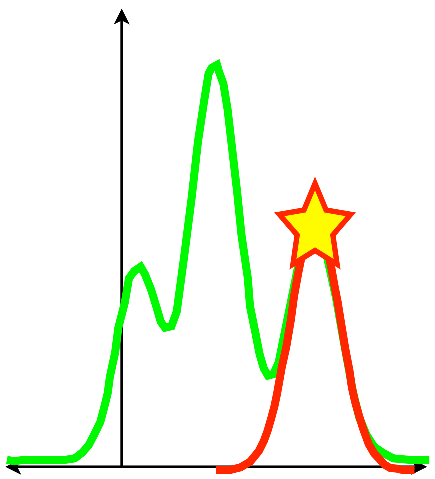
Nonlinear data assimilation
Part II: High dimensional particle filters

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

Particle filters

Variational methods:



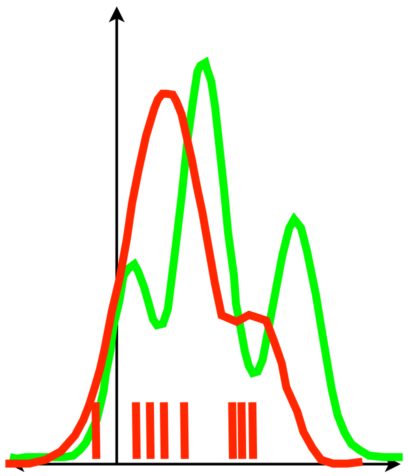
3D-Var

$$\frac{1}{2}\left|C_0^{-\frac{1}{2}}(v_0 - m_o)\right|^2 + \frac{1}{2}\left|\Gamma^{-\frac{1}{2}}(y - h(v_0))\right|^2$$

4D-Var

$$\frac{1}{2}\left|C_0^{-\frac{1}{2}}(v_0 - m_o)\right|^2 + \sum_{j=0}^{J-1}\frac{1}{2}\left|\Gamma^{-\frac{1}{2}}(y_{j+1} - h(v_{j+1}))\right|^2$$

Ensemble Kalman Filter:



prediction

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

analysis

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

observation perturbation

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

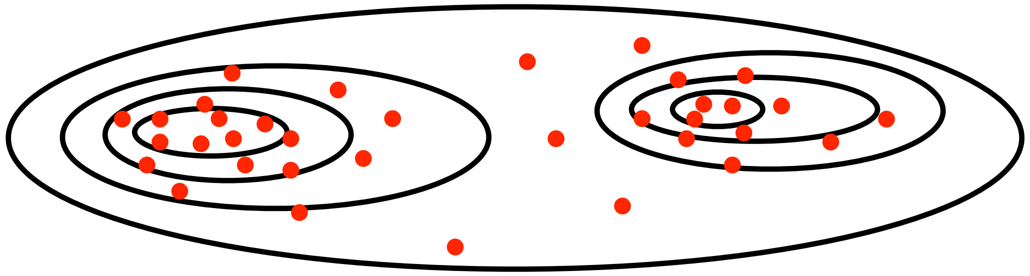
Particle filters

Markov Chain Monte Carlo:

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

Propose a new sample

$$\widehat{v}^{(k)} \sim q(v^{(k-1)}, \widehat{v}^{(k)})$$



Acceptance criteria

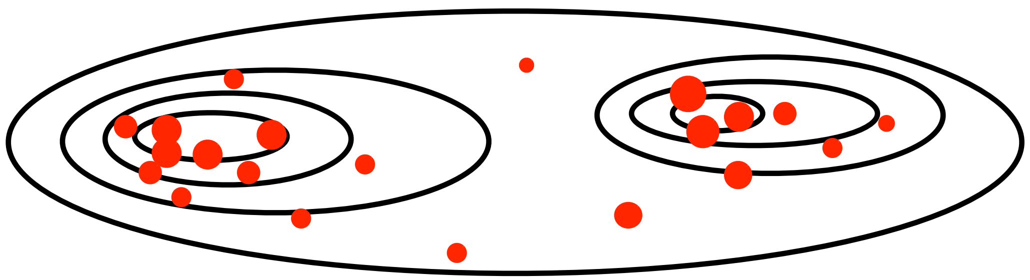
$$a(v^{(k-1)}, \widehat{v}^{(k)}) = 1 \wedge \frac{\rho(v^{(k-1)})q(\widehat{v}^{(k)}, v^{(k-1)})}{\rho\widehat{v}^{(k)})q(v^{(k-1)}, \widehat{v}^{(k)})}$$

Particle filtering:

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

Generate a new sample

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



Weight the sample

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

Particle filters - general formulation

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

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

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

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

SIR, optimal and EnKF as proposal densities

SIR filter:

proposal distribution

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

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

associated weight update

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

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

- Suffers from filter degeneracy even in very low dimensional systems

SIR, optimal and EnKF as proposal densities

Optimal proposal density:

proposal distribution

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

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

associated weight update

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

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

- Improves over SIR - but still fails in high dimensional systems with large numbers of independent observations

SIR, optimal and EnKF as proposal densities

Weighted EnKF:

proposal distribution

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

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

associated weight update

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

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

High dimensional particle filters

- As the dimension of the system and the number of independent observations increase, the variance in the weights also increases. All three schemes suffer from filter degeneracy and so are not applicable to high-dimensional systems.

- More complicated proposal densities are required to reduce the variance in the weights

Particle filters have so far been discussed assuming observations have been available at every time step

Model:
$$v_{j+1} = \Psi(v_j) + \xi_j, \quad j \in \mathbb{N}$$
$$v_0 \sim N(m_0, C_0)$$
$$\xi_j \sim N(0, \Sigma)$$

Observation:
$$y_k = h(v_{lk}) + \eta_k, \quad k \in \mathbb{N}$$
$$\eta \sim N(0, \Gamma)$$

Assume now observations are available only every l time steps

Particle filters

How to determine the new positions and weights of particles at time kl given the weights and positions at time (k-1)l?

$$\{v_{kl}^{(n)}, w_{kl}^{(n)}\}_{n=1}^{N} \mapsto \{v_{(k+1)l}^{(n)}, w_{(k+1)l}^{(n)}\}_{n=1}^{N}$$

First option:

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

$$v_{j+1}^{(n)} \sim \mathbb{Q}(v_{j+1}^{(n)} | v_j^{(n)}, Y_{j+1}), \quad j = kl, \dots, (k+1)l$$

These proposal densities do not have to be the same!

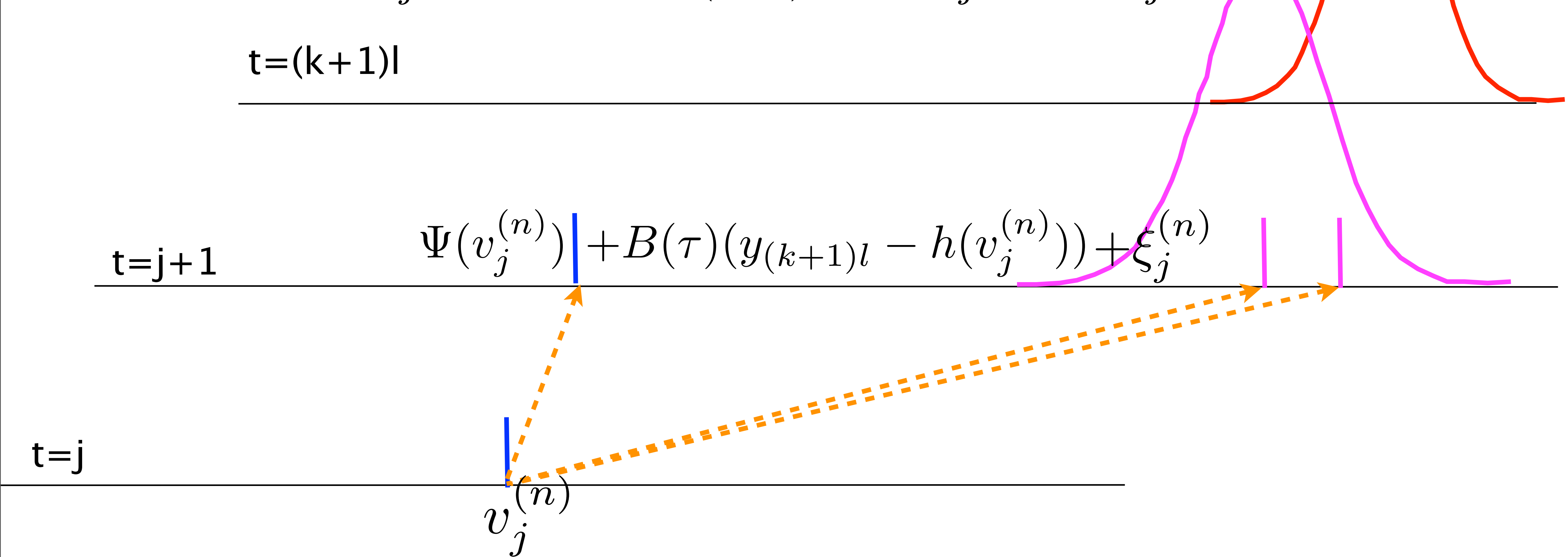
2) The particle is then weighted at observation time by

$$\widehat{w}_{(k+1)l}^{(n)} = \mathbb{P}(y_{k+1} | v_{(k+1)l}^{(n)}) \frac{\mathbb{P}(v_{(k+1)l}^{(n)} | v_{(k+1)l-1}^{(n)})}{\mathbb{Q}(v_{(k+1)l}^{(n)} | v_{(k+1)l-1}^{(n)}, Y_{k+1})}, \dots, \frac{\mathbb{P}(v_{kl+1}^{(n)} | v_{kl}^{(n)})}{\mathbb{Q}(v_{kl+1}^{(n)} | v_{kl}^{(n)}, Y_{k+1})}$$

Equivalent-weights particle filter

Aim: To reduce the variance in the weights by ensuring that all particles are close to the observation at analysis time

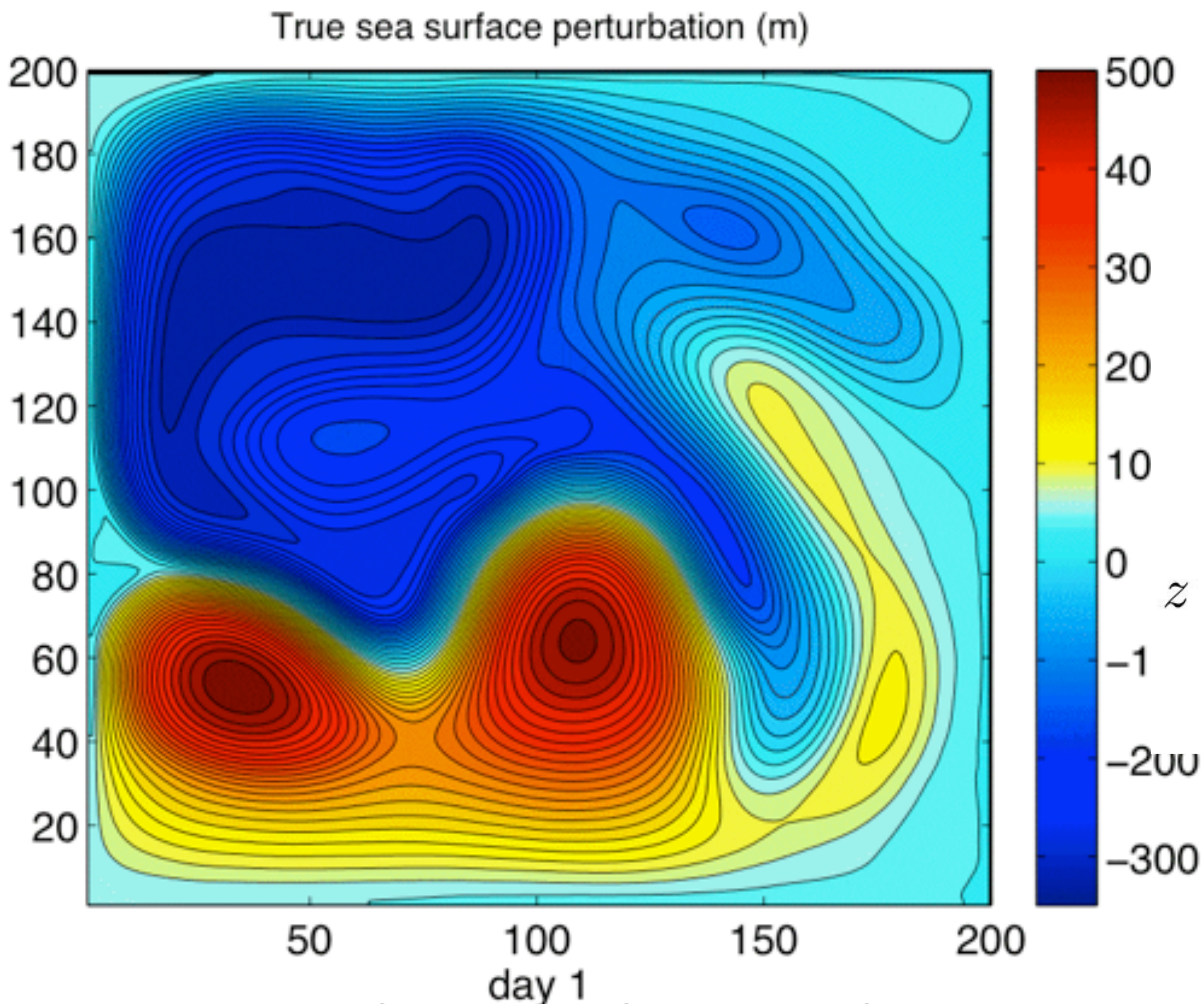
$$v_{j+1} = \Psi(v_j^{(n)}) + B(\tau)(y_{(k+1)l} - h(v_j^{(n)})) + \xi_j^{(n)}$$

