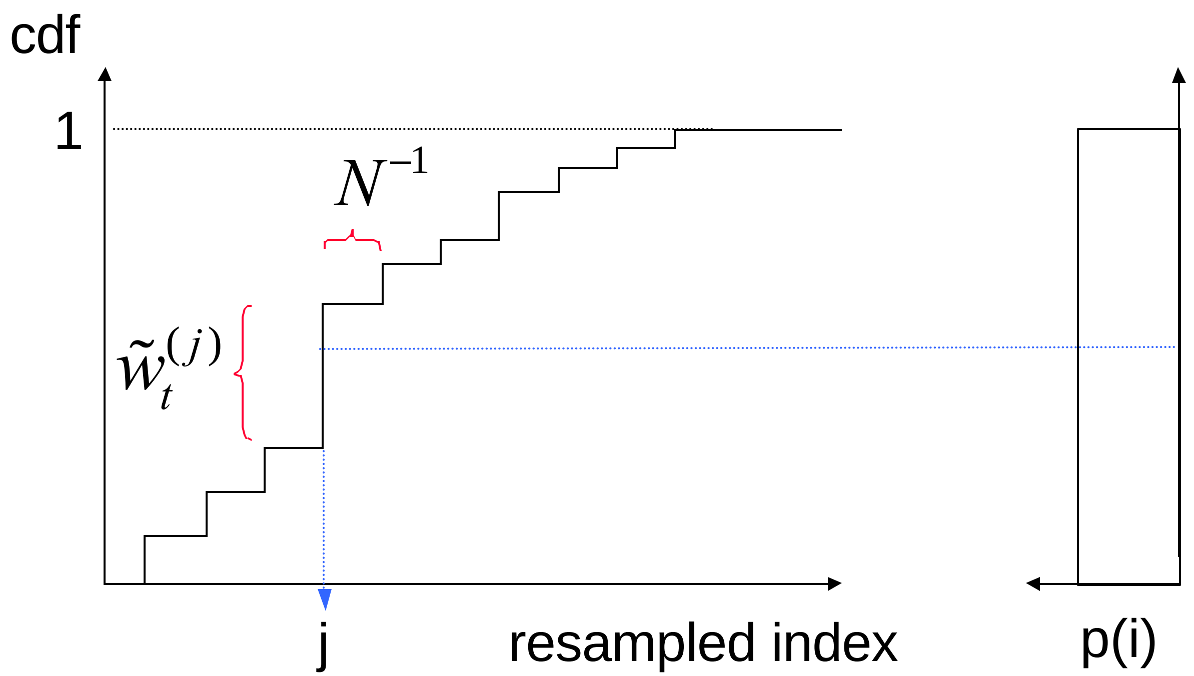
- *Sampling-importance Resampling* (SIR)

# Particle Filters

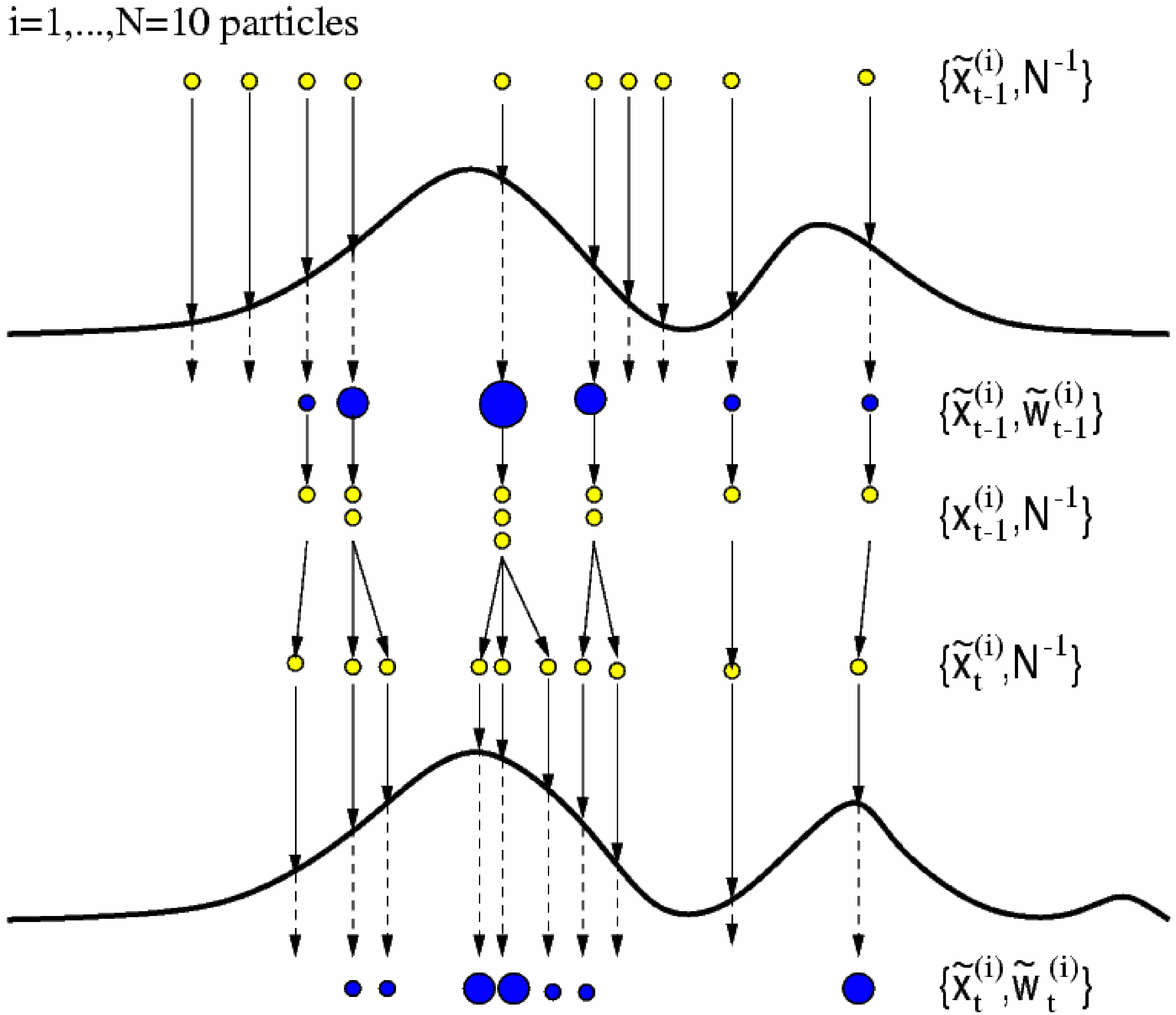
- **Sampling-importance Resampling**
  - Maps the *N* unequally weighted particles into a new set of *N* equally weighted samples.

