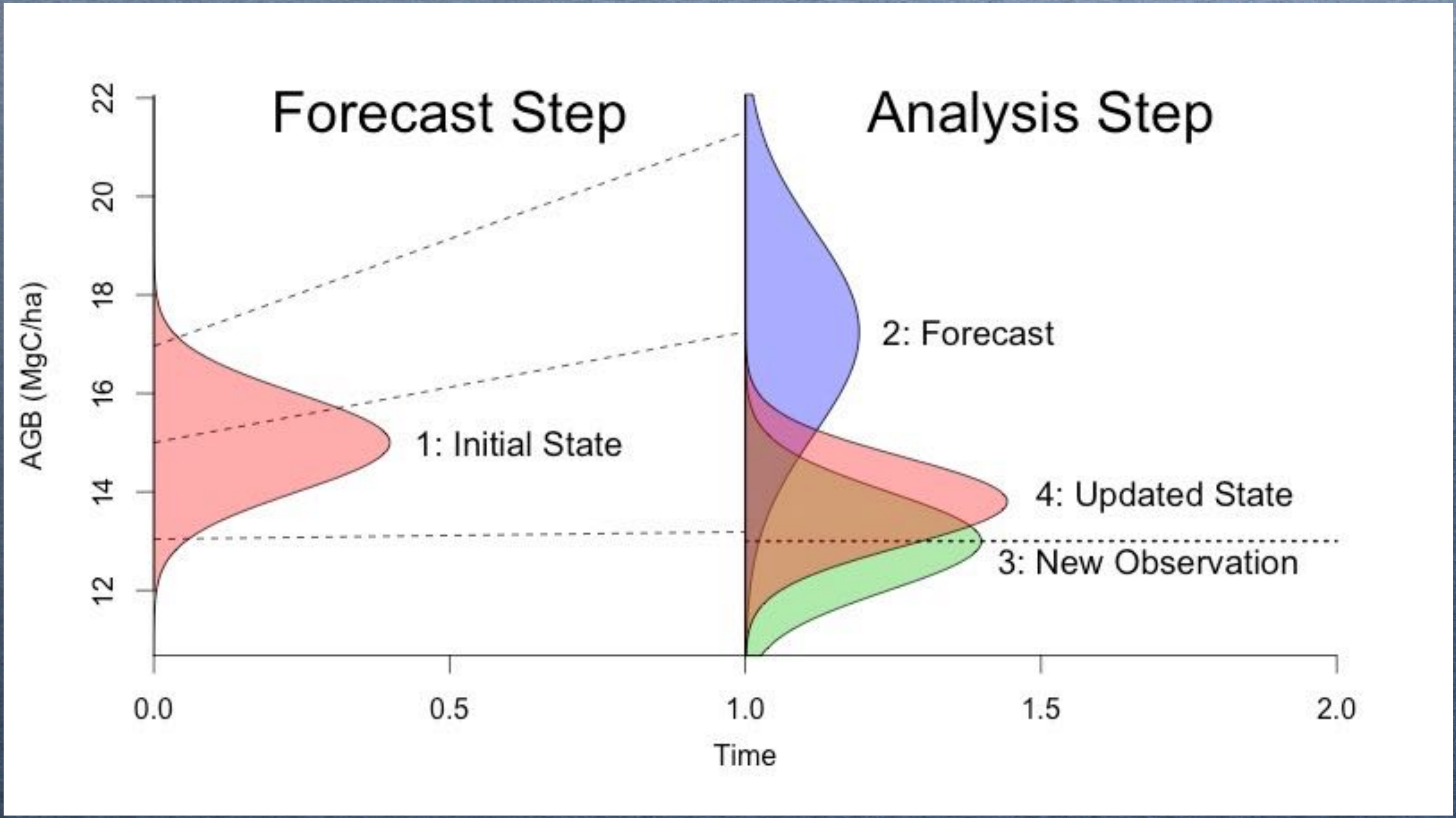
Data Assimilation 2: Monte Carlo Methods

"An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem." John W. Tukey

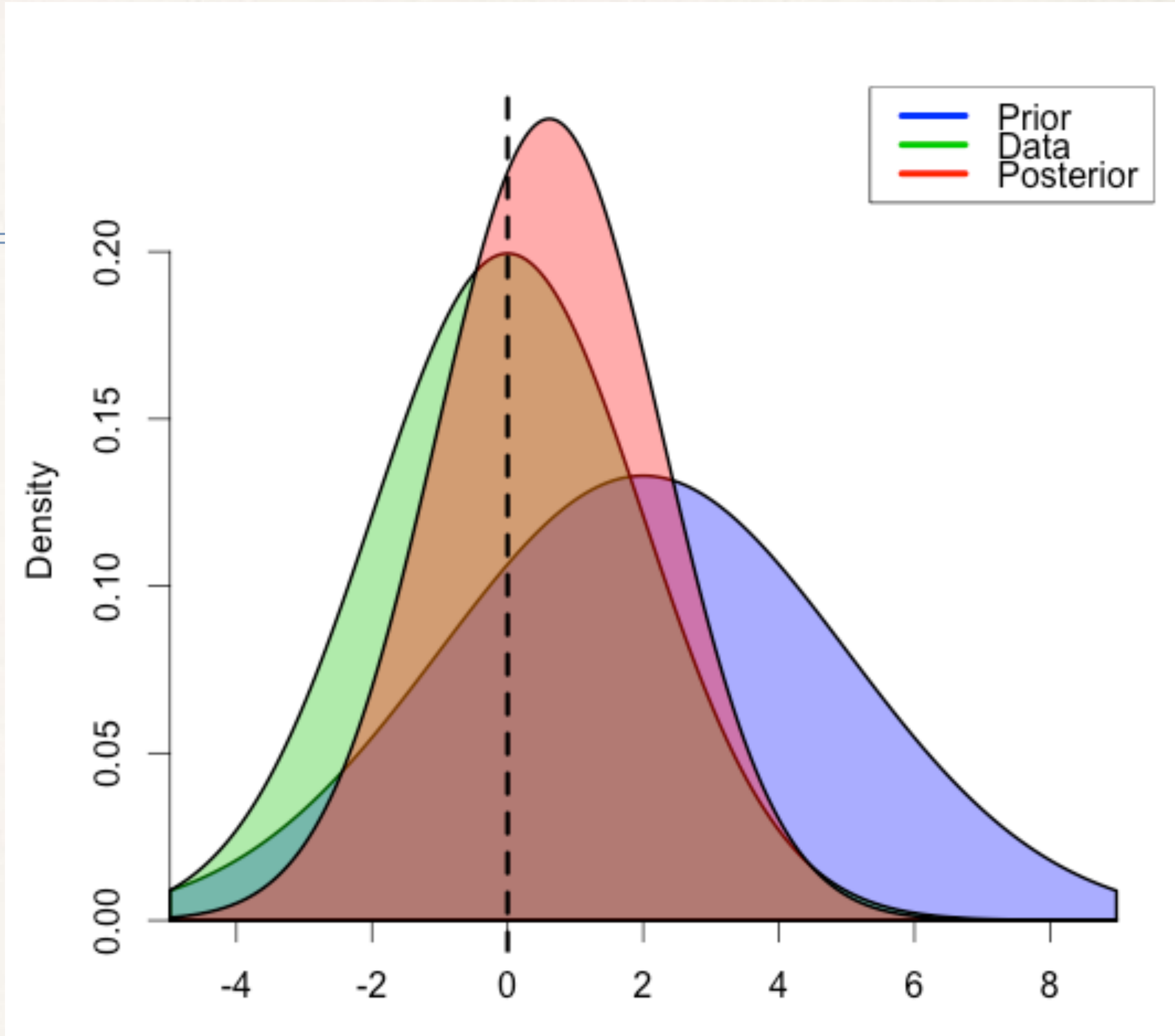


Forecast Cycle

UNCERTAINTY PROPAGATION
APPLIED IN THE FORECAST STEP

		Output	
		Distribution	Moments
Approach	Analytic	Variable Transform	Analytical Moments KF Taylor Series EKF
	Numeric	Monte Carlo PF	Ensemble EnKF

Kalman Analysis



- Forecast:
  Assume $P(X_{t+1}) \sim N(\mu_f, p_f)$
- Observation error:
  Assume $P(Y_{t+1} | X_{t+1}) \sim N(X_{t+1}, r)$
  - Likelihood = Data model
- Assume Y, μ_f, p_f and r are known
- $P(X_{t+1} | Y_{t+1}) \sim N(\mu_a, p_a)$

$$\rho = 1/r \qquad \phi = 1/p_f$$

$$X|Y \sim N\left(\frac{\rho}{n\rho + \phi} n\bar{Y} + \frac{\phi}{n\rho + \phi}\mu_f, n\rho + \phi\right)$$

$$X_a | Y \sim N(Y | HX_a, R)\, N(X_a | \mu_f, P_f)$$

- Solves to be

$$X_a | Y \sim N\Big(\big(H^T R^{-1} H + P_f^{-1}\big)^{-1} \big(H^T R^{-1} Y + P_f^{-1} \mu_f\big), \big(H^T R^{-1} H + P_f^{-1}\big)^{-1} \Big)$$

