Data Assimilation Research Centre

University of Reading

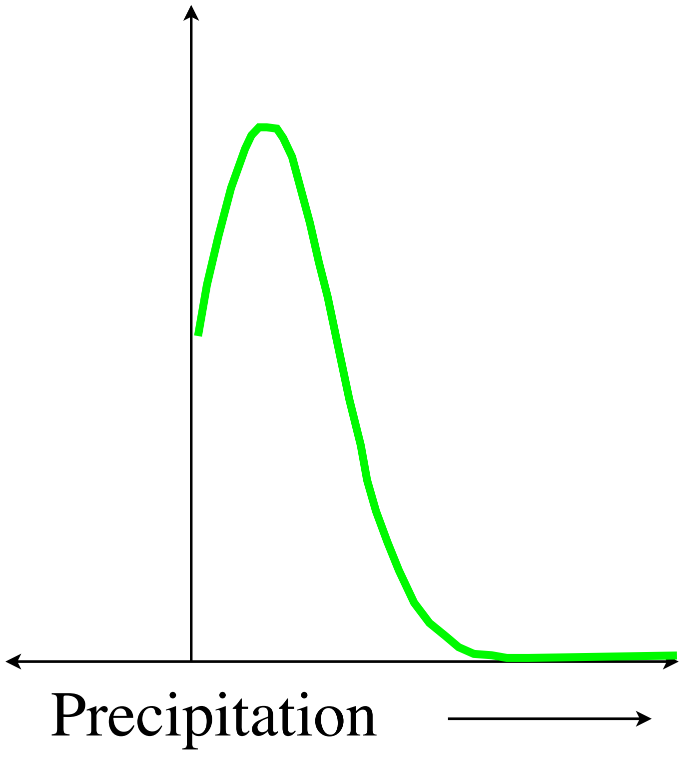
Nonlinear data assimilation
Part I: Particle filters

Melanie Ades, Data Assimilation Research Centre (DARC)
KAUST, Saudi Arabia, February 2014

Why nonlinear data assimilation?

Particle filters are a fully nonlinear data assimilation method, but why are nonlinear schemes required?

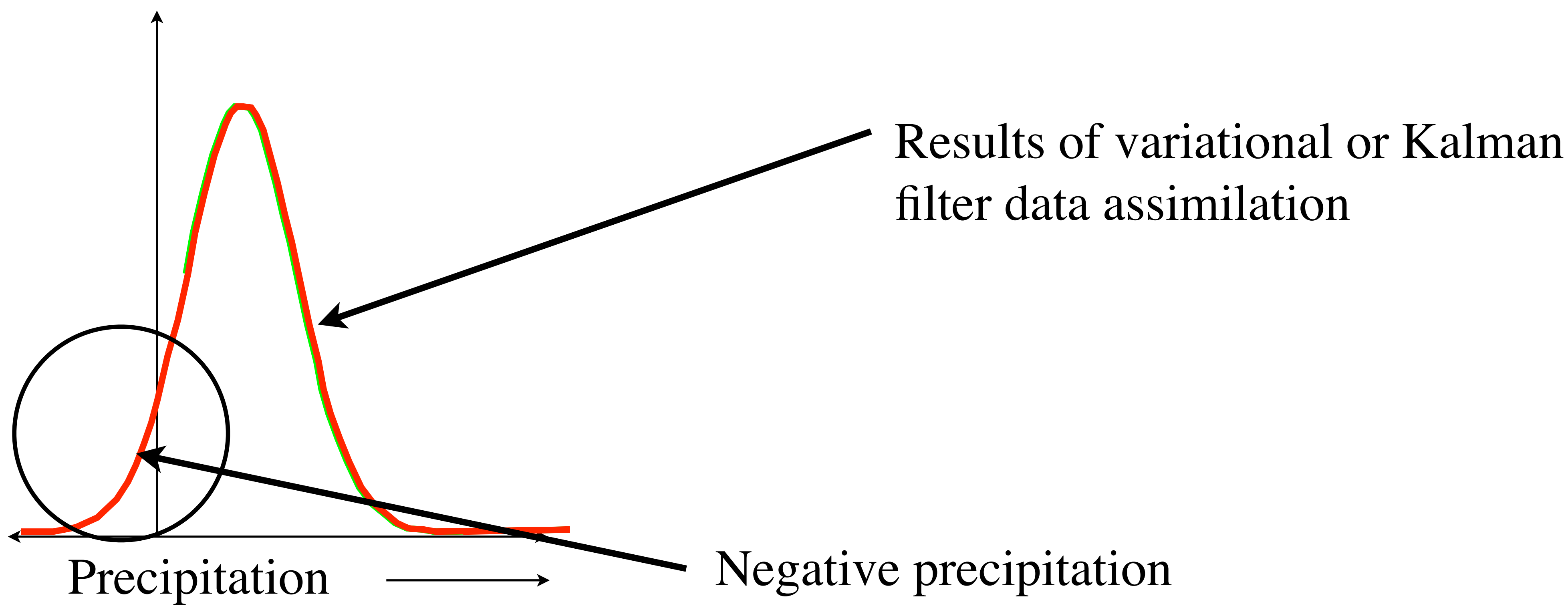
The complexity of models are increasing:



Why nonlinear data assimilation?

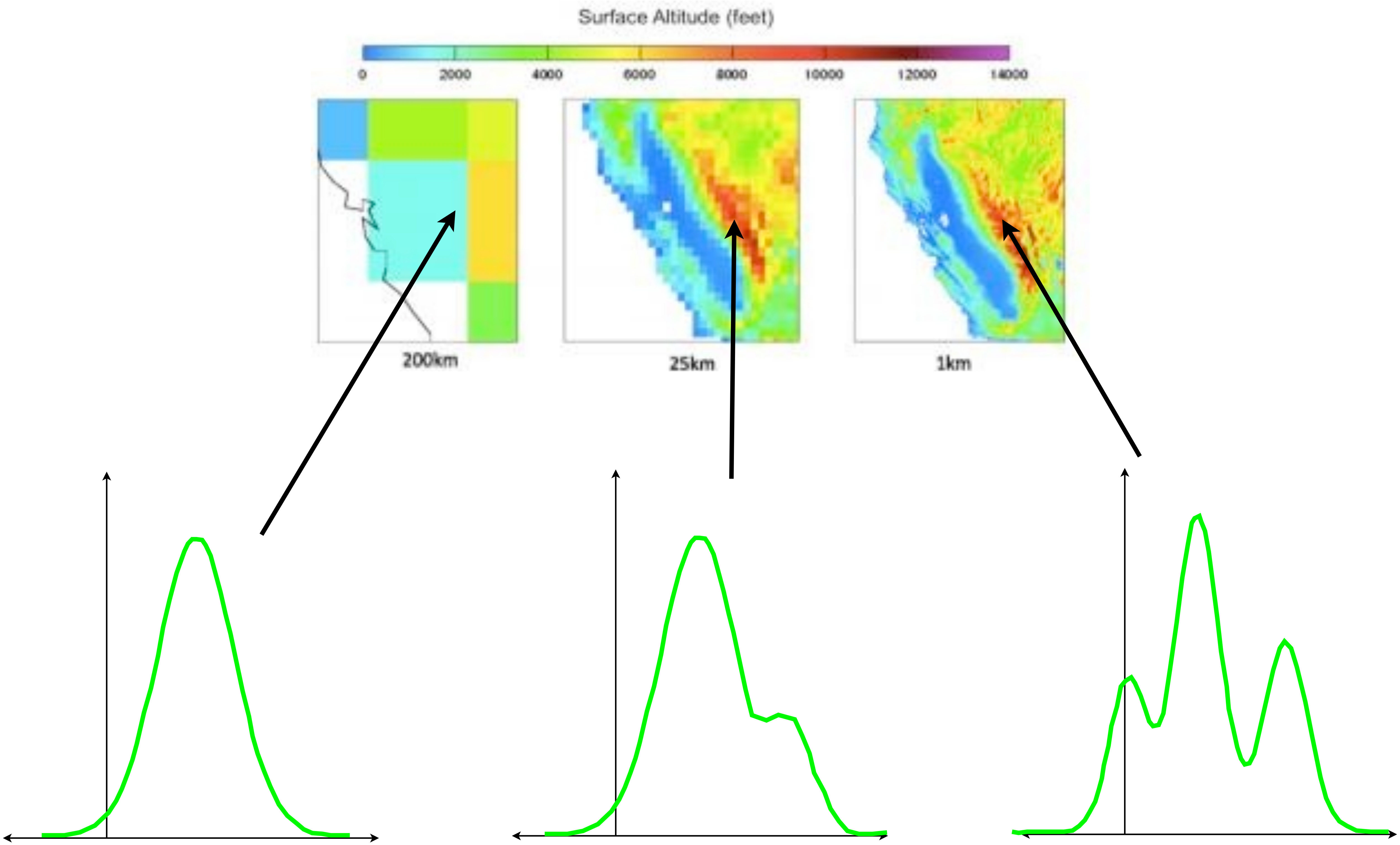
Particle filters are a fully nonlinear data assimilation method, but why are nonlinear schemes required?

The complexity of models are increasing:



Why nonlinear data assimilation?

The resolution of models is also increasing:



Particle filters

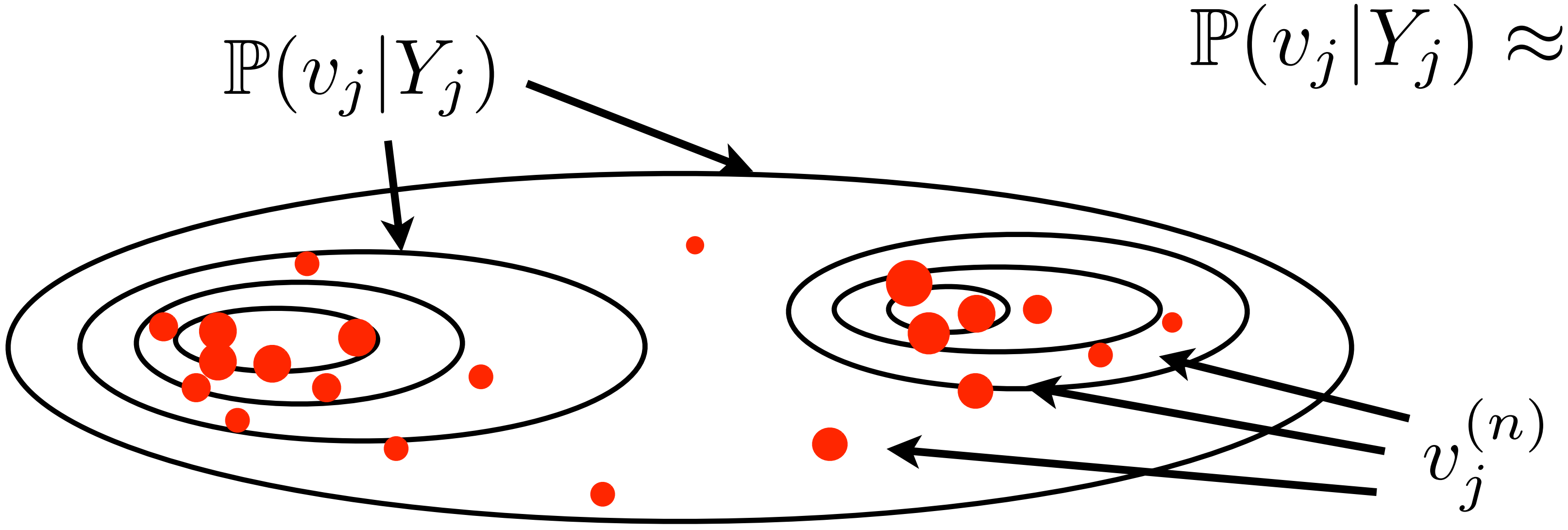
Filtering:

Prediction: $\mathbb{P}(v_{j+1}|Y_j) = \int_{\mathbb{R}^n} \mathbb{P}(v_{j+1}|v_j)\mathbb{P}(v_j|Y_j)dv_j$

Analysis: $\mathbb{P}(v_{j+1}|Y_{j+1}) = \dfrac{\mathbb{P}(y_{j+1}|v_{j+1})\mathbb{P}(v_{j+1}|Y_j)}{\mathbb{P}(y_{j+1}|Y_j)}$

Particle filtering:

$$\mathbb{P}(v_j|Y_j) \approx \sum_{n=1}^{N} w_j^{(n)} \delta(v_j - v_j^{(n)})$$



Particle filters

How to determine the new positions and weights of particles at time j+1 given the weights and positions at time j?

$$\{v_j^{(n)}, w_j^{(n)}\}_{n=1}^N \mapsto \{v_{j+1}^{(n)}, w_{j+1}^{(n)}\}_{n=1}^N$$

Prediction:

$$\mathbb{P}(v_{j+1}|Y_j) = \int_{\mathbb{R}^n} \mathbb{P}(v_{j+1}|v_j)\mathbb{P}(v_j|Y_j)dv_j$$

$$= \int_{\mathbb{R}^n} \frac{\mathbb{P}(v_{j+1}|v_j)}{\mathbb{Q}(v_{j+1}|v_j, Y_{j+1})}\mathbb{Q}(v_{j+1}|v_j, Y_{j+1})\mathbb{P}(v_j|Y_j)dv_j$$

Particle filters

Overall framework, given $\{v_j^{(n)}, w_j^{(n)}\}_{n=1}^N$

1) Sample the new position of each particle by sampling from a proposal probability distribution

$$v_{j+1}^{(n)} \sim \mathbb{Q}(v_{j+1}|v_j^{(n)}, Y_{j+1})$$

2) This new particle is then re-weighted according to the analysis formulae of filtering to give

$$\widehat{w}_{j+1}^{(n)} = w_j^{(n)} \frac{\mathbb{P}(y_{j+1}|v_{j+1}^{(n)})\mathbb{P}(v_{j+1}^{(n)}|v_j^{(n)})}{\mathbb{Q}(v_{j+1}^{(n)}|v_j^{(n)}, Y_{j+1})}$$

