CAHMDA/DAFOH Workshop
Sept 8-12, 2014
Austin, TX

Combined Data Assimilation and Multi-modeling in Seasonal Hydrologic Forecasting:
A More Complete Characterization of Uncertainty

Hamid Moradkhani

Remote Sensing and Water Resources Lab

Civil and Environmental Engineering

Portland State University

- A general lack of information and skillful modeling frameworks leads to forecast products that do not have sufficient ability to be relied upon on an entirely deterministic manner.

- Uncertainty is pervasive throughout hydrologic forecasting.

Difficulties in Hydrologic Predictions

- Space-time variability of climatic inputs
- Heterogeneity of the land surface condition: vegetation, land use, soils, snow extent, etc.
- Selection of one or multiple plausible model/s that can provide reliable and skillful prediction under all circumstances

Global Optimization or Ensemble Inference / Data Assimilation?

Parameter Estimation: Improve estimates of a set of poorly known model parameters leading to a model solution that is close to the measurements.

- All errors in the model are associated with uncertainties in the selected model parameters.
- The model initial conditions, boundary conditions, and the model structure are all exactly known

State Estimation using Data Assimilation: Defined as finding the estimate of the model state that in some weighted measure best fits the observation, the initial and boundary conditions.

Combined Parameter and State Estimation: An improved state estimate and a set of improved model parameters are searched for simultaneously.

Limitations of Most Calibration Techniques

- In the case of insufficient availability of historical data (e.g. ungauged or recently gauged basins) batch calibration cannot be properly applied

- Incapability in investigating the possible temporal variations of model parameters

- Non-uniqueness of solution (Ill-posed inverse problem)

- Mostly provide just a single solution ignoring the uncertainty sources

Some Benefits of Data Assimilation

- Provides a framework for quantifying uncertainty
- Can be used to calibrate models (dual state-parameter estimation framework)
- Can be used to estimate the uncertainty in system states for initializing the forecasts
- Can be used to reduce model uncertainty

Quantifying Uncertainties in Operational Settings

- **Problem:** Current operational streamflow forecasting system does not yet account for all sources of uncertainty
- **Goal:** Move towards a more complete accounting of all sources of uncertainty in forecasting system
  - Meteorological Forcing
  - Model states/parameters (e.g., Moradkhani et al., 2005; AWR; 2012, WRR)
  - Initial Land Surface Condition (DeChant and Moradkhani, 2011, HESS; 2014, JOH)
  - Hydrologic model structure (Parrish et al., 2012, WRR)

Bayesian Inference

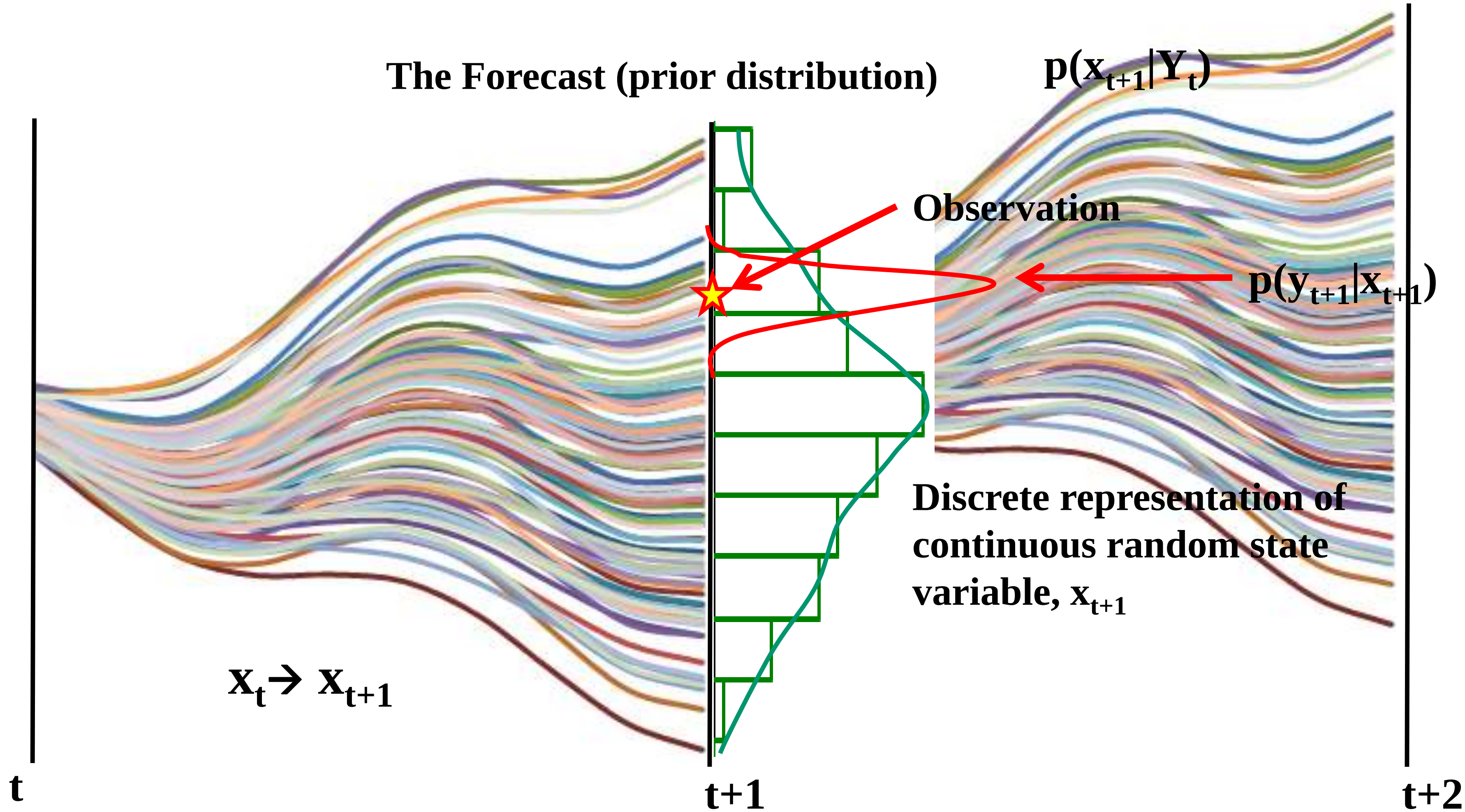
The **Prior Probability describes what you first knew.**
Multiply this by a **term that describes the effect of your new information**, and the result is what you know **after you have taken into account your new information.**

$$\underbrace{P(X,\Theta \mid Data)}_{\text{Posterior Probability}} = \underbrace{P(X,\Theta)}_{\text{Prior Probability}} \frac{\overbrace{P(Data \mid X,\Theta)}^{\text{Likelihood}}}{\underbrace{P(Data)}_{\text{Evidence}}}$$

Ensemble Data Assimilation

- Utiliz ons to reduce the uncertainty in our proba hrough sequential Bayes theorem

$$p(x_{t+1} \mid Y_{t+1}) = \frac{p(y_{t+1} \mid x_{t+1})\, p(x_{t+1} \mid Y_t)}{\int_{x_{t+1}} p(y_{t+1} \mid x_{t+1})\, p(x_{t+1} \mid Y_t)\, dx_{t+1}}$$



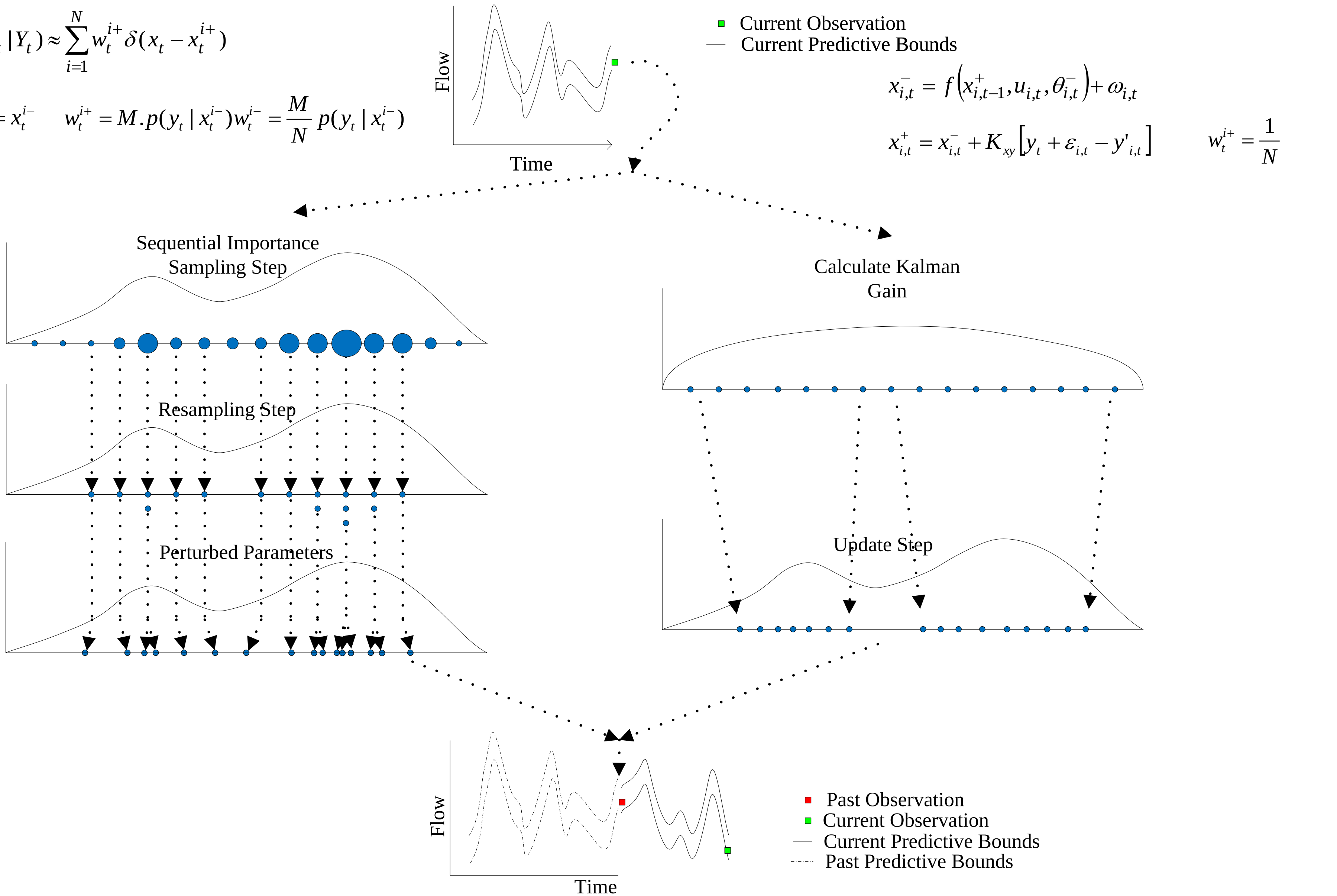
Data Assimilation by the Ensemble Kalman Filter and Particle Filter

$$p(x_t \mid Y_t) \approx \sum_{i=1}^{N} w_t^{i+} \delta(x_t - x_t^{i+})$$

$$x_t^{i+} = x_t^{i-} \qquad w_t^{i+} = M.p(y_t \mid x_t^{i-}) w_t^{i-} = \frac{M}{N} p(y_t \mid x_t^{i-})$$

$$x_{i,t}^{-} = f\left(x_{i,t-1}^{+}, u_{i,t}, \theta_{i,t}^{-}\right) + \omega_{i,t}$$

$$x_{i,t}^{+} = x_{i,t}^{-} + K_{xy}\left[y_t + \varepsilon_{i,t} - y'_{i,t}\right] \qquad w_t^{i+} = \frac{1}{N}$$



Particle Filter

Moradkhani et al. 2005, WRR

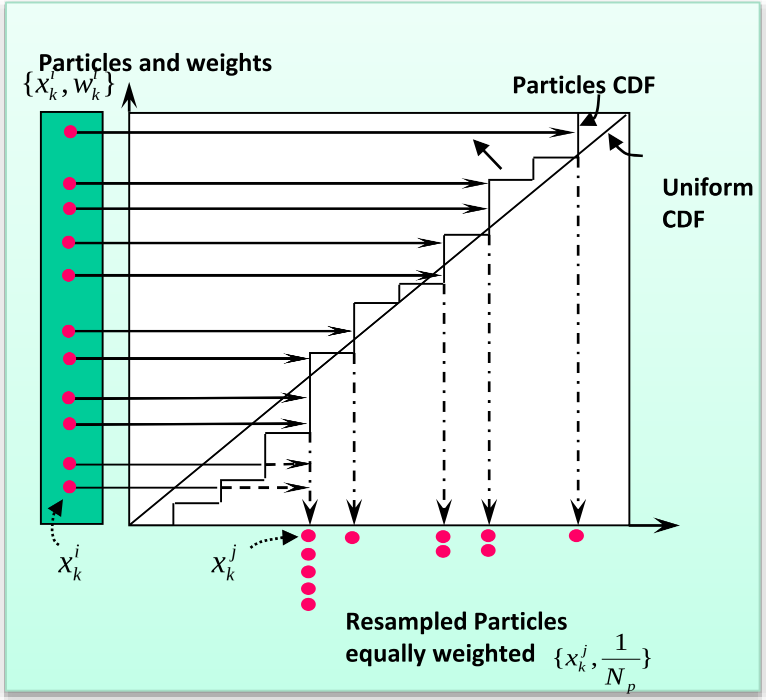
Forecast (Prior) Density

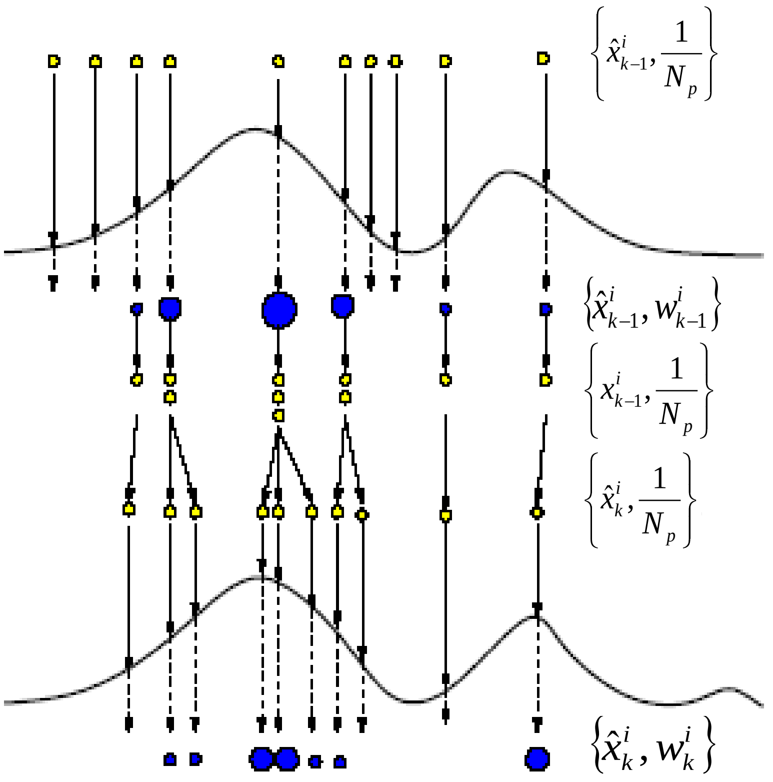
$$p(x_t \mid Y_{t-1}) \approx \sum_{i=1}^{N} w_t^{i-} \delta(x_t - x_t^{i-}) \qquad x_t^{i-} = f(x_{t-1}^{i}, \theta, u_t^{i}) \qquad w_t^{i-} = \frac{1}{N}$$