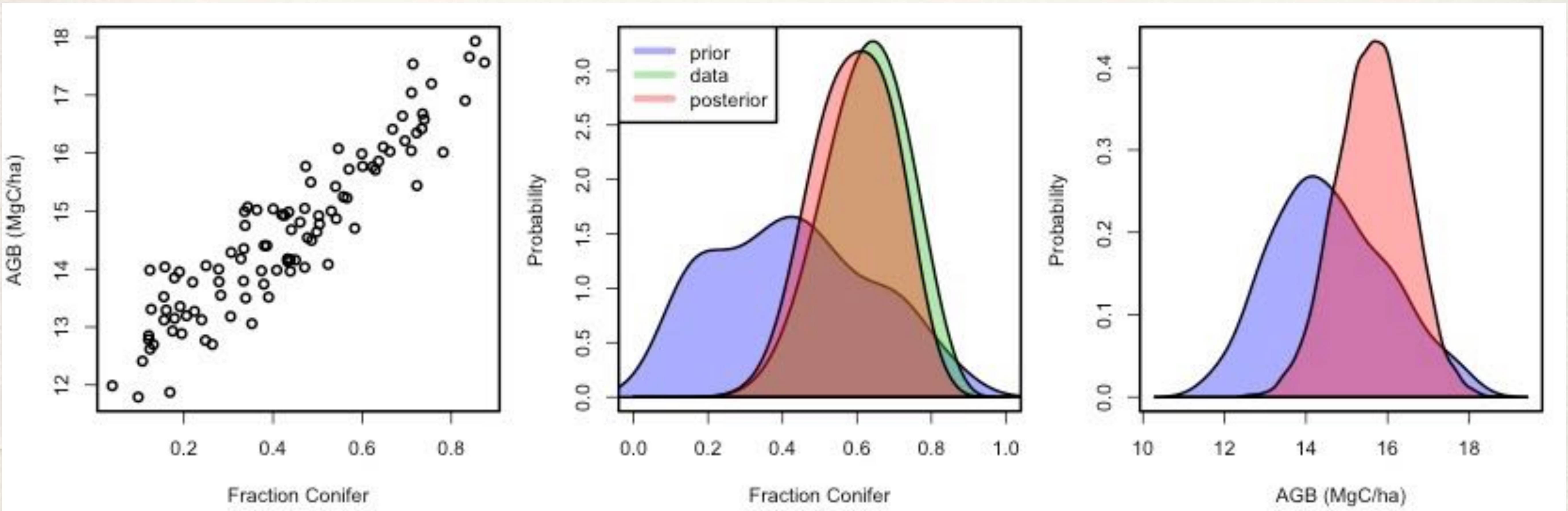
- Mean and variance simplify to

$$E[X_a | Y] = \mu_a = \mu_f + K(Y - H\mu_f)$$

$$Var[X_a | Y] = P_a = (I - KH) P_f$$

$$K = P_{fH}^T (R + H P_f H^T)^{-1}$$

Kalman Gain



AGB (MgC/ha)
Fraction Conifer
prior
data
posterior
Probability
Fraction Conifer
Probability
AGB (MgC/ha)

Forecast Step $X_{t+1} = MX_t + \epsilon$

The posterior distribution of X_{t+1} given X_t is multivariate normal with

$$\mu_{f,t+1} = E[X_{f,t+1} | X_{a,t}] = M_t \mu_{a,t}$$

$$P_{f,t+1} = Var[X_{f,t+1} | X_{a,t}] = Q_t + M_t P_{a,t-1} M_t^T$$

Extended Kalman Filter (EKF)

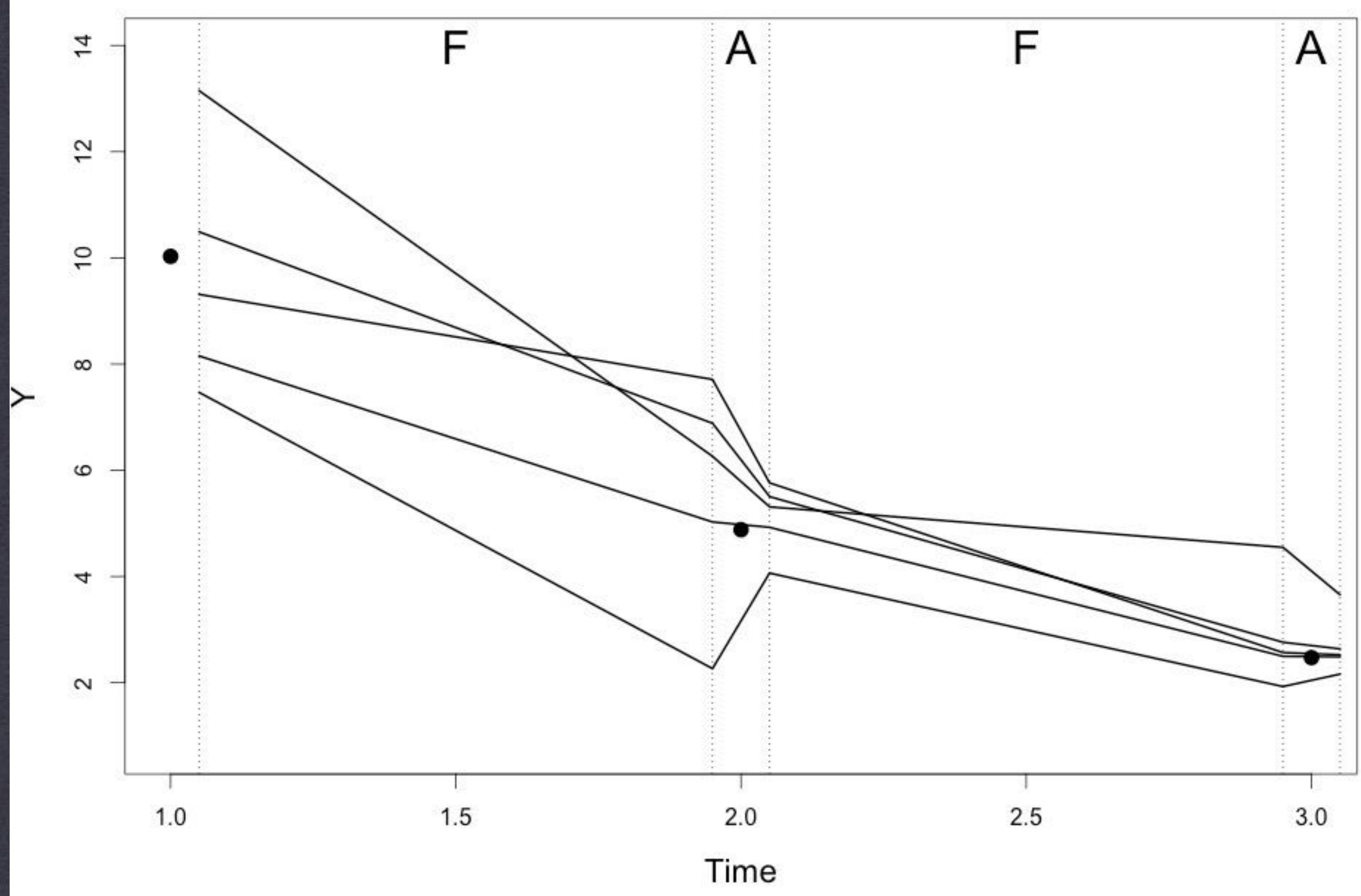
- Addresses **linear** assumption of the Forecast
  - $\mu_f = f(\mu_a)$
- Update variance using a Taylor Series expansion
  - F = Jacobian (df_i / dx_j)
  - $P_f \approx Q + F\, P_a\, F^T$ (was $Q + M\, P_a\, M^T$)
  - Can be extended to higher orders
- Jensen's Inequality: Biased, Normality assumption FALSE

Ensemble Kalman Filter (EnKF)