3) The weights are normalised so that they sum to one

$$w_{j+1}^{(n)} = \frac{\widehat{w}_{j+1}^{(n)}}{\left(\sum_{n=1}^N \widehat{w}_{j+1}^{(n)}\right)}$$

Leading to:

$$\mathbb{P}(v_{j+1}|Y_{j+1}) \approx \mathbb{P}^N(v_{j+1}|Y_{j+1}) := \sum_{n+1}^{N} w_{j+1}^{(n)} \delta(v_{j+1} - v_{j+1}^{(n)})$$

Different particle filters

The key two elements that differ between particle filters are:

proposal distribution

$$v_{j+1}^{(n)} \sim \mathbb{Q}(v_{j+1}^{(n)} | v_j^{(n)}, Y_{j+1})$$

associated weight update

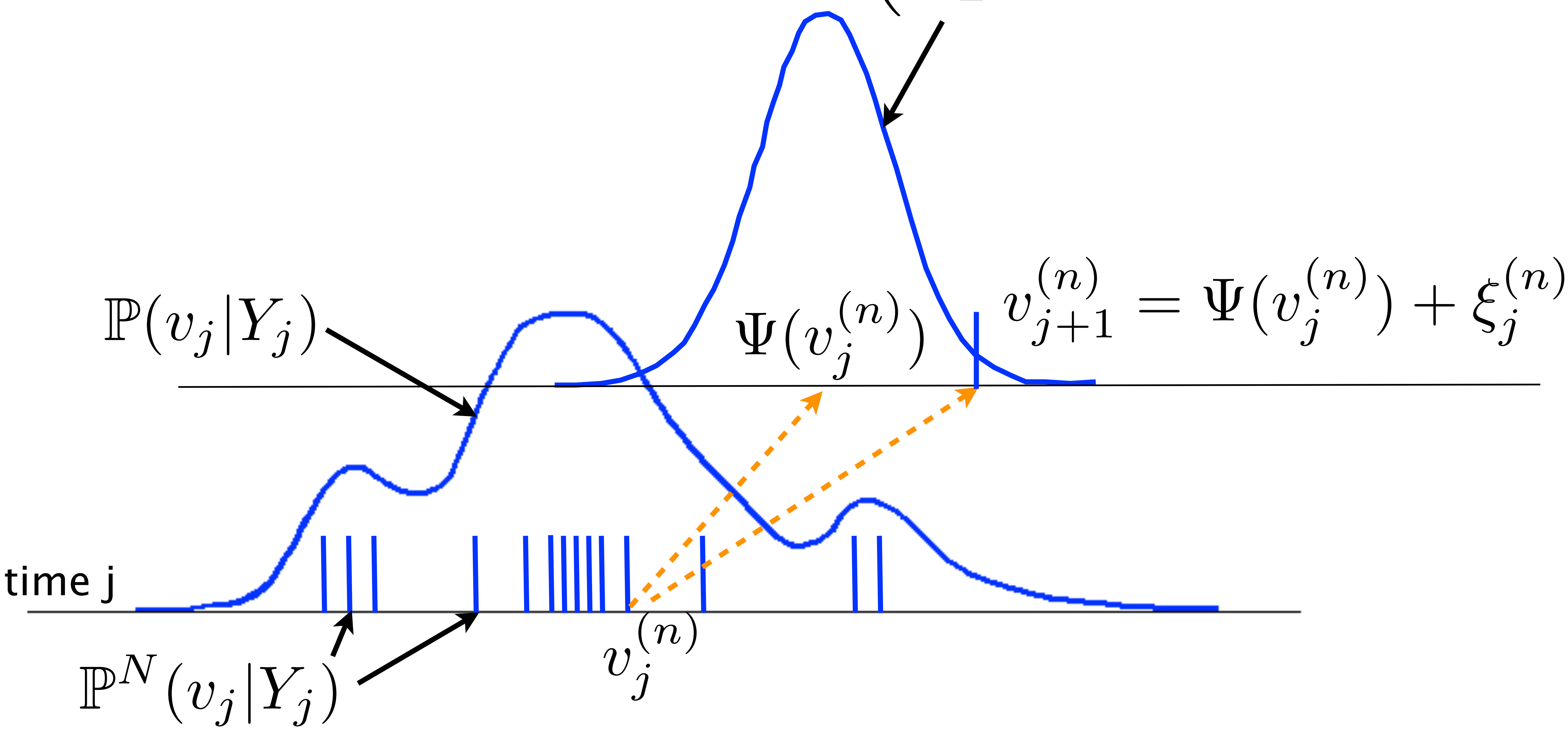
$$\widehat{w}_{j+1}^{(n)} = w_j^{(n)} \frac{\mathbb{P}(y_{j+1} | v_{j+1}^{(n)}) \mathbb{P}(v_{j+1}^{(n)} | v_j^{(n)})}{\mathbb{Q}(v_{j+1}^{(n)} | v_j^{(n)}, Y_{j+1})}$$

Standard particle filter - proposal density

proposal distribution

$$\mathbb{Q}(v_{j+1}|v_j, Y_{j+1}) \equiv \mathbb{P}(v_{j+1}|v_j)$$

$$\mathbb{P}(v_{j+1}|v_j) \propto \exp\left(-\frac{1}{2}\left|\Sigma^{-\frac{1}{2}}(v_{j+1}-\Psi(v_j))\right|^2\right)$$



Standard particle filter - weight update

$$\mathbb{Q}(v_{j+1}^{(n)}|v_j^{(n)}, Y_{j+1}) \equiv \mathbb{P}(v_{j+1}^{(n)}|v_j^{(n)}) \implies$$

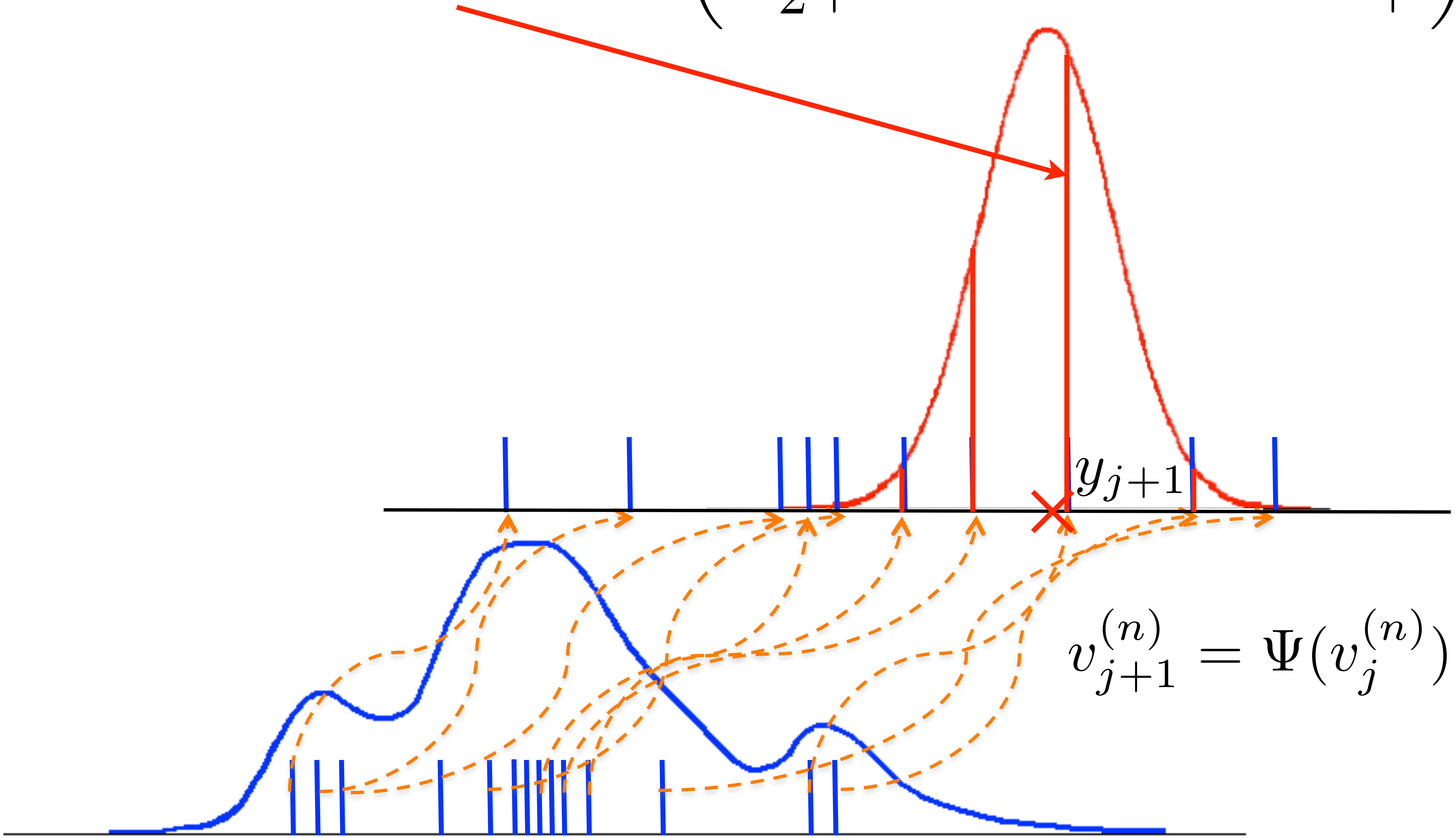
$$\widehat{w}_{j+1}^{(n)} = w_j^{(n)} \frac{\mathbb{P}(y_{j+1}|v_{j+1}^{(n)})\mathbb{P}(v_{j+1}^{(n)}|v_j^{(n)})}{\mathbb{P}(v_{j+1}^{(n)}|v_j^{(n)})}$$

associated weight update

$$\widehat{w}_{j+1}^{(n)} \propto w_j^{(n)}\mathbb{P}(y_{j+1}|v_{j+1}^{(n)})$$

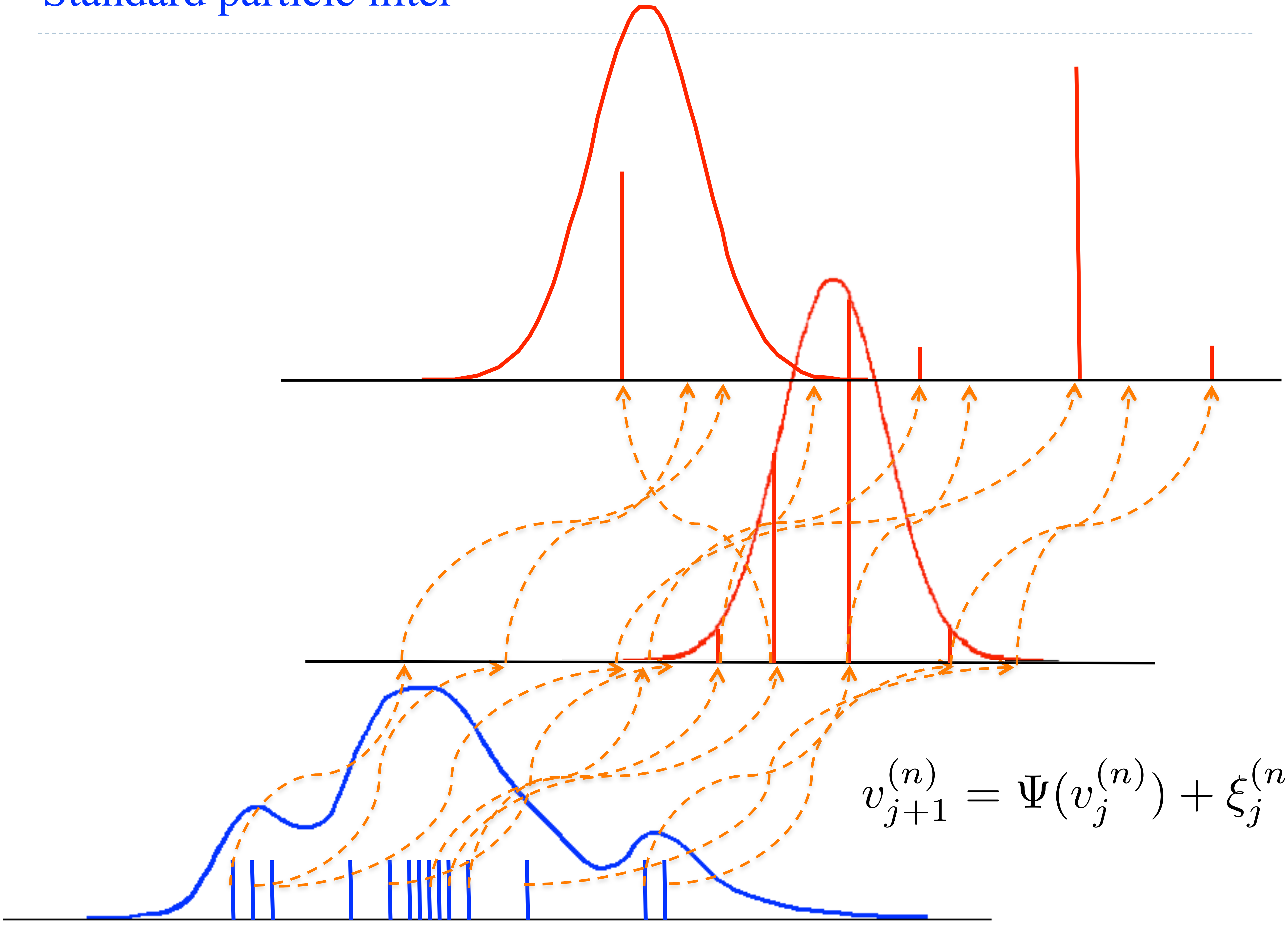
Standard particle filter - weight update

$$\mathbb{P}(y_{j+1}|v_{j+1}^{(n)}) \propto \exp\left(-\frac{1}{2}\left|\Gamma^{-\frac{1}{2}}(y_{j+1} - h(v_{j+1}^{(n)}))\right|^2\right)$$