$$\left\{\mathbf{x}_k^{(i)}, \tilde{w}_k^{(i)}\right\} \rightarrow \left\{\mathbf{x}_k^{(j)}, N^{-1}\right\}$$

  - Method proposed by Gordon, Salmond & Smith (1993) and proven mathematically by Gordon (1994).



# Particle Filters



# Particle Filters

- **Choice of Proposal Distribution**

$$w_k = w_{k-1} \frac{p\left(\mathbf{y}_k \mid \mathbf{x}_k\right) p\left(\mathbf{x}_k \mid \mathbf{x}_{k-1}\right)}{q\left(\mathbf{x}_k \mid \mathbf{X}_{k-1}, \mathbf{Y}_k\right)}$$

critical design issue for successful particle filter

- samples/particles are drawn from this distribution
- used to evaluate importance weights

- Requirements

1) Support of proposal distribution must include support of true posterior distribution, i.e. *heavy-tailed* distributions are preferable.

2) Must include most recent observations.

# Particle Filters

- Most popular choice of proposal distribution does not satisfy these requirements though:

$$q\left(\mathbf{x}_k \mid \mathbf{X}_{k-1}, \mathbf{Y}_k\right) = p\left(\mathbf{x}_k \mid \mathbf{x}_{k-1}\right)$$

[Isard and Blake 96, Kitagawa 96, Gordon et al. 93, Beadle and Djuric 97, Avitzour 95]

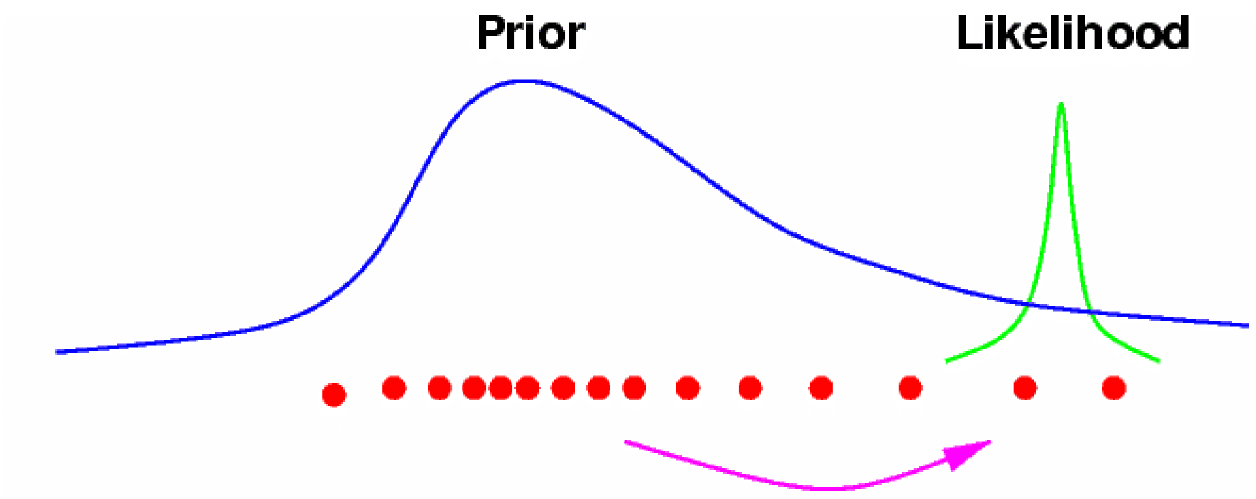
- Easy to implement :

$$w_k = w_{k-1} \frac{p\left(\mathbf{y}_k \mid \mathbf{x}_k\right) p\left(\mathbf{x}_k \mid \mathbf{x}_{k-1}\right)}{p\left(\mathbf{x}_k \mid \mathbf{x}_{k-1}\right)} = w_{k-1} p\left(\mathbf{y}_k \mid \mathbf{x}_k\right)$$

- Does not incorporate most recent observation though !

# Improving Particle Filters

- **Incorporate New Observations into Proposal**




  - Use Gaussian approximation (i.e. Kalman filter) to generate proposal by combining new observation with prior

$$q\left(\mathbf{x}_k \mid \mathbf{X}_{k-1}, \mathbf{Y}_k\right) = p_G\left(\mathbf{x}_k \mid \mathbf{X}_{k-1}, \mathbf{Y}_k\right)$$
$$= \mathcal{N}\left(\hat{\mathbf{x}}_k, \mathrm{cov}[\mathbf{x}_k]\right)$$