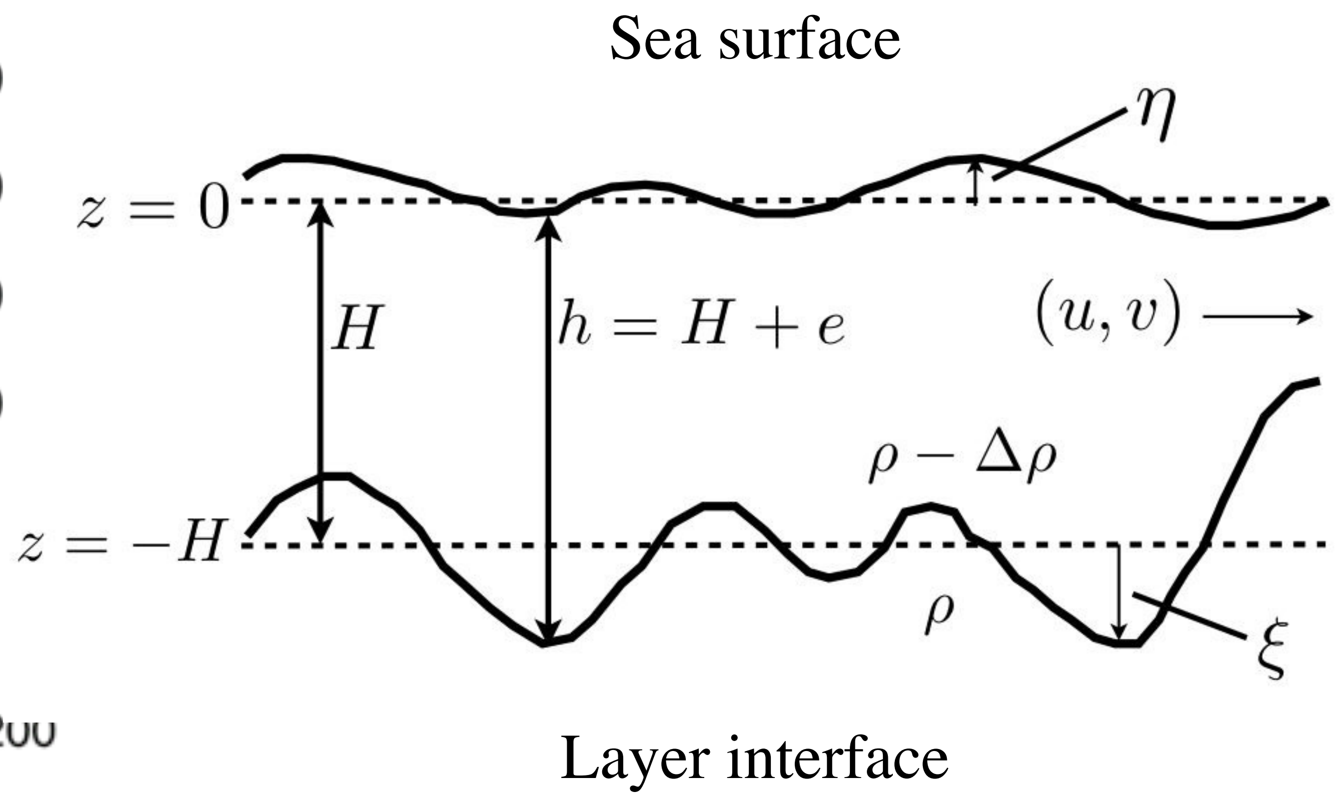


proposal distribution

$$\mathbb{Q}(v_{j+1}|v_j^{(n)}, Y_{k+1}) = N\left(\Psi(v_j^{(n)}) + B(\tau)(y_{(k+1)l} - h(v_j^{(n)})), \Sigma\right)$$

Primitive equation model - what should the relaxation be?





$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv + g'\frac{\partial h}{\partial x} = \frac{\tau^{(x)}}{\rho h} - A\Delta u + d\beta_u$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu + g'\frac{\partial h}{\partial y} = -A\Delta v + d\beta_y$$

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}(hu) + \frac{\partial}{\partial y}(hv) = d\beta_h,$$

Primitive equation model - what should the relaxation be?

If we have observations only of sea surface height, then how will it effect velocity if we only relax towards these observations?

Relaxation term:

$$B(\tau) = bp(\tau)\Sigma H^T \Gamma^{-1}(y_{(k+1)l} - h(v_j^{(n)}))$$

$$\Sigma = U\Sigma_\psi U^T$$

$$U: \quad u = -\frac{\partial \psi}{\partial y}, \quad v = \frac{\partial \psi}{\partial x}, \quad h = \frac{f_0}{g'}\psi$$

Information from h is seen by u and v through the multiplication by Σ

Equivalent-weights particle filter

associated weight update at each intermediate time

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

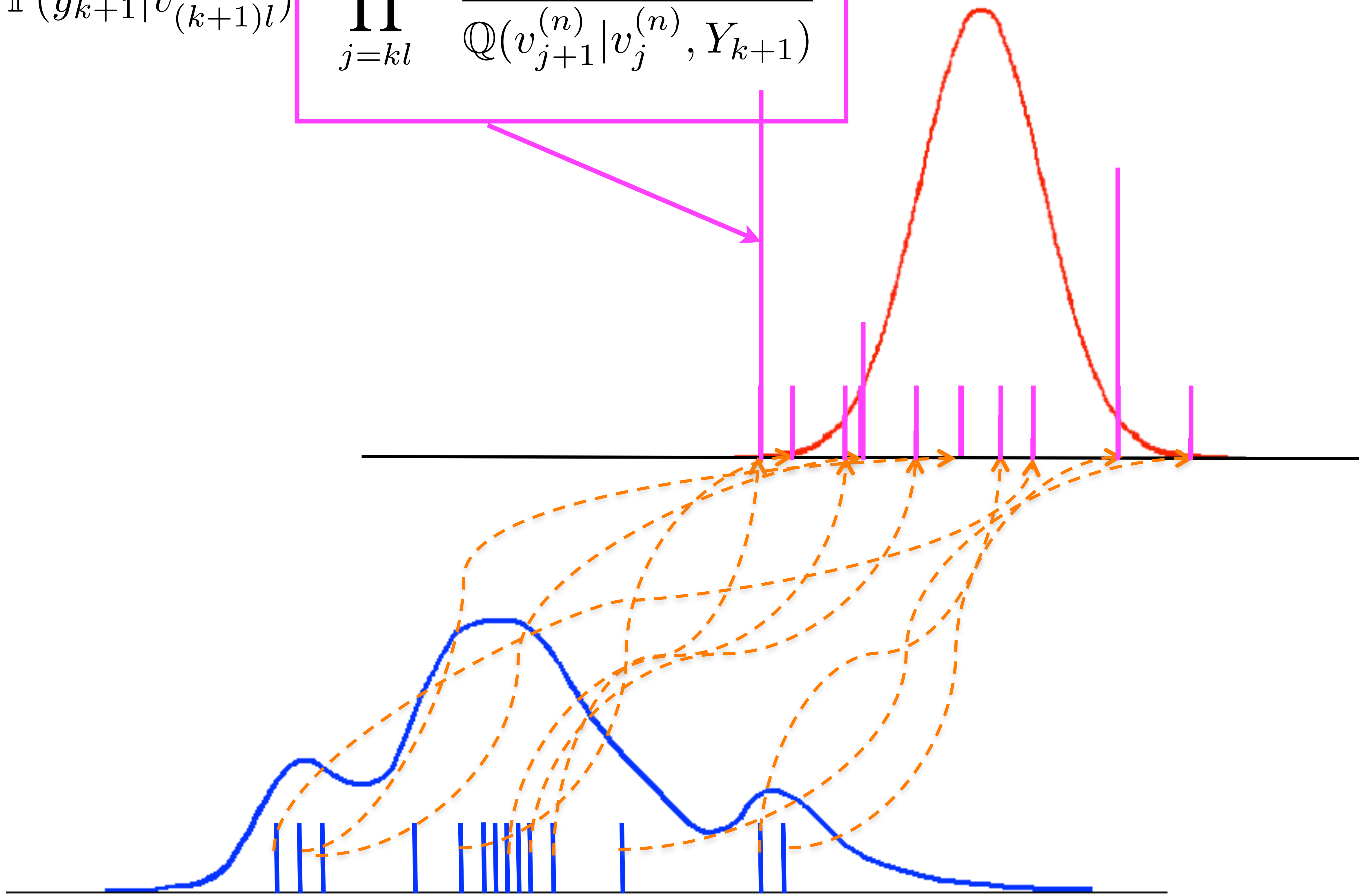
$$\mathbb{P}(v_{j+1}^{(n)}|v_j^{(n)}) \propto \exp\left(-\frac{1}{2}\left|\Sigma^{-\frac{1}{2}}\left(v_{j+1}^{(n)} - \Psi(v_j^{(n)})\right)\right|^2\right)$$

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

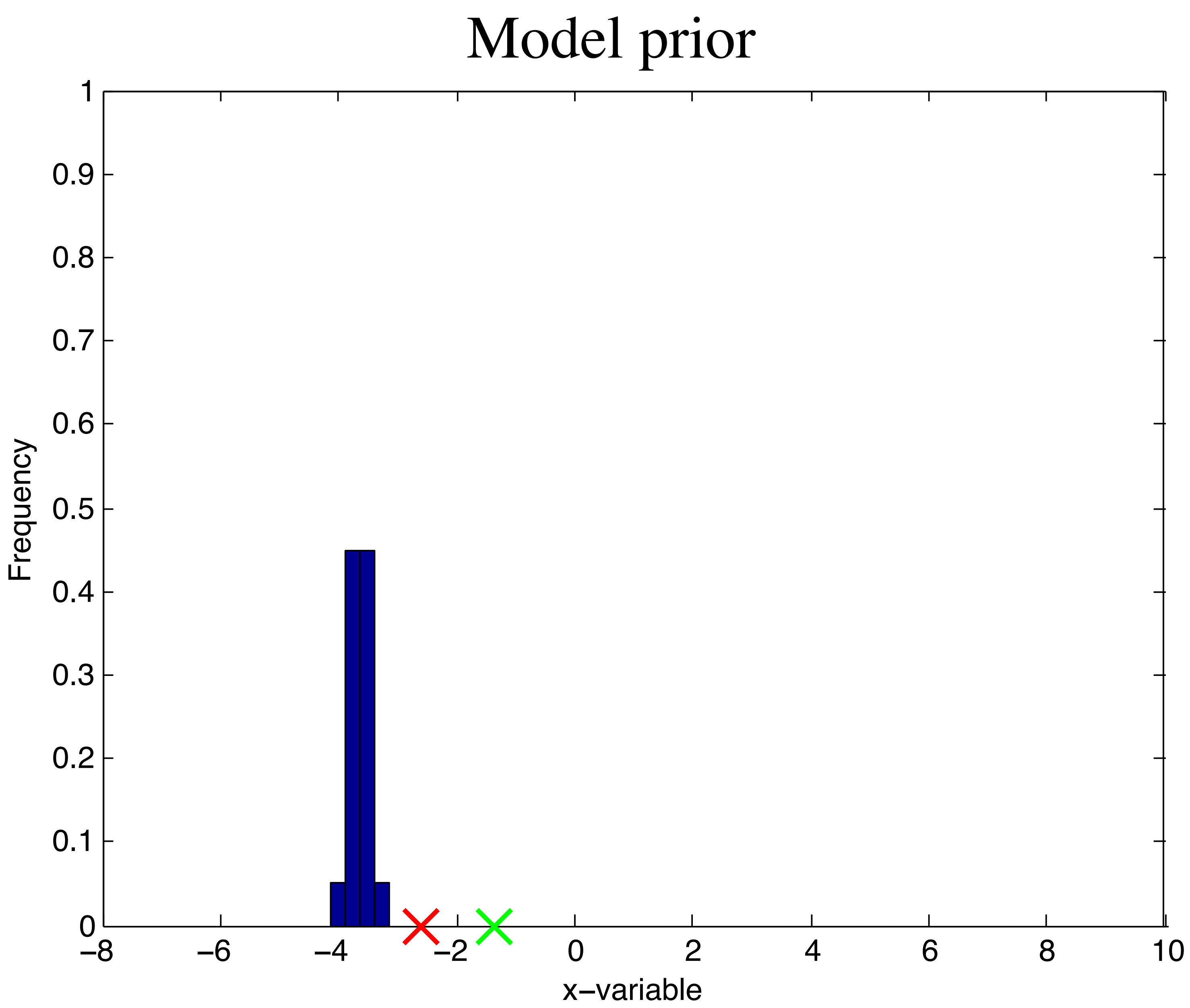
$$\exp\left(-\frac{1}{2}\left|\Sigma^{-\frac{1}{2}}\left(v_{j+1}^{(n)} - \left[\Psi(v_j^{(n)}) + B(\tau)(y_{(k+1)l} - h(v_j^{(n)})\right]\right)\right|^2\right)$$

Filter degeneracy of relaxation proposal density

$$\widehat{w}^{(n)}_{(k+1)l} = \mathbb{P}(y_{k+1}|v^{(n)}_{(k+1)l}) \boxed{\prod_{j=kl}^{(k+1)l-1} \frac{\mathbb{P}(v^{(n)}_{j+1}|v^{(n)}_{j})}{\mathbb{Q}(v^{(n)}_{j+1}|v^{(n)}_{j}, Y_{k+1})}}$$