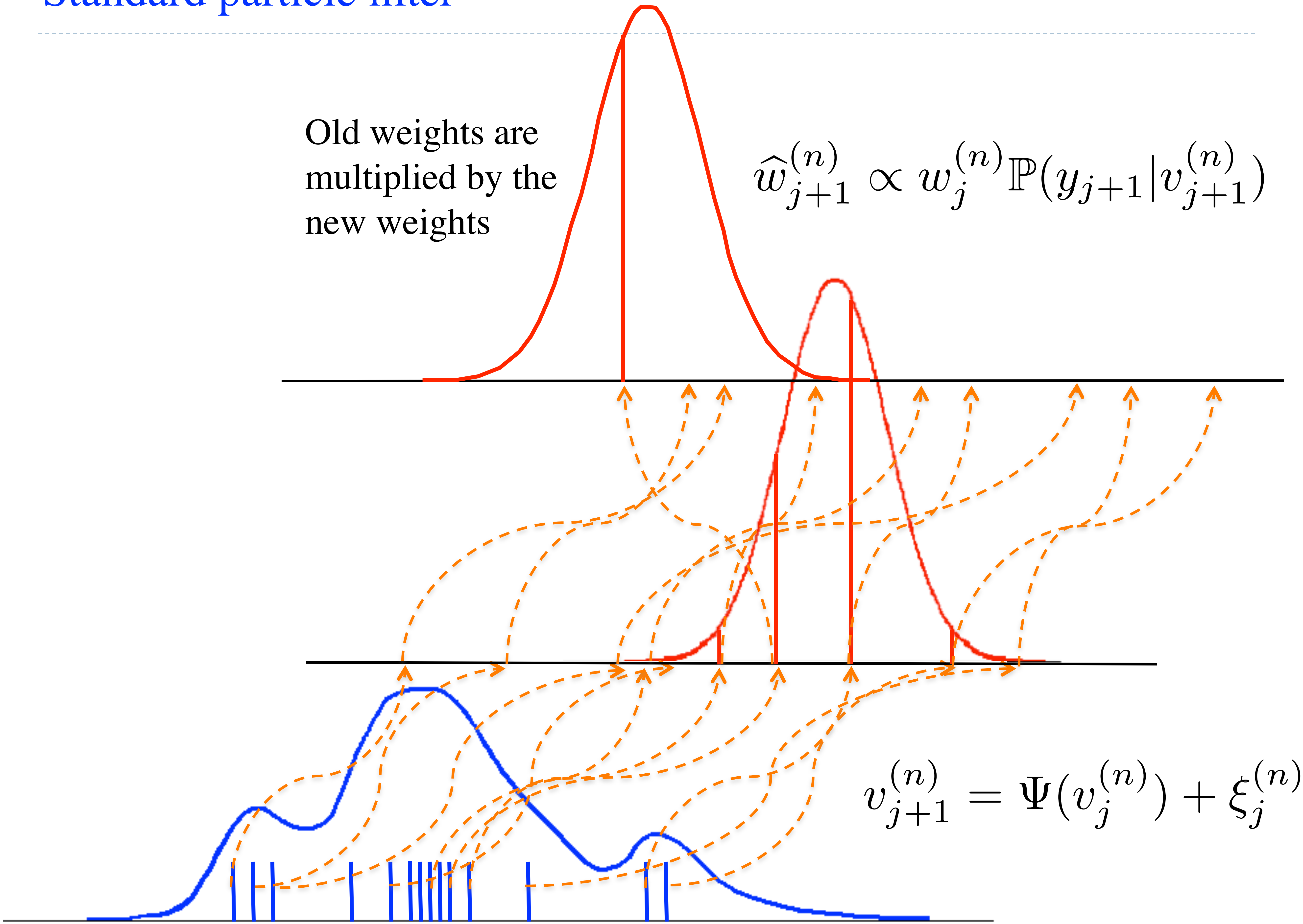
$$v_{j+1}^{(n)} = \Psi(v_j^{(n)}) + \xi_j^{(n)}$$

Standard particle filter

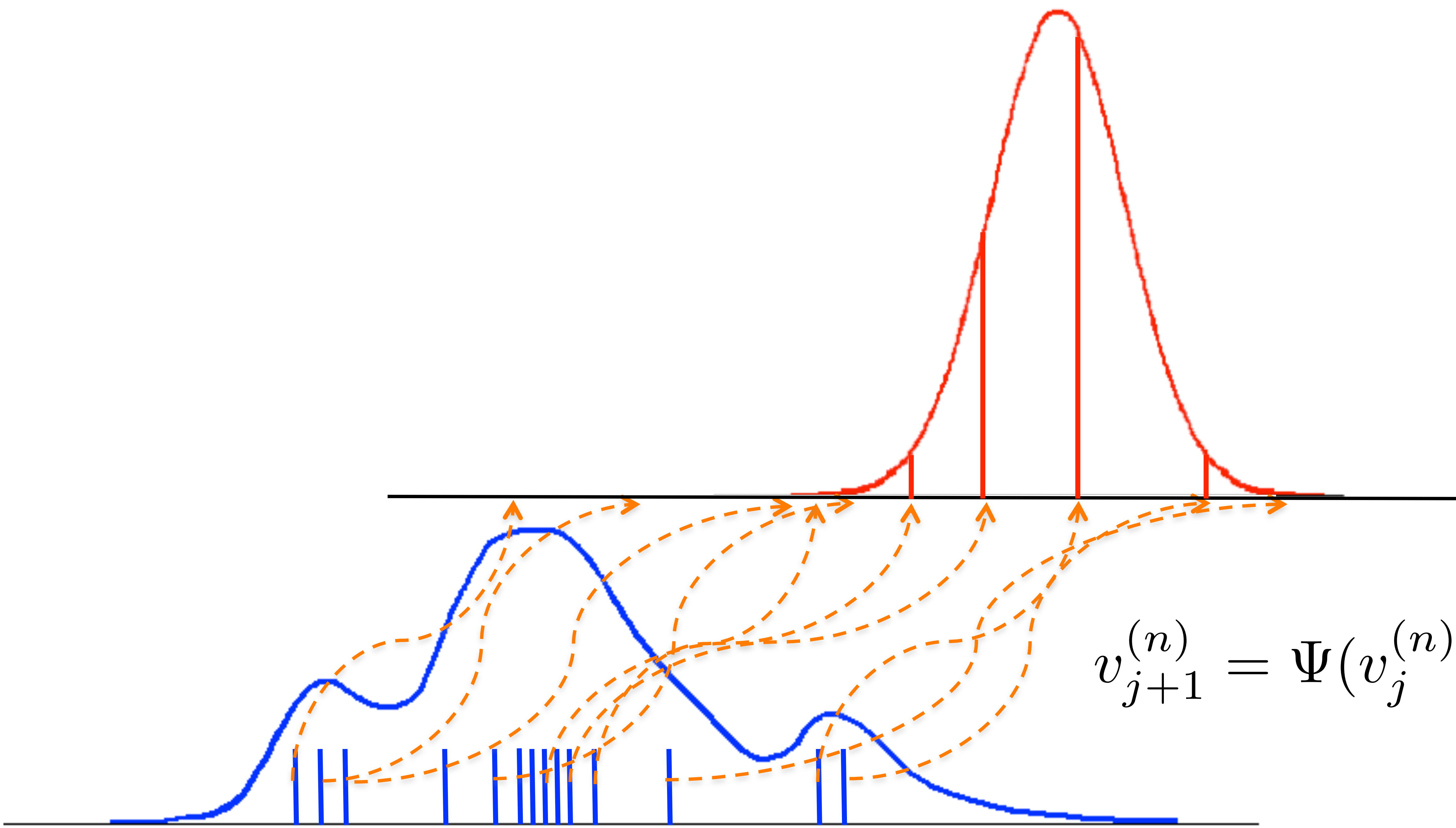


$$v_{j+1}^{(n)} = \Psi(v_j^{(n)}) + \xi_j^{(n)}$$

Standard particle filter

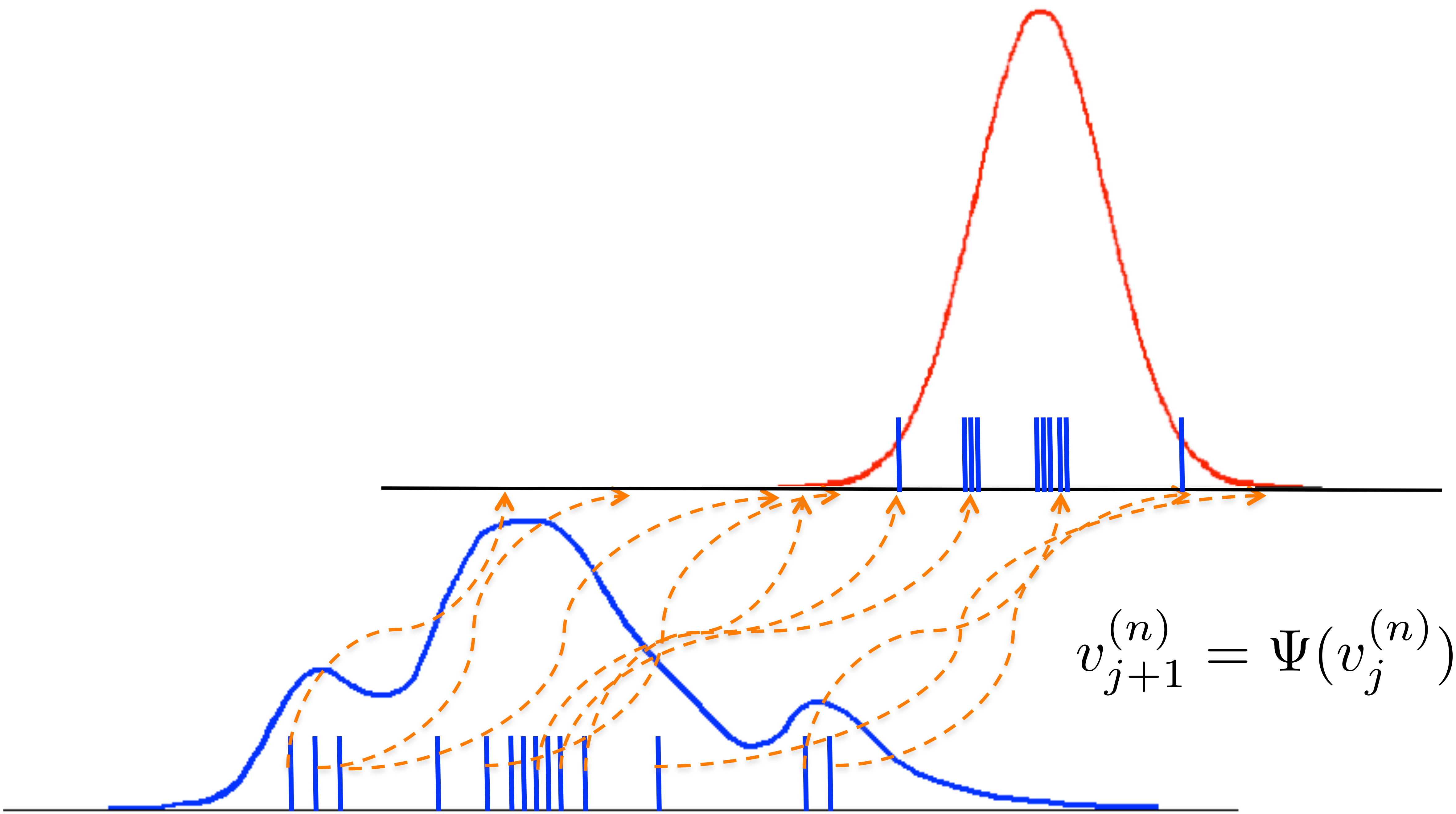


Sequential Importance Resampling (SIR) particle filter



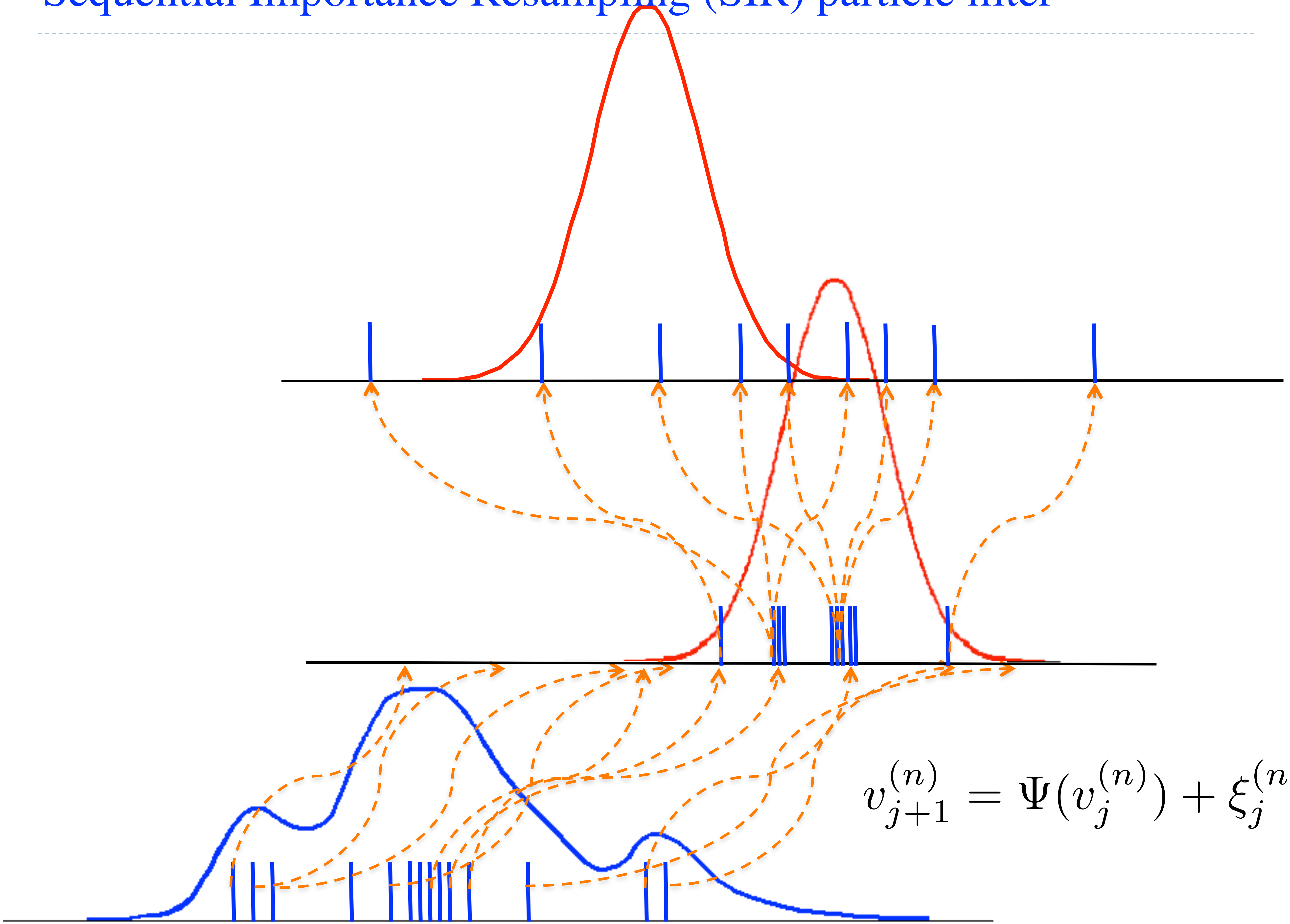
$$v_{j+1}^{(n)} = \Psi(v_j^{(n)}) + \xi_j^{(n)}$$