Analysis (Posterior) Density

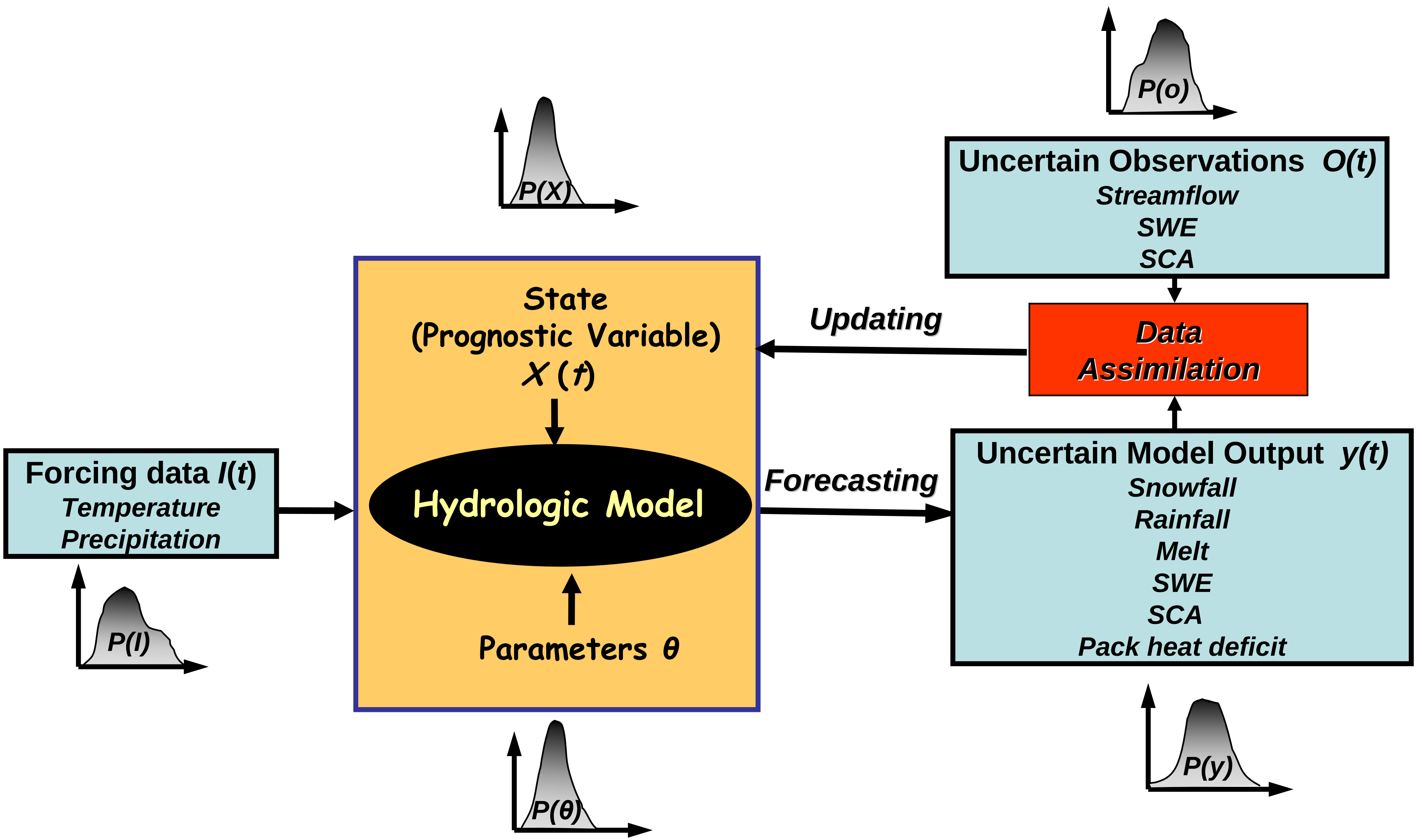
$$p(x_t \mid Y_t) \approx \sum_{i=1}^{N} w_t^{i+} \delta(x_t - x_t^{i+}) \qquad x_t^{i+} = x_t^{i-} \qquad w_t^{i+} = M.p(y_t \mid x_t^{i-}) w_t^{i-} = \frac{M}{N} p(y_t \mid x_t^{i-})$$

Sampling Importance Resampling (SIR)



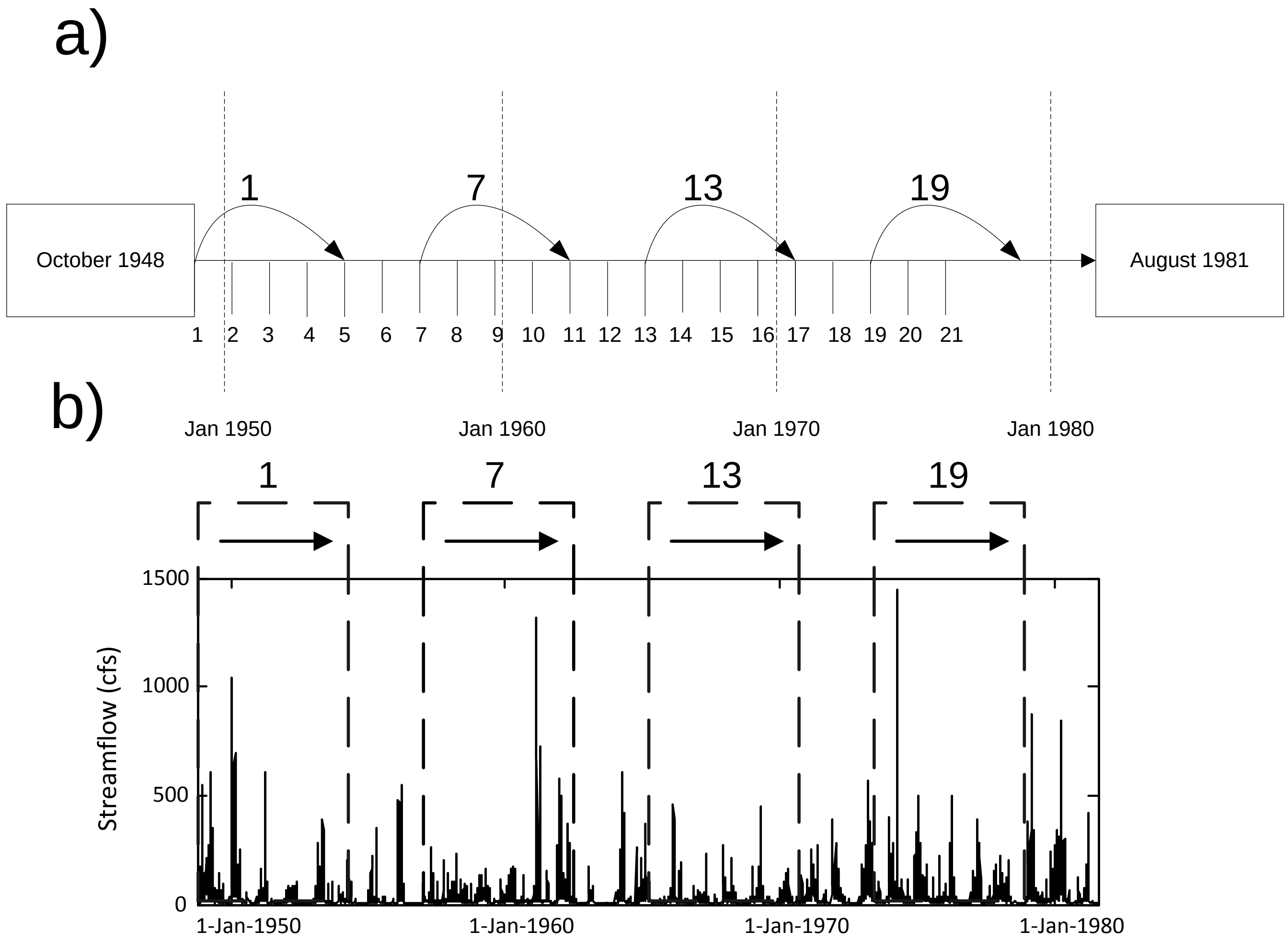


Implementation of Sequential Data Assimilation



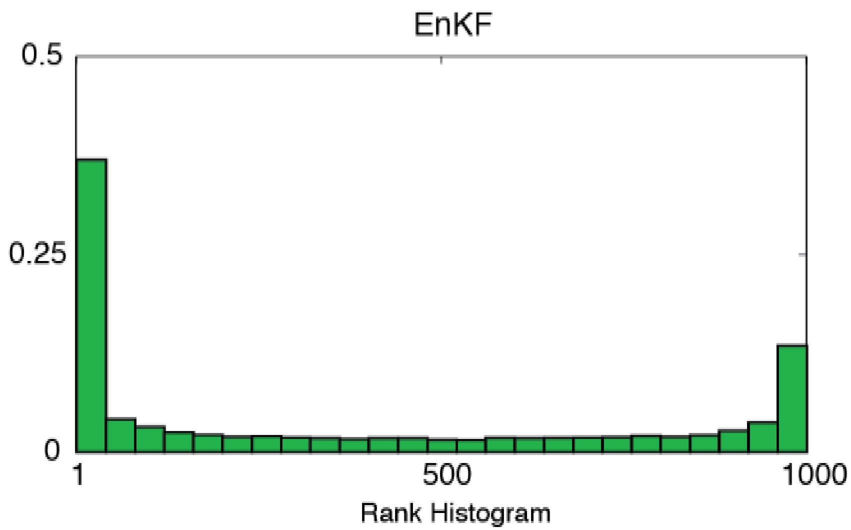
Calibration Replicate Method for DA

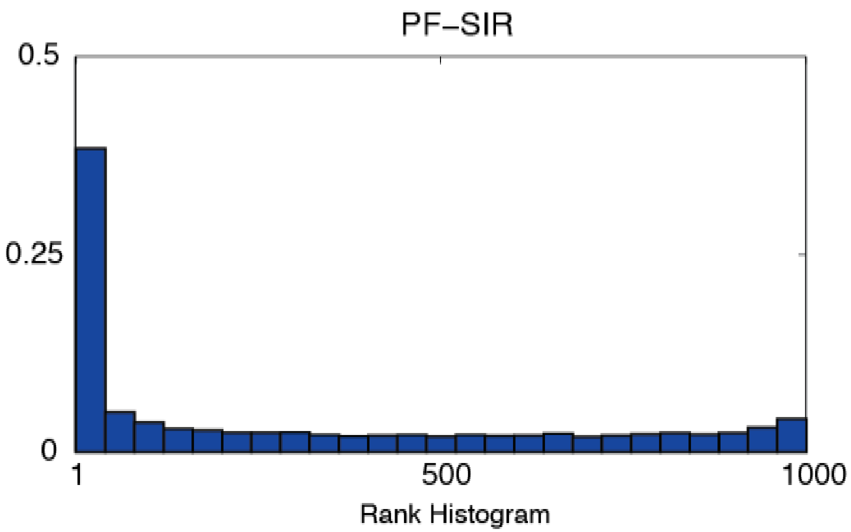
DeChant and Moradkhani (2012), WRR

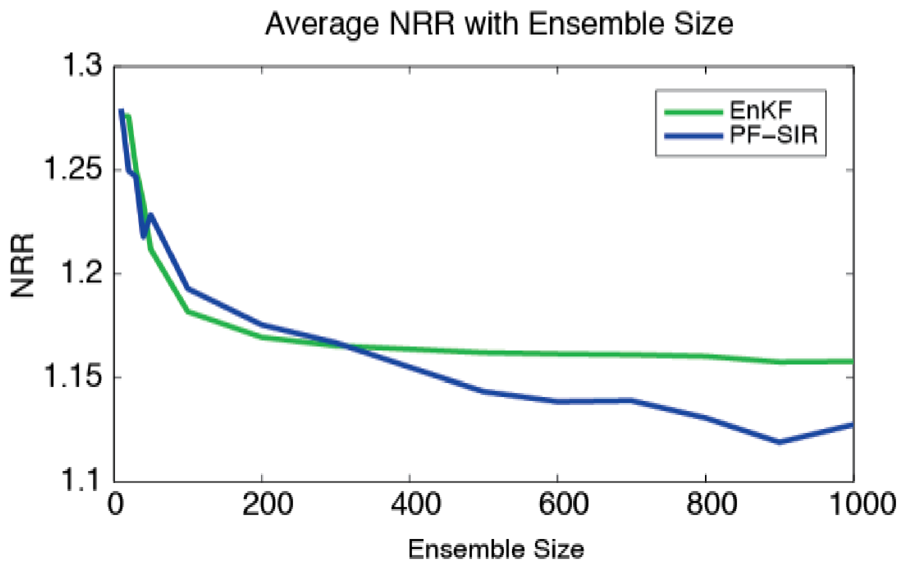


EnKF vs. PF

DeChant and Moradkhani (2012), WRR







EnKF vs. PF

EnKF

- Makes the assumption that errors are Gaussian and states/parameters are linearly correlated with prediction
  - Allows for direct adjustment to states/parameters for characterizing the posterior distribution
  - Not susceptible to sample impoverishment
- Hydrologic modeling problems are typically non-Gaussian
  - Leads to overconfident predictions