# Improving Particle Filters

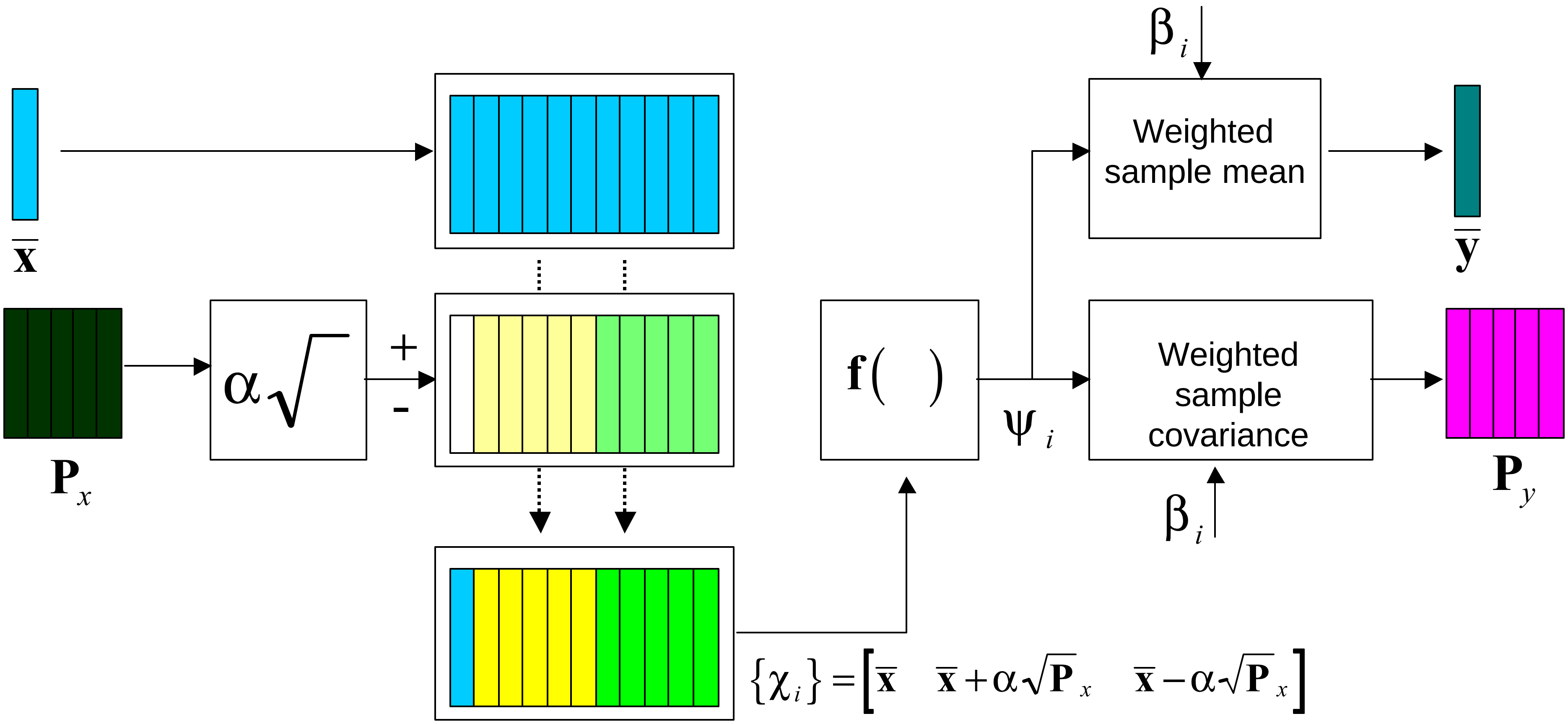
- ***Extented Kalman Filter* Proposal Generation**
  - De Freitas (1998), Doucet (1998), Pitt & Shephard (1999).
  - Greatly improved performance compared to standard particle filter in problems with very accurate measurements, i.e. likelihood very peaked in comparison to prior.
  - In highly nonlinear problems, the EKF tends to be very inaccurate and *underestimates* the true covariance of the state. This violates the distribution support requirement for the proposal distribution and can lead to poor performance and filter divergence.

- **We propose the use of the *Unscented Kalman Filter* for proposal generation to address these problems !**