Sequential Importance Resampling (SIR) particle filter



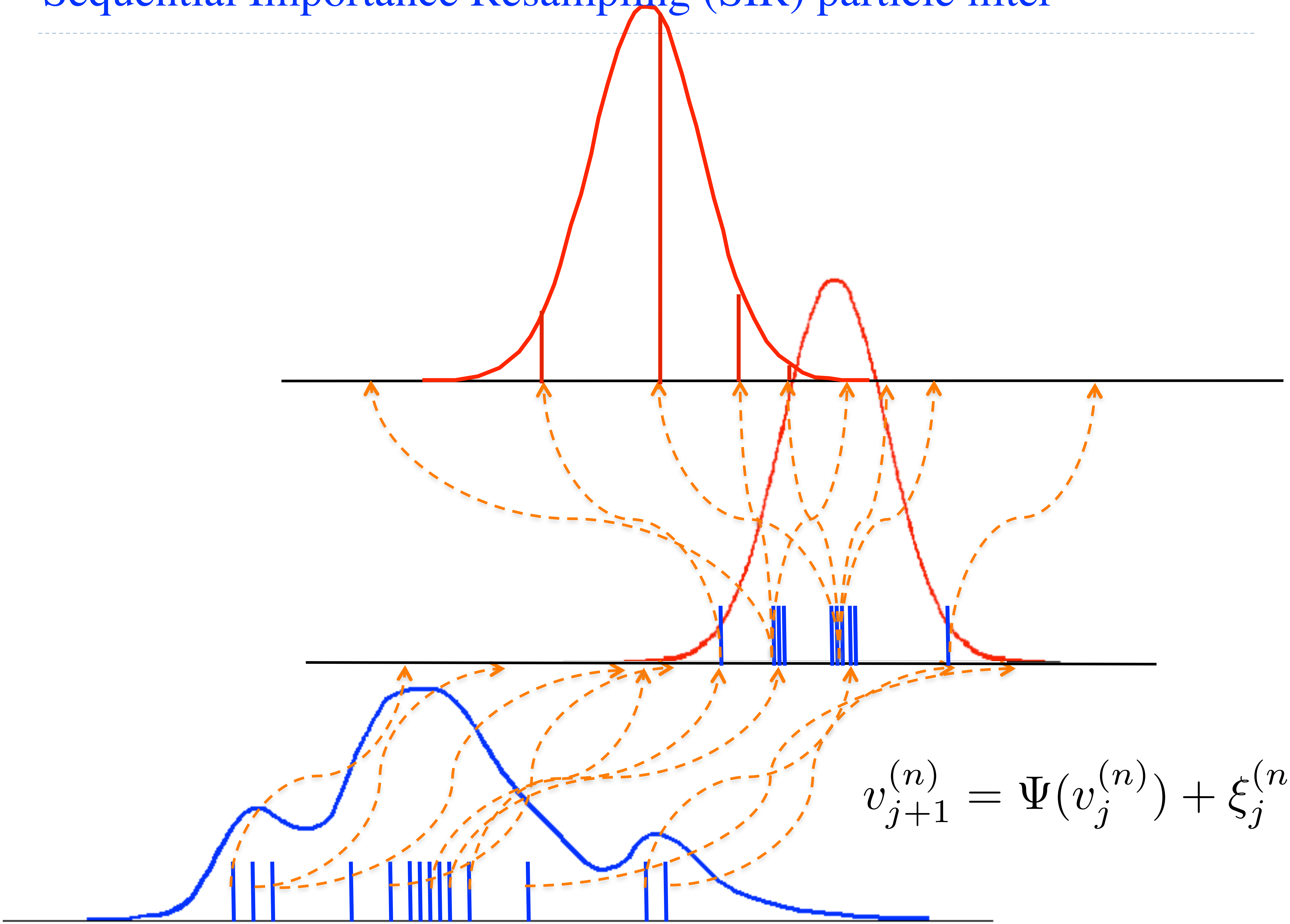
$$v_{j+1}^{(n)} = \Psi(v_j^{(n)}) + \xi_j^{(n)}$$

Sequential Importance Resampling (SIR) particle filter



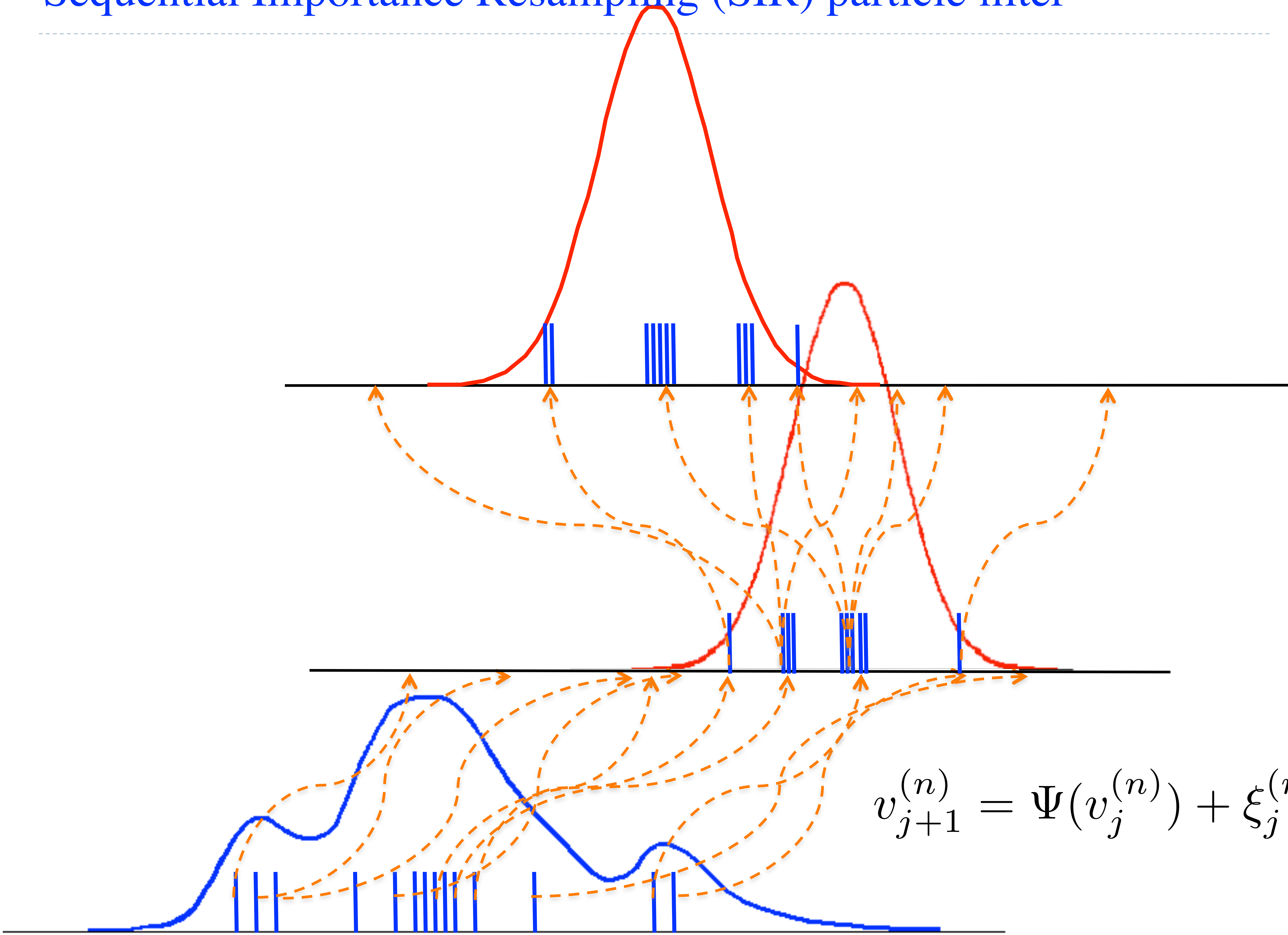
$$v_{j+1}^{(n)} = \Psi(v_j^{(n)}) + \xi_j^{(n)}$$

Sequential Importance Resampling (SIR) particle filter



$$v_{j+1}^{(n)} = \Psi(v_j^{(n)}) + \xi_j^{(n)}$$

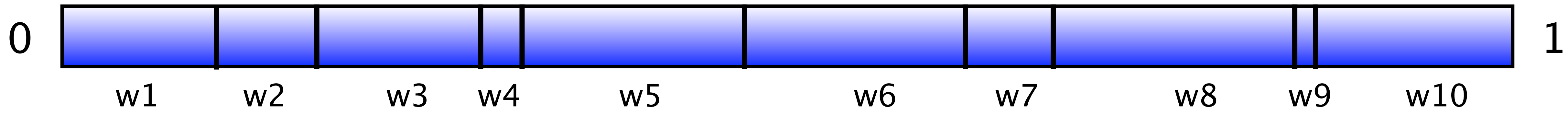
Sequential Importance Resampling (SIR) particle filter



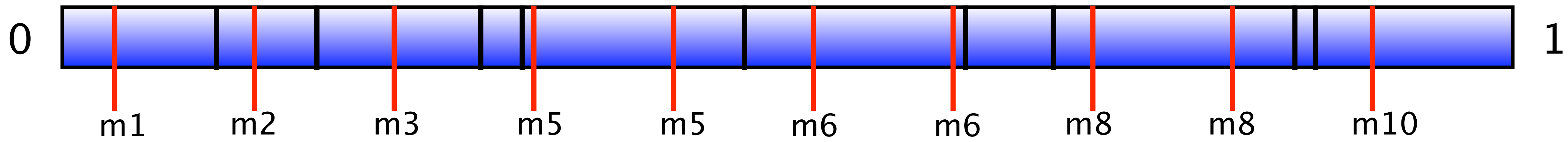
$$v_{j+1}^{(n)} = \Psi(v_j^{(n)}) + \xi_j^{(n)}$$

A simple resampling scheme

1. Put all the weights after each other on the unit interval:



2. Draw a random number from the uniform distribution over [0,1/N], in this case (with 10 members) over [0,1/10].
3. Put that number on the unit interval: its end point is the first member drawn.
4. Add 1/N to the end point: the new end point is our second member drawn. Repeat this until N new members are obtained,



5. In our example we choose m1, m2, m3, m5 twice, m6 twice, m8 twice and m10 and loose m4, m7 and m9 due to small weights