PF

- Most general solution available for data assimilation
  - Theoretically more accurate in the non-Gaussian problems
  - Requires extra attention to avoid sample impoverishment
  - Effective adjustment to parameters is a difficult task
- Results are less overconfident than the EnKF
  - Particle filter approaches optimal solution

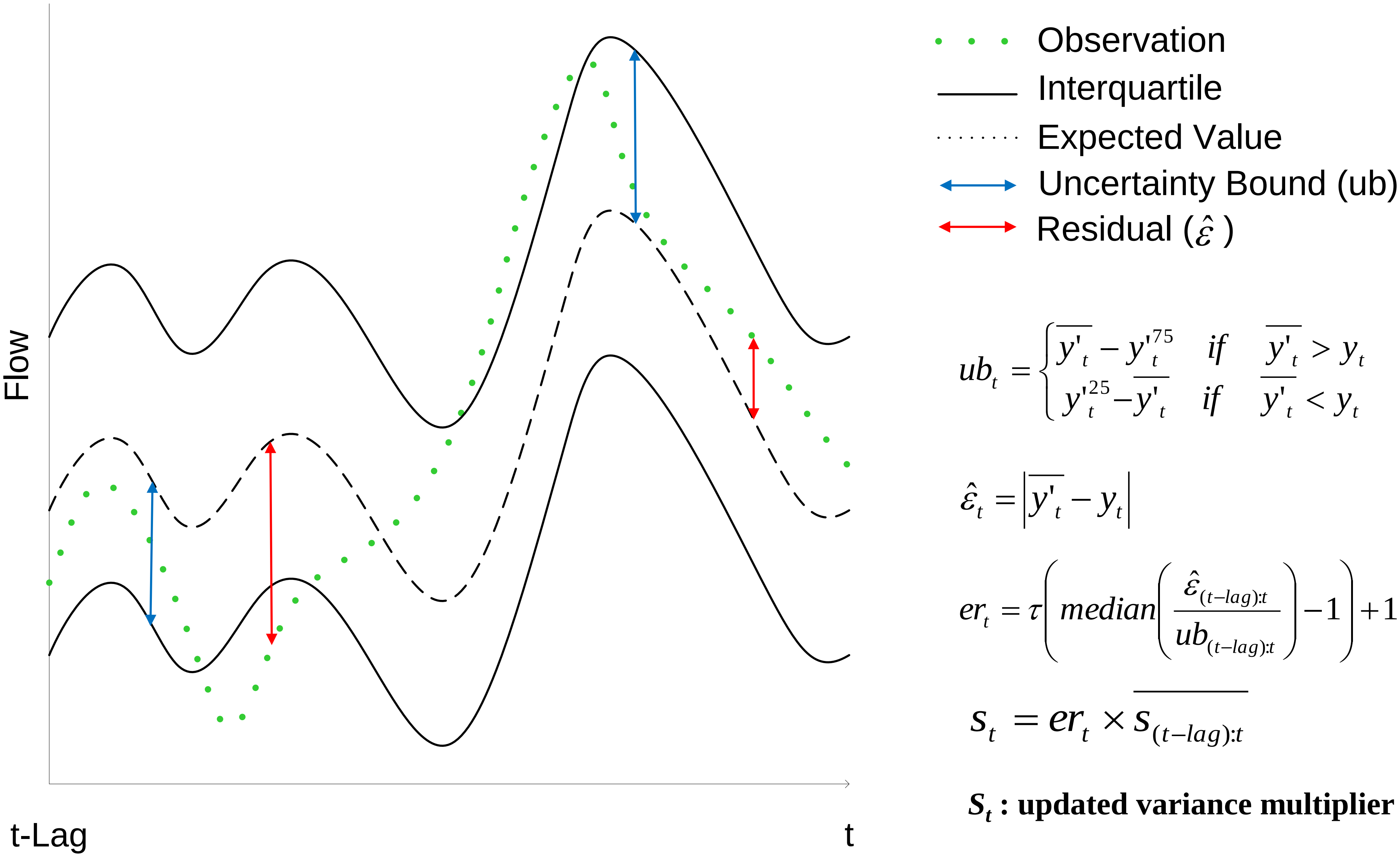
What's New?

- PF is more reliable than EnKF, but still overconfident
  – PF approaches reliable distribution (EnKF does not)
  – Parameter distribution tends to be overconfident
  – Requires large ensemble size to avoid overconfidence
- Need larger parameter moves
  – This is difficult to achieve without moving parameters outside posterior
- Two solutions
  – Automatic tuning of parameter perturbation value
    - Variable Variance Multipliers [*Leisenring and Moradkhani*, 2012; JOH]
  – Ensure parameters remain within posterior
    - Markov Chain Monte Carlo step can reject poor parameter moves [*Moradkhani et al.*, 2012, WRR]

Variable Variance Multipliers

Leisenring and Moradkhani (2012), JOH

- **Automatic tuning of parameter perturbation value**



$$ub_t = \begin{cases} \overline{y'_t} - y'^{75}_t & \text{if } \overline{y'_t} > y_t \\ y'^{25}_t - \overline{y'_t} & \text{if } \overline{y'_t} < y_t \end{cases}$$

$$\hat{\varepsilon}_t = \left| \overline{y'_t} - y_t \right|$$

$$er_t = \tau \left(median\left(\frac{\hat{\varepsilon}_{(t-lag):t}}{ub_{(t-lag):t}} \right) - 1 \right) + 1$$

$$s_t = er_t \times \overline{s_{(t-lag):t}}$$

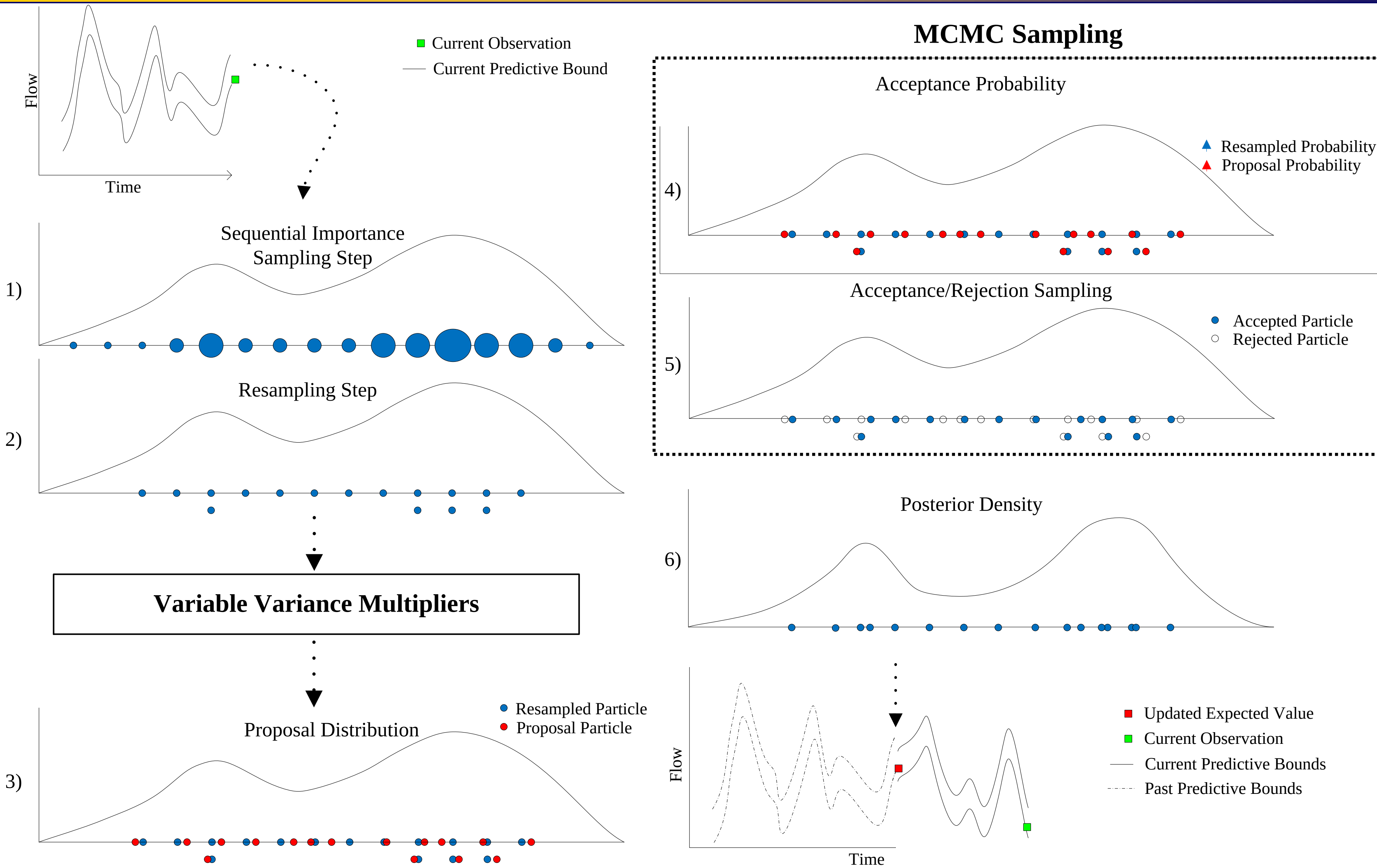
S_t : updated variance multiplier

Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method

WATER RESOURCES RESEARCH, 2012

Combining Particle Filter with MCMC

Moradkhani et al. (2012), WRR



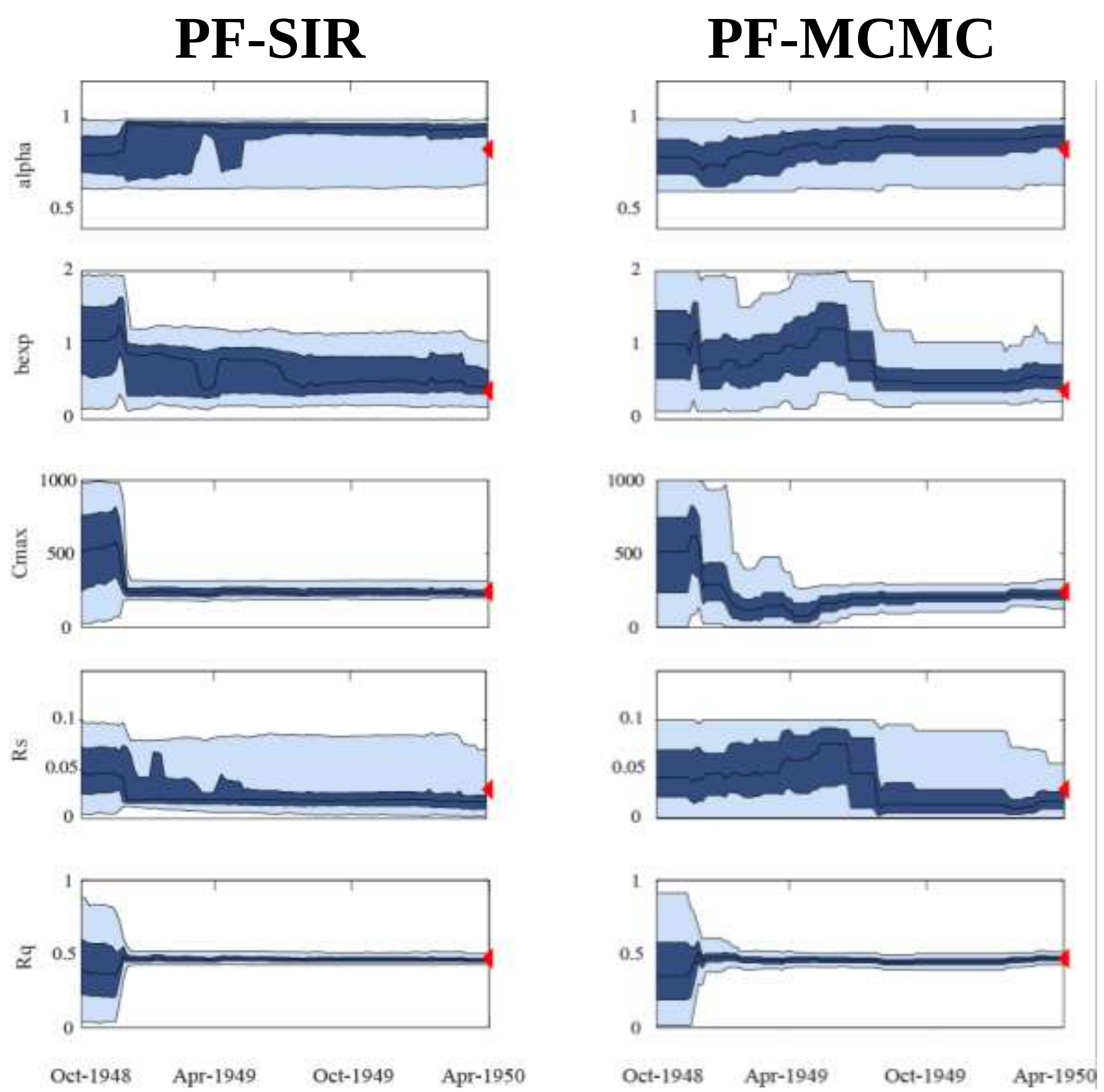
Extra Considerations

- Must create effective proposal distribution
  - $\theta^{p}{}_{i,t} = \theta^{+}{}_{i,t-1} + \varepsilon_{i,t-1} \qquad \varepsilon_{i,t-1} \sim N\left(0, sVar\left(\theta^{-}{}_{i,t-1}\right)\right)$
  - Tune jump rate “s” with VVM methodology to ensure wide enough proposal distribution
- Proposal parameter probability is not readily available
  - Requires assumption about filtering posterior to include all prior information $\left(P\left(\theta_{i,t} | y_{1:t-1}\right)\right)$
  - Here Gaussian assumption for simplicity
  - $\mu_t = \sum_{i=1}^{N} w^{+}{}_{i,t-1}\, \theta^{-}{}_{i,t-1}$
  - $\sigma^2{}_t = \sum_{i=1}^{N} w^{+}{}_{i,t-1}\left(\theta^{-}{}_{i,t-1} - \mu_t\right)^2$
- Posterior is proportional to product of likelihood and prior
  - $P\left(\theta_{i,t} | y_{1:t}\right) \propto L(y'_t - y_t | R_k) * N\left(\theta_{i,t}, \mu_t, \sigma^2{}_t\right)$