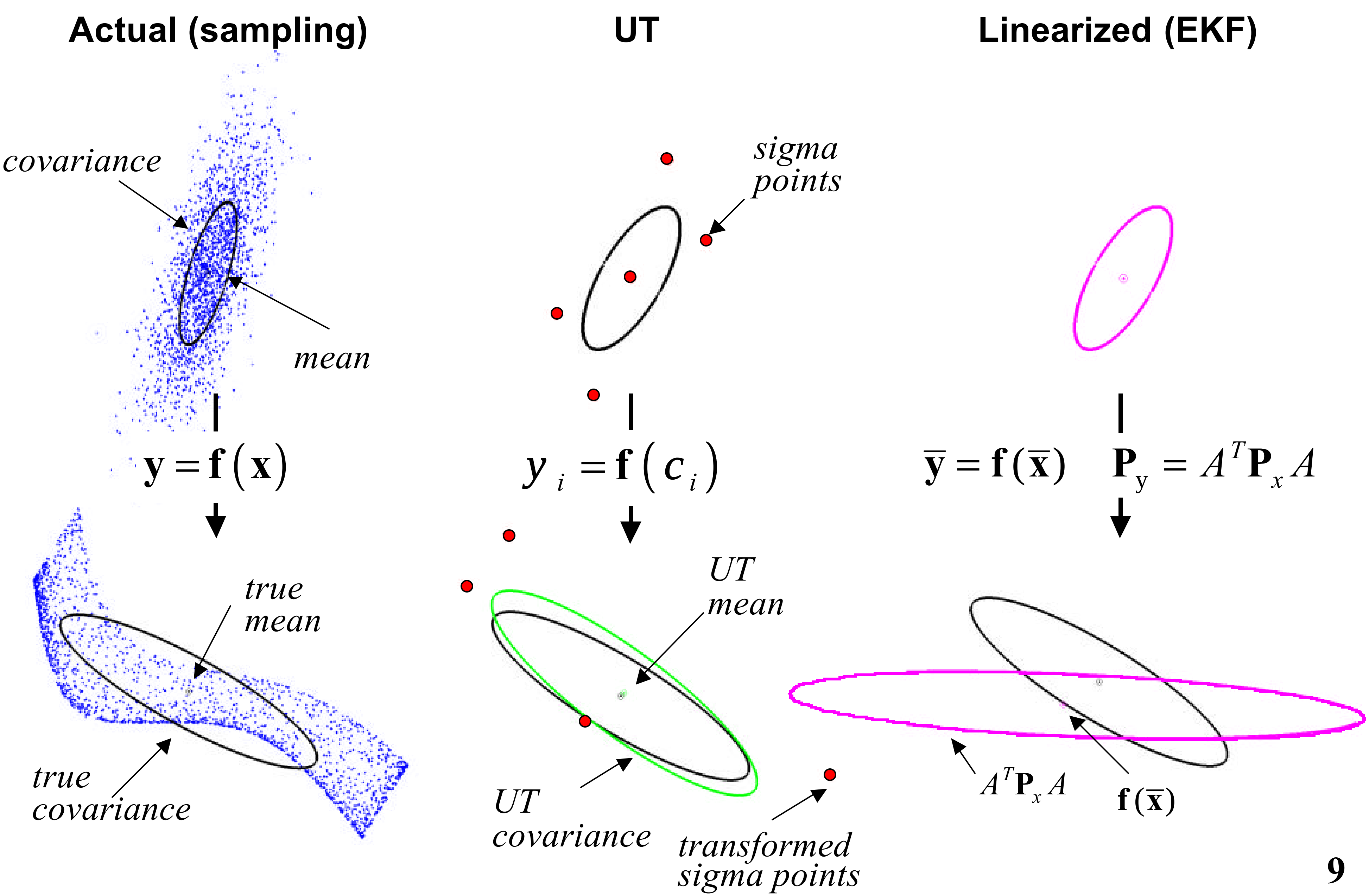
# Improving Particle Filters

- ***Unscented Kalman Filter*** **Proposal Generation**
  - UKF is a recursive MMSE estimator based on the Unscented Transformation (UT).
  - UT : Method for calculating the statistics of a RV that undergoes a nonlinear transformation (Julier and Uhlmann 1997)
  - UT/UKF : - accurate to 3rd order for Gaussians
    - higher order errors scaled by choice of transform parameters.
  - More accurate estimates than EKF (Wan, van der Merwe, Nelson 2000)
  - Have some control over higher order moments, i.e. kurtosis, etc. ⟶ heavy tailed distributions !

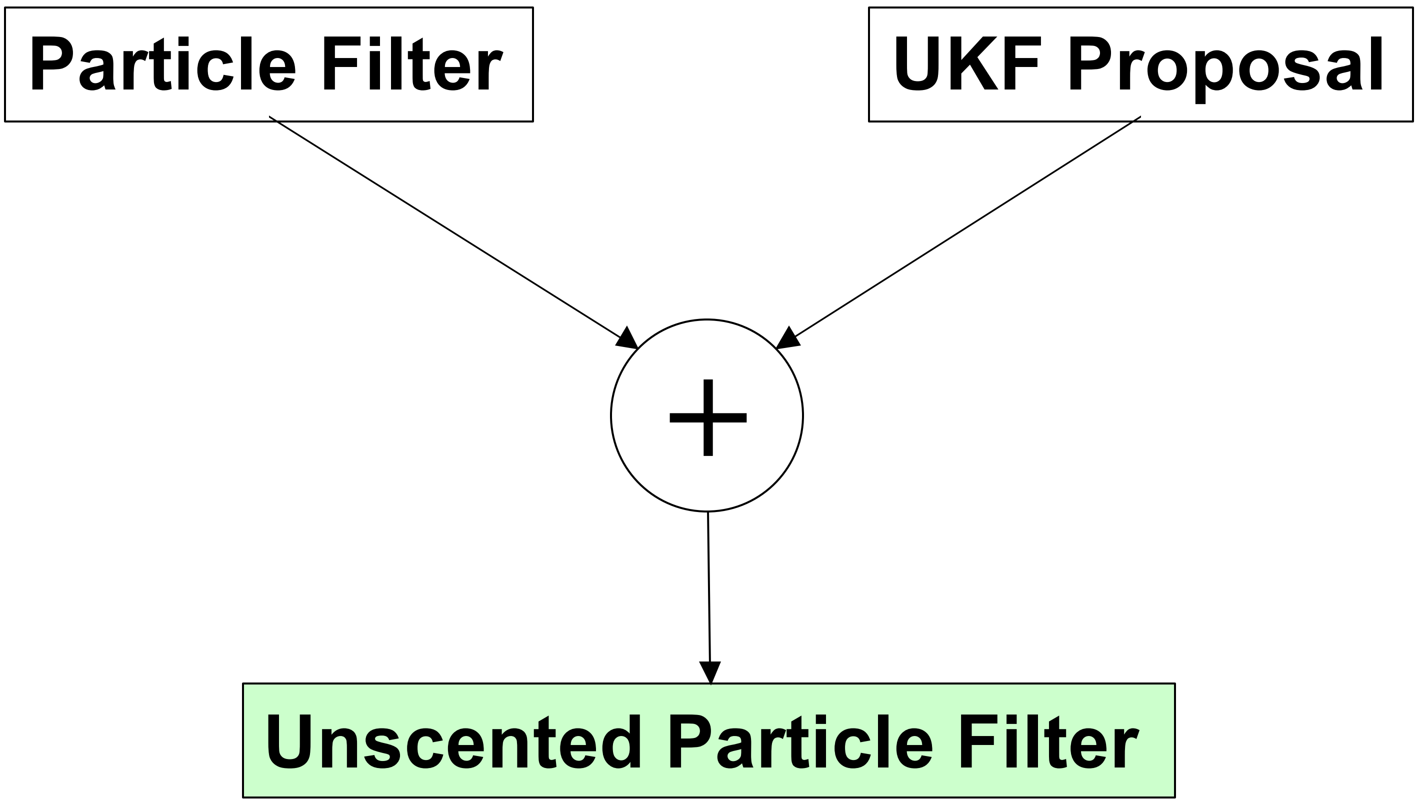
# Unscented Transformation



# The Unscented Transformation



# Unscented Particle Filter



# Experimental Results

- **Synthetic Experiment**
  - Time-series
    - process model :

$$x_{k+1} = 1 + \sin(wp\,k) + f\,x_k + v_k$$

process noise (Gamma)

