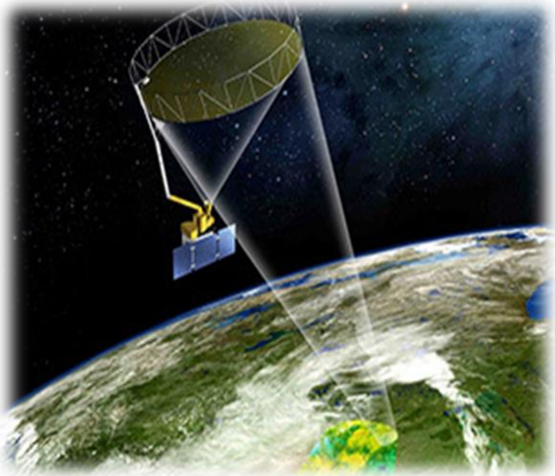
Recent Developments In Evolutionary Data Assimilation And Model Uncertainty Estimation For Hydrologic Forecasting

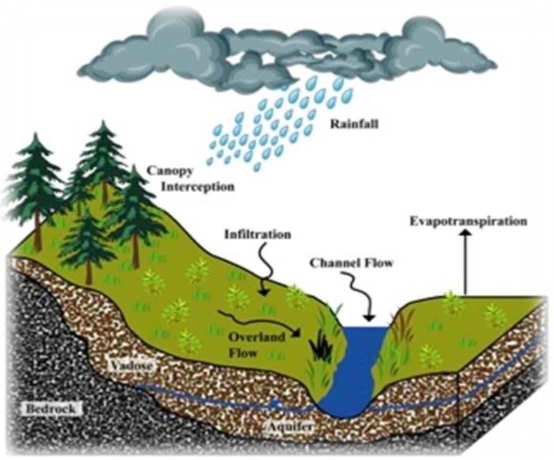
Hamid Moradkhani

Center for Complex Hydrosystems Research

Department of Civil, Construction and Environmental Engineering

THE UNIVERSITY OF ALABAMA





Uncertainties in Hydrologic Modeling

1) Meteorological forcing

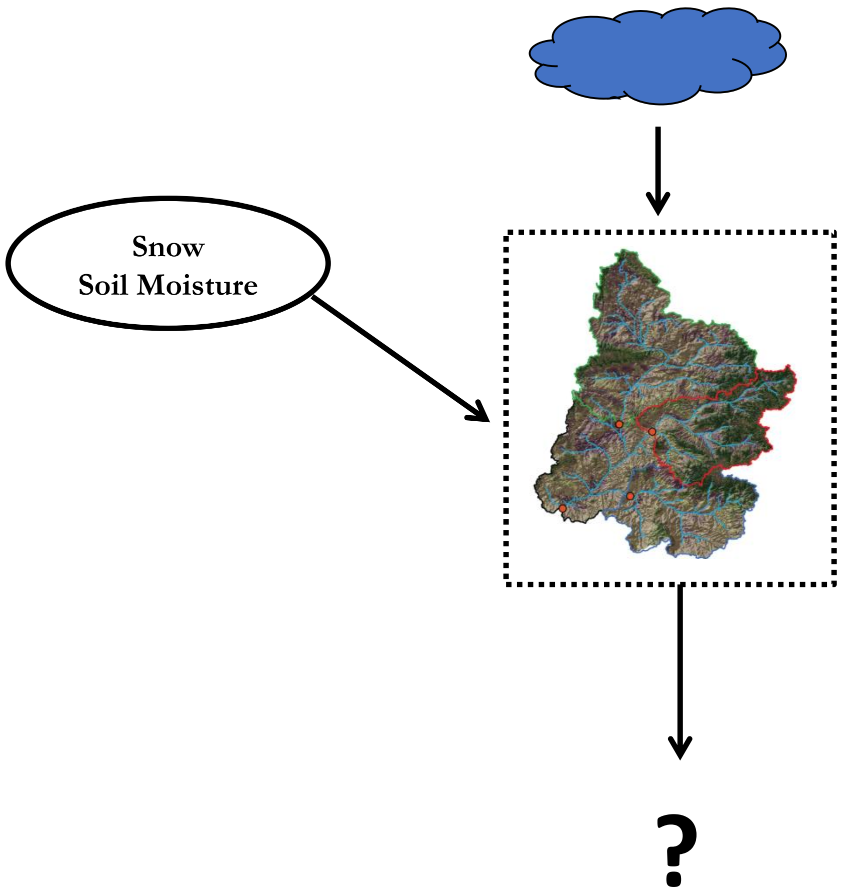
- Earth's chaotic atmosphere makes forecasting unreliable at extended lead times