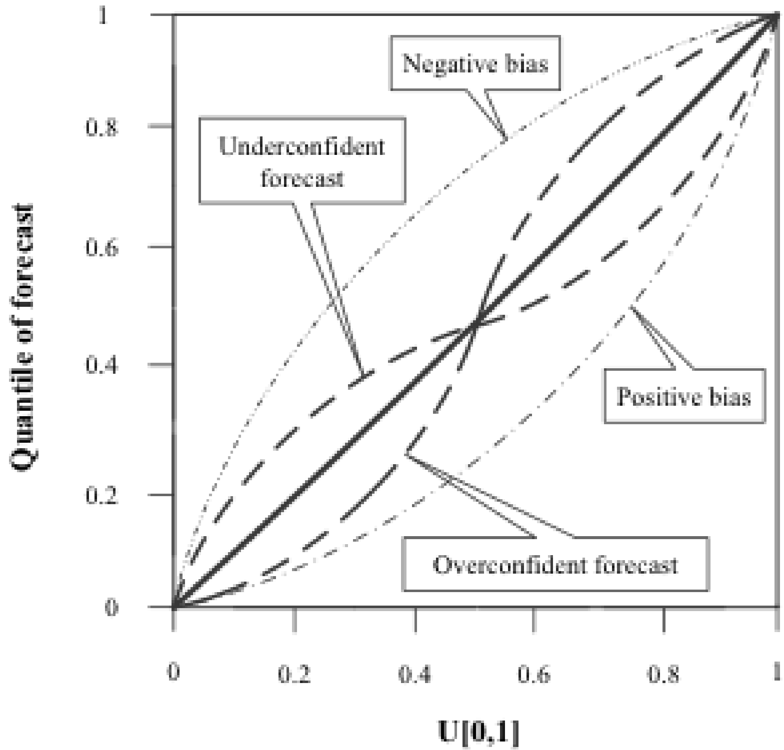
Experiment

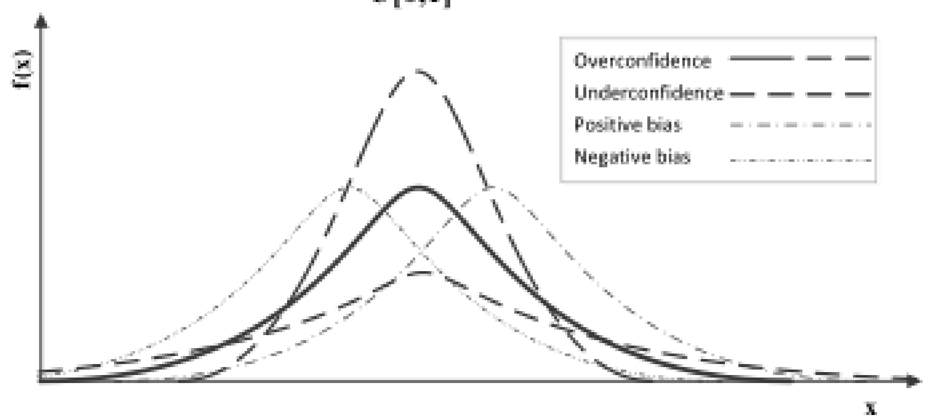
- Perform state-parameter estimation with PF-SIR, and PF-MCMC with VVM
  - Use time-lagged replicates [*DeChant and Moradkhani*, 2012] to increase the number of calibration runs
  - Perform experiments in both a calibration and validation phase
    - Calibration tests streamflow prediction during parameter estimation
    - Validation uses stochastic parameter estimates from calibration
- Perform experiments with HyMod model
  - Data from the leaf river basin
  - Validate with one day ahead prediction of streamflow