    - nonstationary observation model :

$$y_k = \begin{cases} f\,x_k^2 + n_k & k \le 30 \\ f\,x_k - 2 + n_k & k > 30 \end{cases}$$

measurement noise (Gaussian)

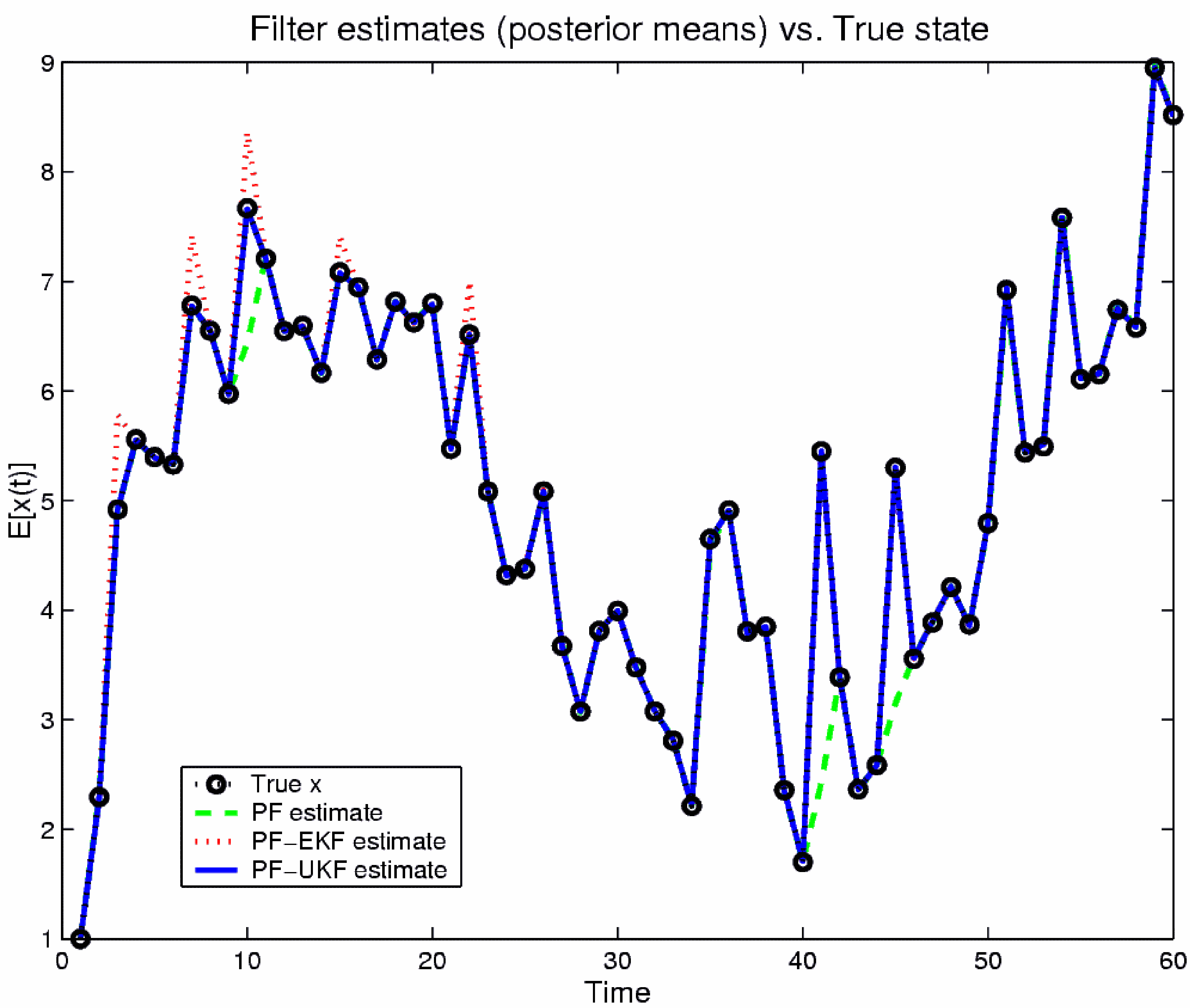
# Experimental Results

- **Synthetic Experiment : (100 independent runs)**

| Filter | MSE | |
|---|---|---|
| | mean | variance |
| Extended Kalman Filter (EKF) | 0.374 | 0.015 |
| Unscented Kalman Filter (UKF) | 0.280 | 0.012 |
| Particle Filter : generic | 0.424 | 0.053 |
| Particle Filter : EKF proposal | 0.310 | 0.016 |
| *Unscented Particle Filter* | 0.070 | 0.006 |

# Experimental Results

- **Synthetic Experiment**



# Experimental Results

- **Pricing Financial Options**
  - Options : financial derivative that gives the holder the right (but not obligation) to do something in the future.
    - Call option : - allow holder to *buy* an underlying cash product
      - at a *specified future date* ("maturity time")
      - for a *predetermined price* ("strike price")
    - Put option : - allow holder to *sell* an underlying cash product
  - *Black Scholes* partial differential equation
    - Main industry standard for pricing options

$$\frac{\partial f}{\partial t} + rS\frac{\partial f}{\partial S} + \tfrac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} = rf$$