Sequential Importance Resampling (SIR) particle filter

1. Set $j=0$ and $\mathbb{P}^N(v_0|Y_0)=\mathbb{P}(v_0)$

2. Draw $v_j^{(n)} \sim \mathbb{P}^N(v_j|Y_j)$ (resample)

3. Set $w_j^{(n)} = 1/N, \quad n = 1, \ldots, N$

4. Draw $\widehat{v}_{j+1}^{(n)} \sim \mathbb{P}(\widehat{v}_{j+1}|v_j^{(n)})$

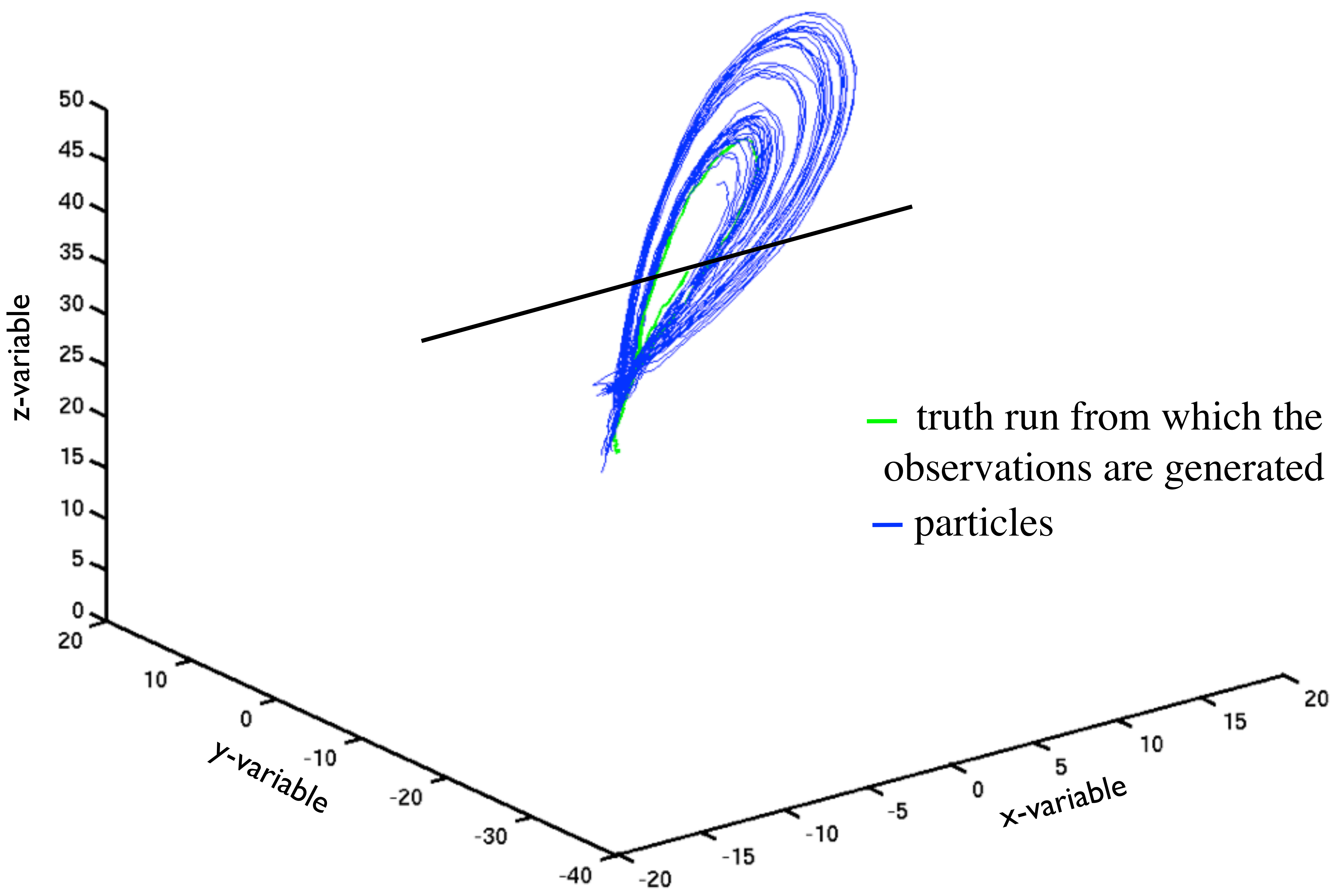
$$\text{i.e. Set } \widehat{v}_{j+1}^{(n)} = \Psi(v_j^{(n)}) + \xi_j^{(n)}, \quad \xi_j^{(n)} \sim N(0, \Sigma)$$

5. Calculate $w_{j+1}^{(n)} = \mathbb{P}(y_{j+1}|v_{j+1}^{(n)}) / \left(\sum_{n=1}^{N} \mathbb{P}(y_{j+1}|v_{j+1}^{(n)}) \right)$

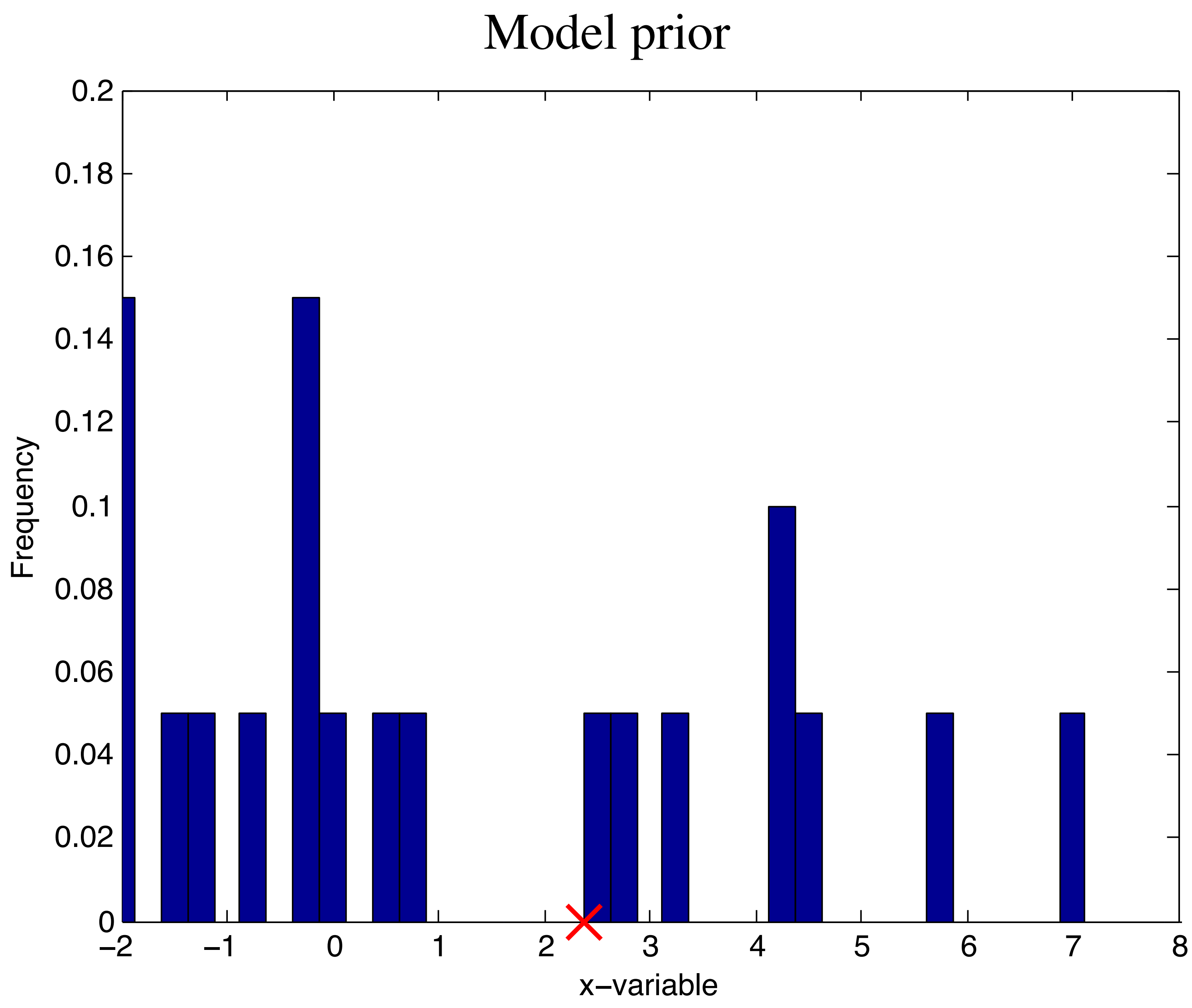
$$\text{where } \mathbb{P}(y_{j+1}|v_{j+1}^{(n)}) \propto \exp\left(-\frac{1}{2} \left| \Gamma^{-\frac{1}{2}} (y_{j+1} - h(v_{j+1}^{(n)})) \right|^2 \right)$$