2) Initial condition (states)

- Land surface hydrological conditions are highly variable spatially (e.g., snow and soil moisture)

3) Hydrologic model

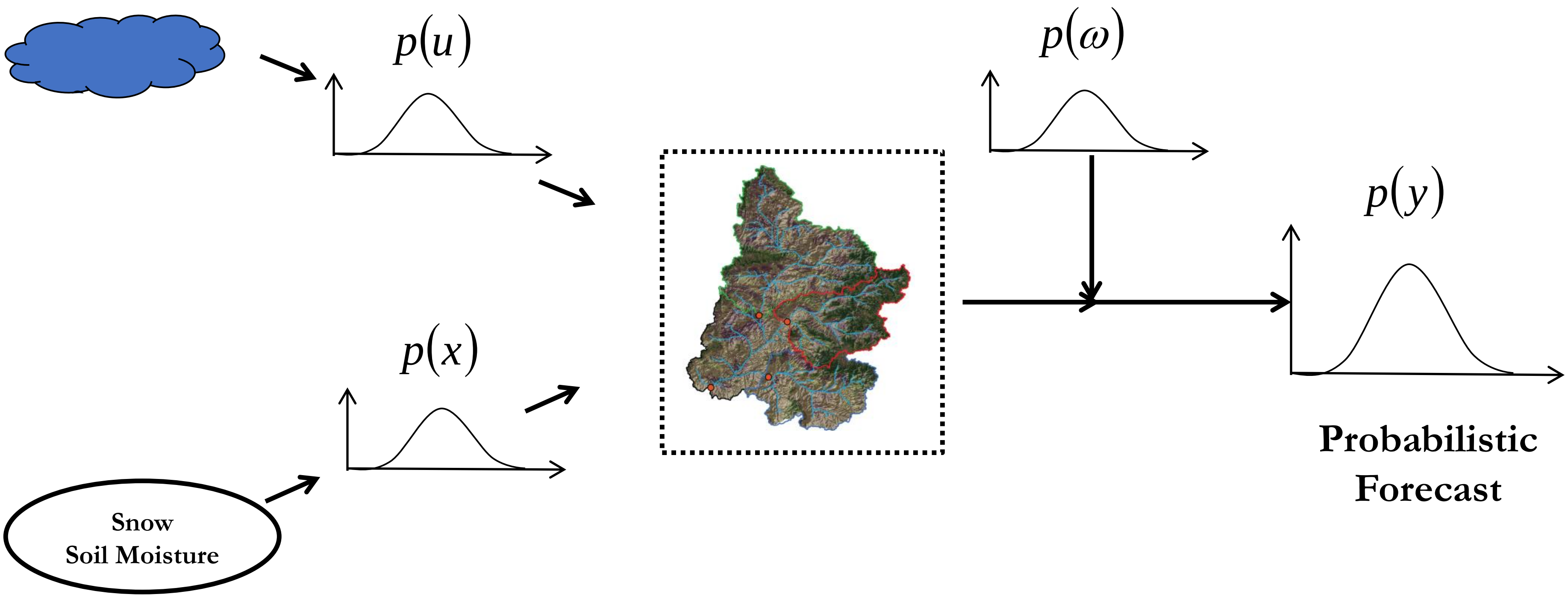
- Hydrologic models are simplifications to land surface processes