Synthetic Experiment

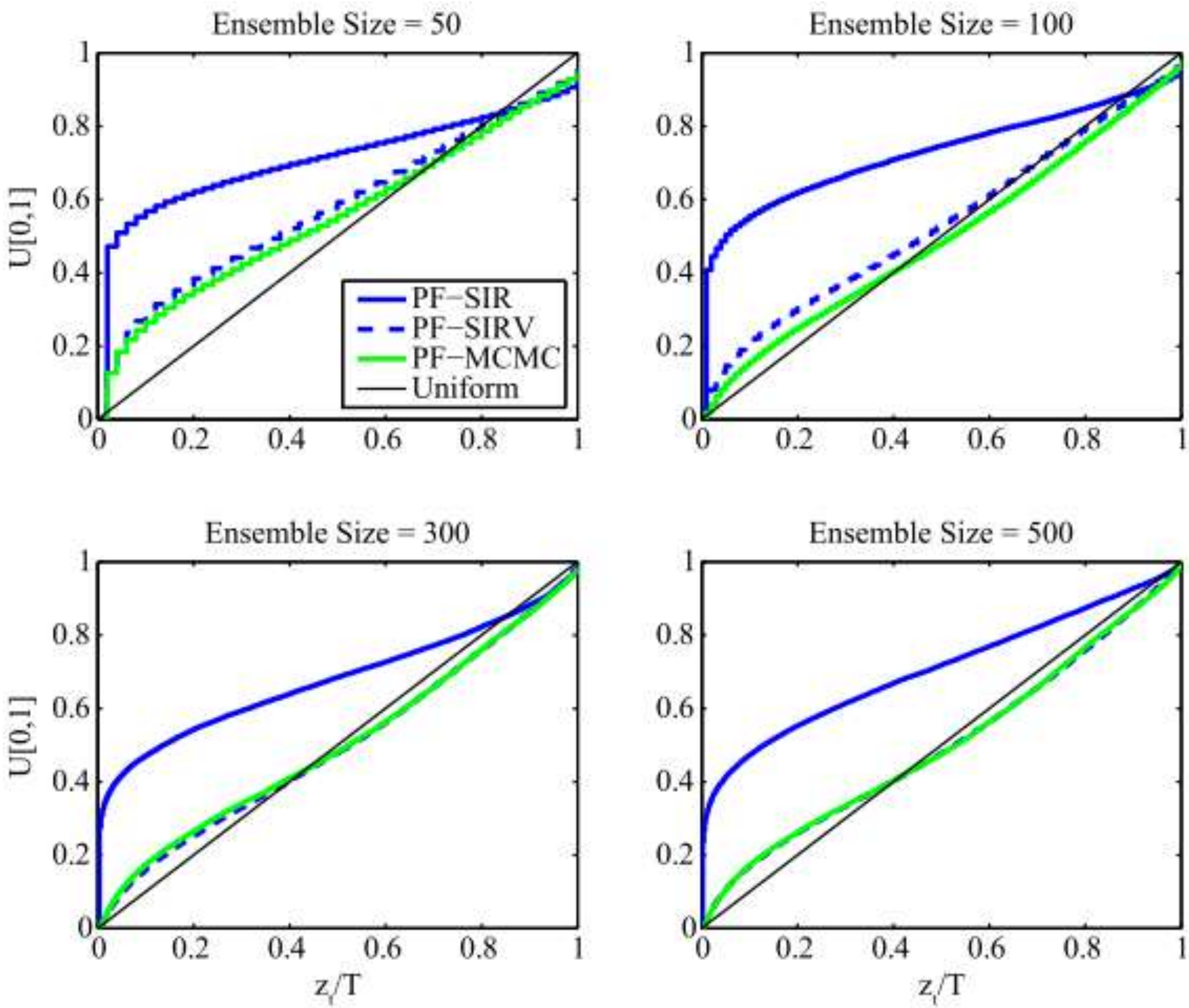


Verification using QQ plot

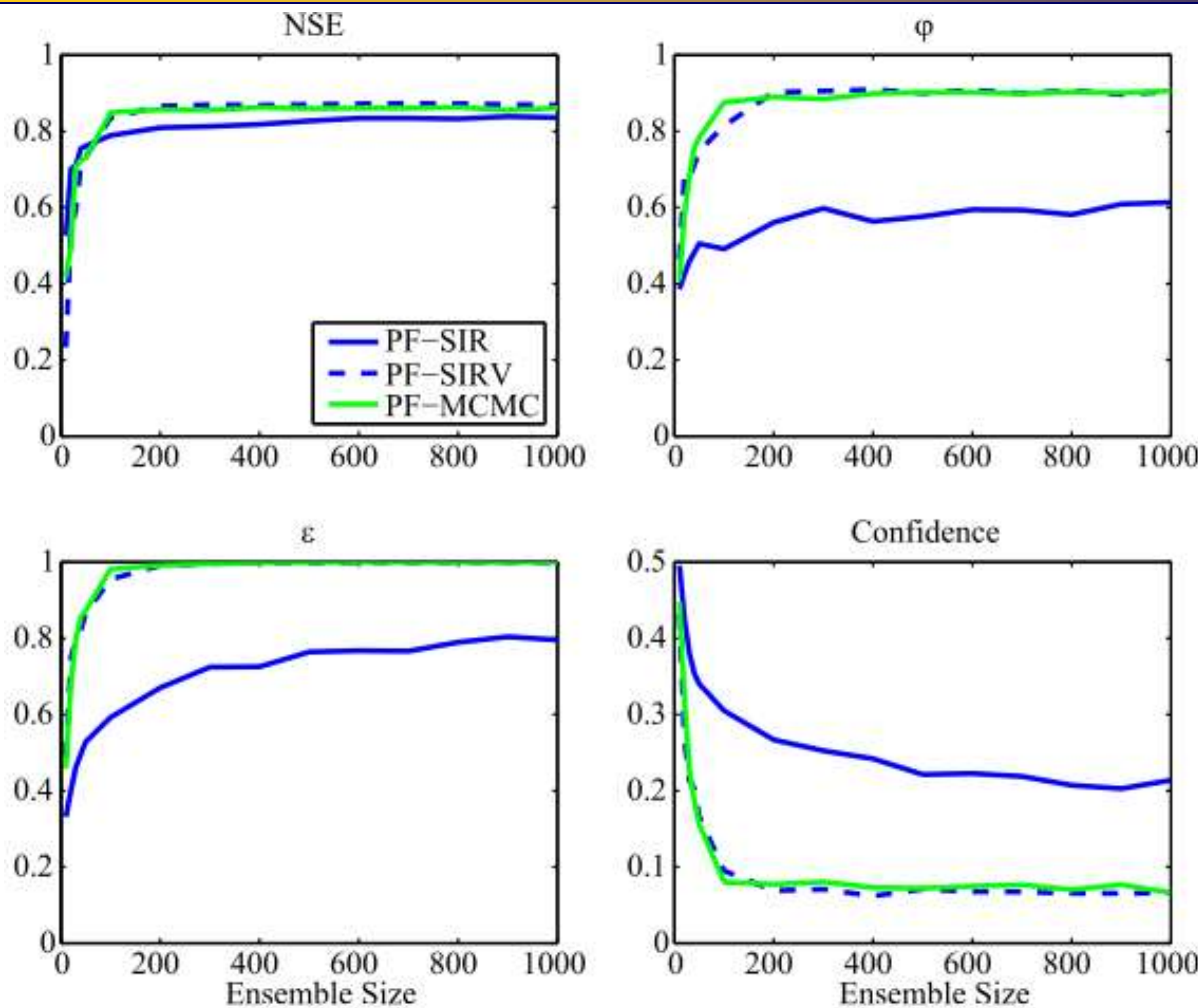




Predictive QQ plots of the three filters for state-parameter estimation

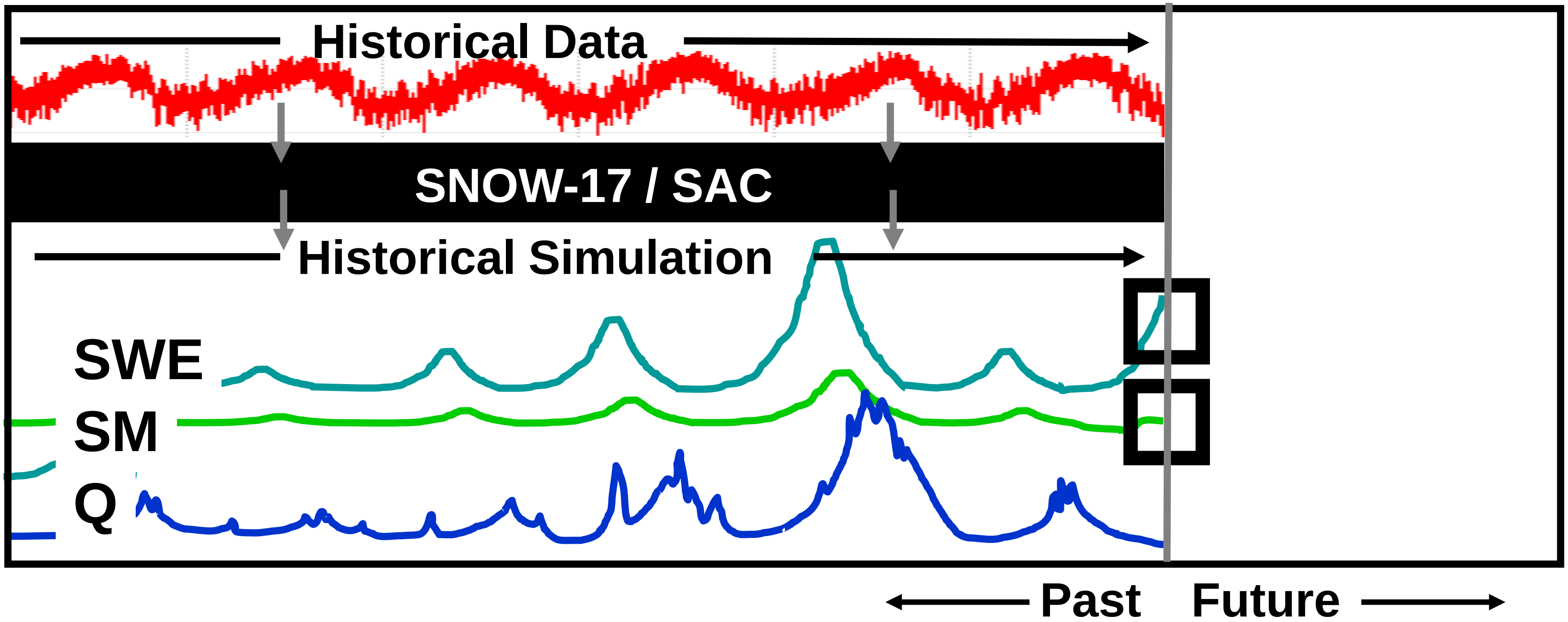


Performance Measures in Real Streamflow Data Assimilation



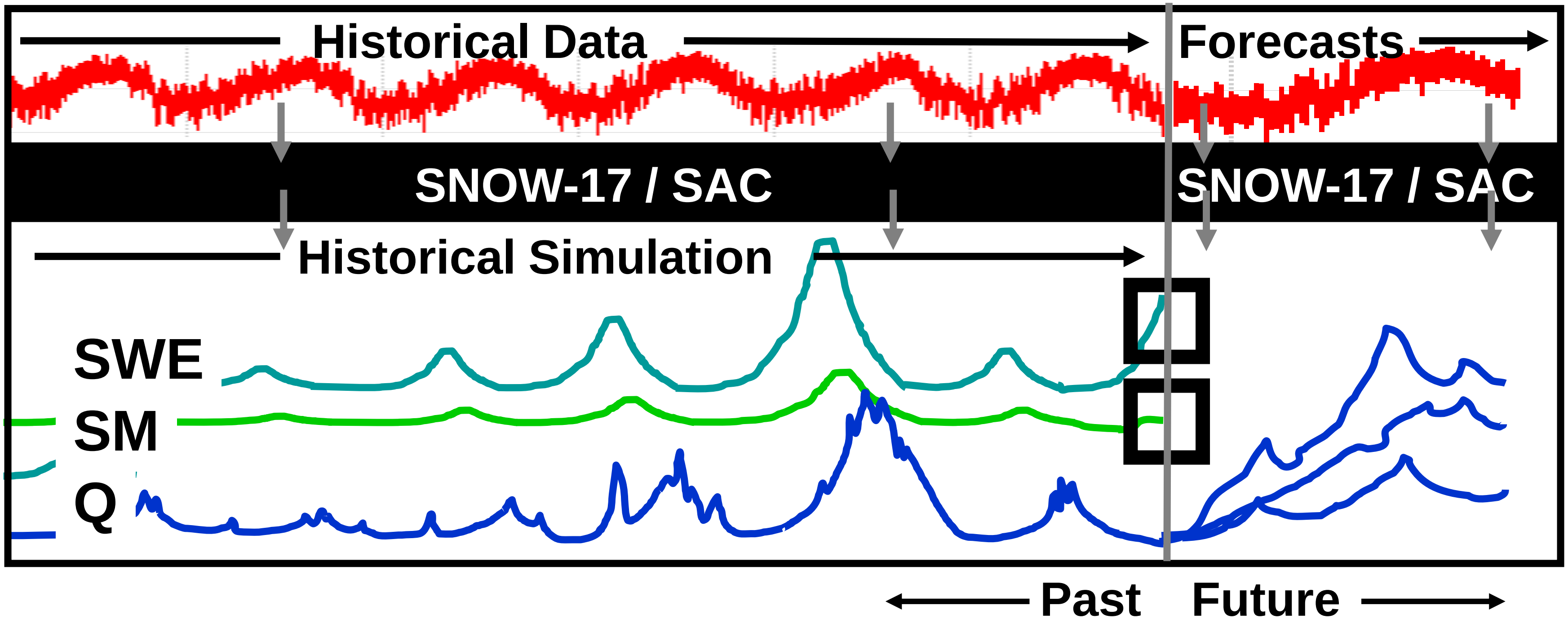
Typical Streamflow Forecasting Method…

Courtesy of M. Clark



1. Run hydrologic model up to the start of the forecast period to estimate basin initial conditions;

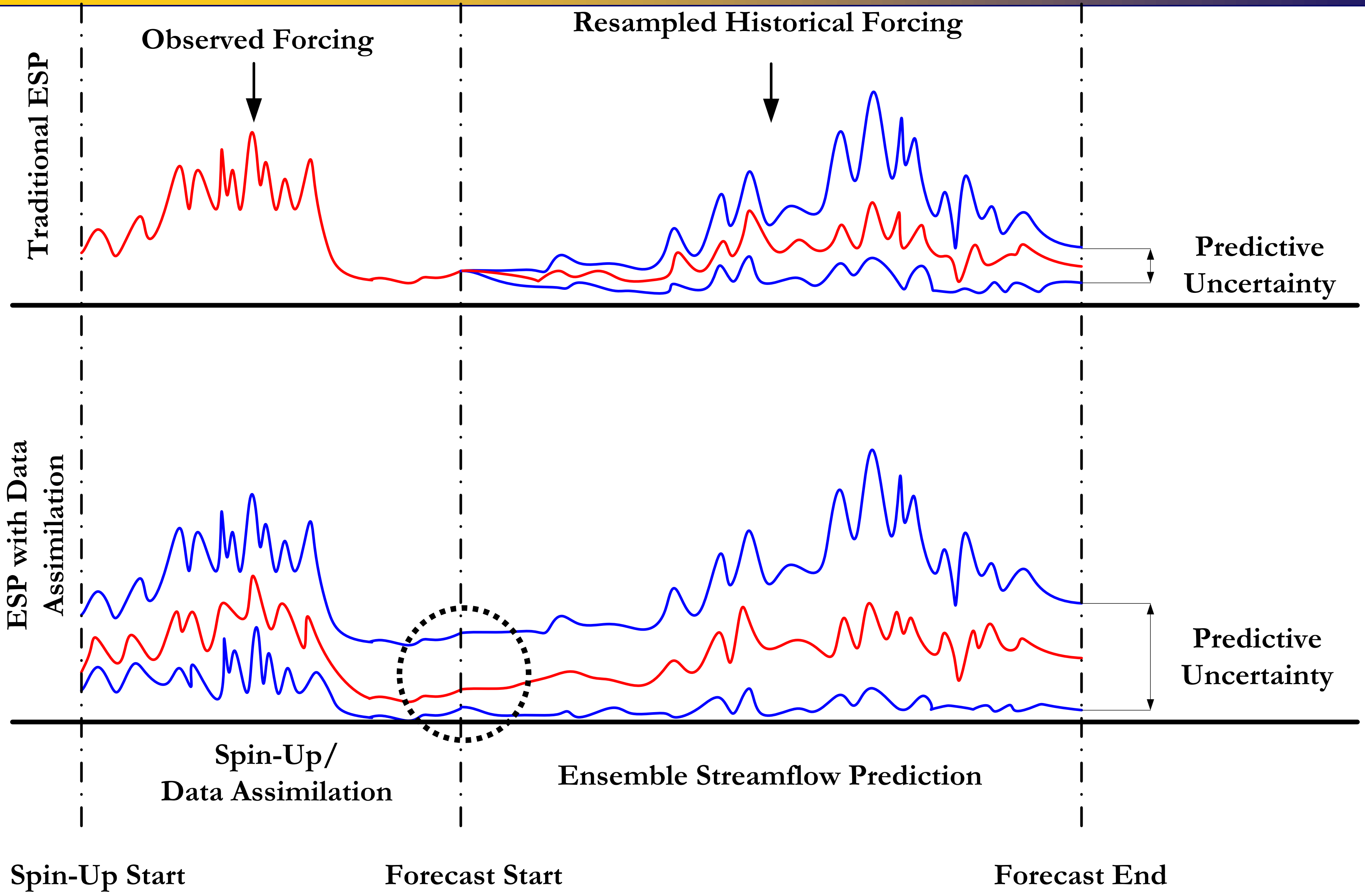
Typical Streamflow Forecasting Method…



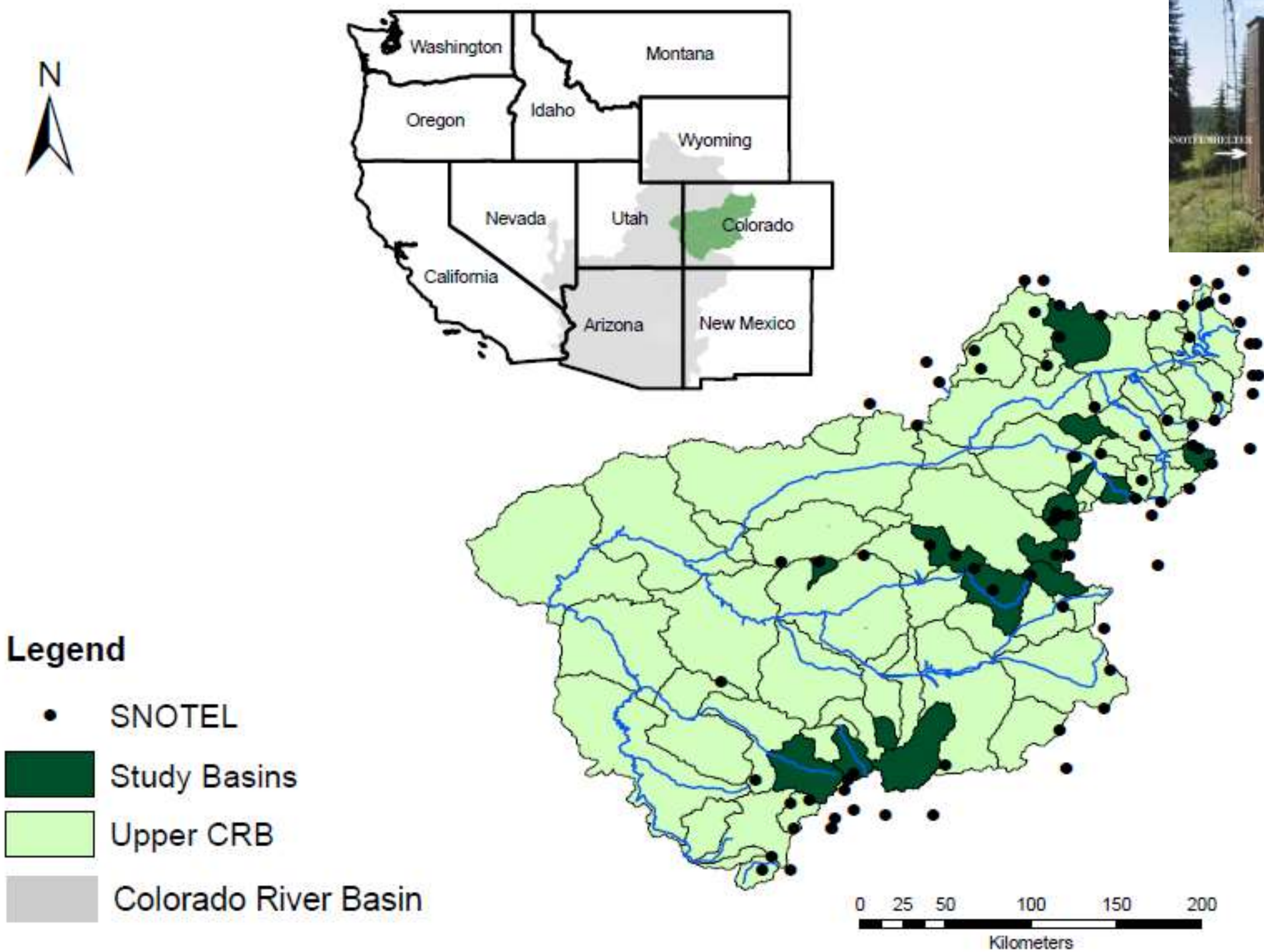
1. Run hydrologic model up to the start of the forecast period to estimate basin initial conditions;
2. Run hydrologic model into the future, using an ensemble of local-scale weather and climate forecasts.