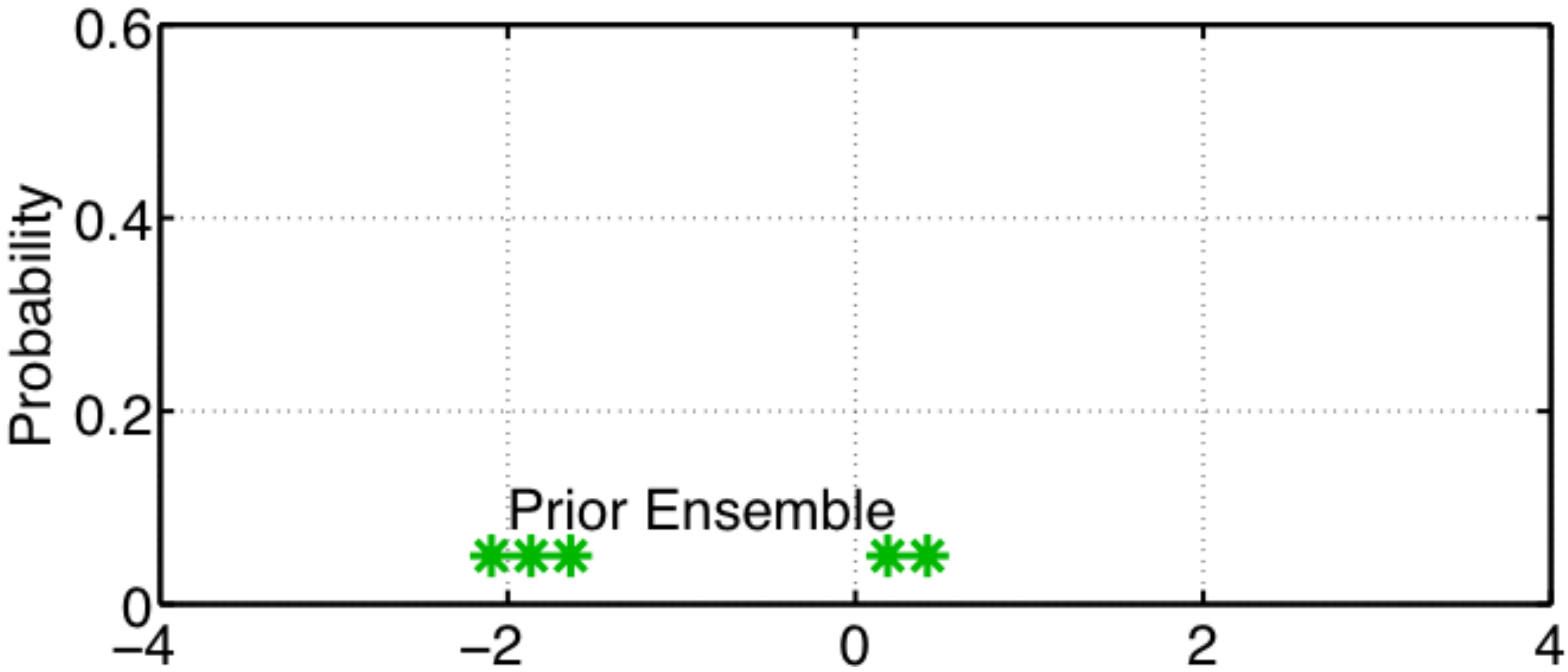
- Analysis identical to KF
- Uses Monte Carlo samples to approximate Forecast distribution
  - Draw m samples from the Analysis posterior
  - Run process model + process error for sample