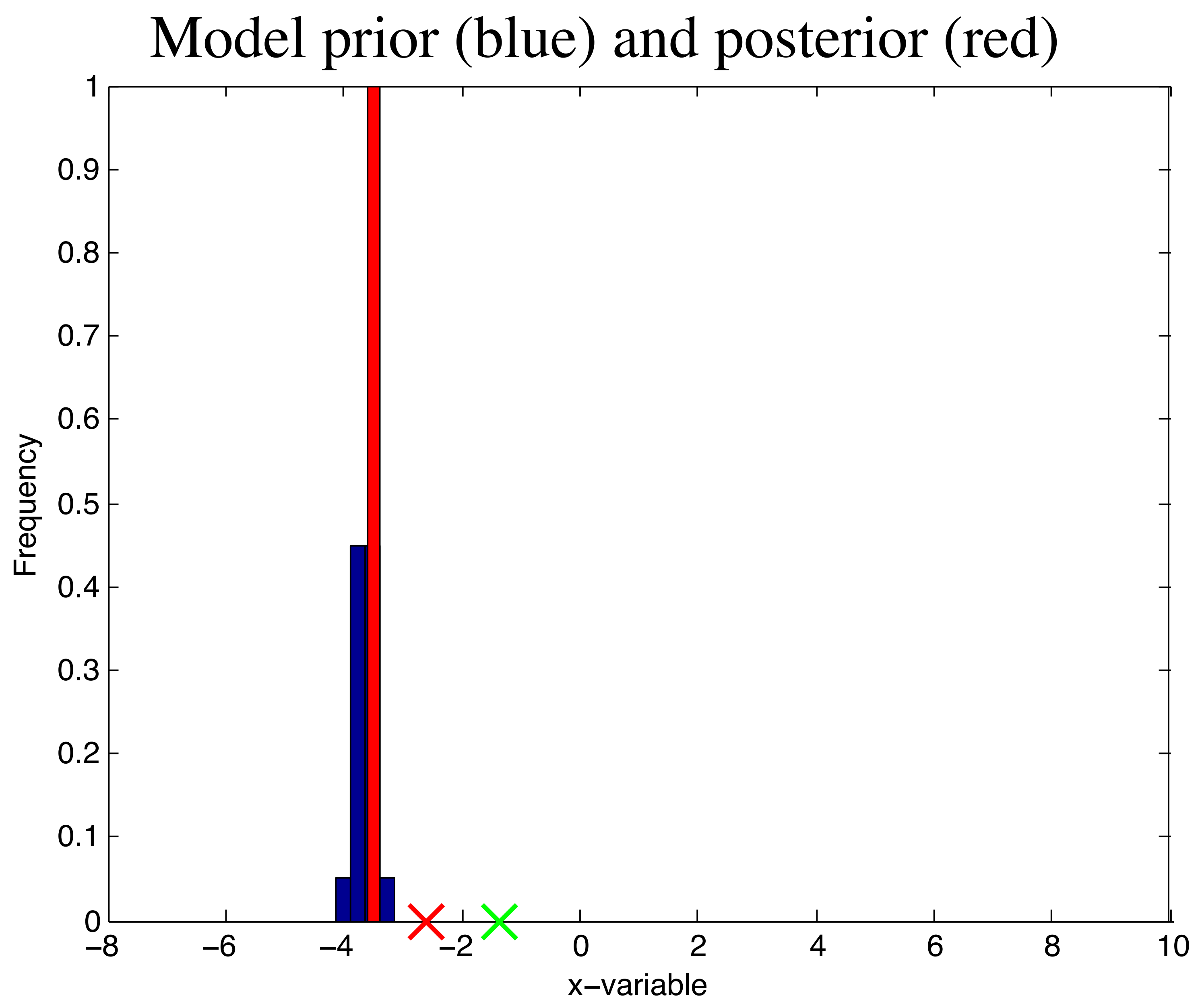


Lorenz 63 - filter degenerate if only relaxation proposal density is used



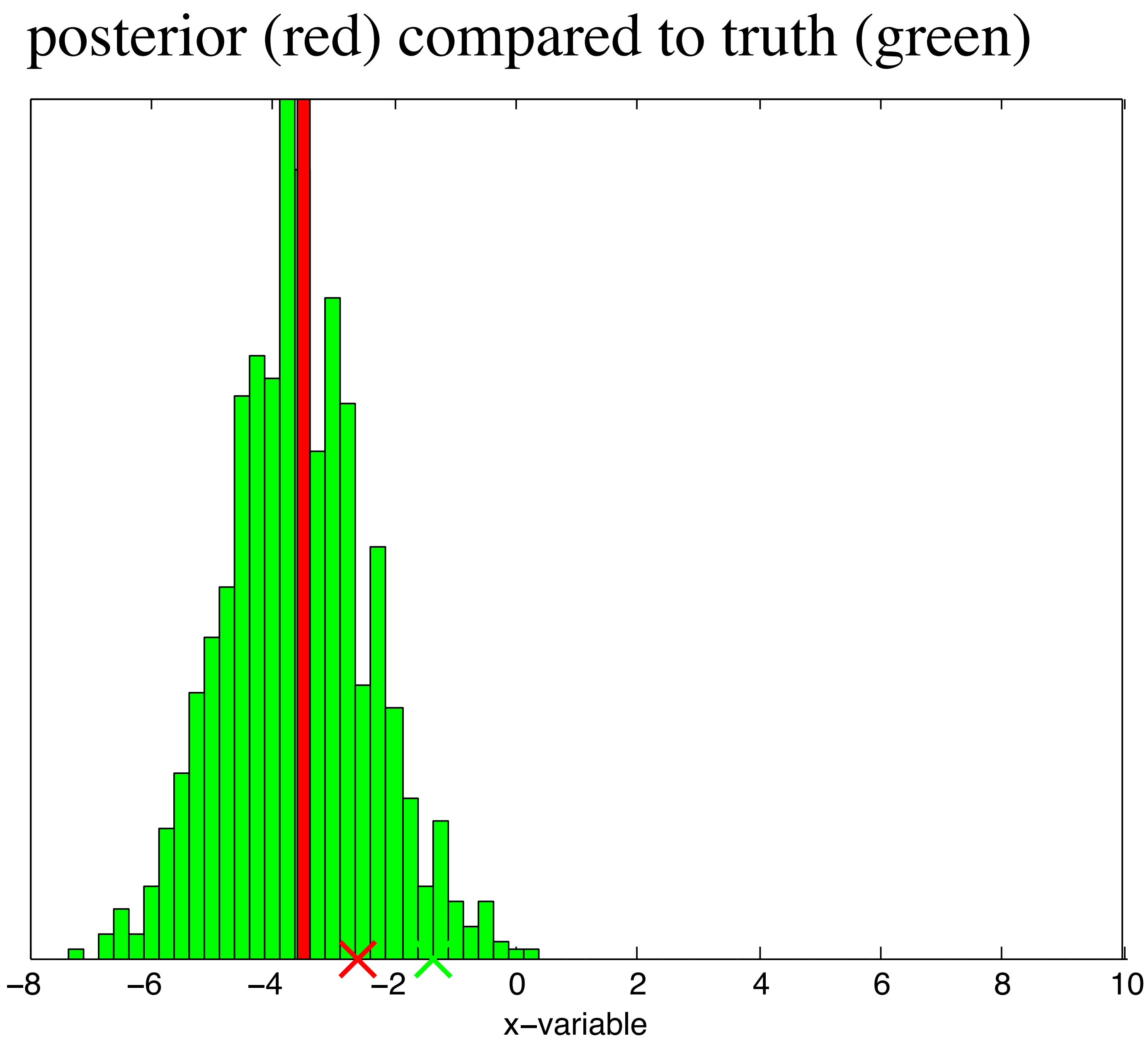
- observation
- truth

Lorenz 63 - filter degenerate if only relaxation proposal density is used



- observation
- truth

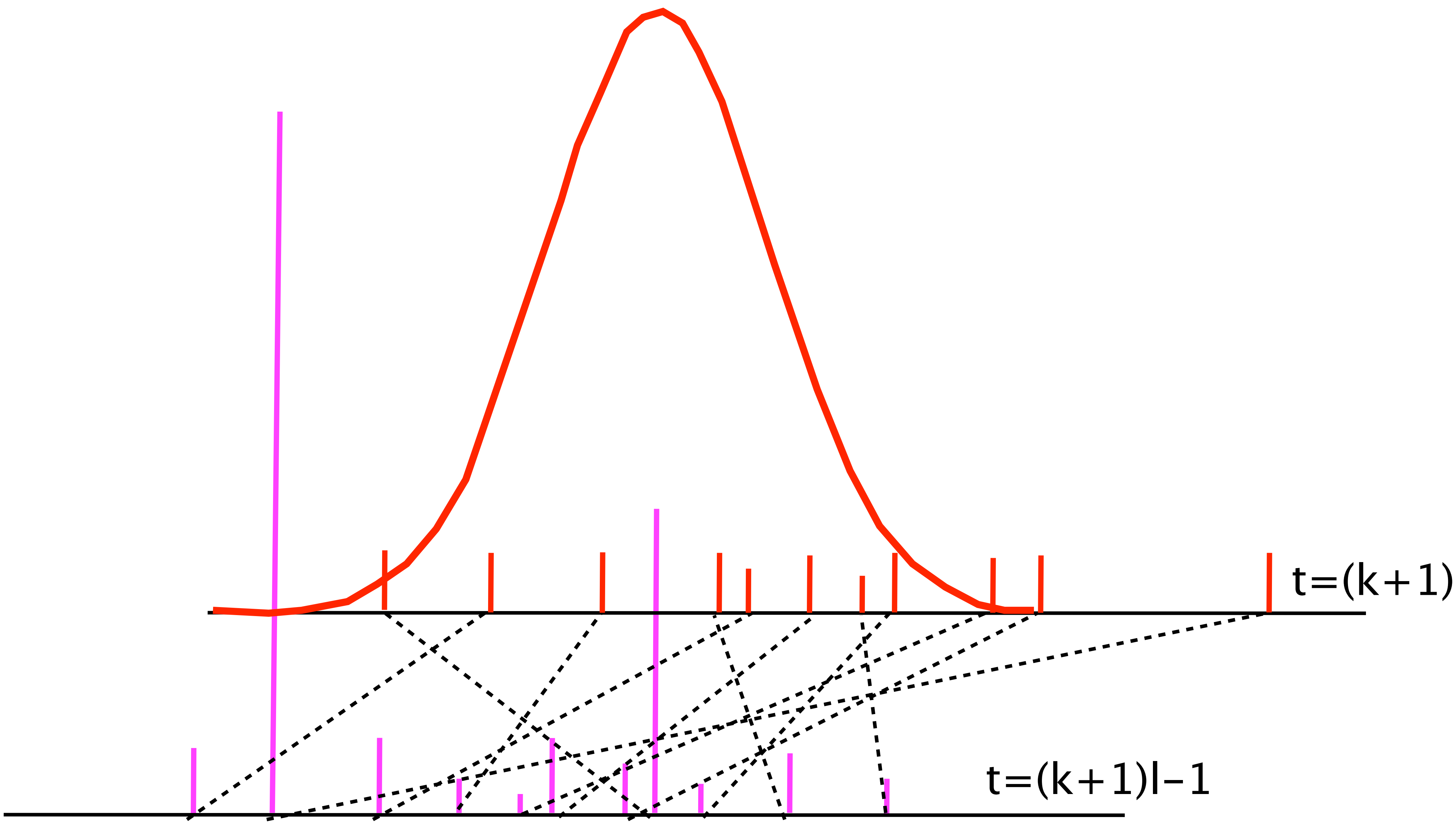
Lorenz 63 - filter degenerate if only relaxation proposal density is used



Filter degeneracy of relaxation proposal density

Particle filters: Choose proposal density -> Calculate weight

Final proposal density: Choose weight -> Calculate proposal density



What should the weight be?

$$\widehat{w}_{(k+1)l}^{(n)} = \frac{\mathbb{P}(y_{k+1}|v_{(k+1)l}^{(n)})\mathbb{P}(v_{(k+1)l}^{(n)}|v_{(k+1)l-1}^{(n)})}{\mathbb{Q}(v_{(k+1)l}^{(n)}|v_{(k+1)l-1}^{(n)}, Y_{k+1})} \prod_{j=kl}^{(k+1)l-2} \widehat{w}_j^{(n)}$$

Optimal proposal density:

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

However, already leads to filter degeneracy in high dimensional systems without accounting for the additional weight from the relaxation proposal densities

Equivalent-weights proposal density:

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

$$\zeta_j^{(n)} \sim (1-\epsilon)\widetilde{U}_k(-\gamma_U, \gamma_U) + \epsilon N(0, \gamma_N^2 \Sigma)$$

mixture density

Equivalent-weights particle filter

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

Relaxation proposal density:

2. If $j+1 \neq kl$ then draw $v_{j+1}^{(n)} \sim \mathbb{Q}_{\mathrm{relax}}(v_{j+1}|v_j^{(n)}, y_{kl})$

$$\text{i.e. Set } \widehat{v}_{j+1}^{(n)} = \Psi(v_j^{(n)}) + B(\tau)\left(y_{kl} - h(v_j^{(n)})\right) + \xi_j^{(n)}$$

3. Calculate $w_{j+1}^{(n)} = \dfrac{\mathbb{P}(v_{j+1}^{(n)}|v_j^{(n)})}{\mathbb{Q}_{\mathrm{relax}}(v_{j+1}^{(n)}|v_j^{(n)}, y_{kl})}$