6. $j+1 \mapsto j$ and return to step 2

Lorenz 63 system of equations - 20 particles

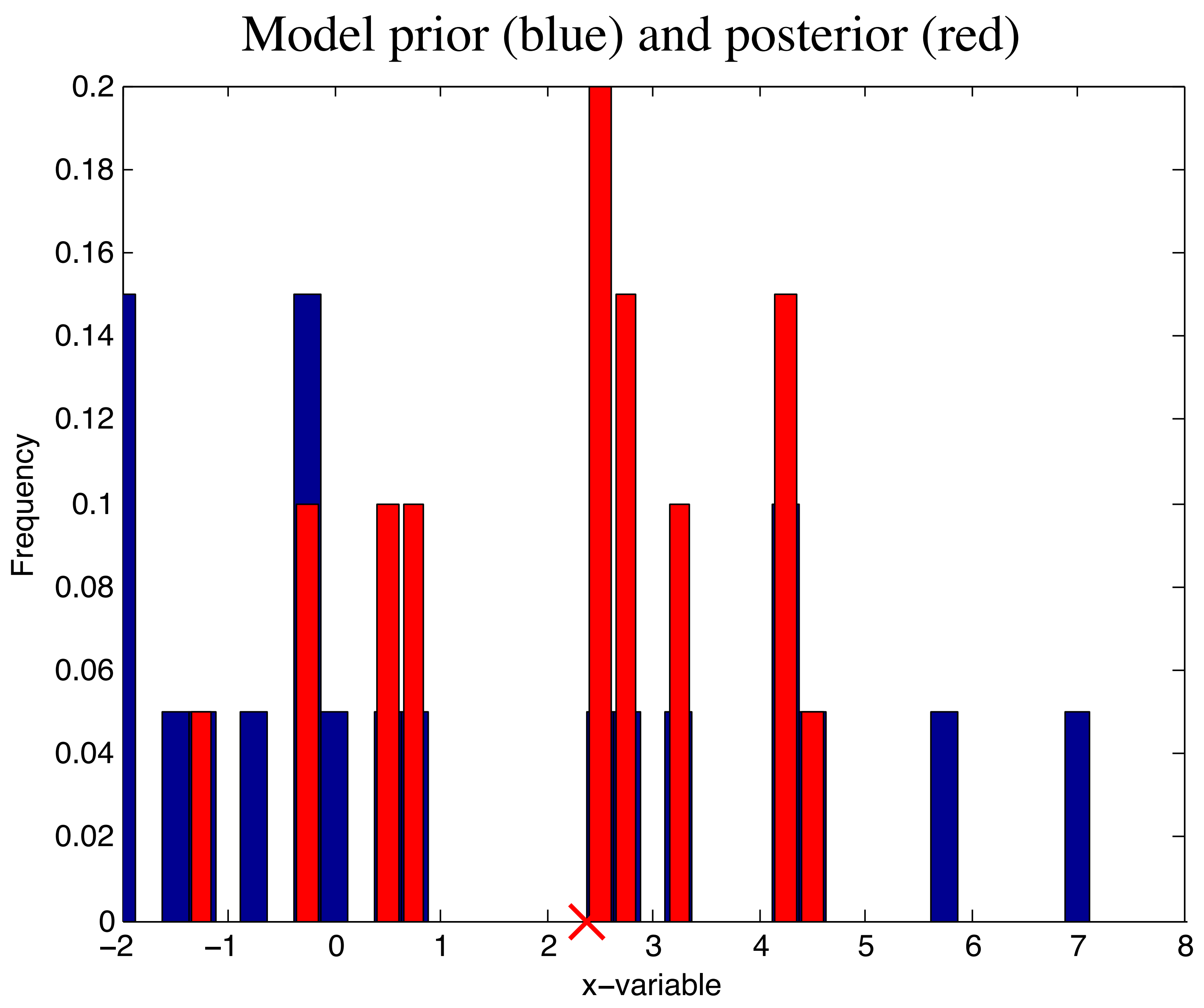


Lorenz 63 system of equations



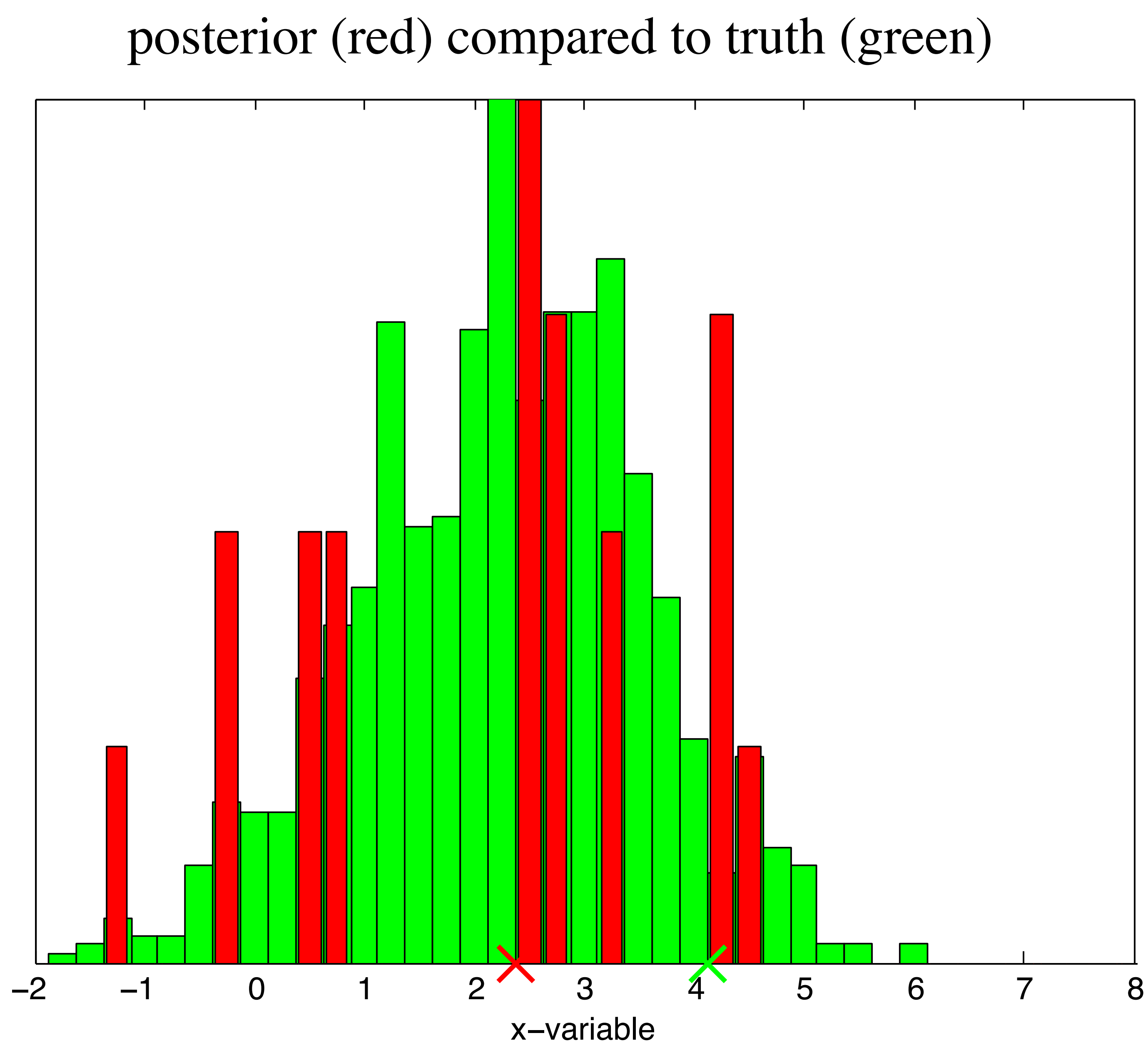
- observation

Lorenz 63 system of equations



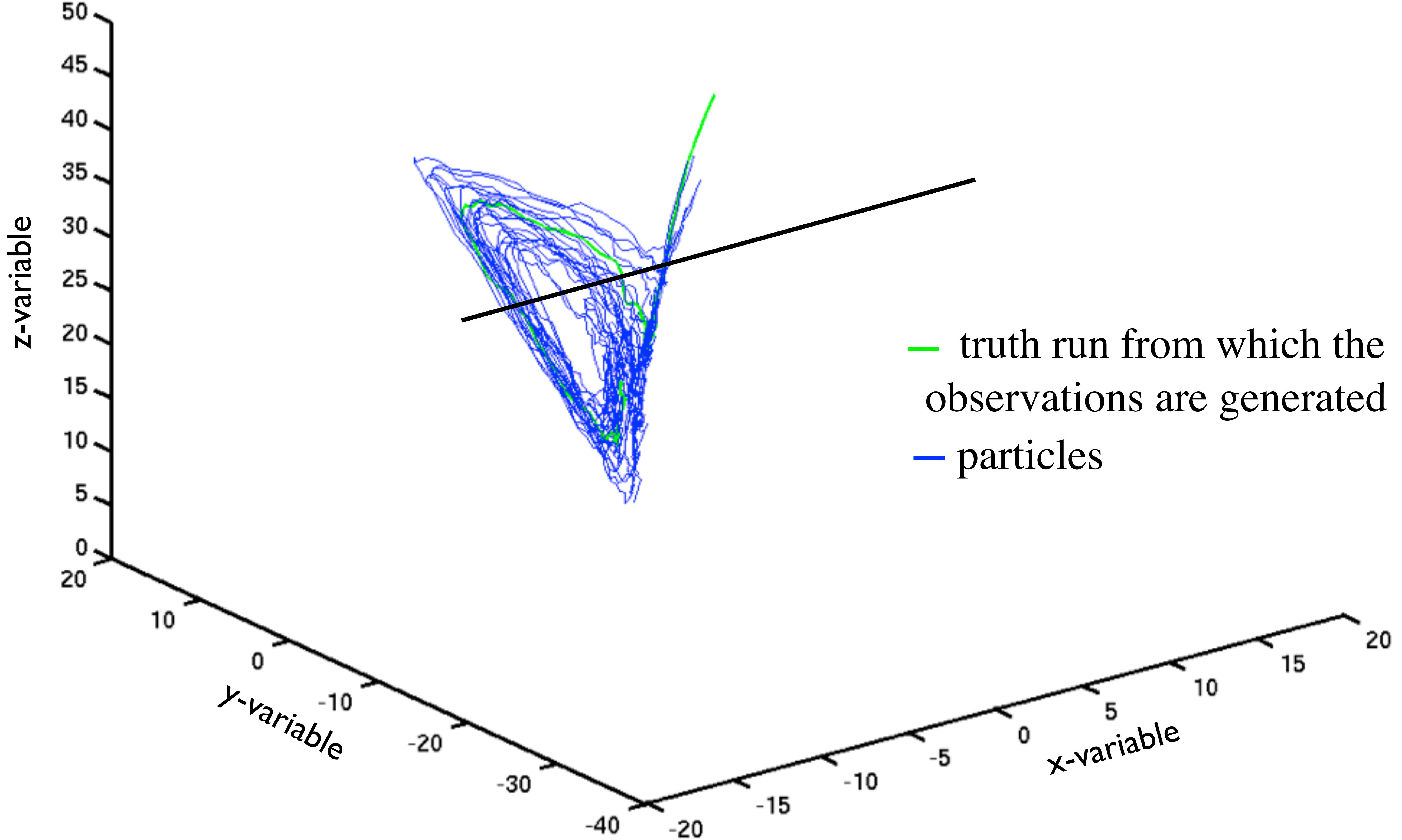
✕ - observation

Lorenz 63 system of equations

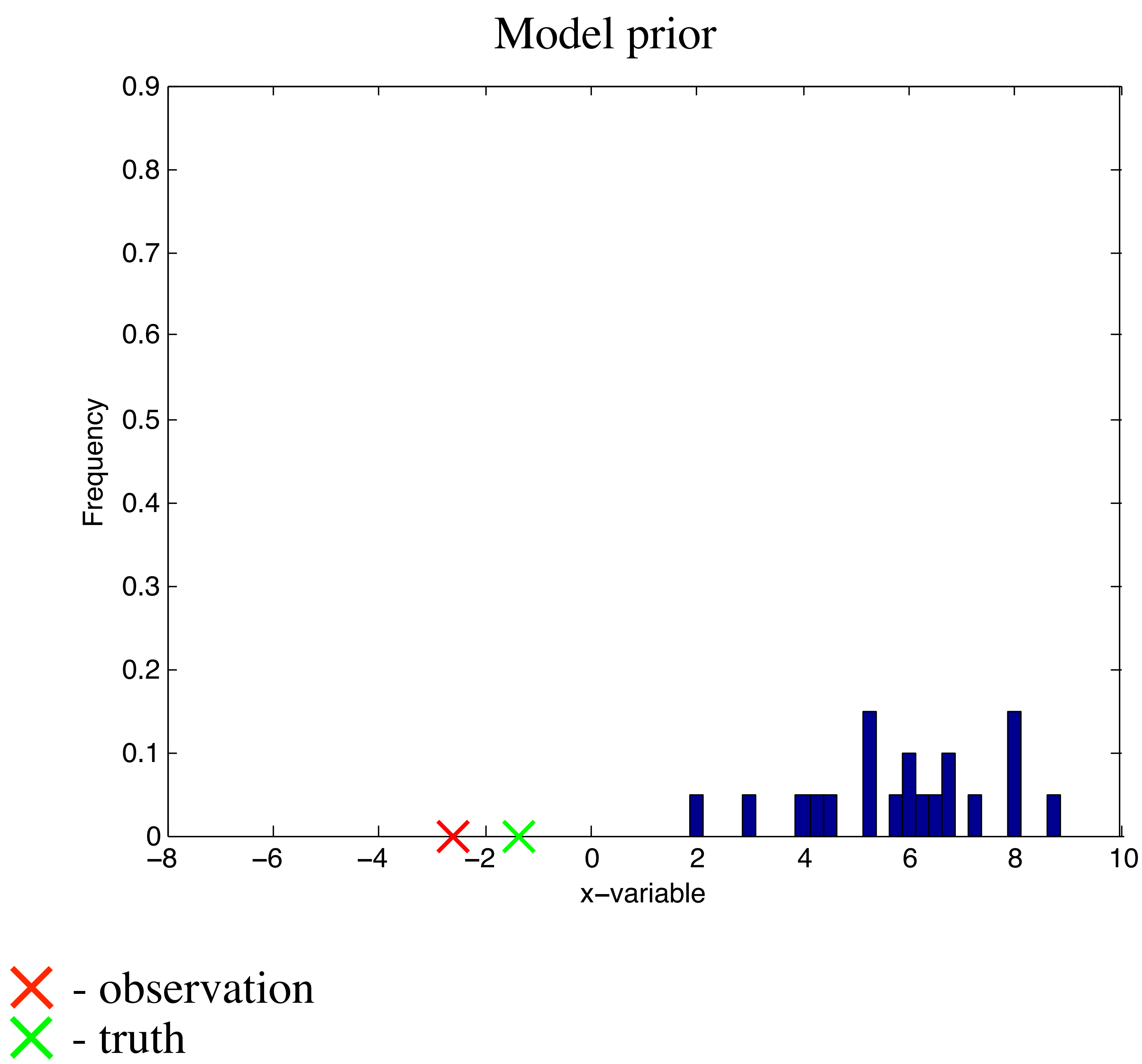


green truth posterior pdf comes from a 1000 particle run of the standard particle filter

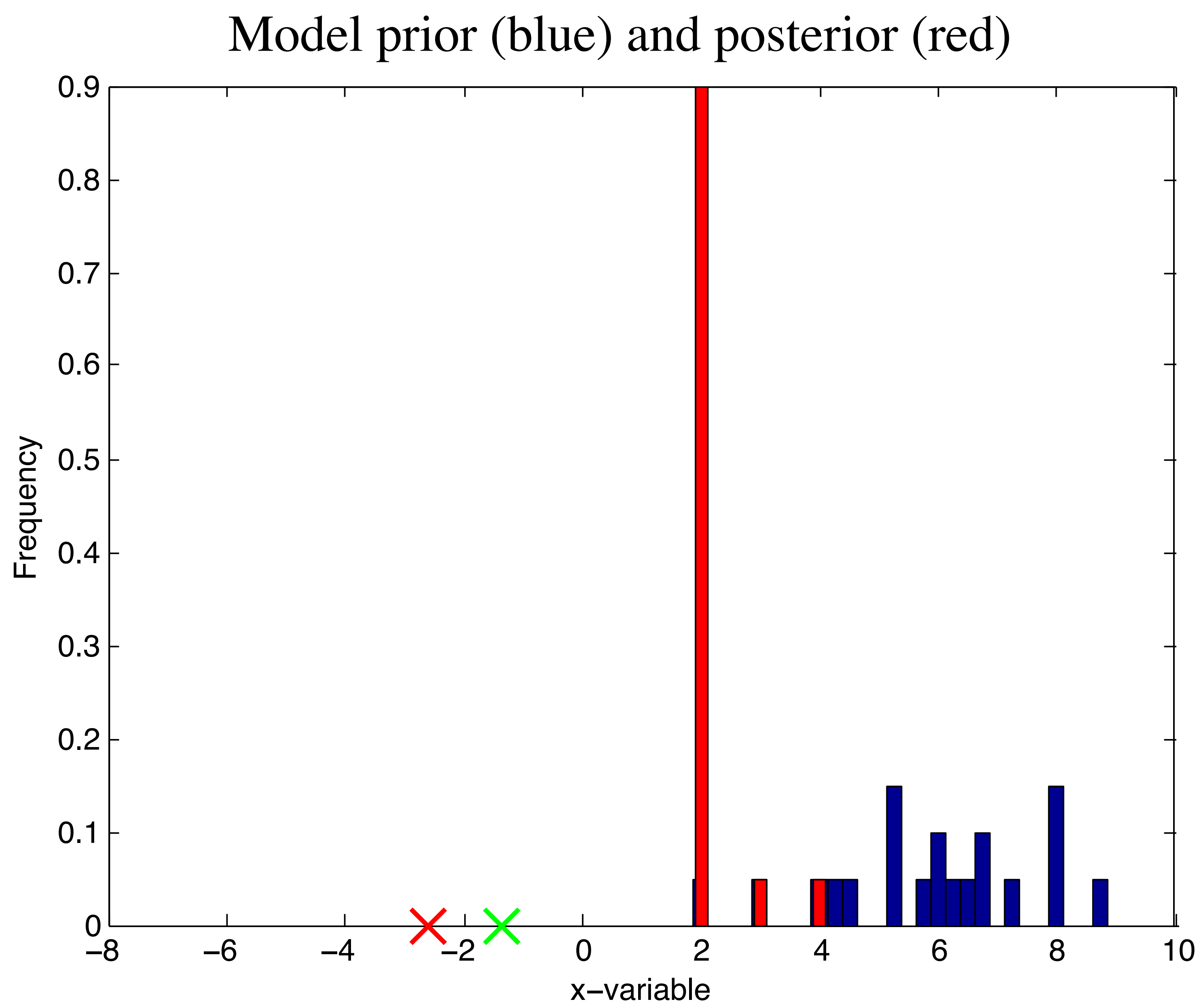
Lorenz 63 system of equations - 20 particles



Standard Particle Filter fails - filter degeneracy



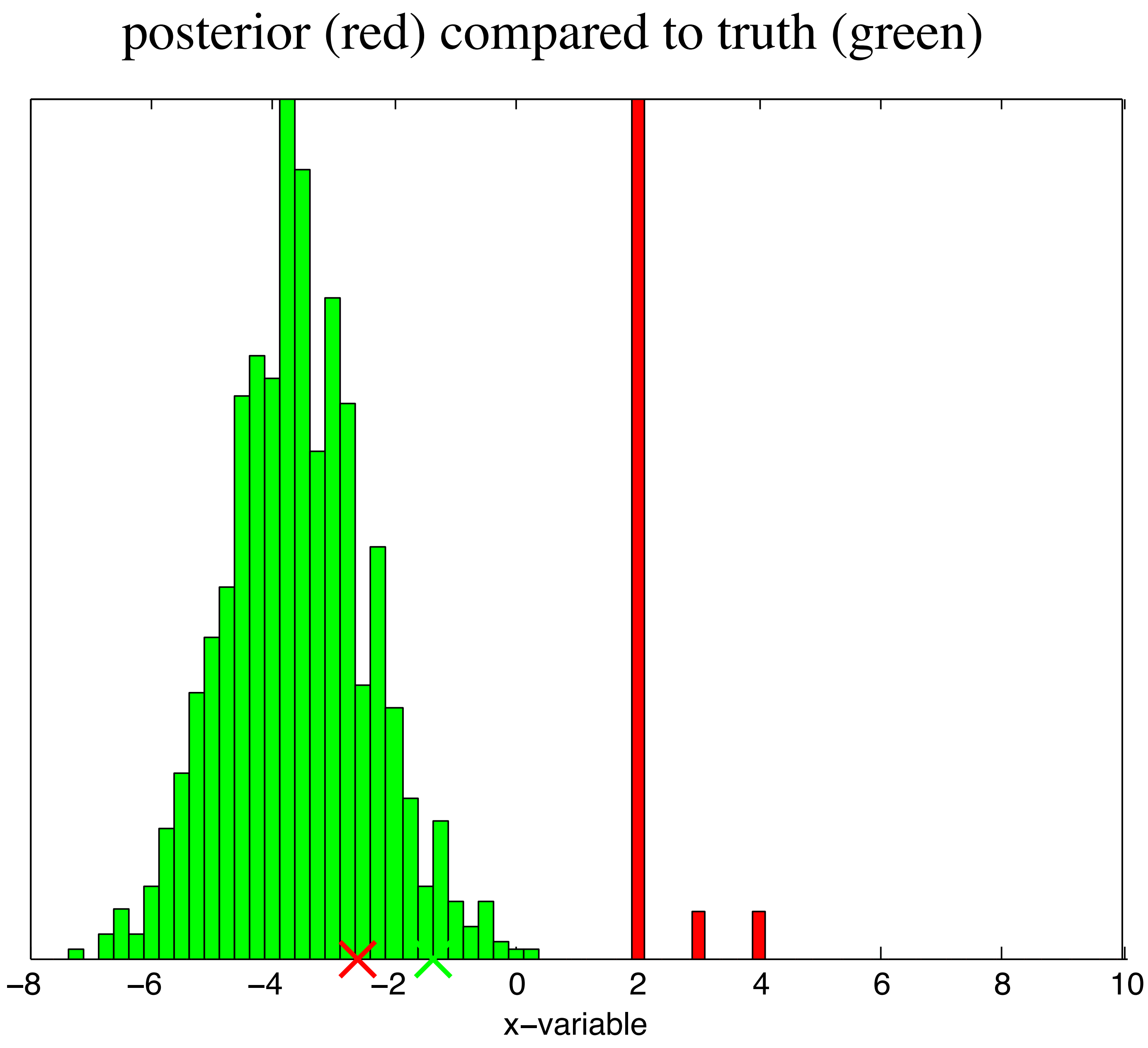
Standard Particle Filter fails - filter degeneracy