$$\mu_{f,t+1} = \frac{1}{m} \sum X_{f,i}$$

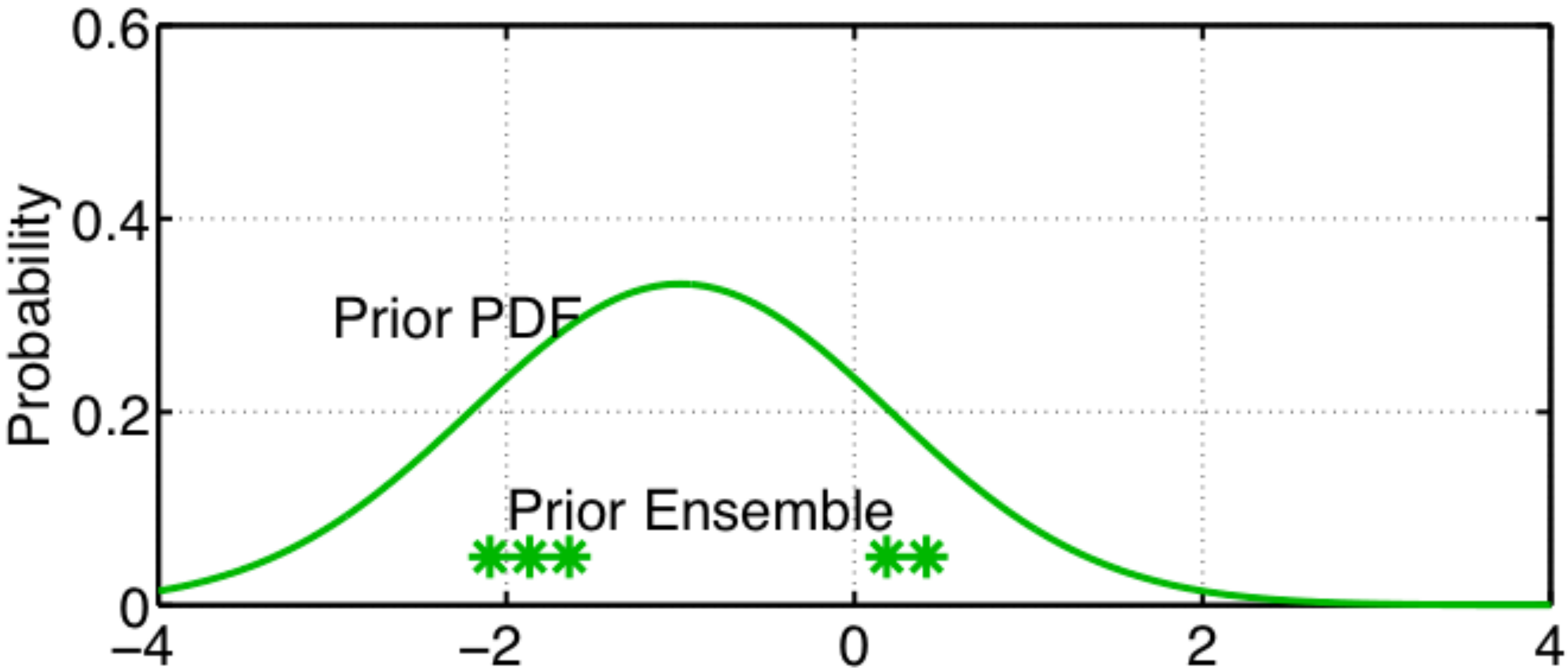
$$P_{f,t+1} = COV[X_{f,i}]$$


F
A
F
A
Y
14
12
10
8
6
4
2
1.0
1.5
2.0
2.5
3.0
Time

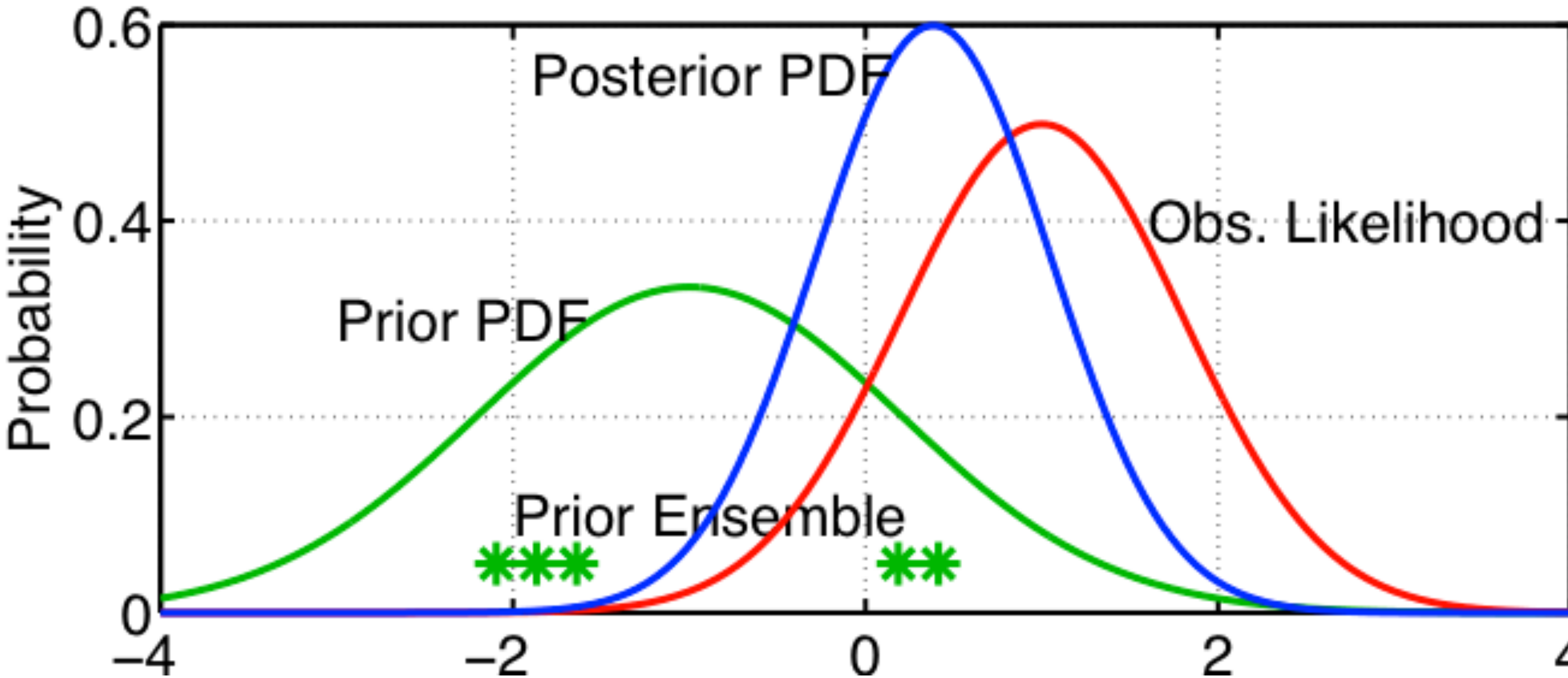
Ensemble adjustment (Kalman) filter



Ensemble adjustment (Kalman) filter



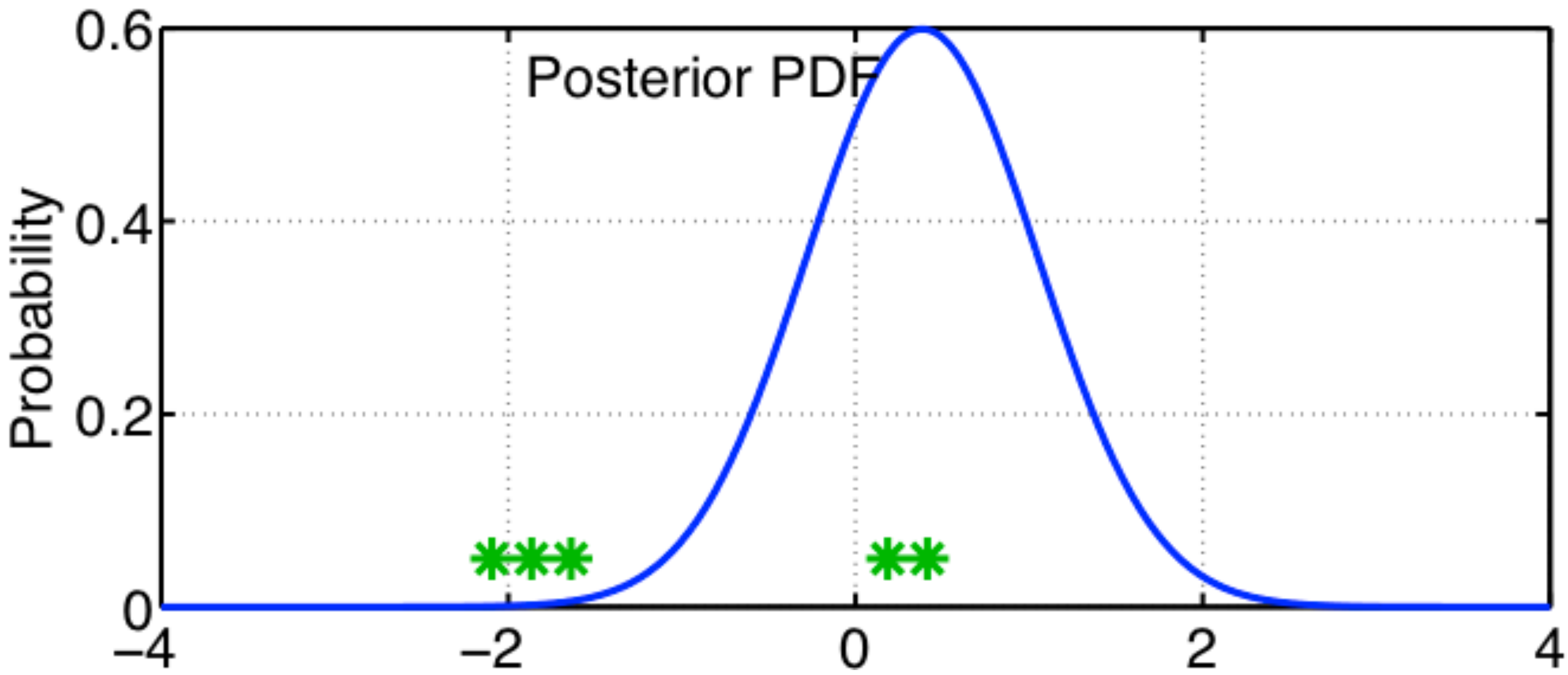
Ensemble adjustment (Kalman) filter



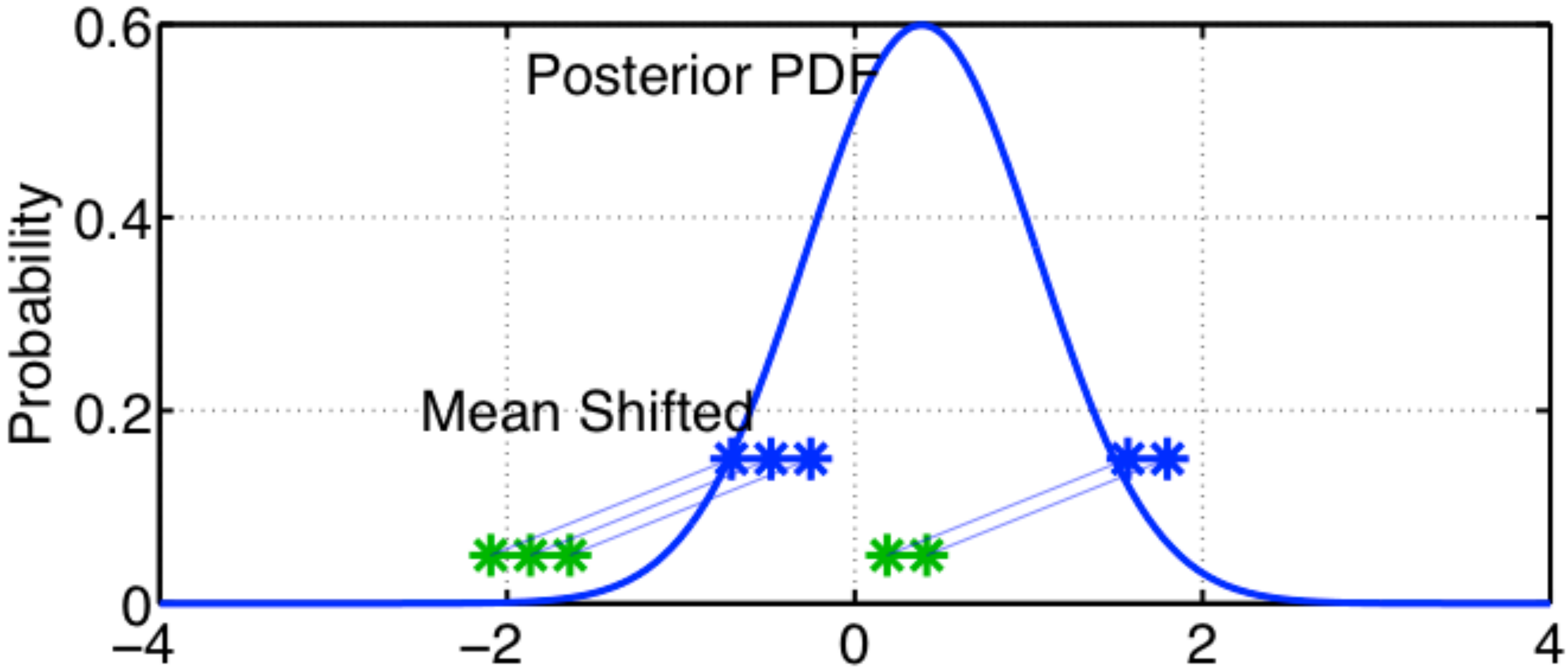
observation likelihood

posterior distribution

Ensemble adjustment (Kalman) filter

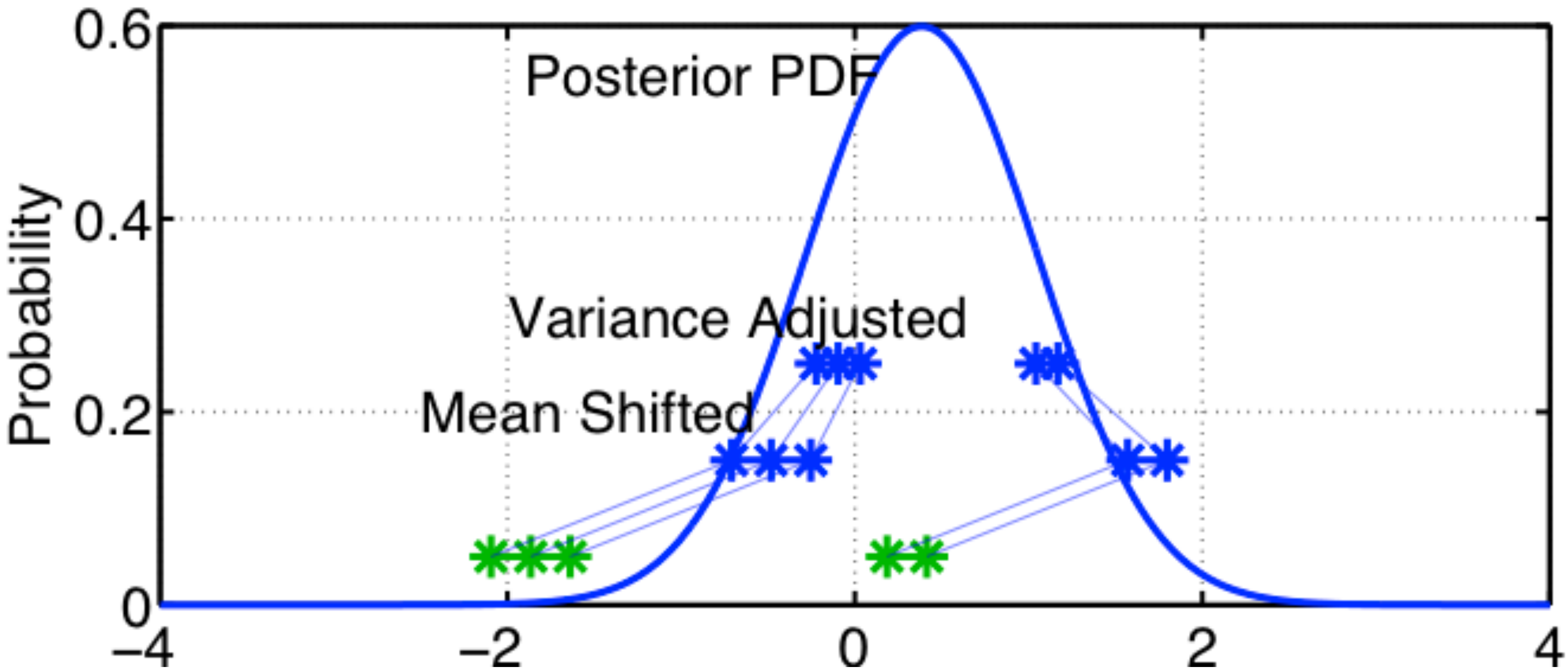


Ensemble adjustment (Kalman) filter



prior posterior

Ensemble adjustment (Kalman) filter



posterior

Ensemble Adjustment

- Alt to resampling analysis posterior, nudge current ensemble
- Useful when other uncert & latent states
- SVD: $P = VLV^{-1}$
- Normalize: $Z_i = \sqrt{L_f}^{-1} V_f^{-1} * (X_{i,f} - \mu_f)$
- Update: $X_{i,a} = V_a \sqrt{L_a} Z_i + \mu_a$

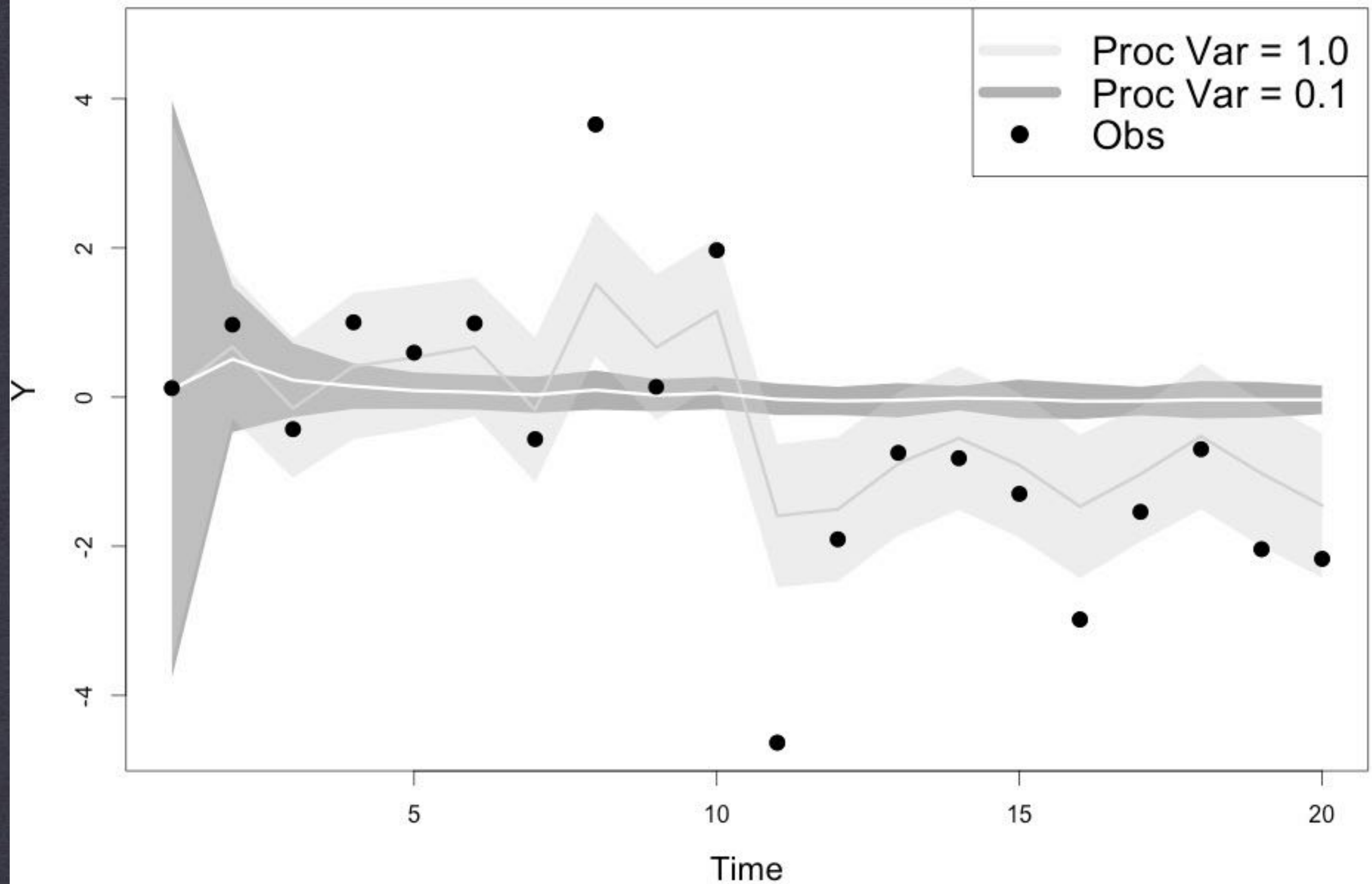
EnKF pro/con

- Nonlinear
- Existing code: No Jacobian
- Simple to implement, understand
- Sample size chosen based on power analysis
  - Con: larger than Analytical methods