Equivalent-weights particle filter

Equivalent-weights proposal density:

4. If $j+1=kl$ then draw $v_{kl}^{(n)} \sim \mathbb{Q}_{\text{equiv}}(v_{kl}|v_{kl-1}^{(n)}, y_{kl})$

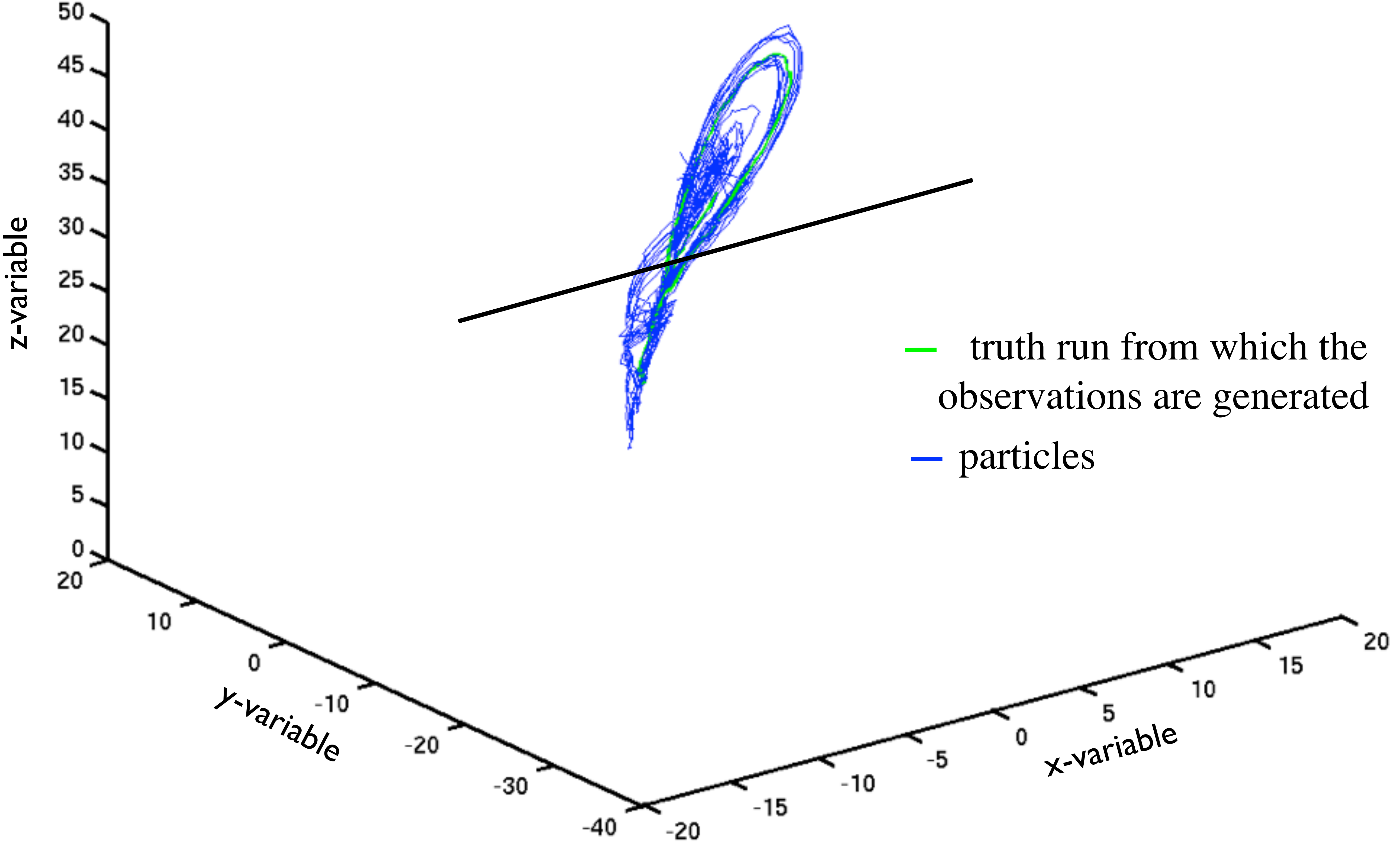
i.e. Set $v_{j+1}^{(n)} = \Psi(v_j^{(n)}) + \alpha^{(n)} \Sigma H^T \left(H\Sigma H^T + \Gamma\right)^{-1} \left(y_{j+1} - H\Psi(v_j^{(n)})\right) + \zeta_j^{(n)}$

5. Calculate $\widehat{w}_{kl}^{(n)} = \dfrac{\mathbb{P}(y_{kl}|v_{kl}^{(n)})\mathbb{P}(v_{kl}^{(n)}|v_{kl-1}^{(n)})}{\mathbb{Q}_{\text{equiv}}(v_{kl}^{(n)}|v_{kl-1}^{(n)}, y_{kl})} \prod_{j=(k-1)l}^{kl-1} w_j^{(n)}$

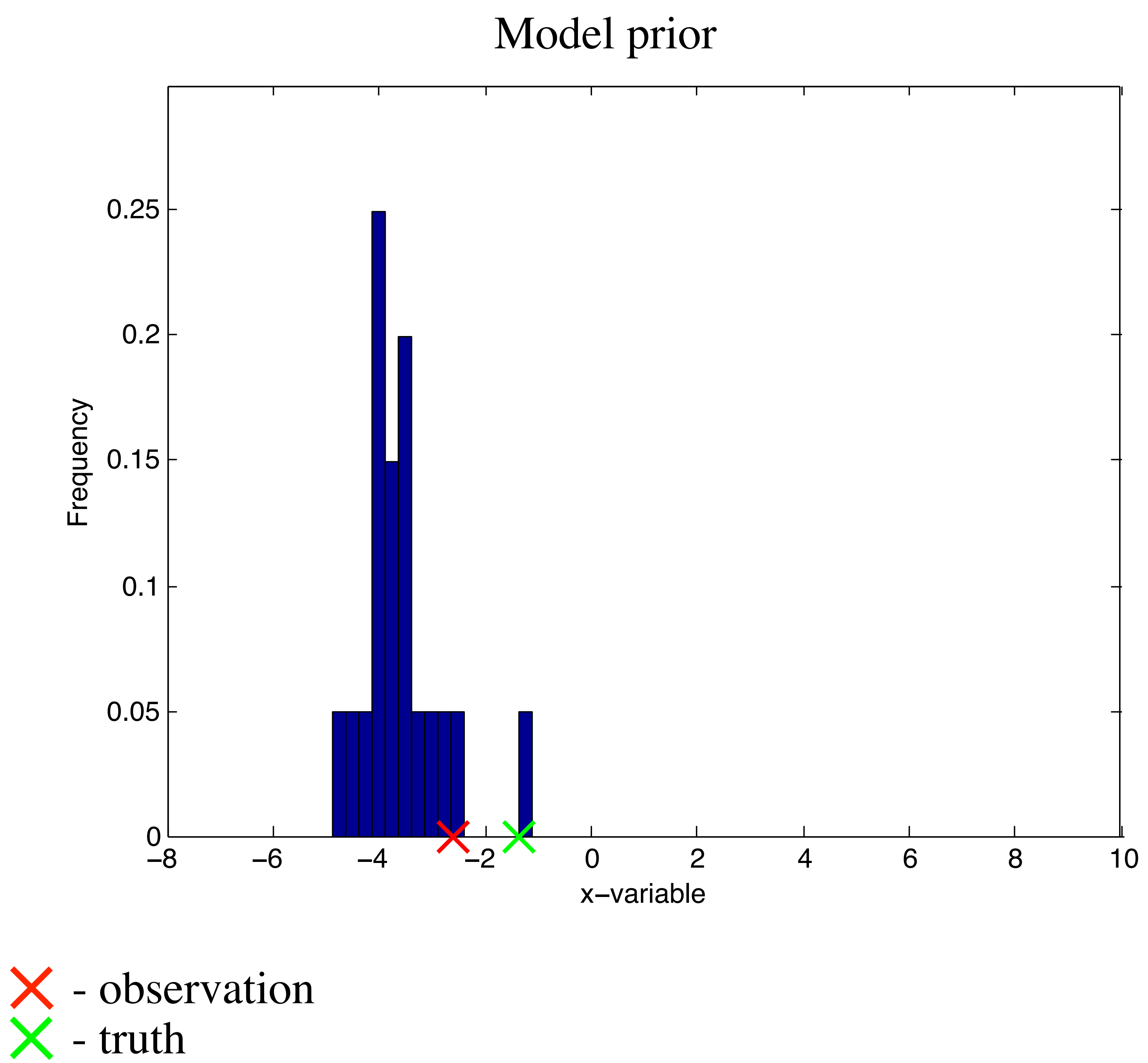
6. Calculate $w_{kl}^{(n)} = \mathbb{P}(y_{kl}|v_{kl}^{(n)}) / \left(\sum_{n=1}^{N} \mathbb{P}(y_{kl}|v_{kl}^{(n)})\right)$

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

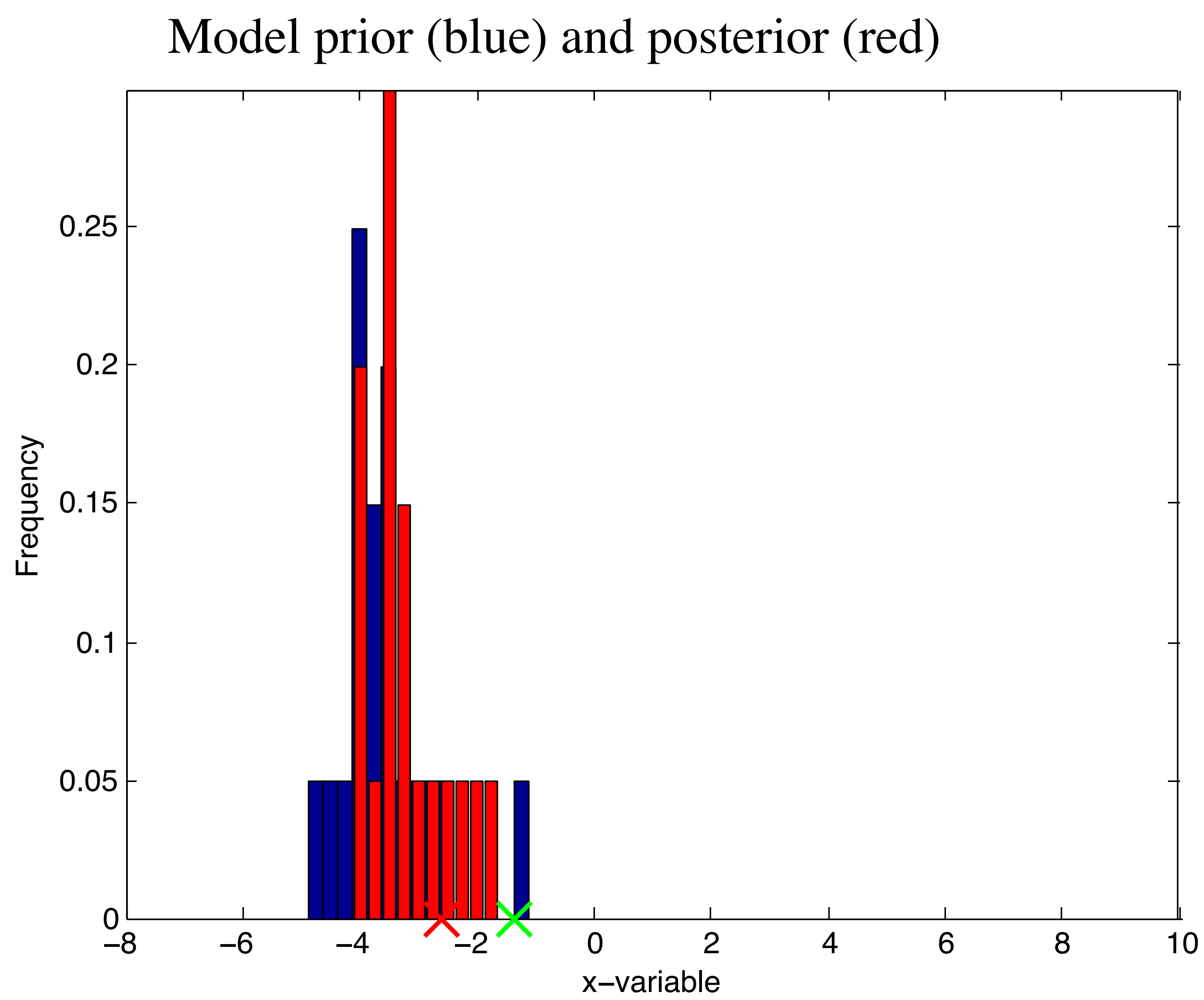
Lorenz 63 - equivalent-weights particle filter