**option value** ($f$)

**risk-free interest rate** ($r$)

**volatility of cash product** ($\sigma$)

**value of underlying cash product** ($S$)

# Experimental Results

- **Pricing Financial Options**
  - Black & Scholes (1973) derived the following pricing solution:

$$C = S\mathcal{N}_c(d_1) - Xe^{-rt_m}\mathcal{N}_c(d_2)$$

$$P = -S\mathcal{N}_c(-d1) + Xe^{-rt_m}\mathcal{N}_c(-d_2)$$

$$d_1 = \frac{\ln(S/X) + (r + \frac{1}{2}s^2)t_m}{s\sqrt{t_m}}$$

$$d_2 = d_1 - s\sqrt{t_m}$$

$$\mathcal{N}_c(.) = \text{cumulative normal distribution}$$

# Experimental Results

- **Pricing Financial Options**
  - State-space representation to model system for particle filters
    - Hidden states : $r$ , $s$
    - Output observations: $C$ , $P$
    - Known control signals: $t_m$ , $S$
  - Estimate call and put prices over a 204 day period on the FTSE-100 index.
    - Performance : normalized square error for one-step-ahead predictions

$$NSE = \sqrt{\sum_k \left(\mathbf{y}_k - \hat{\mathbf{y}}_k\right)}$$

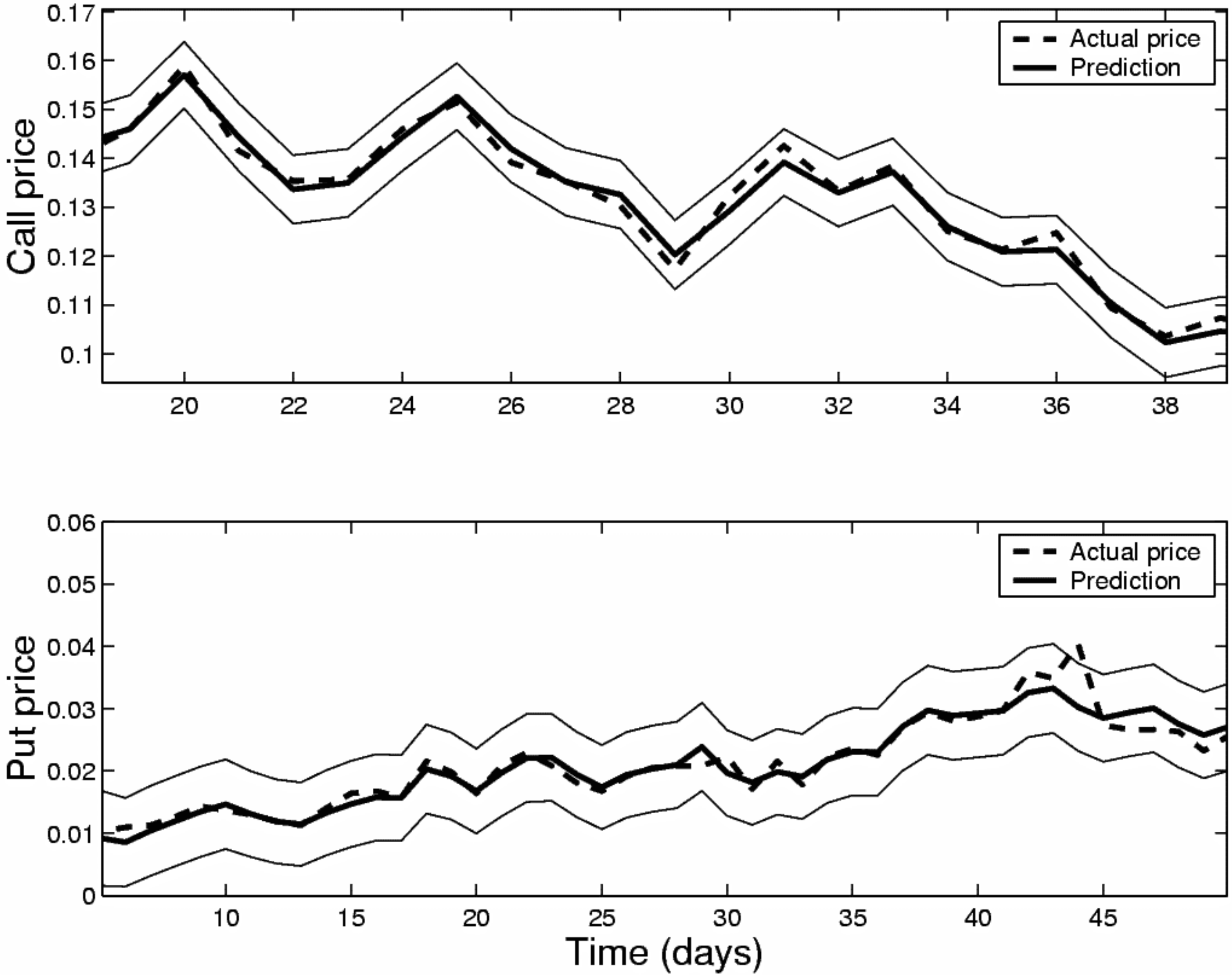
# Experimental Results

- **Options Pricing Experiment : (100 independent runs)**

| Option Type | Algorithm | NSE | |
|---|---|---|---|
| | | mean | var |
| **Call** | Trivial | 0.078 | 0.000 |
| | Extended Kalman Filter (EKF) | 0.037 | 0.000 |
| | Unscented Kalman Filter (UKF) | 0.037 | 0.000 |
| | Particle Filter : generic | 0.037 | 0.000 |
| | Particle Filter : EKF proposal | 0.092 | 0.508 |
| | *Unscented Particle Filter* | *0.009* | *0.000* |
| **Put** | Trivial | 0.035 | 0.000 |
| | Extended Kalman Filter (EKF) | 0.023 | 0.000 |
| | Unscented Kalman Filter (UKF) | 0.023 | 0.000 |
| | Particle Filter : generic | 0.023 | 0.000 |
| | Particle Filter : EKF proposal | 0.024 | 0.007 |
| | *Unscented Particle Filter* | *0.008* | *0.000* |