- Simpler to add other sources of uncert. (e.g. driver)
- Moments OK on Jensen's Inequality
- Normal, but violates Normality
  - Analysis not hard to generalize (Likelihood * Prior) but unlikely to have an analytical sol'n

Localization

- All KF flavors involve matrix inversion
- Cheaper if correlation matrix is sparse
- Often assume correlations beyond some distance are zero
  - avoids spurious correlations
  - distance need not be physical

FILTER DIVERGENCE



Filter Divergence

- Practitioners of DA in atm sci frequently worry about model variance collapsing to zero
- Model then ignores (diverges from) data
- Process error is TUNED [BAD]
- Ecology is far less chaotic
  - Occasionally, convergence is right answer
  - In others, indicates misspecified process model or partitioning of process error

No KF variant can estimate process and observation errors

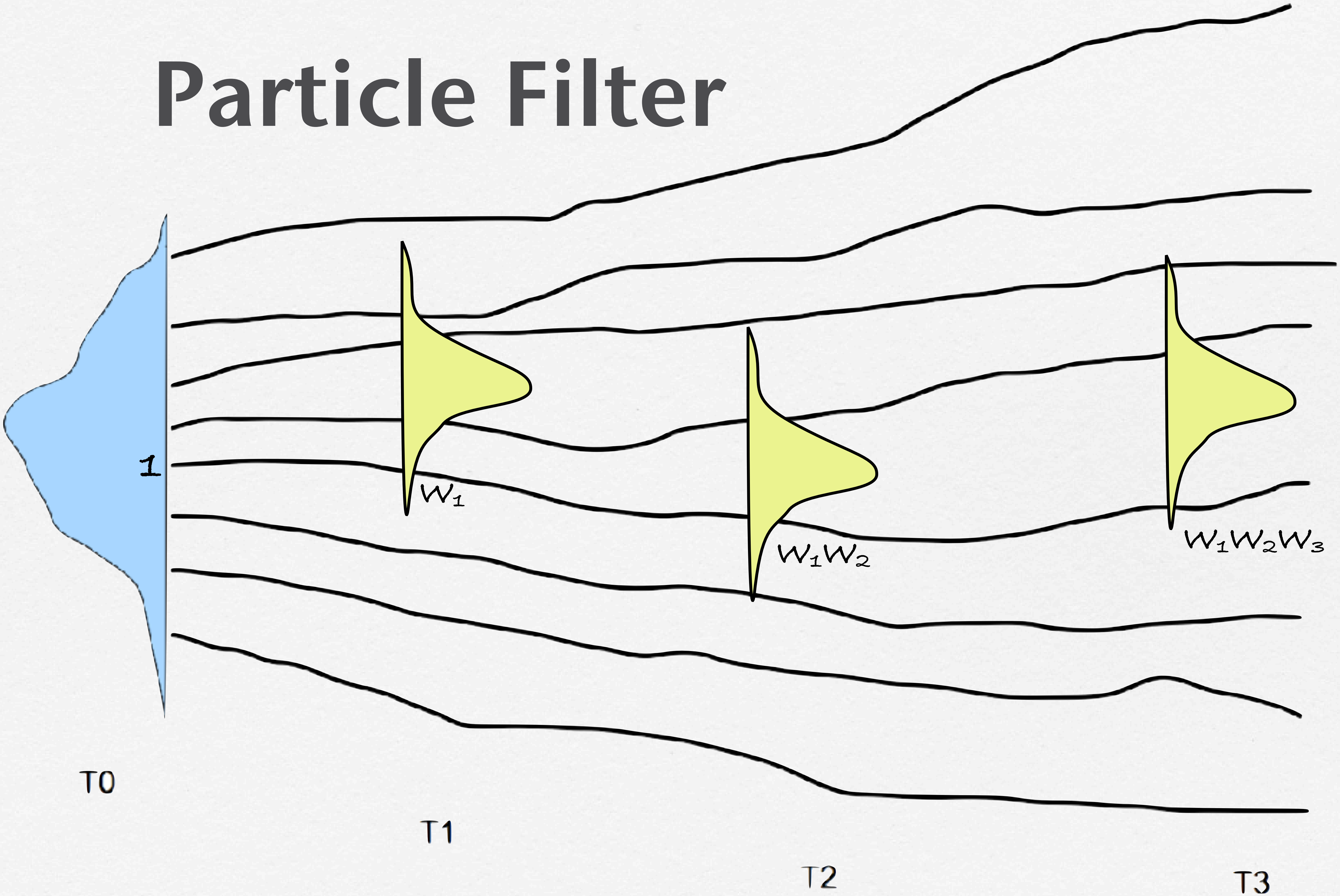
Random Walk State Space

$$P(X_t, \tau_{obs}, \tau_{proc} \mid Y_t) \propto N(Y_t \mid X_t, \tau_{obs}) \times N(X_t \mid X_{t-1}, \tau_{proc})\, \Gamma(\tau_{proc})\, \Gamma(\tau_{obs})$$

What if we forecast with a large Monte Carlo sample?