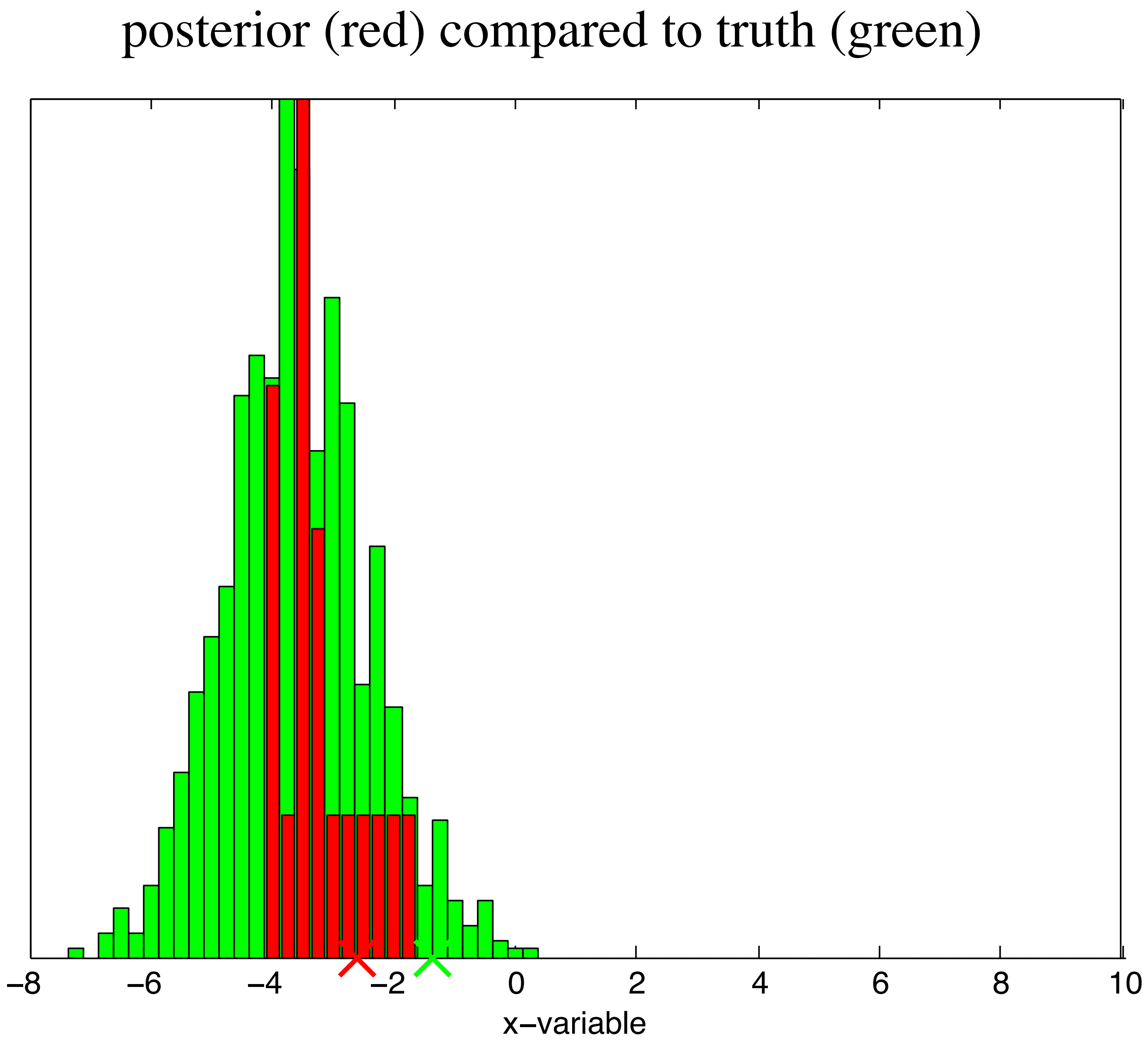
Equivalent weights particle filter - spread seen in model prior distribution



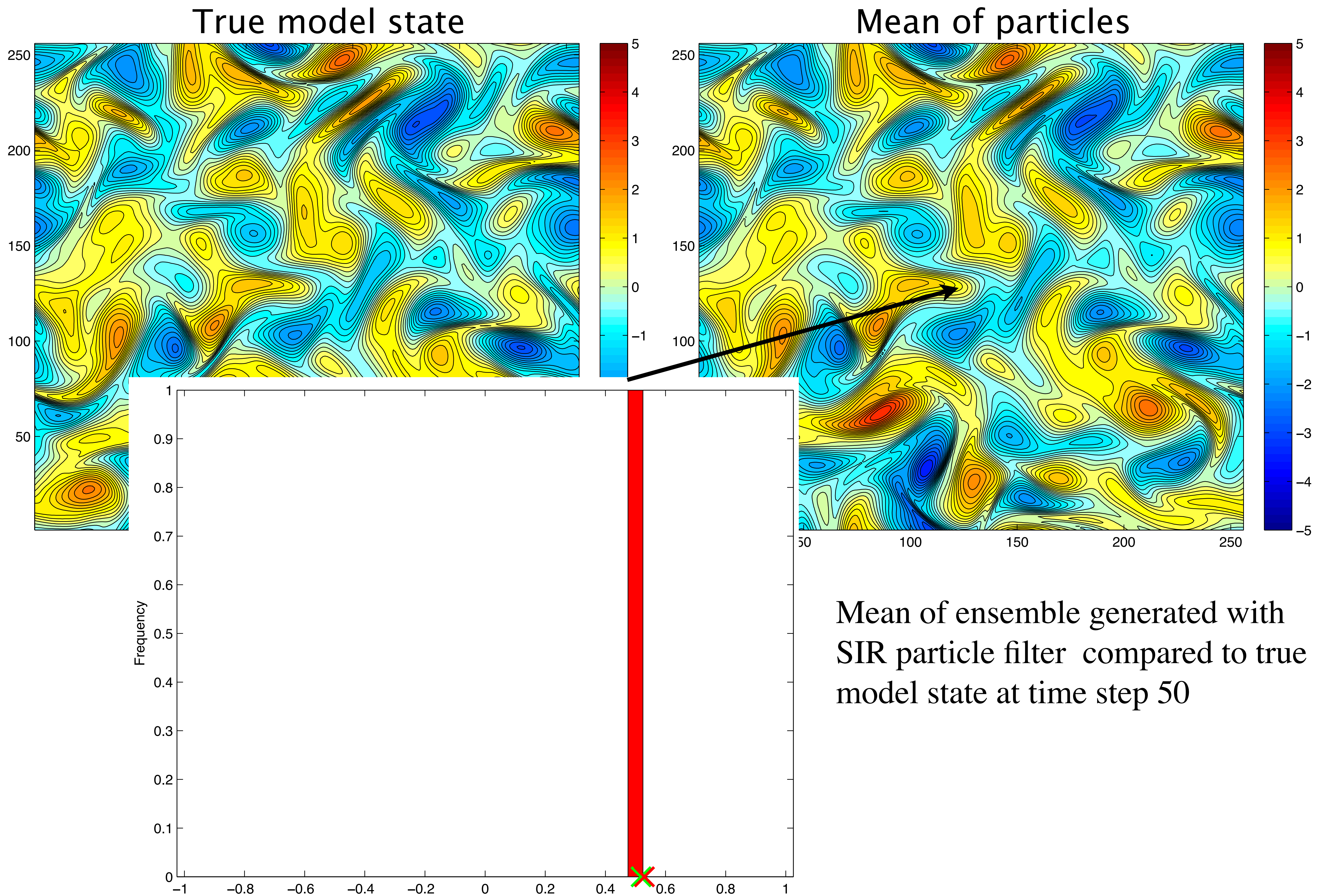
Equivalent weights particle filter - spread also seen in posterior distribution



Equivalent weights particle filter - both spread and mode of truth captured?



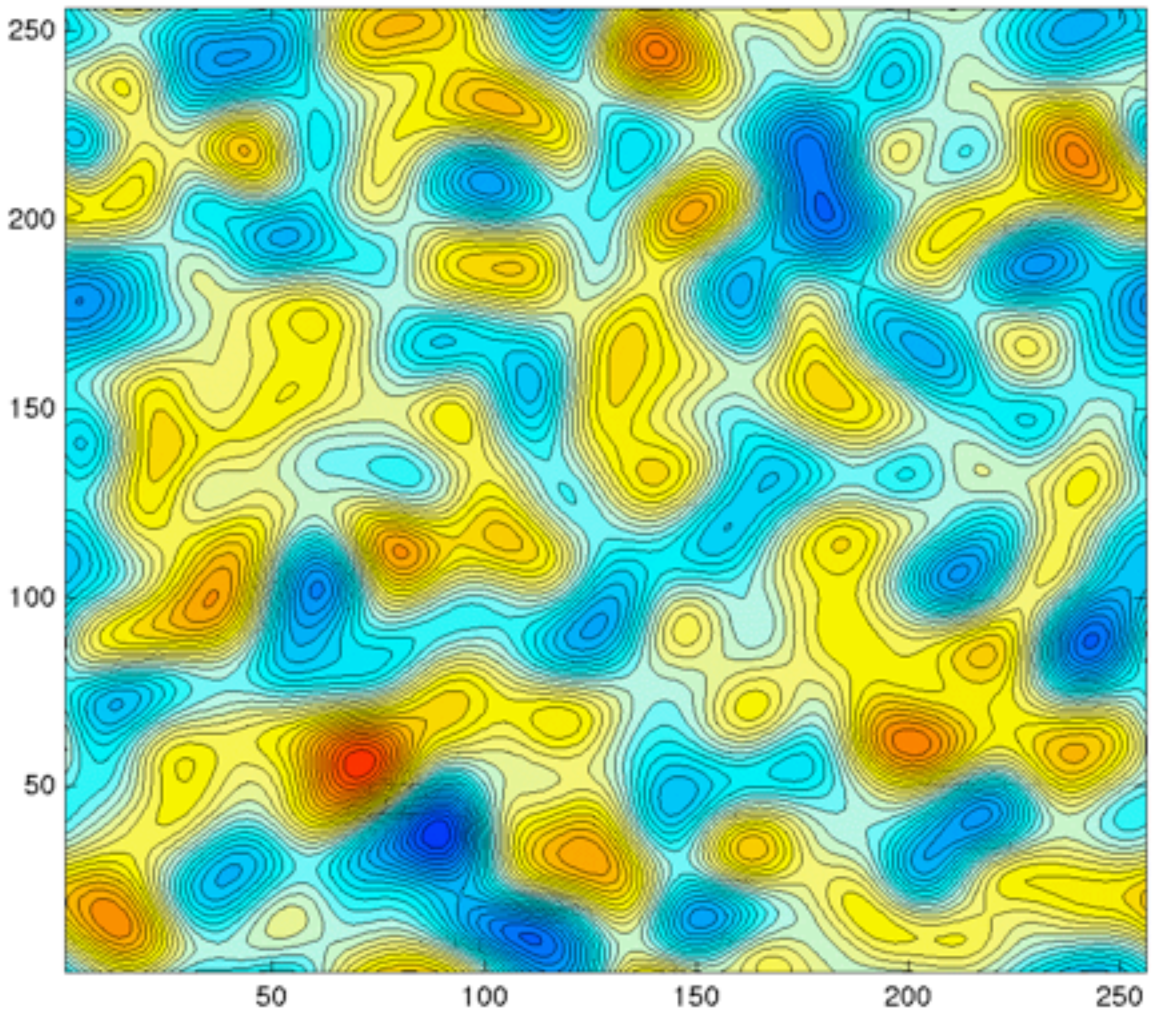
Standard Particle Filter - filter degeneracy is evident



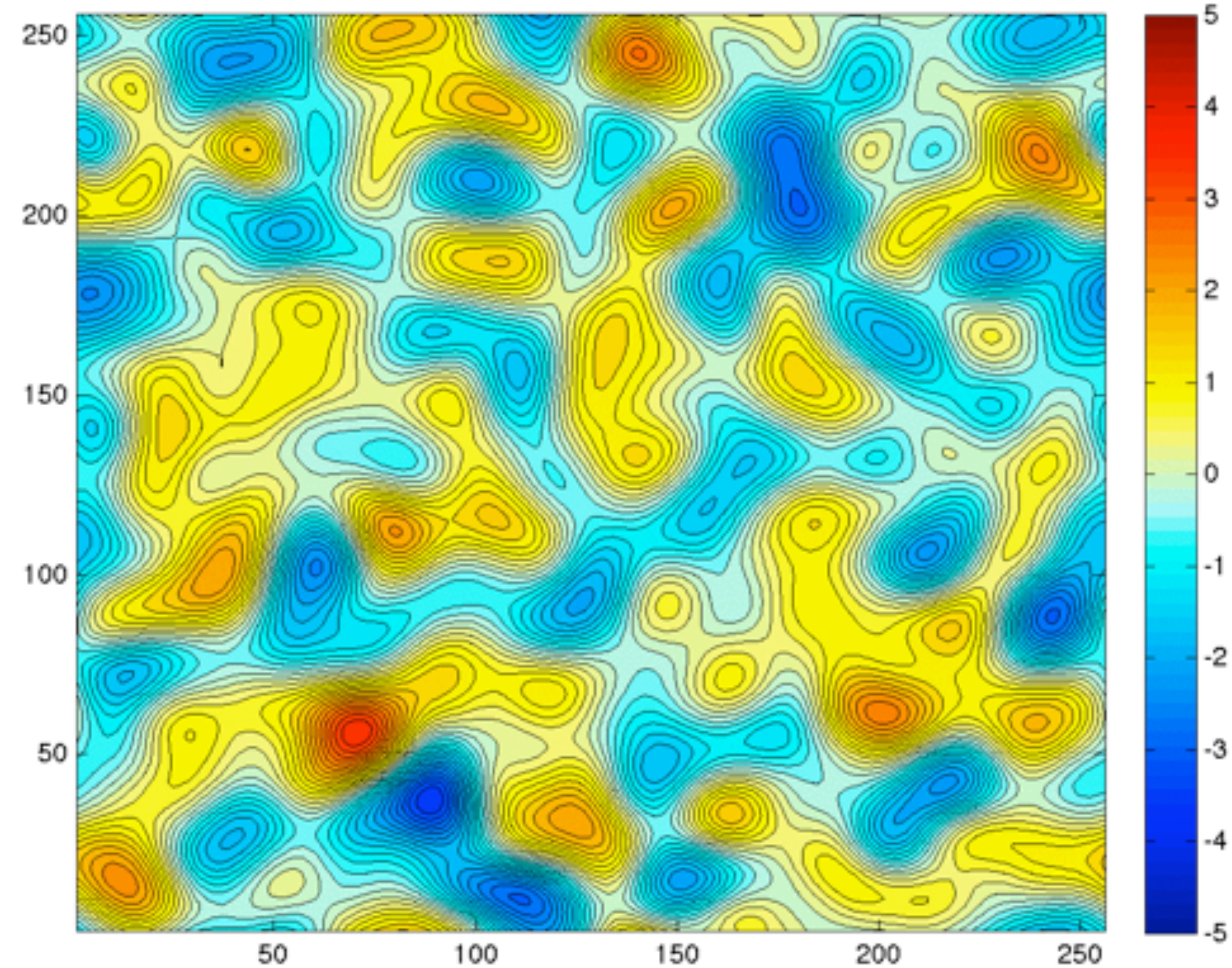
Mean of ensemble generated with SIR particle filter compared to true model state at time step 50