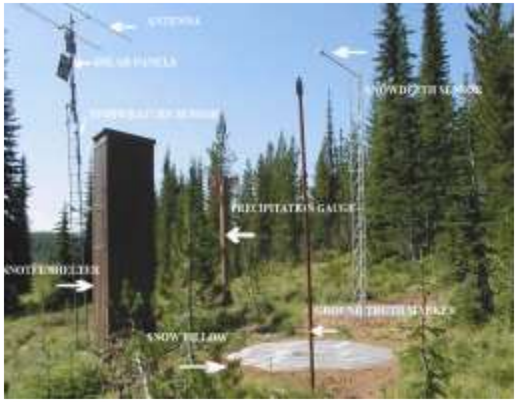
→ *Approach ignores uncertainty in initial conditions as well as uncertainty in the land model used to produce the forecast*

Combining Data Assimilation and ESP

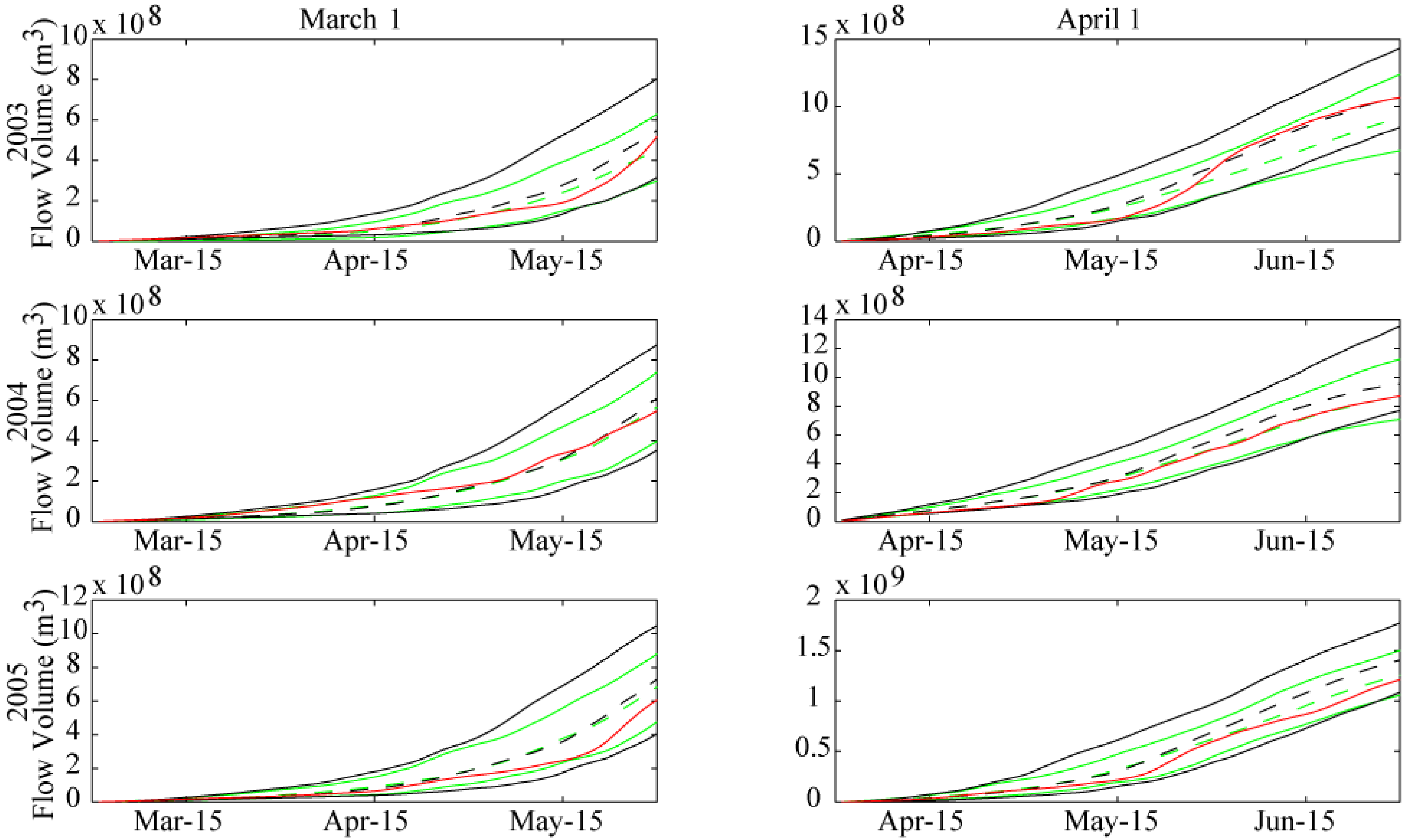


Study Area-Upper Colorado River Basin





Seasonal Cumulative Streamflow Prediction

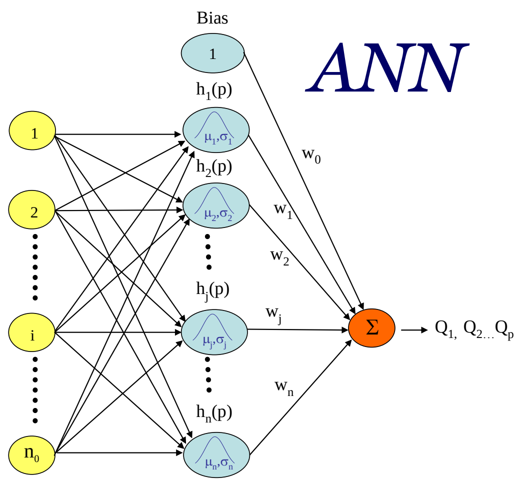


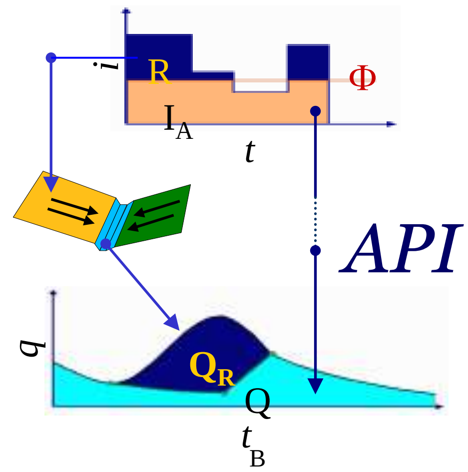
More complete Characterization of Uncertainty by Data Assimilation and Multimodeling

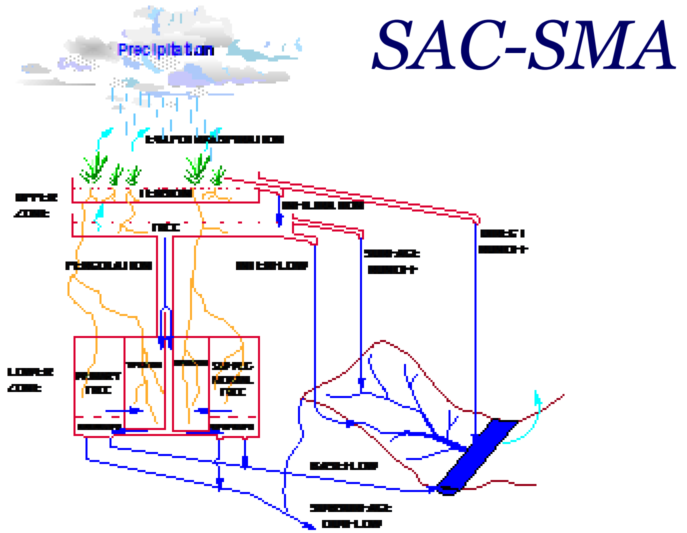
Parrish, M., H. Moradkhani, and C.M. DeChant (2012), Towards Reduction of Model Uncertainty: Integration of Bayesian Model Averaging and Data Assimilation, *Water Resources Research*, 48, W03519, doi:10.1029/2011WR011116.

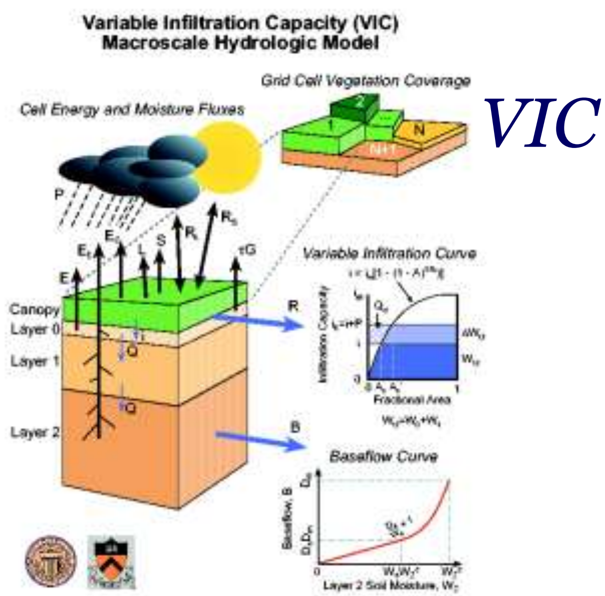
DeChant C.M., and H. Moradkhani (2014), Toward a Reliable Prediction of Seasonal Forecast Uncertainty: Addressing Model and Initial Condition Uncertainty with Ensemble Data Assimilation and Sequential Bayesian Combination, *Journal of Hydrology*, special issue on Ensemble Forecasting and data assimilation, DOI: 10.1016/j.jhydrol.2014.05.045.

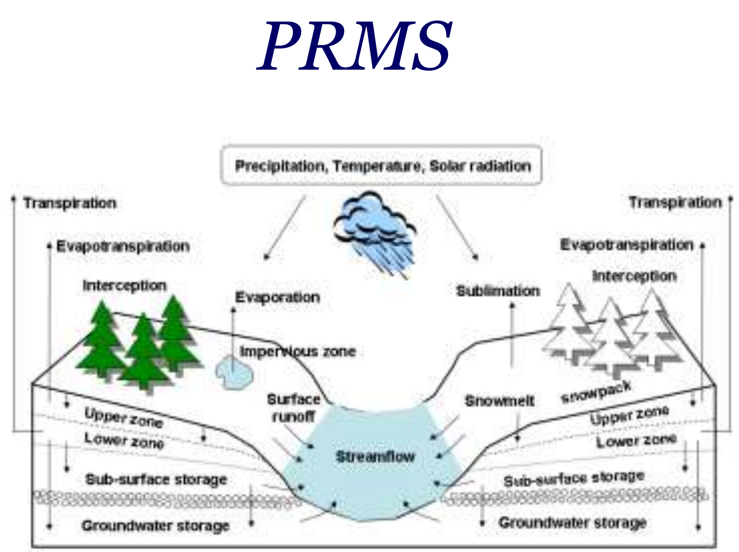
Hydrologic Models of Different Complexities…