- Can eliminate distributional assumptions!
- Can eliminate Normal x Normal Analysis
- How to do Analysis step when prior is a sample, not an equation?

Particle Filter

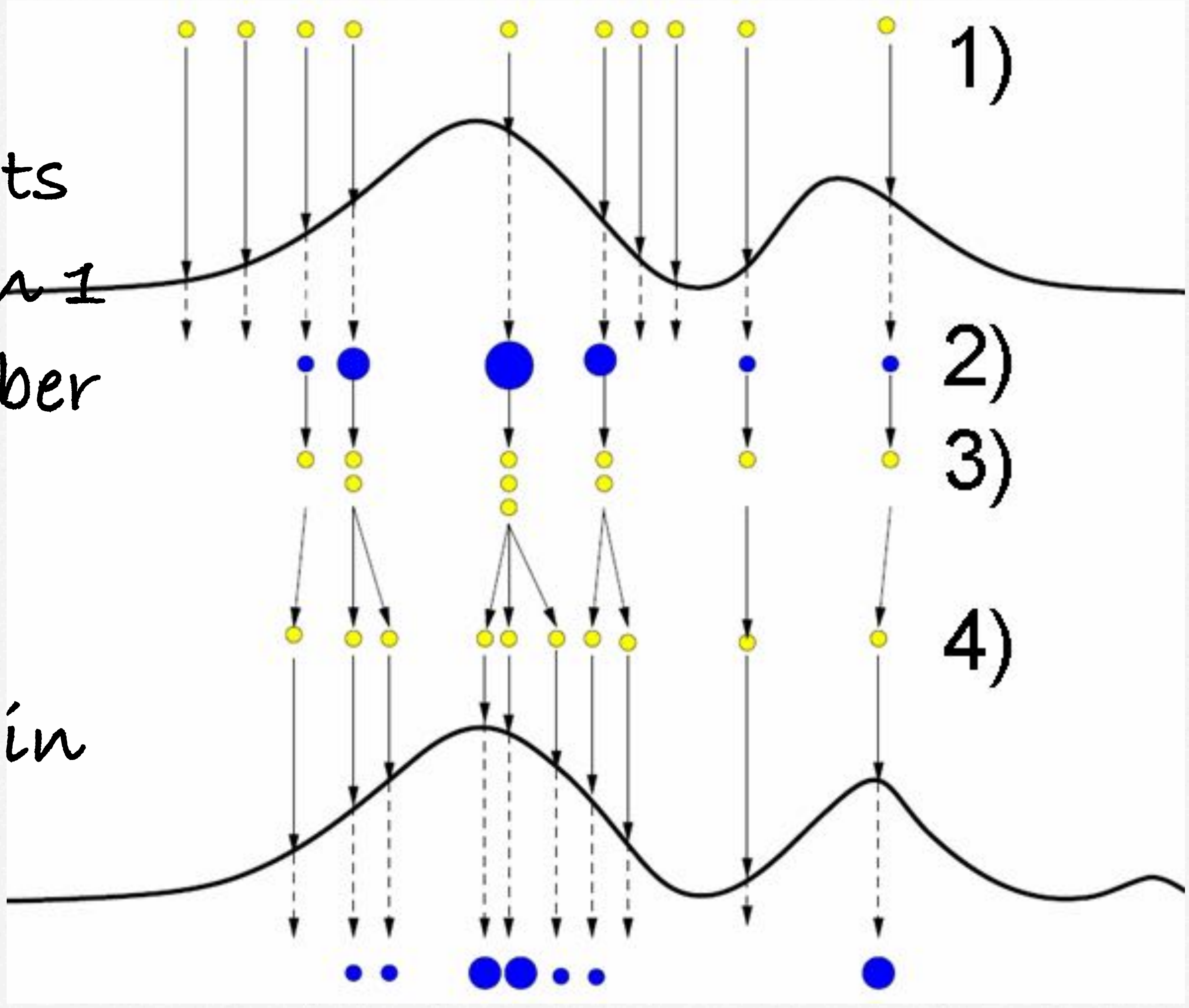


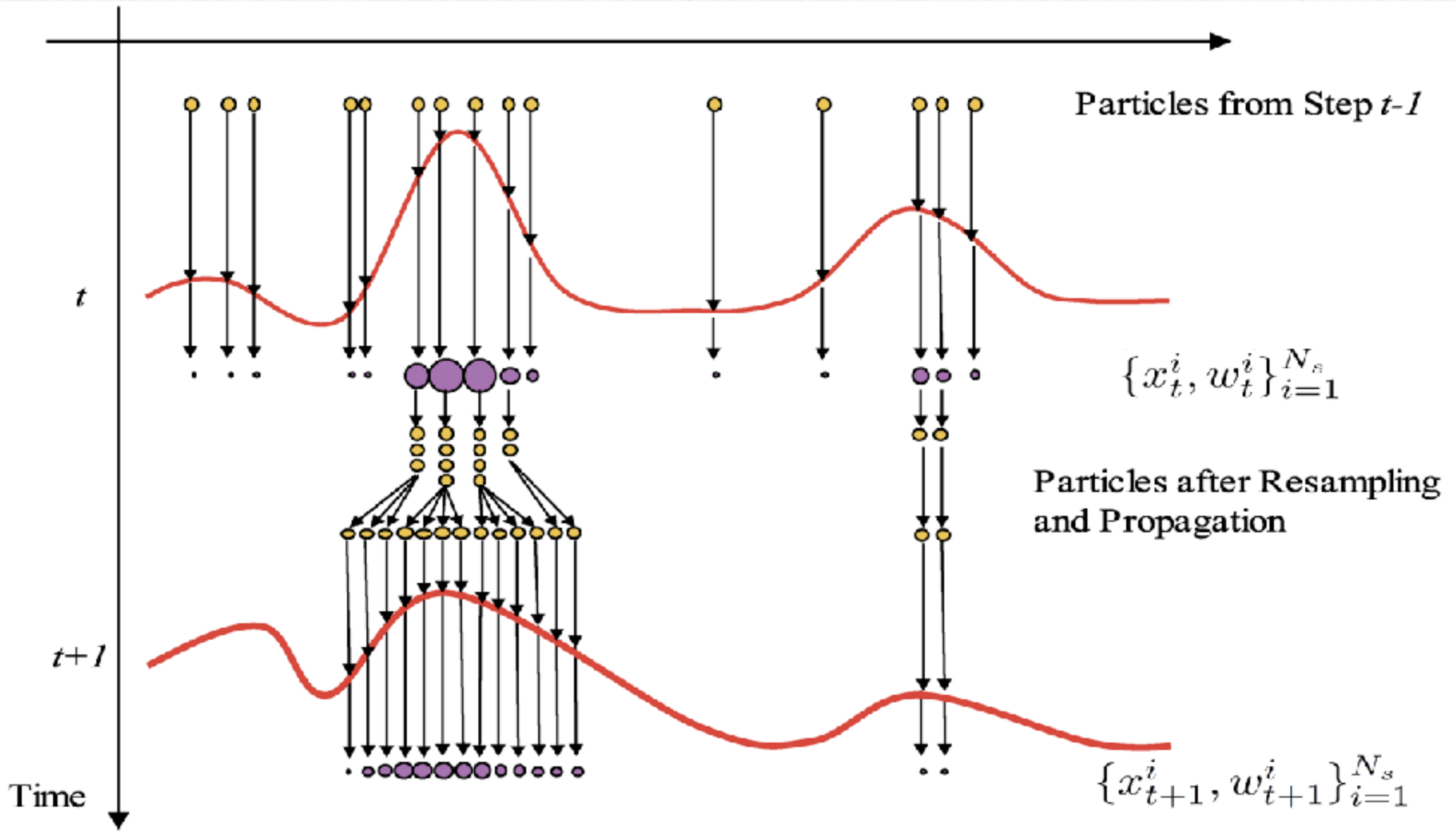
Particle Filter

- Weights provided by the likelihood
  - posterior ∝ likelihood X prior
- Estimates based on weighted mean, variance, CI, etc.
- a.k.a. Sequential Monte Carlo

Resampling PF

- Problem: weights can converge on 1 ensemble member
- Solution: resampling & split to maintain a distribution





Particles from Step t-1
t
$\{x_t^i, w_t^i\}_{i=1}^{N_s}$
Particles after Resampling and Propagation
$t+1$
$\{x_{t+1}^i, w_{t+1}^i\}_{i=1}^{N_s}$
Time

When to resample?