Equivalent weights particle filter - 1/4 observations over the full state

True model state



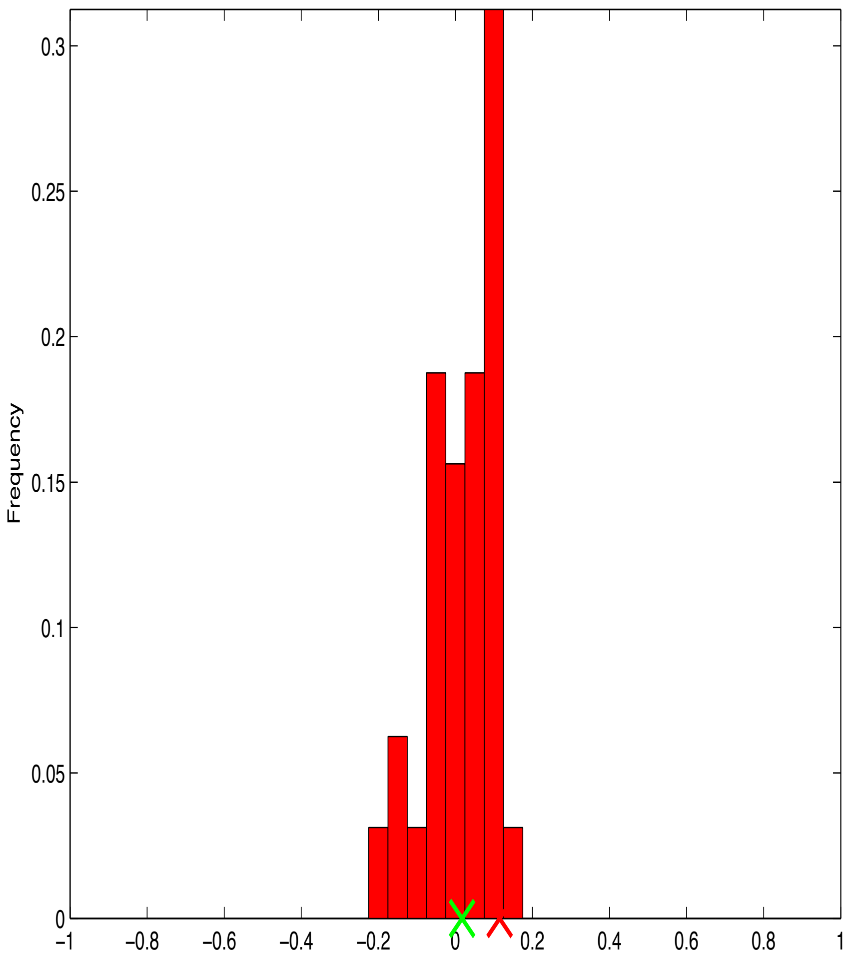
Mean of particles



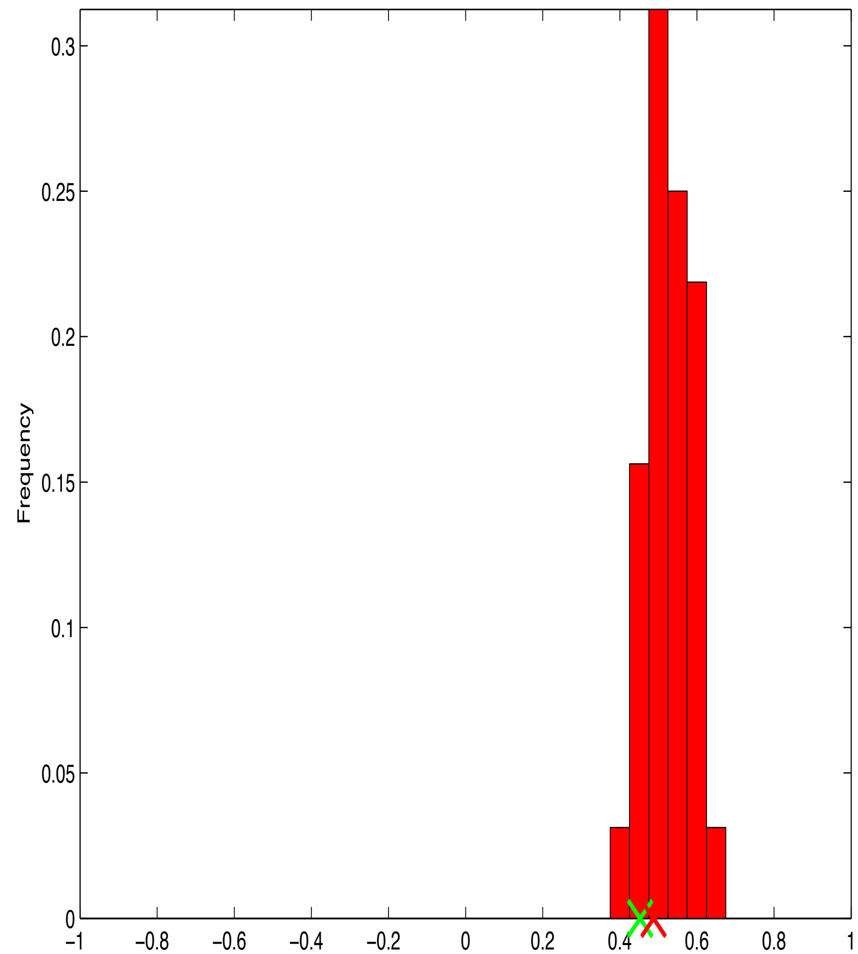
- Every other variable is observed
- 32 particles
- Observations every 50 timesteps
- 1200 time steps

Spread now seen in the posterior pdfs

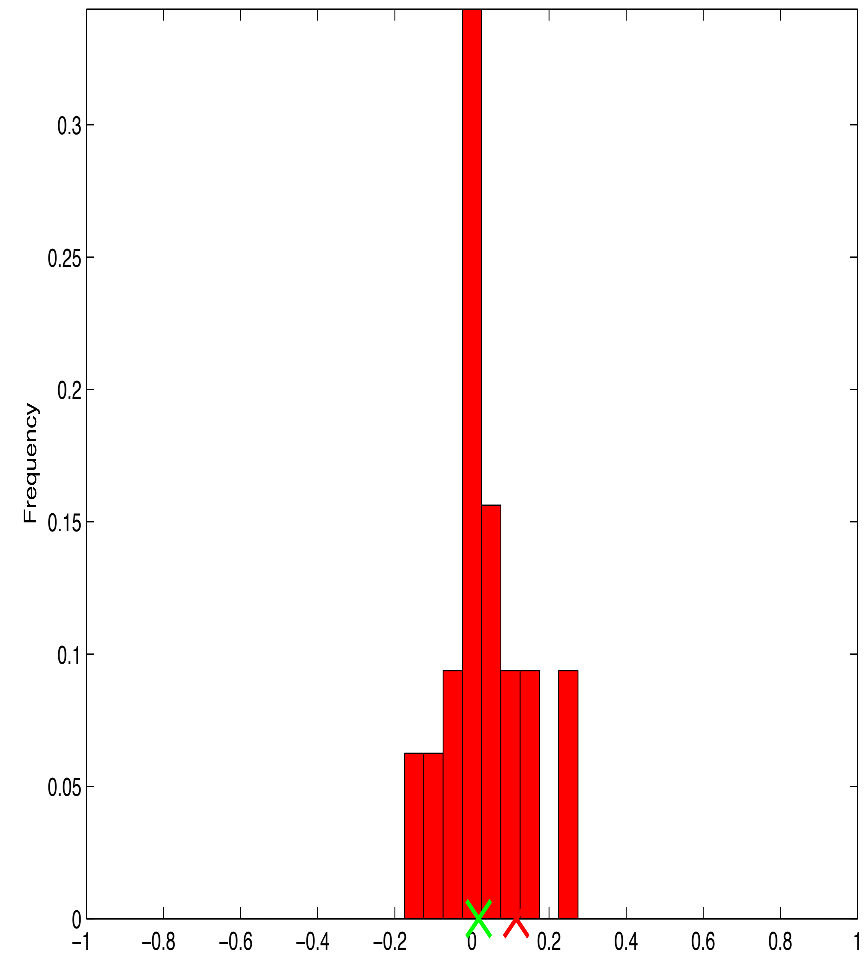
Marginal posterior pdfs at time step 1150
Observations of every other variable



variable (93,65)



variable (137,193)

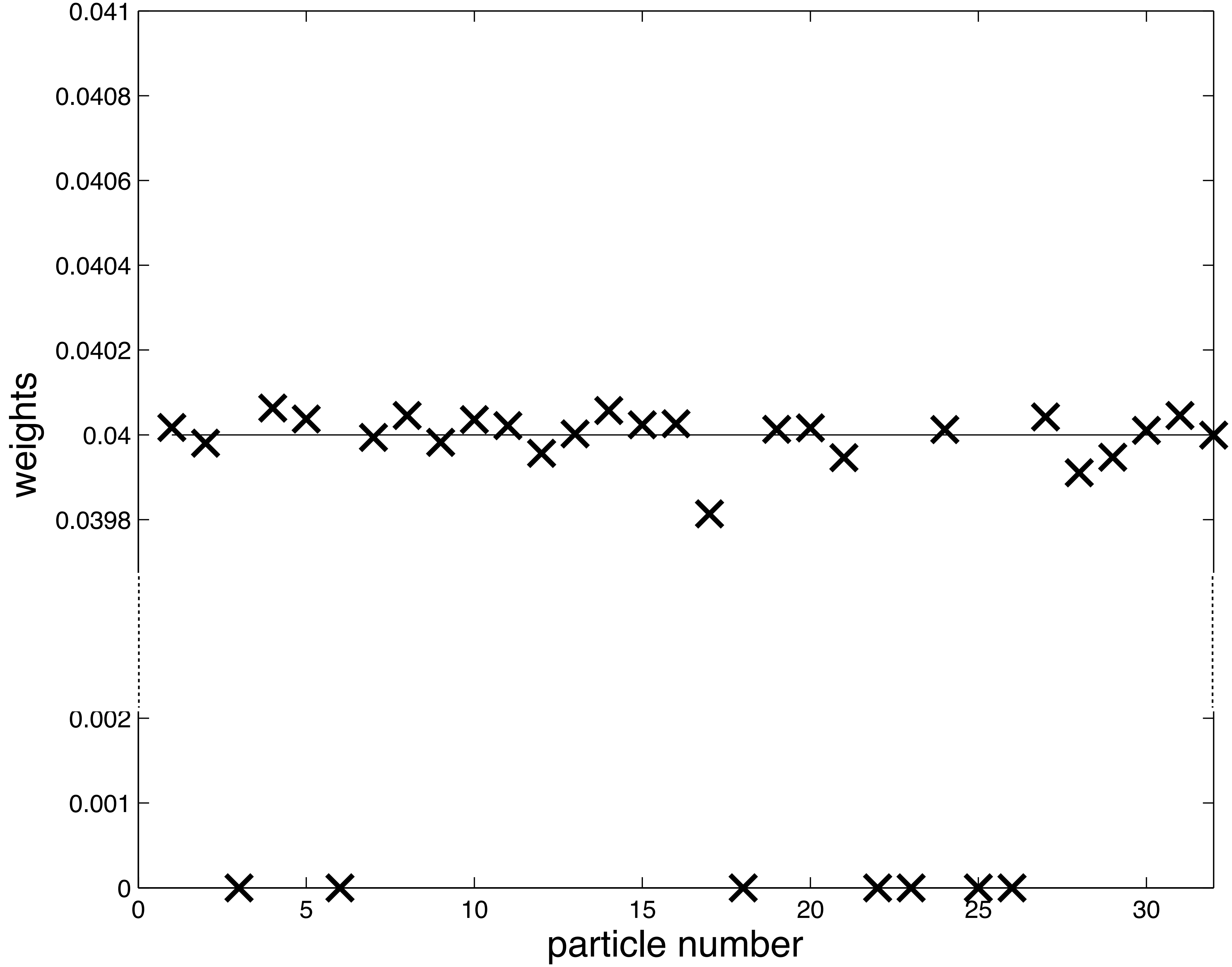


variable (177,65)