- observation


- truth

Standard Particle Filter fails - filter degeneracy



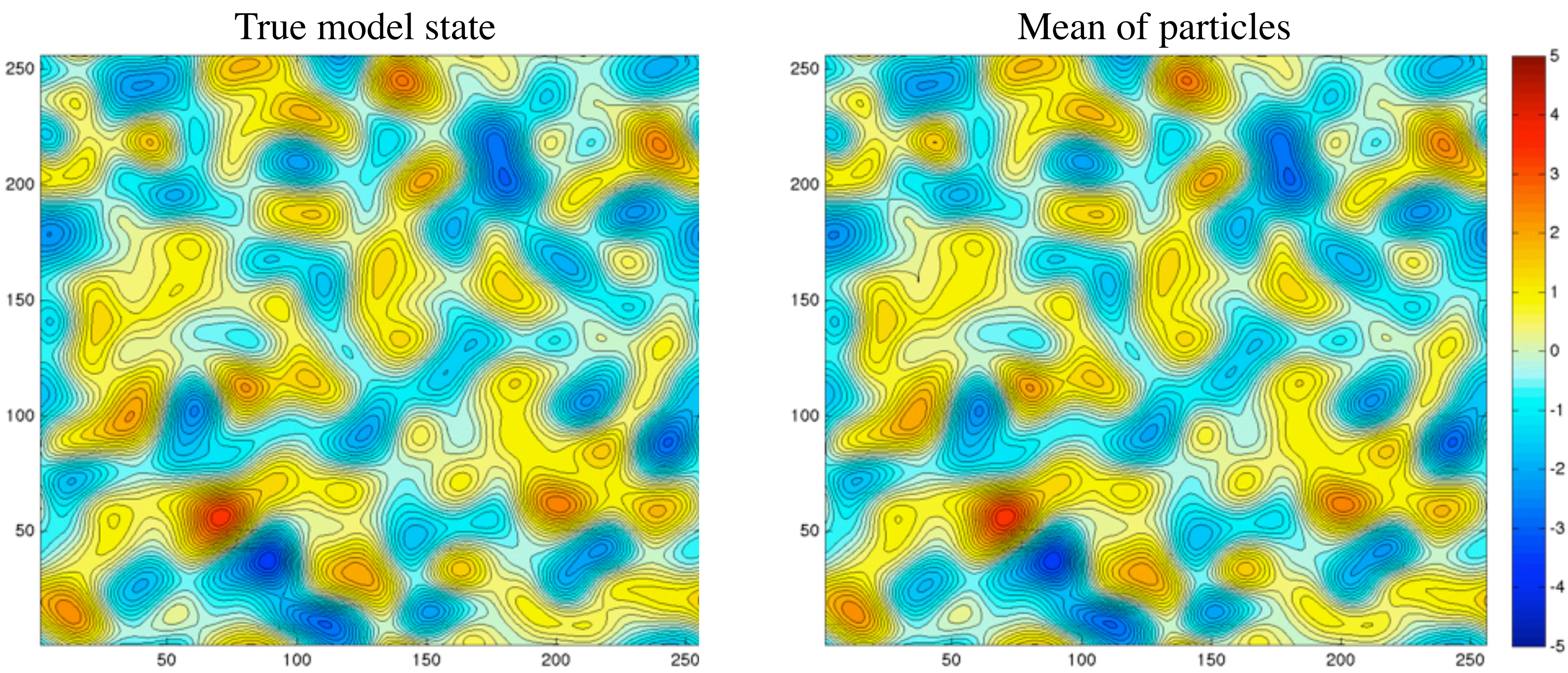
High dimensional systems

- Barotropic vorticity:

$$\frac{Dq}{Dt} = \frac{\partial q}{\partial t} + u.\nabla q = 0$$

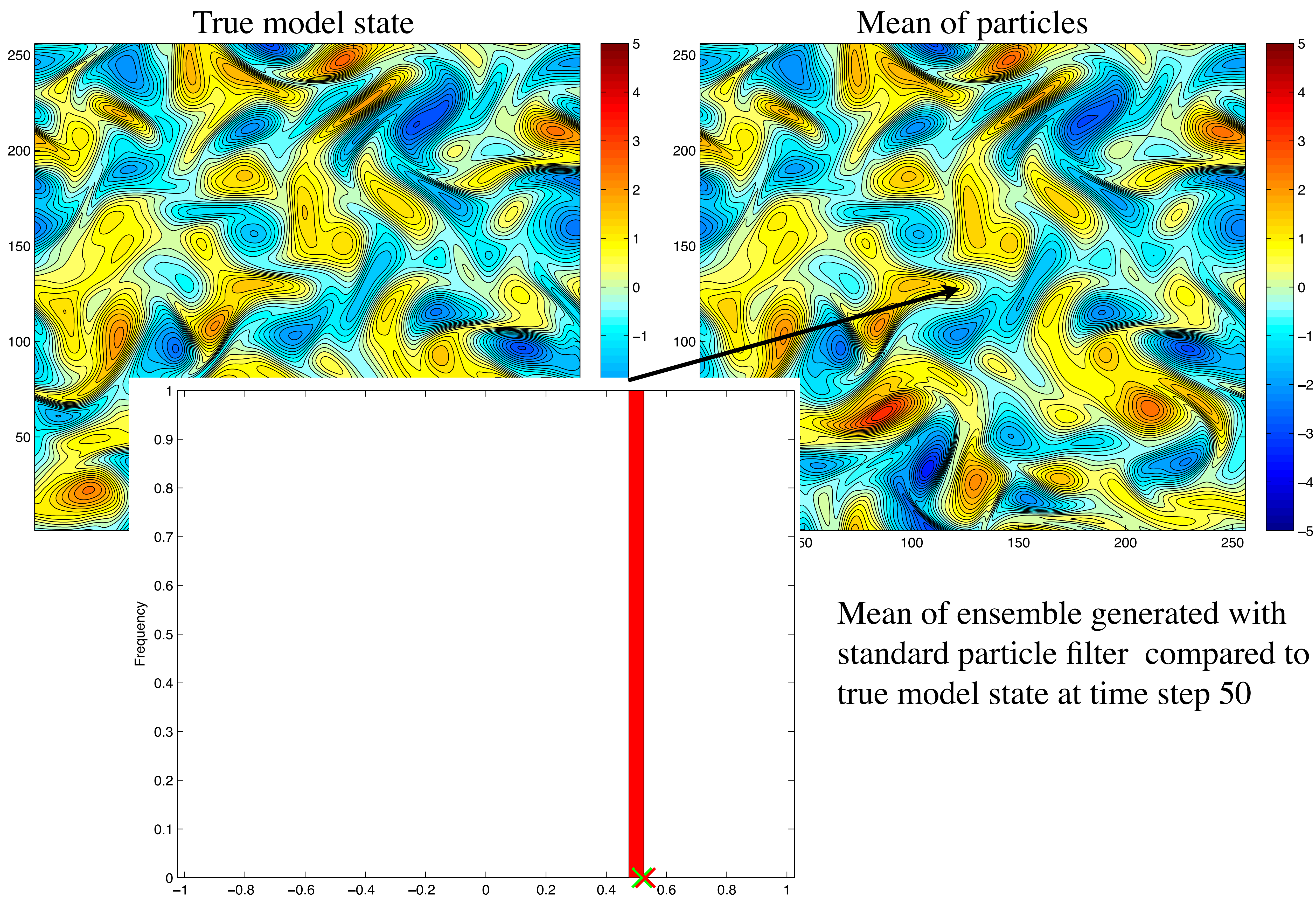
- 256 by 256 grid - 65,536 variables
- Doubly periodic boundary conditions
- Semi-langrangian time stepping scheme
- Twin experiments
- Observations every 50 time steps - decorrelation time of 42
- 32 particles

Mean of SIR filter fails to capture truth



- Every variable is observed
- 1200 time steps

SIR Filter - filter degeneracy is evident



Mean of ensemble generated with standard particle filter compared to true model state at time step 50

A closer look at the weights

Assume particle 1 is at 0.1 standard deviations s of M independent observations. Then its weight will be:

$$\widehat{w}_{j+1}^{(1)} \propto \exp\left(-\frac{1}{2}\left|\Gamma^{-\frac{1}{2}}(y_{j+1} - h(v_{j+1}^{(1)}))\right|^2\right) = \exp(-0.005M)$$

Assume particle 2 is at 0.2 standard deviations s of M independent observations. Then its weight will be:

$$\widehat{w}_{j+1}^{(2)} \propto \exp\left(-\frac{1}{2}\left|\Gamma^{-\frac{1}{2}}(y_{j+1} - h(v_{j+1}^{(2)}))\right|^2\right) = \exp(-0.02M)$$

A closer look at the weights

The ratio of the weights is: $$\frac{\widehat{w}_{j+1}^{(2)}}{\widehat{w}_{j+1}^{(1)}} = \exp(-0.015M)$$

So for M=2 the ratio of the two weights is:

$$\frac{\widehat{w}_{j+1}^{(2)}}{\widehat{w}_{j+1}^{(1)}} = \exp(-0.03) \approx 0.1$$

But for M=1000 the ratio of the two weights is:

$$\frac{\widehat{w}_{j+1}^{(2)}}{\widehat{w}_{j+1}^{(1)}} = \exp(-15) \approx 3 \times 10^{-7}$$

Conclusion: the number of independent observations is responsible for the degeneracy in particle filters

Different particle filters

proposal distribution

$$v_{j+1}^{(n)} \sim \mathbb{Q}(v_{j+1}^{(n)} | v_j^{(n)}, Y_{j+1})$$

associated weight update

$$\widehat{w}_{j+1}^{(n)} = w_j^{(n)} \frac{\mathbb{P}(y_{j+1} | v_{j+1}^{(n)}) \mathbb{P}(v_{j+1}^{(n)} | v_j^{(n)})}{\mathbb{Q}(v_{j+1}^{(n)} | v_j^{(n)}, Y_{j+1})}$$

Proposal densities can be chosen to try and reduce this variance in the weights.

Optimal proposal density

proposal distribution

$$\mathbb{Q}(v_{j+1}|v_j^{(n)}, Y_{j+1}) \equiv \mathbb{P}(v_{j+1}|v_j^{(n)}, y_{j+1})$$

for: $$\mathbb{P}(v_{j+1}|v_j^{(n)}) \propto \exp\left(-\frac{1}{2}\left|\Sigma^{-\frac{1}{2}}(v_{j+1} - \Psi(v_j^{(n)}))\right|^2\right)$$

$$\mathbb{P}(y_{j+1}|v_{j+1}^{(n)}) \propto \exp\left(-\frac{1}{2}\left|\Gamma^{-\frac{1}{2}}(y_{j+1} - h(v_{j+1}^{(n)}))\right|^2\right)$$

this equates to (assuming a linear observation operator):

$$v_{j+1}^{(n)} = \Psi(v_j^{(n)}) + \Sigma H^T \left(H\Sigma H^T + \Gamma\right)^{-1}\Big(y_{j+1} - H\Psi(v_j^{(n)})\Big) + \zeta_j^{(n)}$$

$$\zeta_j^{(n)} \sim N(0, P), \quad P^{-1} = \Sigma^{-1} + H^T\Gamma^{-1}H$$

This is similar to Kalman gain matrix K but using model error rather than prior error

Optimal proposal density - weight update

$$\mathbb{Q}(v_{j+1}^{(n)}|v_j^{(n)}, Y_{j+1}) = \mathbb{P}(v_{j+1}^{(n)}|v_j^{(n)}, y_{j+1}) = \frac{\mathbb{P}(y_{j+1}|v_{j+1}^{(n)})\mathbb{P}(v_{j+1}^{(n)}|v_j^{(n)})}{\mathbb{P}(y_{j+1}|v_j^{(n)})}$$