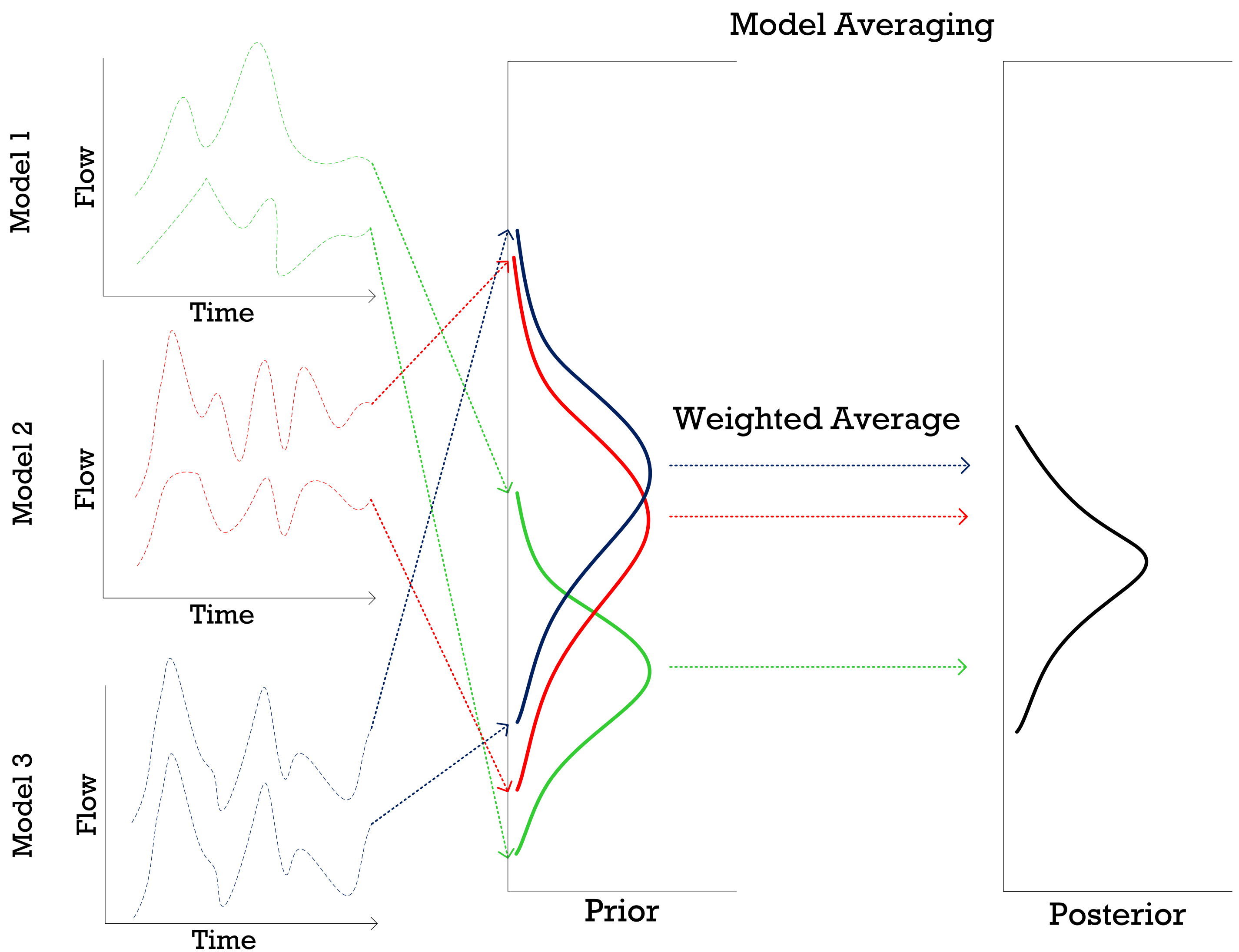




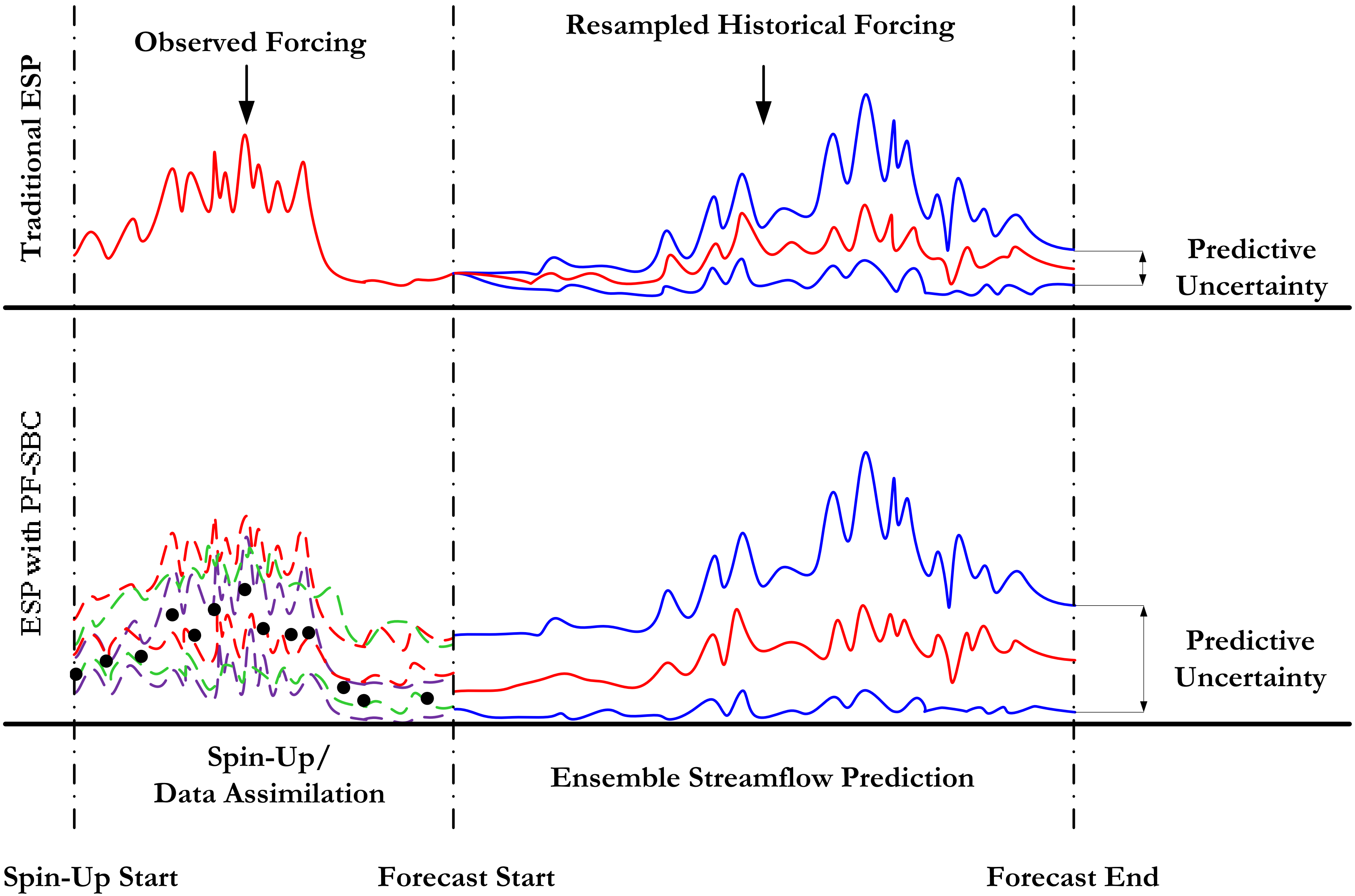




Sequential Bayesian Combination



Combining PF-SBC with ESP



Modeling Cases

- Two Models

  1) Variable Infiltration Capacity (VIC)

     - Physically-based distributed model

  2) National Weather Service (NWS) models

     - Conceptual semi-distributed models

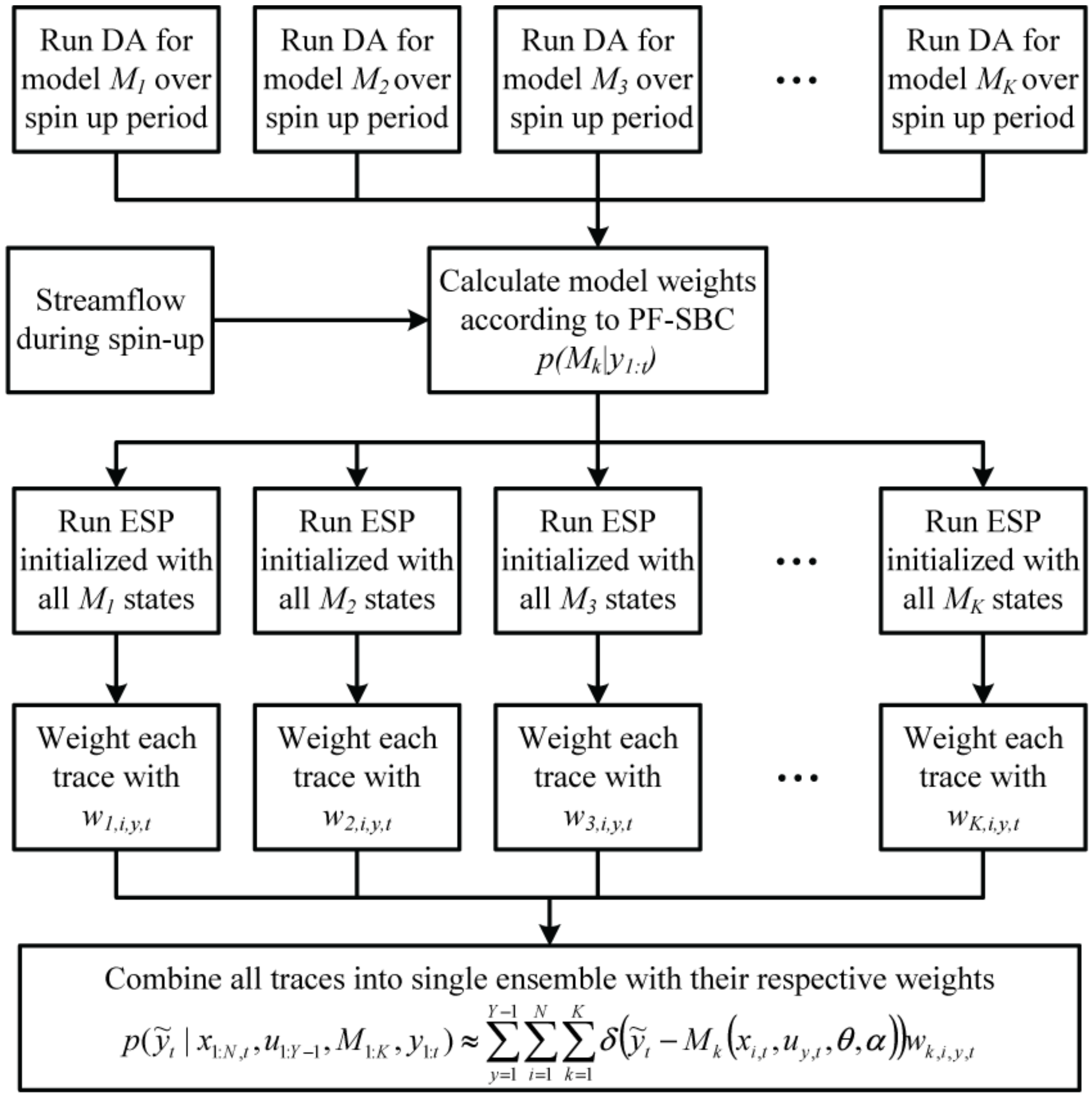
- Three cases for forecast spin-up

  1) Open Loop (no assimilation)

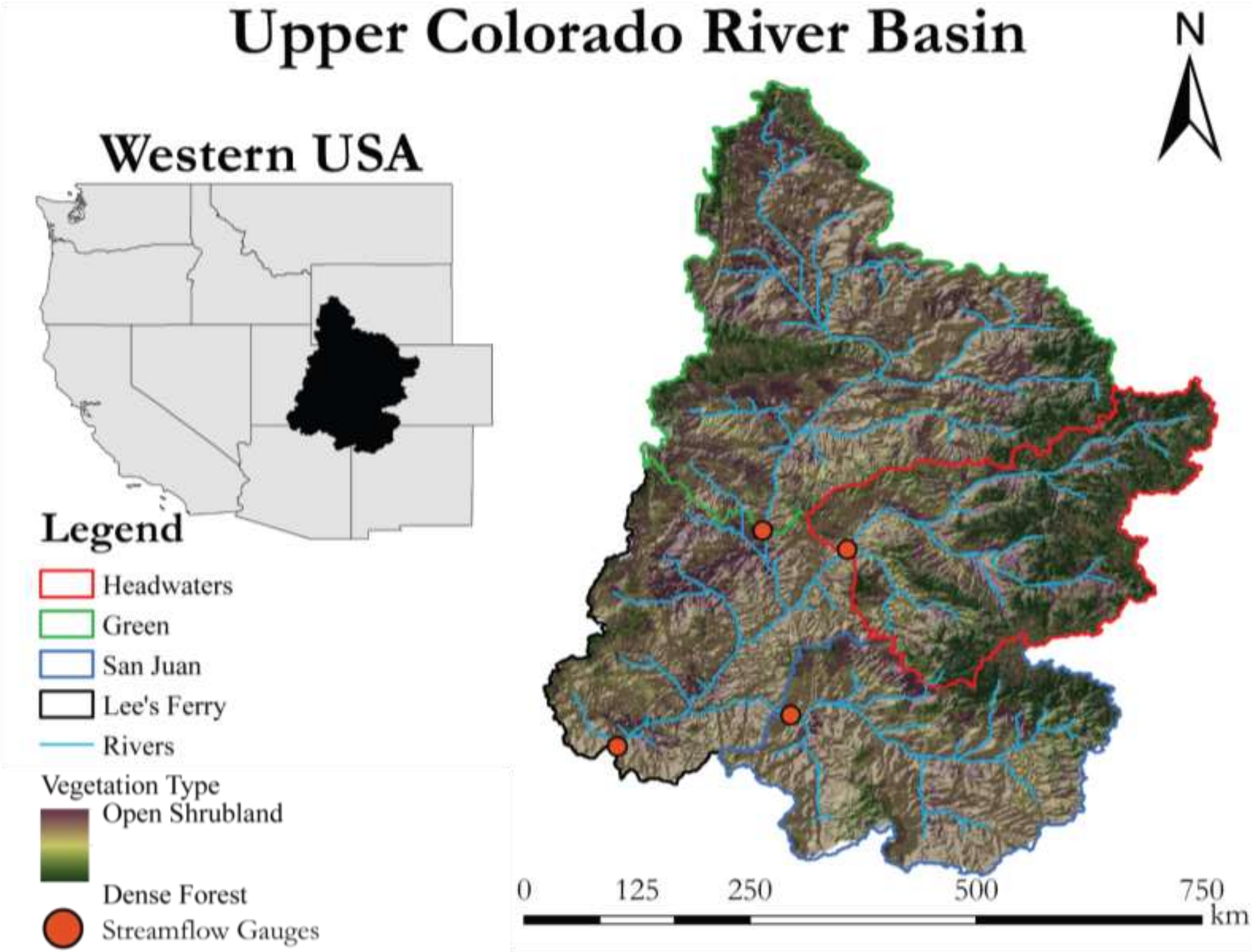
  2) Passive Microwave Brightness Temperature (TB)

  3) Land Surface Temperature (LST) with TB

Combination of DA, Multi-modeling and ESP



Study Area



Upper Colorado River Basin
N
Western USA
Legend
Headwaters
Green
San Juan
Lee's Ferry
Rivers
Vegetation Type
Open Shrubland
Dense Forest
Streamflow Gauges
0
125
250
500
750
km