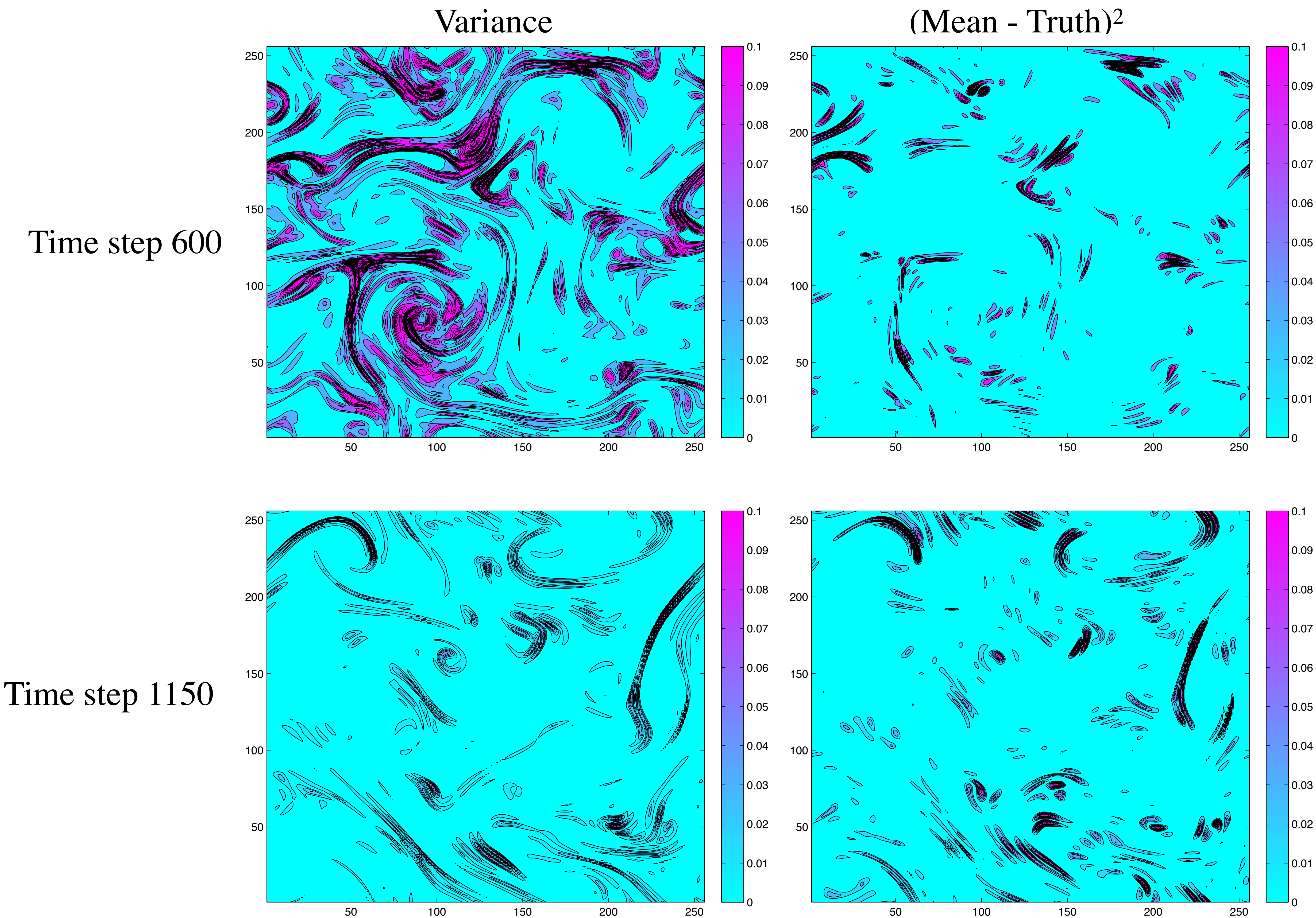
More work required to really assess whether these posterior representations match the true posterior pdf

The majority of particles have equivalent weights

Weights of particles before resampling at time step 1150

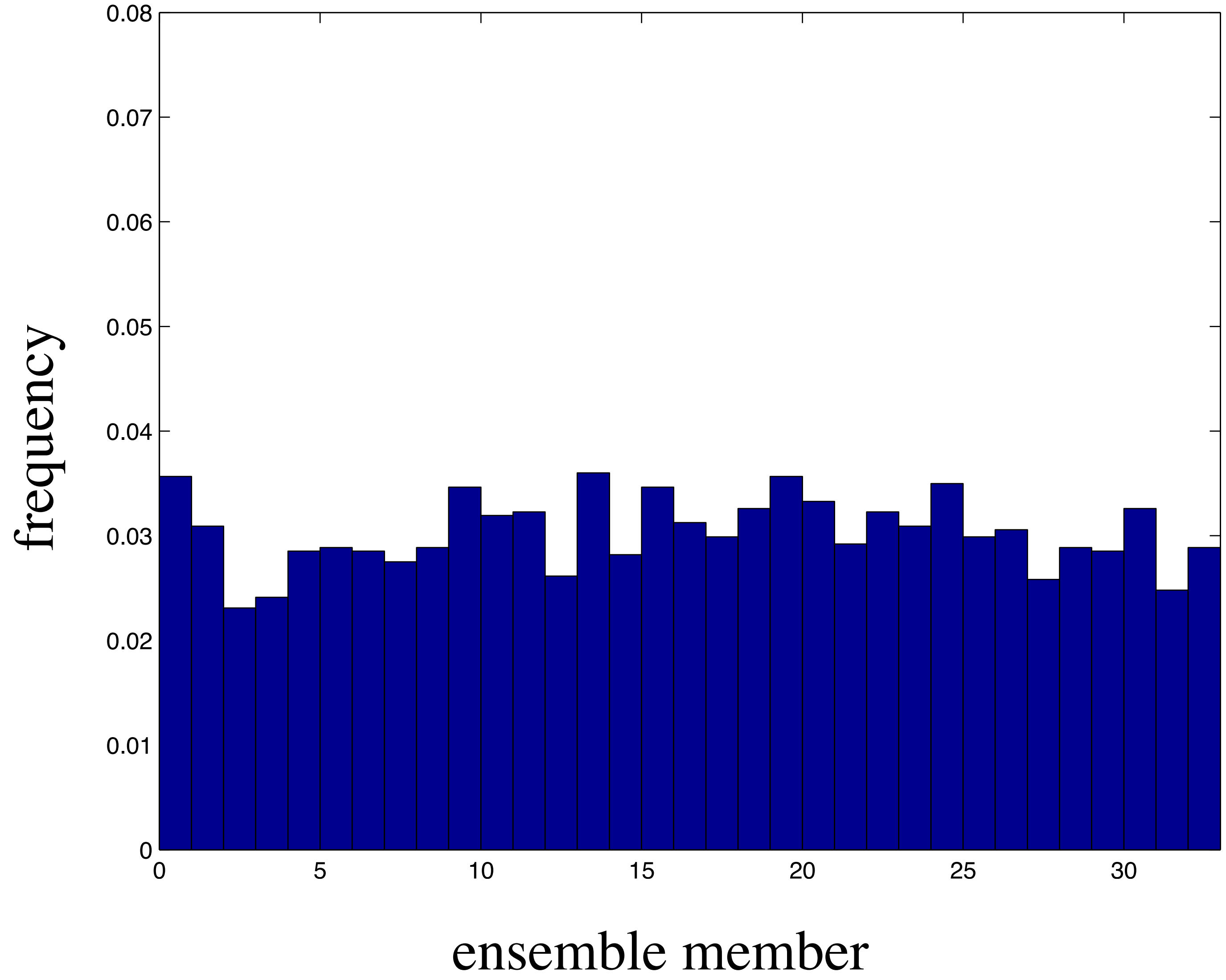


Similar patterns in spread - 1/4 obs



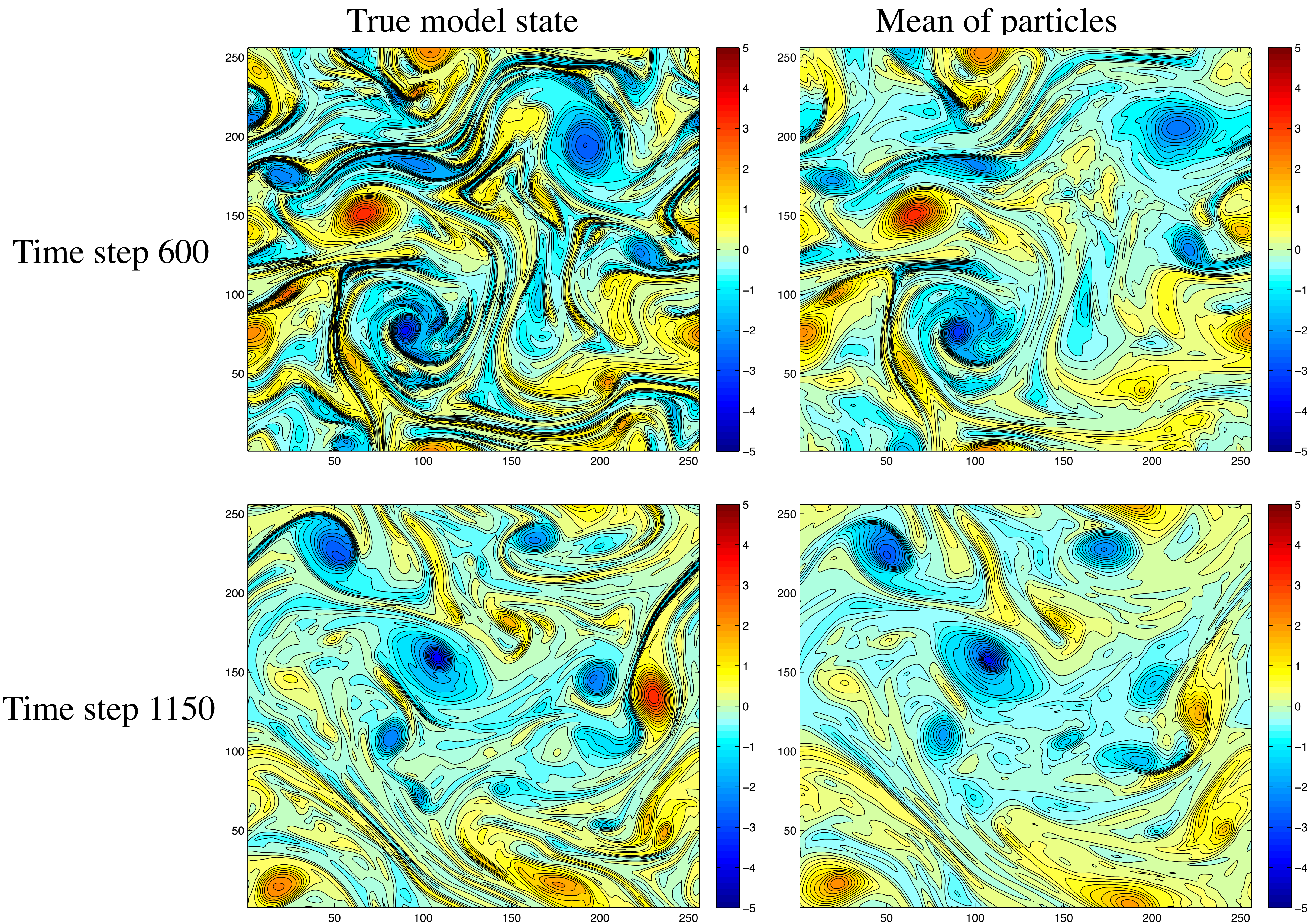
Good spread in the ensemble judged over all time steps

Observations of every other variable

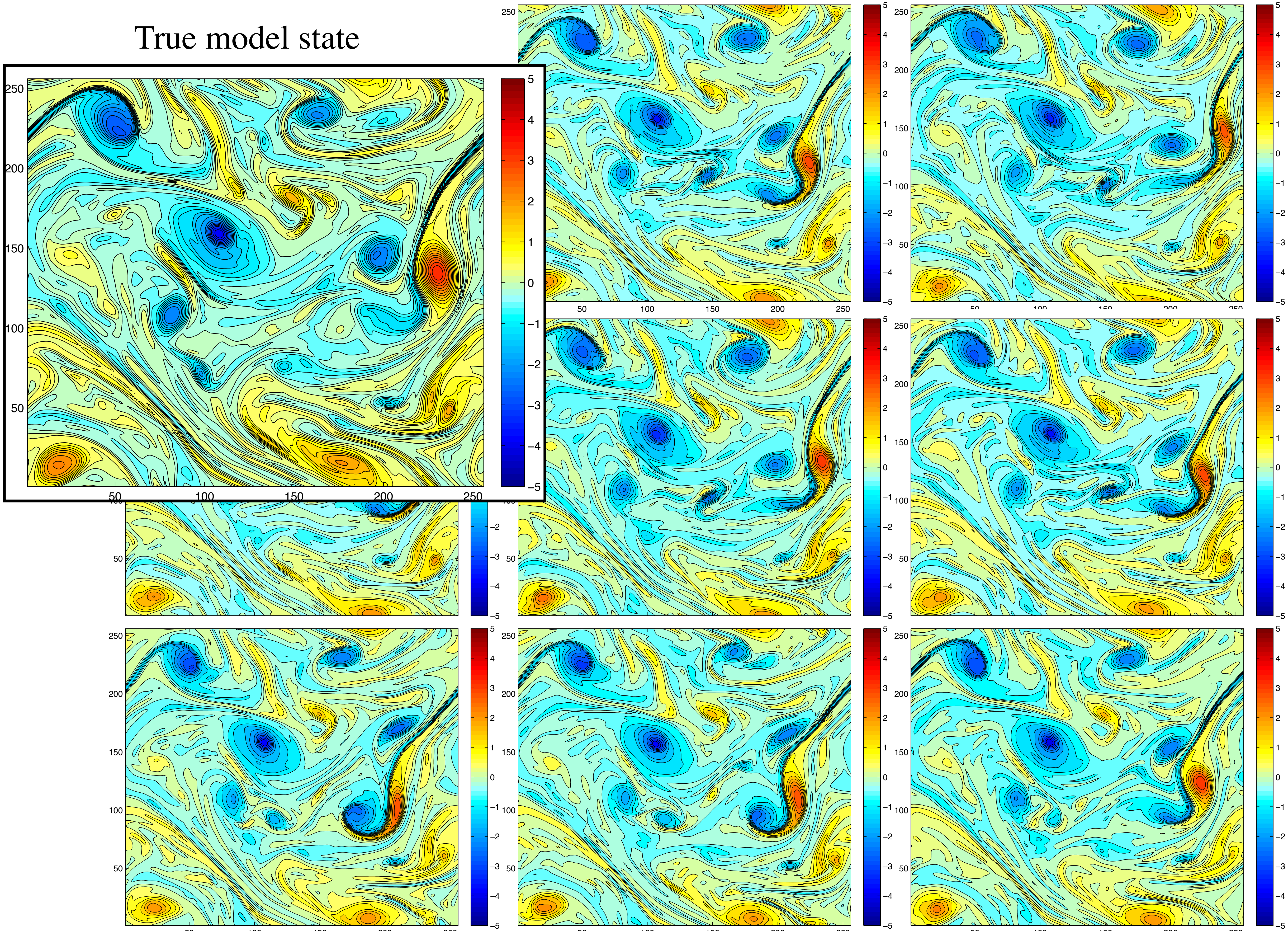


Rank histogram - observation error perturbed model variables compared to observation