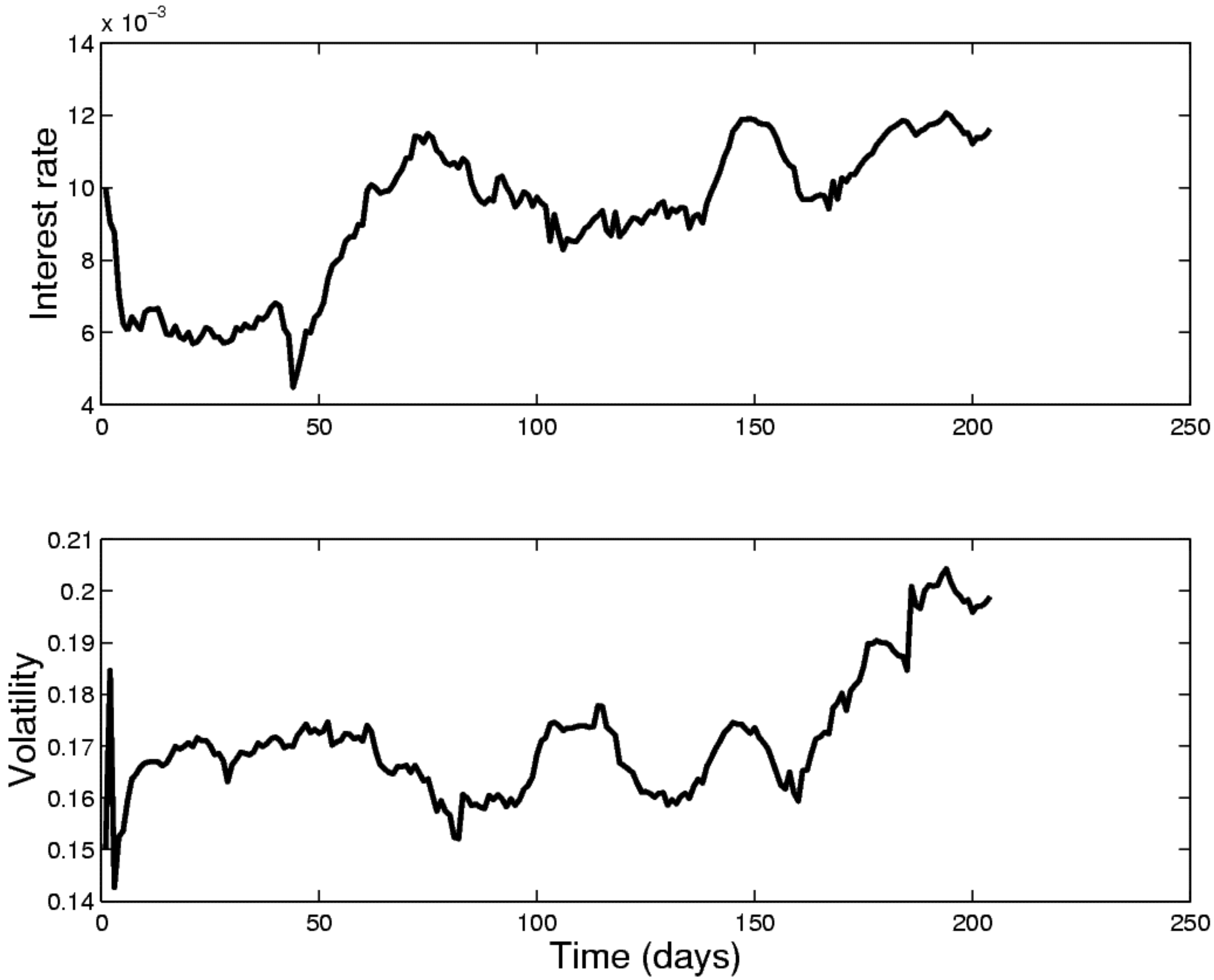
# Experimental Results

- **Options Pricing Experiment : UPF one-step-ahead predictions**



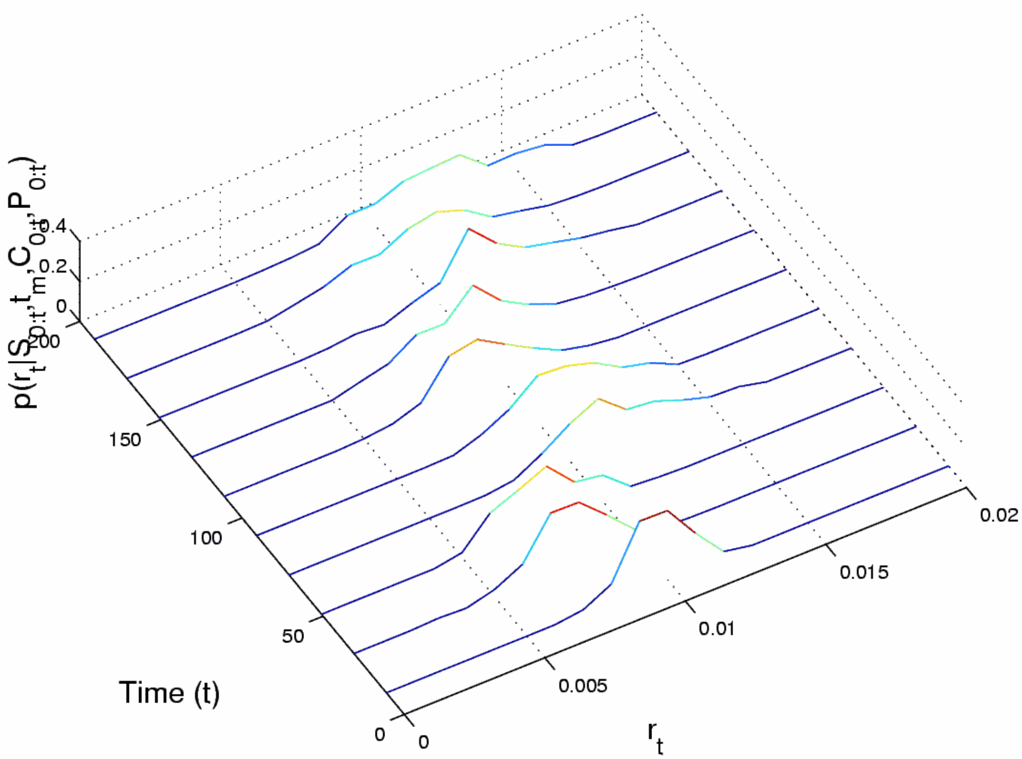
# Experimental Results

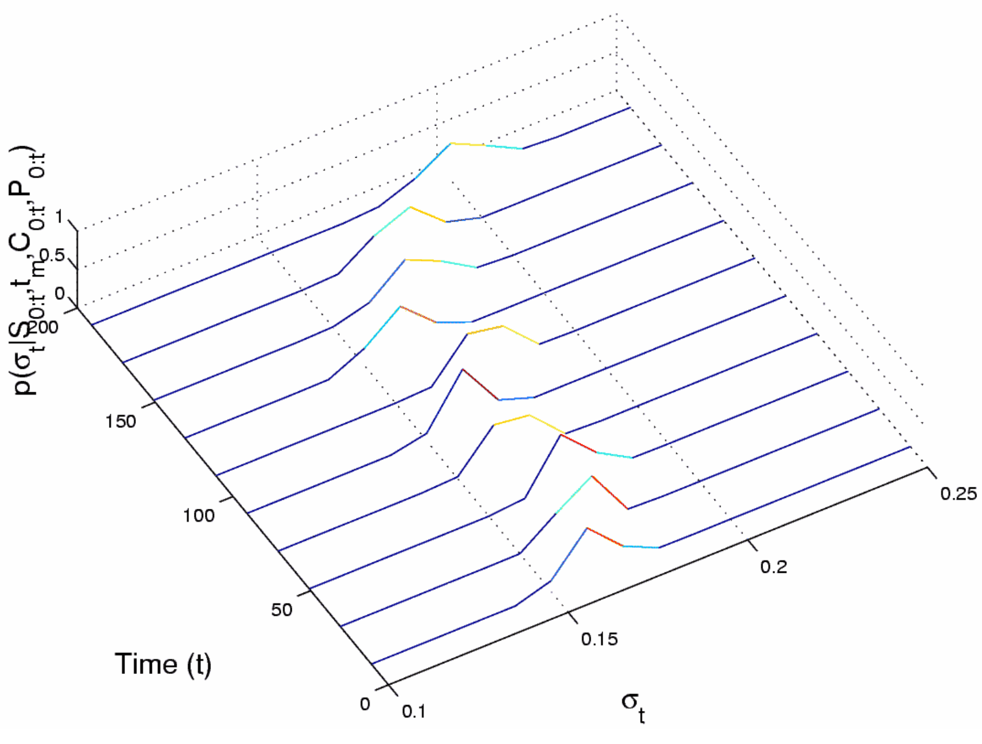
- **Options Pricing Experiment : Estimated interest rate and volatility**



# Experimental Results

- **Options Pricing Experiment : Probability distributions of implied interest rate and volatility**

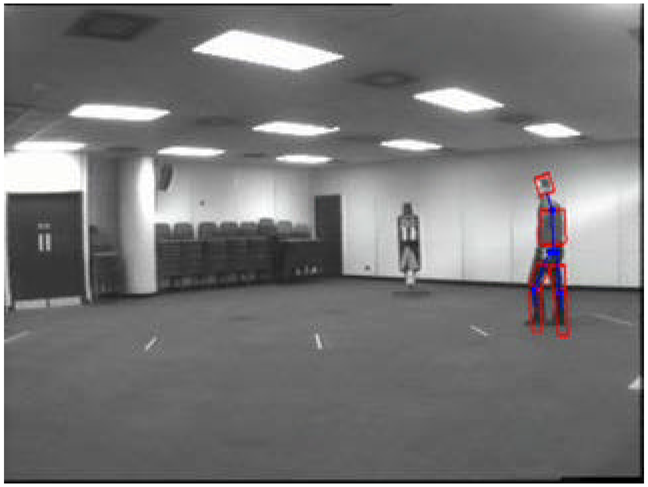


# Particle Filter Demos

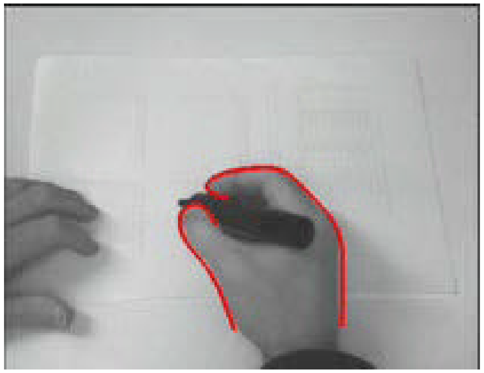
- *Visual Dynamics Group*, Oxford. (Andrew Blake)

**Tracking agile motion**



**Tracking motion against camouflage**

**Mixed state tracking**



# Conclusions

- **Particle filters allow for a practical but complete representation of posterior probability distribution.**

- **Applicable to general nonlinear, non-Gaussian estimation problems where standard Gaussian approximations fail.**

- **Particle filters rely on importance sampling, so the proper choice of proposal distribution is very important:**

  - Standard approach (i.e. transition prior proposal) fails when likelihood of new data is very peaked (accurate sensors) or for heavy-tailed noise sources.

  - EKF proposal : Incorporates new observations, but can diverge due to inaccurate and inconsistent state estimates.

  - *Unscented Particle Filter* : UKF proposal
    - More consistent and accurate state estimates.
    - Larger support overlap, can have heavier tailed distributions.
    - Theory predicts and experiments prove ***significantly better*** performance.

**The End**