Linked to the use of the Kalman gain, the choice of new model state is the maximum a-posterior of these two distributions

$$\widehat{w}_{j+1}^{(n)} = w_j^{(n)} \frac{\mathbb{P}(y_{j+1}|v_{j+1}^{(n)})\mathbb{P}(v_{j+1}^{(n)}|v_j^{(n)})}{\mathbb{Q}(v_{j+1}^{(n)}|v_j^{(n)}, Y_{j+1})}$$

$$= w_j^{(n)}\mathbb{P}(y_{j+1}|v_j^{(n)})$$

weight does not depend on new model state

associated weight update

$$\widehat{w}_{j+1}^{(n)} \propto w_j^{(n)}\mathbb{P}(y_{j+1}|v_j^{(n)})$$

Optimal proposal density - weight update

sample of random error from stated distribution has no effect on the weight

$$v_{j+1}^{(n)} = \Psi(v_j^{(n)}) + \Sigma H^T \left(H\Sigma H^T + \Gamma\right)^{-1} \left(y_{j+1} - H\Psi(v_j^{(n)})\right) + \zeta_j^{(n)}$$

$$\mathbb{P}(y_{j+1}|v_j^{(n)}) \propto \exp\left(-\frac{1}{2}\left|(H\Sigma H^T + \Gamma)^{-\frac{1}{2}}(y_{j+1} - H\Psi(v_j^{(n)}))\right|^2\right)$$

weight is the maximum weight it is possible
for a particle to achieve given its position $v_j^{(n)}$

The variance in the weights is therefore the variance in the maximum weight
it is possible for each particle to achieve

Optimal proposal density

1. Set $j = 0$ and $\mathbb{P}^N(v_0|Y_0) = \mathbb{P}(v_0)$

2. Draw $v_j^{(n)} \sim \mathbb{P}^N(v_j|Y_j)$ (resample)

3. Set $w_j^{(n)} = 1/N, \quad n = 1, \ldots, N$

4. Draw $\widehat{v}_{j+1}^{(n)} \sim \mathbb{P}(\widehat{v}_{j+1}|v_j^{(n)}, y_{j+1})$

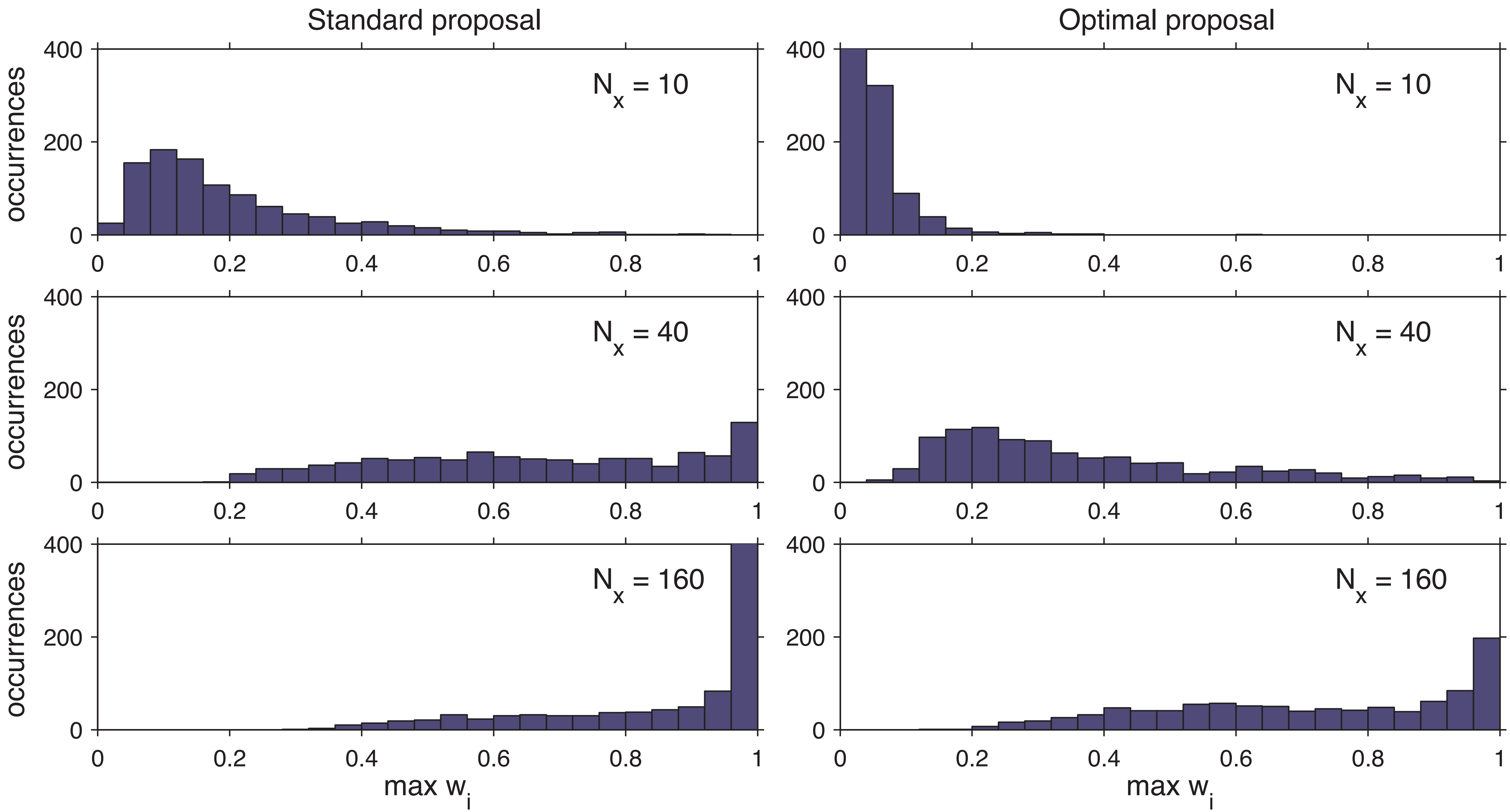
i.e. Set $v_{j+1}^{(n)} = \Psi(v_j^{(n)}) + \Sigma H^T \left(H\Sigma H^T + \Gamma\right)^{-1} \left(y_{j+1} - H\Psi(v_j^{(n)})\right) + \zeta_j^{(n)}$

5. Calculate $w_{j+1}^{(n)} = \mathbb{P}(y_{j+1}|v_j^{(n)}) / \left(\sum_{n=1}^{N} \mathbb{P}(y_{j+1}|v_j^{(n)})\right)$

$$\text{where } \mathbb{P}(y_{j+1}|v_j^{(n)}) \propto \exp\left(-\frac{1}{2}\left|(H\Sigma H^T + \Gamma)^{-\frac{1}{2}}(y_{j+1} - H\Psi(v_j^{(n)}))\right|^2\right)$$

6. $j + 1 \mapsto j$ and return to step 2

Optimal proposal density - variance of weights



('Particle filters, the 'optimal' proposal and high dimensional systems' Snyder, 2012)

Optimal proposal improves on SIR filter, but filter degeneracy still occurs in high dimensional systems

Ensemble Kalman Filter as proposal

EnKF:

$$\text{prediction} \quad \widehat{v}_{j+1}^{(n)} = \Psi(v_j^{(n)}) + \xi_j^{(n)}, \quad \xi_j^{(n)} \sim N(0, \Sigma)$$

$$\text{analysis} \quad v_{j+1}^{(n)} = (I - K_{j+1}H)\widehat{v}_{j+1}^{(n)} + K_{j+1}y_{j+1}^{(n)}$$

$$\text{observation perturbation} \quad y_{j+1}^{(n)} = y_{j+1} + \eta_{j+1}^{(n)}, \quad \eta_{j+1}^{(n)} \sim N(0, \Gamma)$$

$$v_{j+1}^{(n)} = \underbrace{\Psi(v_j^{(n)}) + K_{j+1}(y_{j+1} - H(\Psi(v_j^{(n)})))}_{\text{Deterministic} = \mu_{j+1}^{(n)}} + \underbrace{(I - K_{j+1}H)\xi_j^{(n)} + K_{j+1}\eta_{j+1}^{(n)}}_{\text{Stochastic}}$$

$$Q = (I - K_{j+1}H)\Sigma(I - K_{j+1}H)^T + K_{j+1}\Gamma K_{j+1}$$

proposal distribution

$$\boxed{\mathbb{Q}(v_{j+1}^{(n)} | v_j^{(n)}, Y_{j+1}) \propto \exp\left(-\frac{1}{2}\left|Q^{-\frac{1}{2}}(v_{j+1}^{(n)} - \mu_{j+1}^{(n)})\right|^2\right)}$$

Ensemble Kalman Filter as proposal

Kalman gain: $K_{j+1} = \widehat{C}_{j+1} H^T \left(H \widehat{C}_{j+1} H^T + \Gamma \right)^{-1}$

Approximated using the ensemble

$$\widehat{C}_{j+1} = \frac{1}{N-1} \sum_{n=1}^{N} (\widehat{v}_{j+1}^{(n)} - \widehat{m}_{j+1})(\widehat{v}_{j+1}^{(n)} - \widehat{m}_{j+1})^T$$

$$\widehat{m}_{j+1} = \frac{1}{N} \sum_{n=1}^{N} \widehat{v}_{j+1}^{(n)}$$

So the proposal used to update each particle actually depends on all the other particle positions. However, for an infinitely large ensemble of particles, the Kalman gain will depend only on the system of model equations via the prediction step.

Ensemble Kalman Filter as proposal - weight update

No longer have a simplification in the weights (as in the optimal proposal density) and so need to directly calculate the different constituent parts of the weight for each particle

$$\widehat{w}_{j+1}^{(n)} = w_j^{(n)} \frac{\mathbb{P}(y_{j+1}|v_{j+1}^{(n)})\mathbb{P}(v_{j+1}^{(n)}|v_j^{(n)})}{\mathbb{Q}(v_{j+1}^{(n)}|v_j^{(n)}, Y_{j+1})}$$

$$\mathbb{Q}(v_{j+1}^{(n)}|v_j^{(n)}, Y_{j+1}) \propto \exp\left(-\frac{1}{2}\left|Q^{-\frac{1}{2}}(v_{j+1}^{(n)} - \mu_{j+1}^{(n)})\right|^2\right)$$

$$\mathbb{P}(v_{j+1}^{(n)}|v_j^{(n)}) \propto \exp\left(-\frac{1}{2}\left|\Sigma^{-\frac{1}{2}}(v_{j+1}^{(n)} - \Psi(v_j^{(n)}))\right|^2\right)$$

$$\mathbb{P}(y_{j+1}|v_{j+1}^{(n)}) \propto \exp\left(-\frac{1}{2}\left|\Gamma^{-\frac{1}{2}}(y_{j+1} - h(v_{j+1}^{(n)}))\right|^2\right)$$

$$\widehat{w}_{j+1}^{(n)} \propto w_j^{(n)} \exp\left(-\frac{1}{2}\left|\Gamma^{-\frac{1}{2}}(y_{j+1} - h(v_{j+1}^{(n)}))\right|^2 - \frac{1}{2}\left|\Sigma^{-\frac{1}{2}}(v_{j+1}^{(n)} - \Psi(v_j^{(n)}))\right|^2 + \frac{1}{2}\left|Q^{-\frac{1}{2}}(v_{j+1}^{(n)} - \mu_{j+1}^{(n)})\right|^2\right)$$

Ensemble Kalman Filter as proposal

1. Set $j = 0$ and $\mathbb{P}^N(v_0|Y_0) = \mathbb{P}(v_0)$

2. Draw $v_j^{(n)} \sim \mathbb{P}^N(v_j|Y_j)$ (resample)

3. Set $w_j^{(n)} = 1/N, \quad n = 1, \ldots, N$

Weighted EnKF now ensures the true posterior is represented but unfortunately filter degeneracy still occurs.

4. Set

$$v_{j+1}^{(n)} = \Psi(v_j^{(n)}) + K_{j+1}(y_{j+1} - H(\Psi(v_j^{(n)}))) + (I - K_{j+1}H)\xi_j^{(n)} + K_{j+1}\eta_{j+1}^{(n)}$$

5. Calculate $w_{j+1}^{(n)} = \widehat{w}_{j+1}^{(n)} / \left(\sum_{n=1}^{N} \widehat{w}_{j+1}^{(n)} \right)$

where $$\widehat{w}_{j+1}^{(n)} \propto \exp\left(-\frac{1}{2} \left| \Gamma^{-\frac{1}{2}}(y_{j+1} - h(v_{j+1}^{(n)})) \right|^2 - \frac{1}{2} \left| \Sigma^{-\frac{1}{2}}(v_{j+1}^{(n)} - \Psi(v_j^{(n)})) \right|^2 + \frac{1}{2} \left| Q^{-\frac{1}{2}}(v_{j+1}^{(n)} - \mu_{j+1}^{(n)}) \right|^2 \right)$$

6. $j + 1 \mapsto j$ and return to step 2

Summary

SIR particle filter

- The proposal density is the model transition density and so the model equations are used to propagate each particle forward in time
- The weight is calculated based on the likelihood, so the distance of each particle to the observation

Optimal proposal density

- The new model state is sampled from a proposal which is the density of possible new model states given the old model state and the observation
- The weight is then calculated based on the maximum weight a particle could achieve given the old model state and the observation

Weighted ensemble Kalman filter

- The new model state is generated using the Ensemble Kalman equations
- The weight then must be calculated directly through all three constituent parts

Although the optimal proposal density improves on the SIR filter, all three schemes still suffer from filter degeneracy. This is because of the difficulty in sampling from the high probability region of the posterior in high dimensions with large numbers of independent observations.

References

SIR particle filter

Doucet AN, Freitas D, Gordon N. 2001. *Sequential Monte-Carlo Methods in Practice*. Springer-Verlag

Robert CP, Cassela G. 2004. *Monte-Carlo Statistical Methods*. Springer-Verlag

Optimal proposal density

Doucet AS, Godsill S, Andrieu C. 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing* **10**: 197-208

Snyder C, Bengtsson T, Bickel P, Anderson J. 2008. Obstacles to high-dimensional particle filtering. *Monthly Weather Review* **136**: 4629-4640

Weighted ensemble Kalman filter

Papadakis N, Memin E, Cuzol A, Gengembre N. 2010. Data assimilation with the weighted ensemble Kalman filter. *Tellus A* **62**: 673-697

Van Leeuwen PJ. 2009. Particle filtering in geophysical systems. *Monthly Weather Review* **137**: 4089-4114