- Too often: loose particles through drift
- Not enough: converges (degeneracy), poor distribution
- Typically resample when effective sample size, $1/\mathrm{sum}(W^2)$, drops below some threshold (e.g. $N/2$)
- NOTE: At resample, weights reset to 1!!

Particle Filter pro/con

- Con:
  - Computation!
- Pros:
  - Simple to implement
  - General, Flexible
  - Can evaluate all params
  - Parallelizable

Kernel Smoothing

- Parameters lack process error, subject to degeneracy
- Can be resampled from kernel smoother = continuous approx of joint PDF
- Req choice of smoothing/bandwidth
- Even better if M-H accept/reject proposed Moves
- Global, Gaussian smoothing

$$\theta_i^* = \bar{\theta} + h(\theta_i - \bar{\theta}) + \epsilon_i\sqrt{1-h^2}$$

$$e_i \sim MVN(0,\bar{\Sigma})$$

h=1 no smoothing
h=0 redraw iid

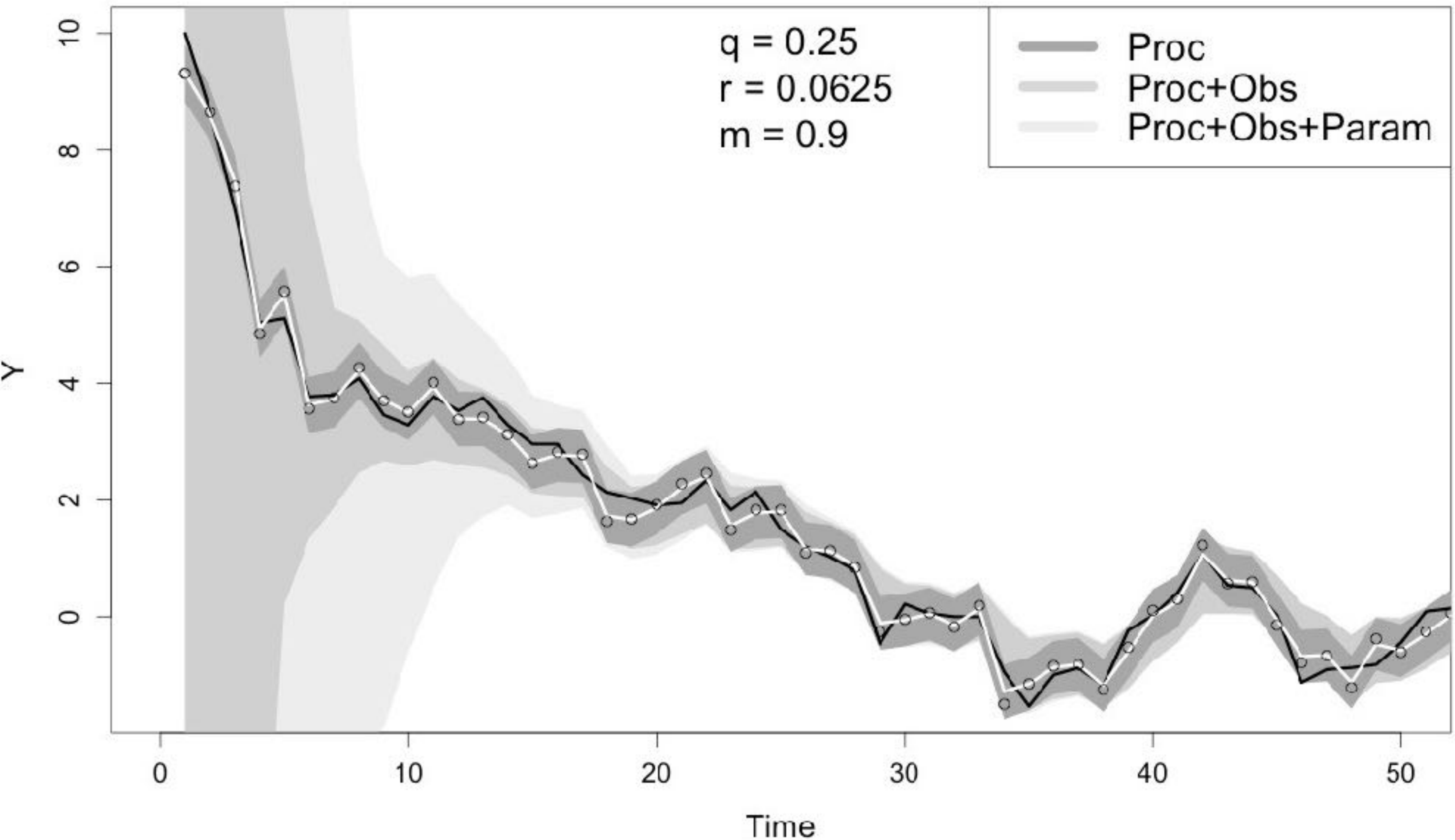
UNCERTAINTY PROPAGATION
APPLIED IN THE FORECAST STEP

		Output	
		Distribution	Moments
Approach	Analytic	Variable Transform	Analytical Moments KF Taylor Series EKF
	Numeric	Monte Carlo PF	Ensemble EnKF

WHAT ABOUT THE <u>ANALYSIS</u> STEP?

What about MCMC?

- Option 1: Refit full State-Space Model
- Option 2: Just update forecast from State-Space
  - A: treat priors (forecast & params)as samples -> PF
  - B: approximate priors w/ dist'n


q = 0.25
r = 0.0625
m = 0.9
Proc
Proc+Obs
Proc+Obs+Param
Y
Time
0
2
4
6
8
10
0
10
20
30
40
50

```
LinStateSpace <- "
model{
  X0 ~ dnorm(mu0,pa0)
  q  ~ dgamma(aq,bq)
  r  ~ dgamma(ar,br)
  m  ~ dnorm(m0,s)

  for(i in 1:nt){
    Y[i] ~ dnorm(X[i],r)
    X[i] ~ dnorm(mu[i],q)

  }
  mu[1] <- X0
  for(i in 2:nt){
    mu[i] <- m*X[i-1]
  }

}"
```

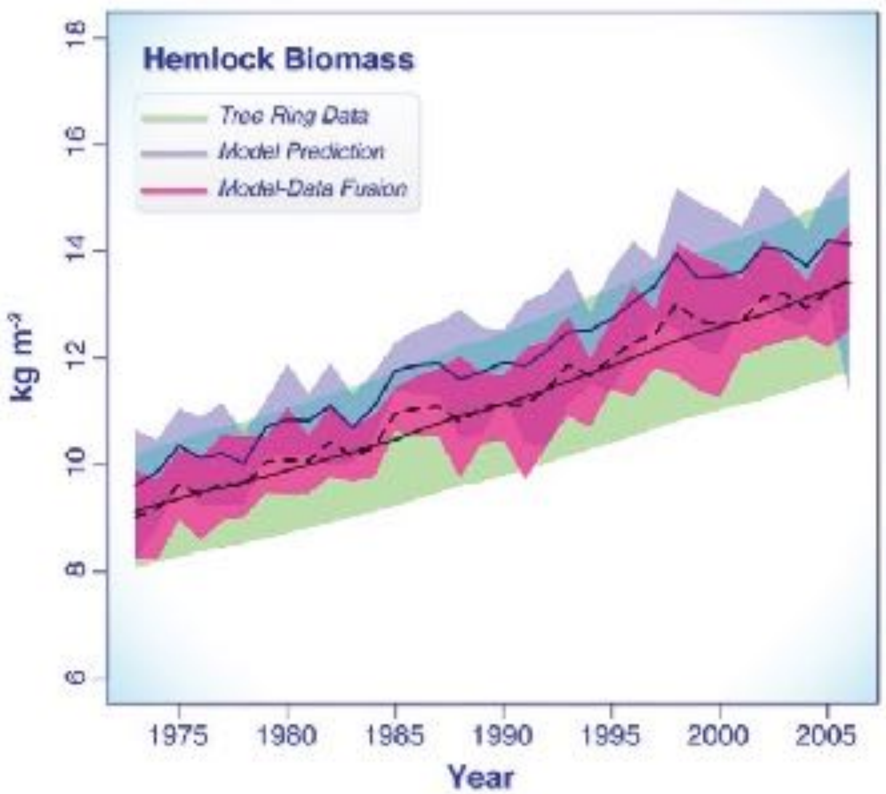
for (i in 1:nt)