Equivalent weights particle filter - 1/4 observations over half the state

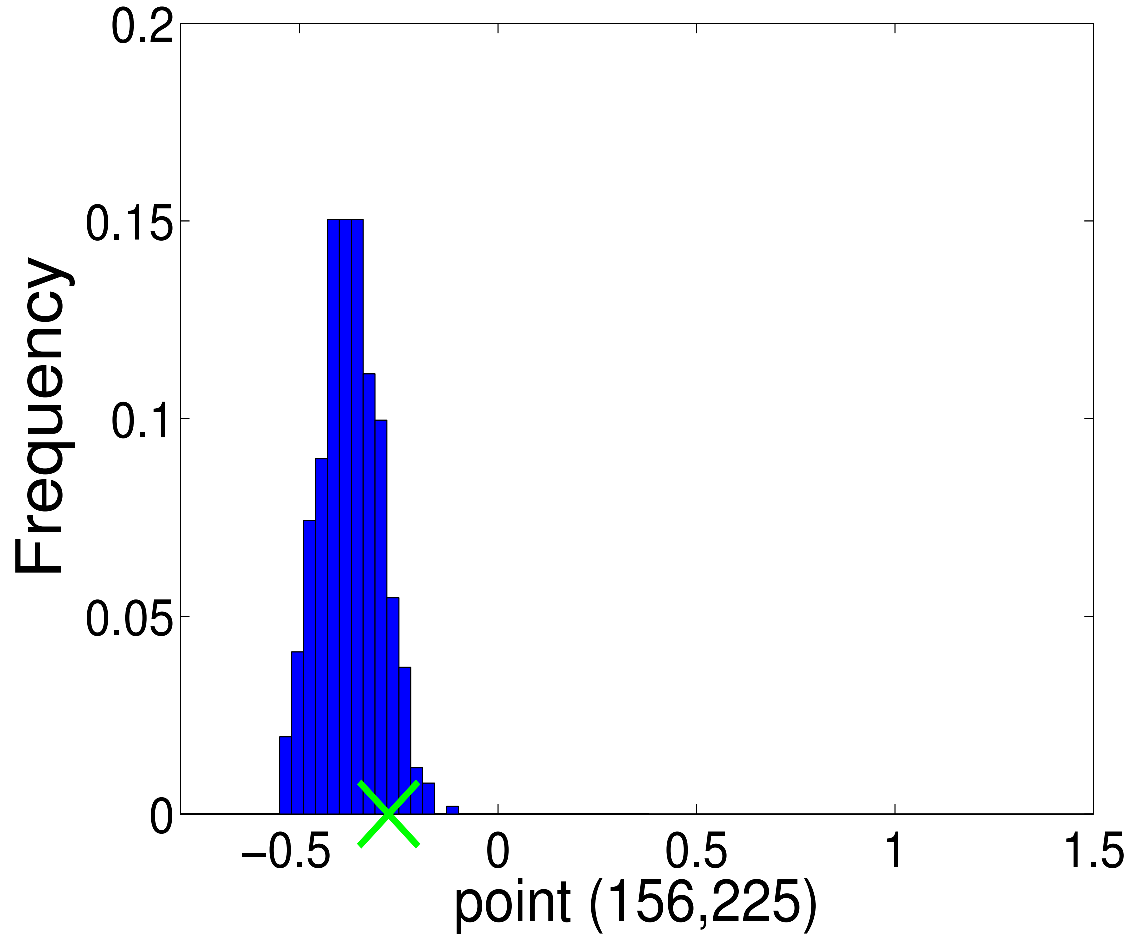


Individual particles -1/4 observations over half the state - time 1150

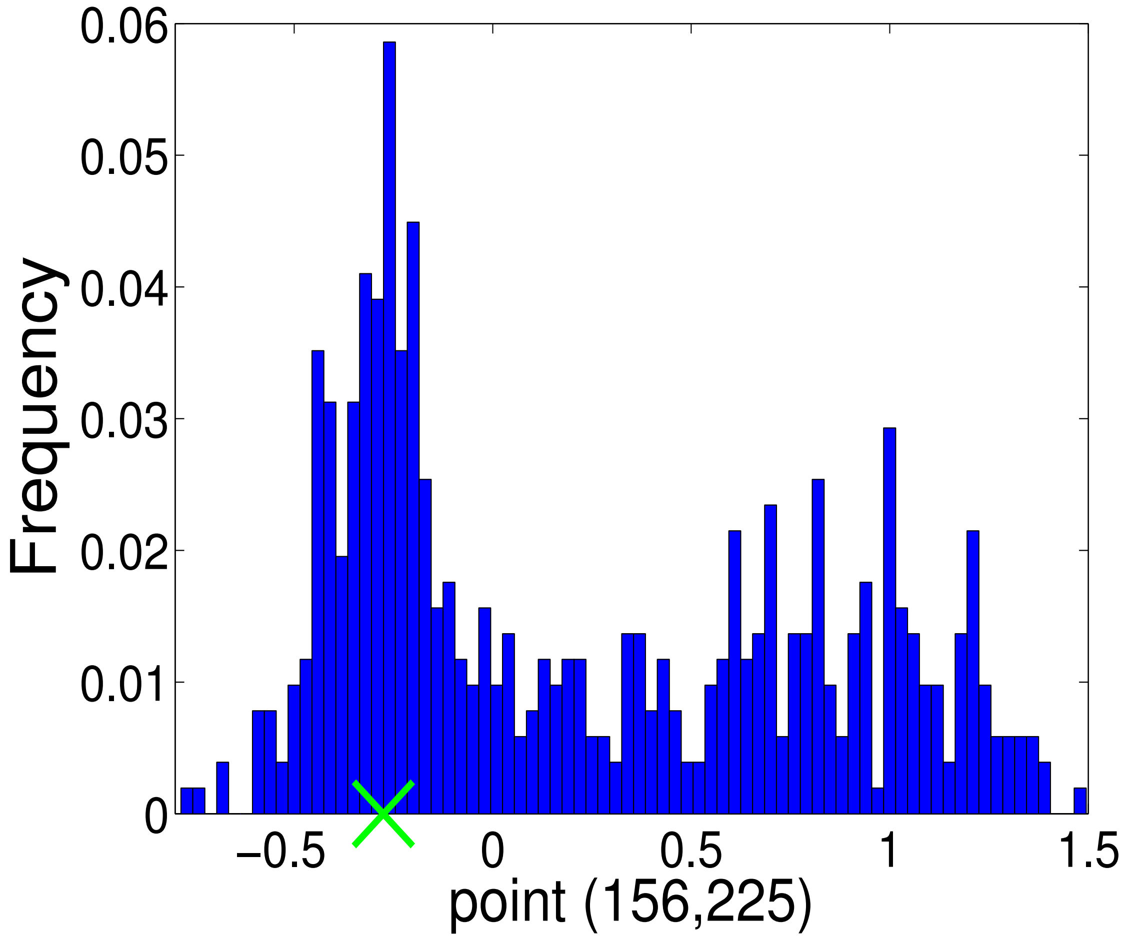


Multi-modal behaviour captured in posterior pdfs

If observed.....



If unobserved.....



Implicit particle filter

How to determine the new positions and weights of particles at time (k+1)l given the weights and positions at time kl?

$$\{v_{kl}^{(n)}, w_{kl}^{(n)}\}_{n=1}^{N} \mapsto \{v_{(k+1)l}^{(n)}, w_{(k+1)l}^{(n)}\}_{n=1}^{N}$$

Second option:

Related to the optimal proposal density

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

However, generalised to trajectories in time incorporating several observations

$$v^{(n)} = (v_1^{(n)}, \ldots, v_{kl}^{(n)}) = \mathrm{argmin}_{v^{(n)}} I(v_1^{(n)}, \ldots, v_{kl}^{(n)}; Y_k) + (\zeta_1^{(n)}, \ldots, \zeta_{kl}^{(n)})$$

$$I(v_1^{(n)}, \ldots, v_{kl}^{(n)}; Y_k) = \sum_{i=1}^{k} \frac{1}{2} \left|\Gamma^{-\frac{1}{2}}(y_i - H(v_{il}^{(n)}))\right|^2 + \sum_{j=1}^{kl-1} \frac{1}{2} \left|\Sigma^{-\frac{1}{2}}(v_{j+1}^{(n)} - \Psi(v_j^{(n)}))\right|^2$$

4D-Var with fixed initial condition

Implicit particle filter

optimal proposal density

$$w_{kl}^{(n)} \propto \exp\left(-\frac{1}{2}\left|(H\Sigma H^T + \Gamma)^{-\frac{1}{2}}(y_k - H\Psi(v_{kl-1}^{(n)}))\right|^2\right)$$

Weight over one time step depends on maximum weight given $v_{kl-1}^{(n)}$

Implicit particle filter

$$w_{kl}^{(n)} \propto \exp\left(-\frac{1}{2}\min_{v^{(n)}} I(v_1^{(n)}, \ldots, v_k l^{(n)}; Y_k)\right)|J(v^{(n)})|$$

Weight for trajectory depends on MAP estimate and a Jacobian which depends on the way the random error is chosen

Variance in weights depends on the variance of the maximum weights possible given the initial conditions

Combine ideas of equivalent-weights and variational methods

Equivalent-weights with 4D-Var

$$w_{kl}^{(n)} \propto \exp\left(-\frac{1}{2}\left[I(v_1^{(n)}, \ldots, v_k l^{(n)}; Y_k) - I^{target}\right]\right) |J(v^{(n)})|$$

Variance in weights now only depends on the way the random error is chosen

Trajectory is no longer chosen to minimise the cost function but rather to ensure it holds a specific value

$$I(v_1^{(n)}, \ldots, v_k l^{(n)}; Y_k) = I^{target}$$

Trajectory no longer has the maximum weight possible given the initial conditions but a chosen percentage of particles will now all have significance when representing the posterior

Summary

Particle filtering:

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

Disadvantages:

- high variance in the weights leads to filter degeneracy, where only one particle has any significance in the posterior.
- filter degeneracy linked to the dimension of the system or the number of independent observations.
- because of this they are difficult to apply in lot of real life high dimensional applications.

Advantages:

- fully nonlinear, so allow a representation of a potentially multi-modal posterior pdf.
- no need for an estimate of the state covariance, only model error covariance required.

High dimensional particle filters use the free choice of proposal density to try and sample particles from the posterior whose weights do not have high variance

Summary

Optimal proposal density → Filter degeneracy still occurs in high dimensional systems

EnKF as proposal density → Filter degeneracy still occurs in high dimensional systems

Implicit particle filter →

- Good results seen, depending on the method for choosing the random error
- Further research required to determine whether filter degeneracy really is avoided with high numbers of observations

Equivalent-weights particle filter →

- Filter degeneracy is avoided by formulation
- Promising results that the posterior representation matches the true pdf, but further research required to fully determine this

References

Equivalent-weights particle filter

Van Leeuwen PJ. 2010. Nonlinear data assimilation in geoscience: an extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society* **136**: 1991-1999

Ades M, Van Leeuwen PJ. 2013. An exploration of the equivalent weights particle filter. *Quarterly Journal of the Royal Meteorological Society* **139**: 820-840, doi:10.1002/qj.1995

Van Leeuwen PJ, Ades M. 2013. Efficient fully nonlinear data assimilation for geophysical fluid dynamics. *Computers and Geosciences* **55**: 16-27, doi:10.1016/j.cageo.2012.04.015

Implicit particle filter

Morzfeld M, Tu X, Atkins E, Chorin AJ. 2012. A random map implementation of implicit filters. *Journal of Computational Physics* **231**: 2049-2066

Chorin AJ, Morzfeld M, Tu X. 2010. Interpolation and iteration for nonlinear filters. *Communications in Applied Mathematics and Computational Science* **5**: 221-240

Atkins E, Morzfeld M, Tu X, Chorin AJ. 2013. Implicit particle methods and their connection with variational data assimilation. *Monthly Weather Review* **141**: 1786-1803