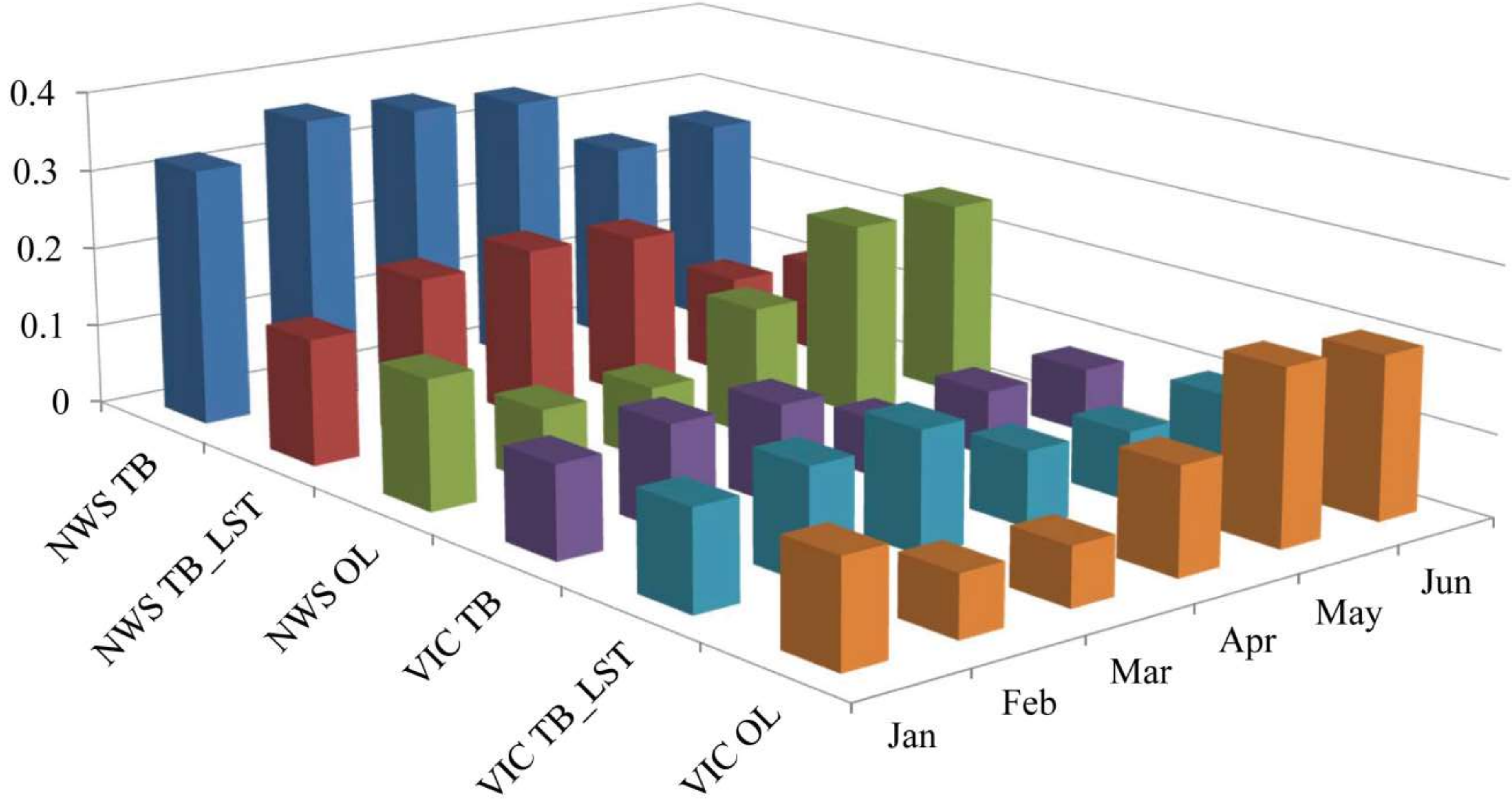
Water Supply Forecasting Experiment

- Generate multiple seasonal (3-month) volumetric streamflow forecasts for the Upper Colorado River Basin
  - Start dates on the 1st and 15th of January through June
  - Years of study from 2003 through 2008
  - 72 total streamflow forecasts for each location

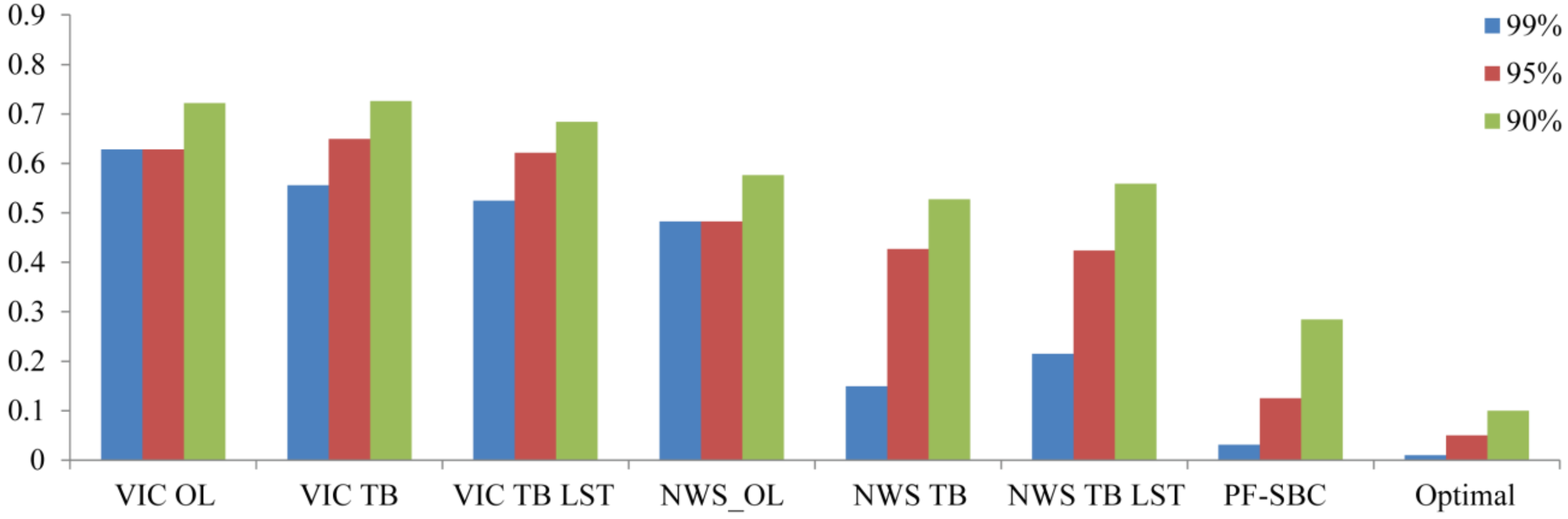
- Does the DA reduce the overconfidence related to ignoring initial condition uncertainty?

- Can Sequential Bayesian Combination improve the accounting of model uncertainty?

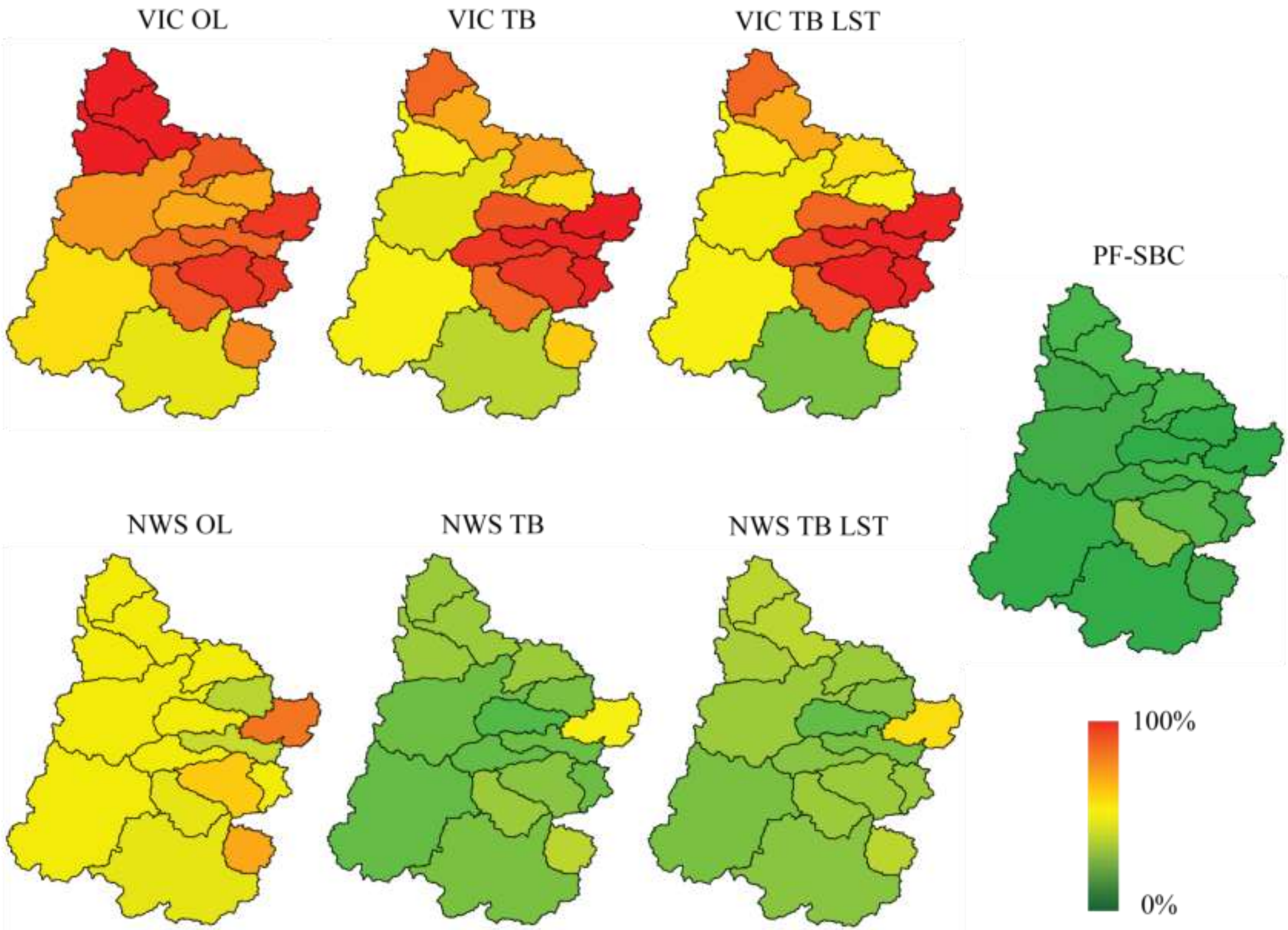
Sequential Bayesian Combination Weights



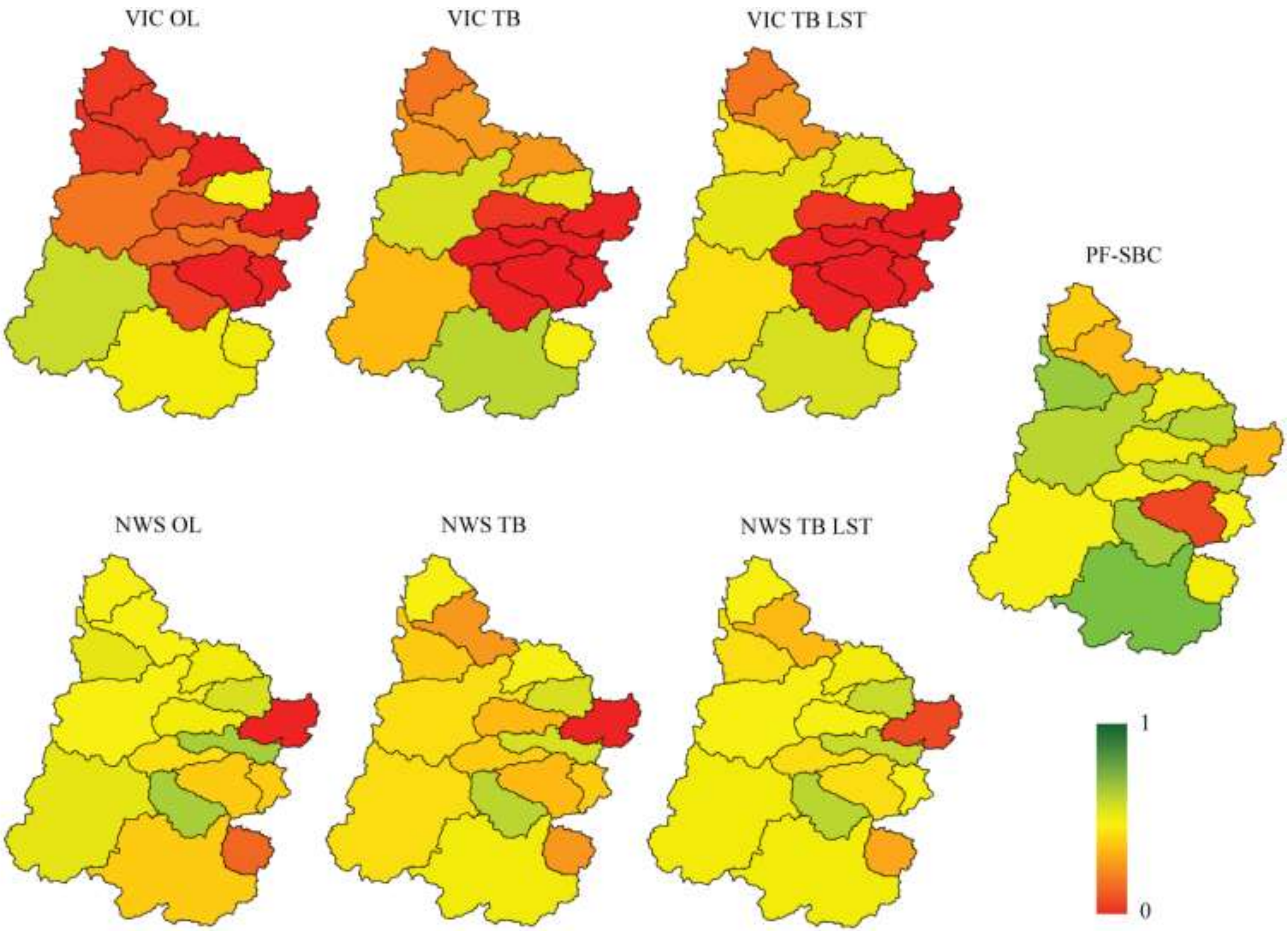
Exceedance Ratios



Spatially Distributed 99% Exceedance Ratio



Reliability



Water Supply Forecasting Conclusions

- ESP produces overconfident seasonal streamflow forecasts
  - Incomplete accounting of all uncertainty sources
- Data assimilation generally improves reliability, but remains overconfident
  - Model uncertainty not effectively managed
- DA-SBC leads to a most reliable forecasts
  - Model itself is a major source of uncertainty!