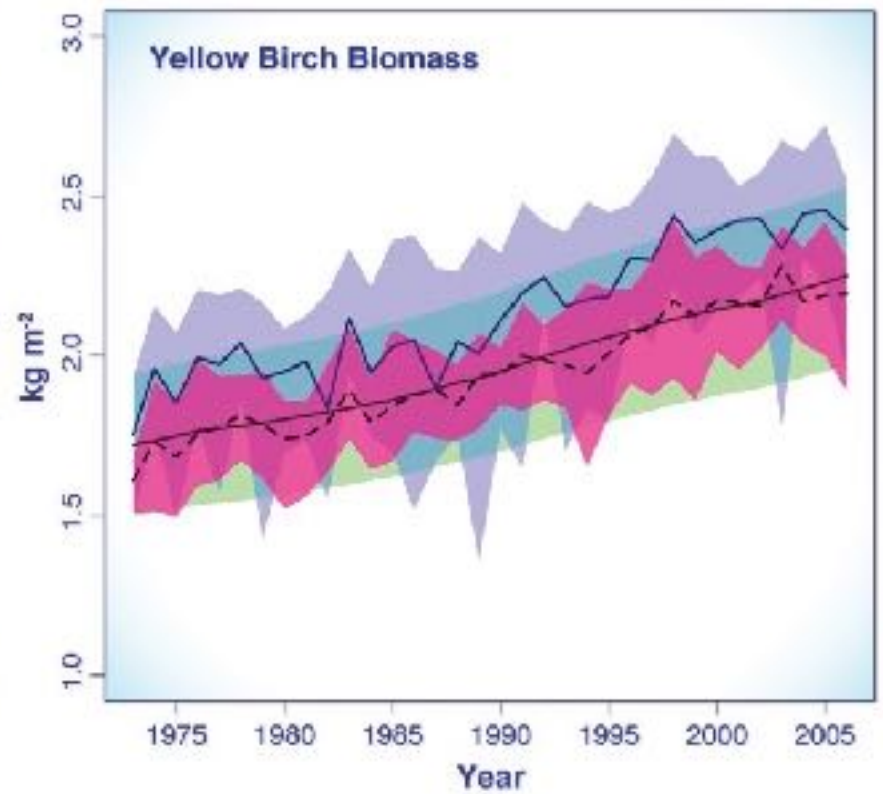
```
LinFilterParam <- "
model{
  ## Priors
  X.ic ~ dnorm(mu0,pa0)
  q  ~ dgamma(aq,bq)
  r  ~ dgamma(ar,br)
  m ~ dnorm(mu.m,tau.m)

  ## Forecast
  mu <- m*X.ic
  X  ~ dnorm(mu,q)

  ## Analysis
  Y  ~ dnorm(X,r)
}"
```

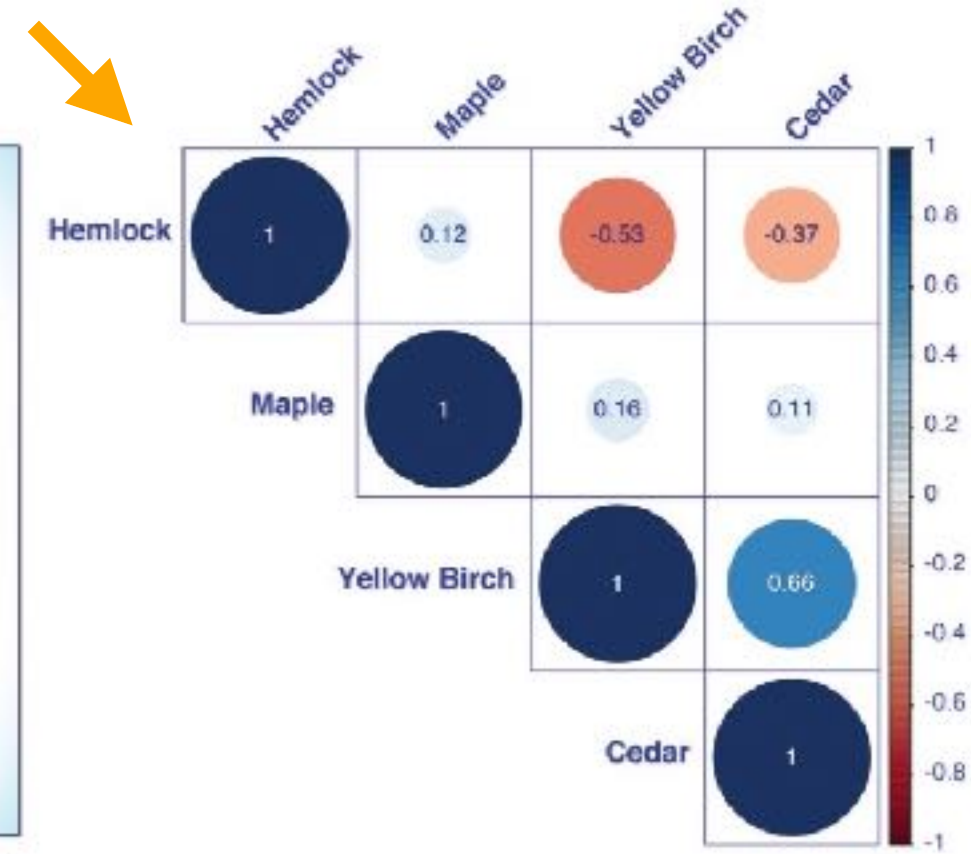
priors set based on previous posteriors

GENERALIZED ENSEMBLE FILTER

Estimated Process Error



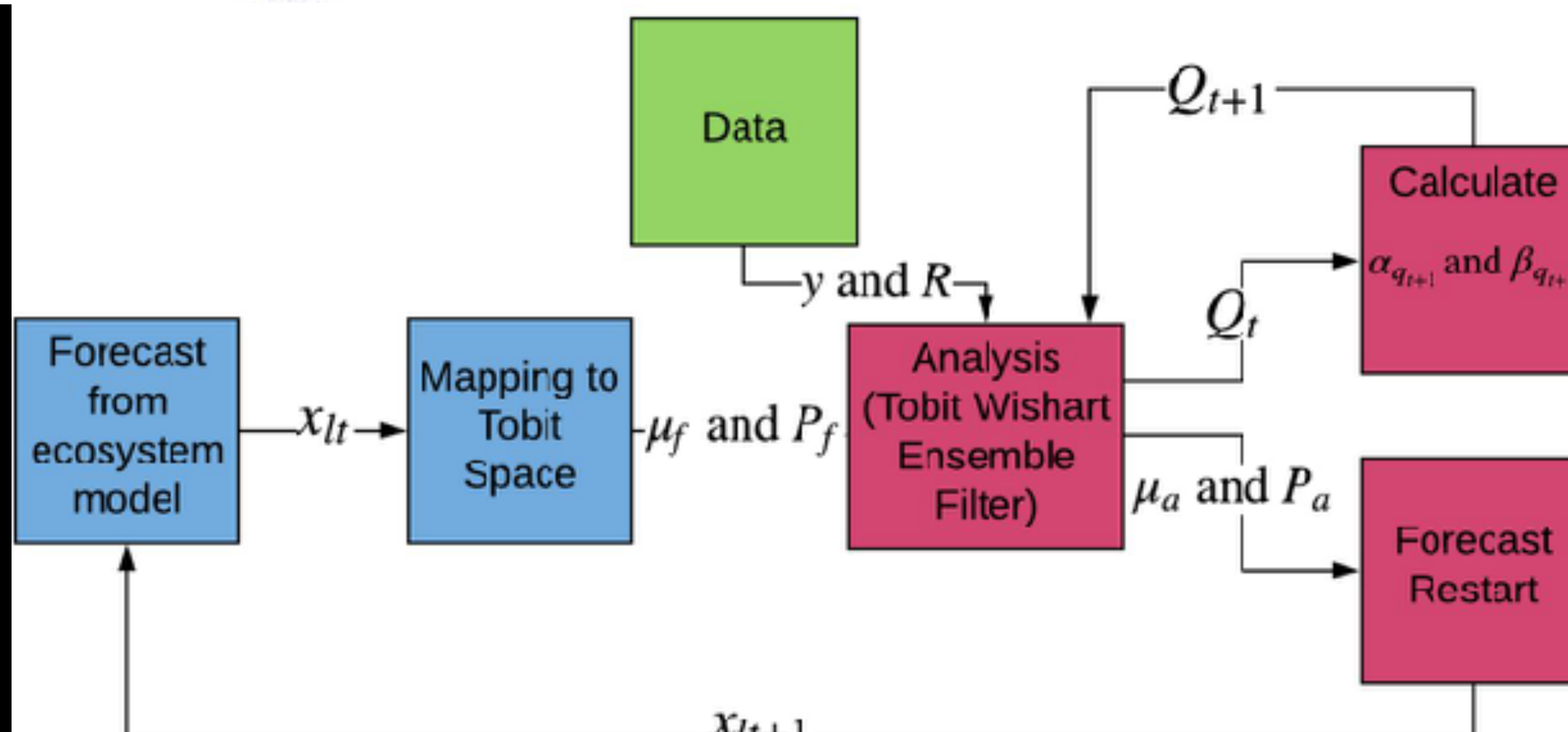




Multivariate Tobit

- Range restrictions
- Zero inflated

Raiho et al 2020 BioRxiv



Take Homes

- Iterative Forecast-Analysis Cycle (Data Assimilation) allow us to continually confront models with data
- All DA variants are forward-only special cases of State Space model
- Forecast Step: Standard DA methods map to uncertainty propagation axes
- Analysis step: Do not feel constrained by Kalman, Take Assumptions into your own hands, MCMC often viable