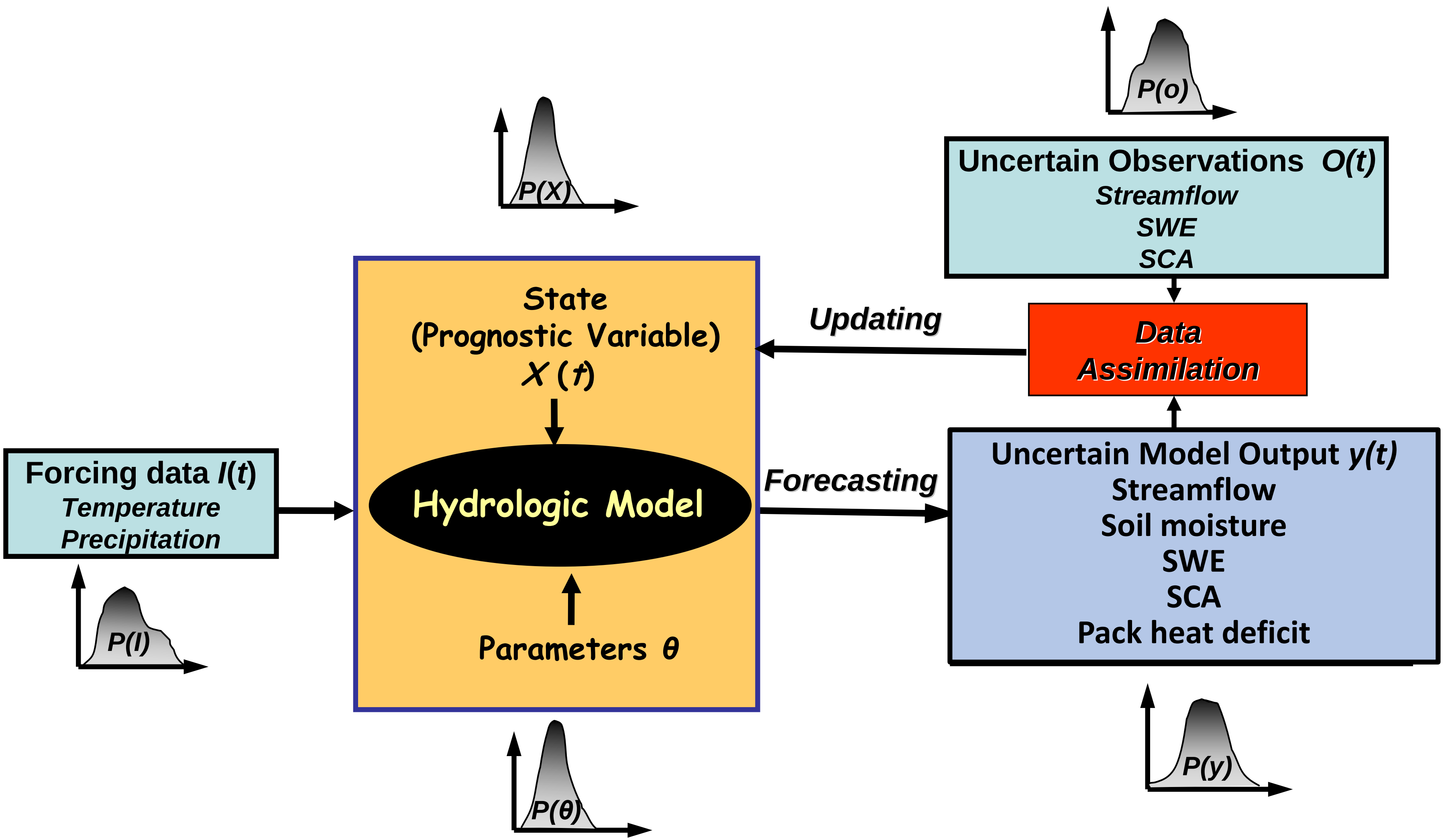
Quantifying Uncertainty

- Requires the formulation of a probabilistic model

$$p(y) = f(p(x), p(u), q) + p(w)$$



Sequential Data Assimilation – the Generic Framework

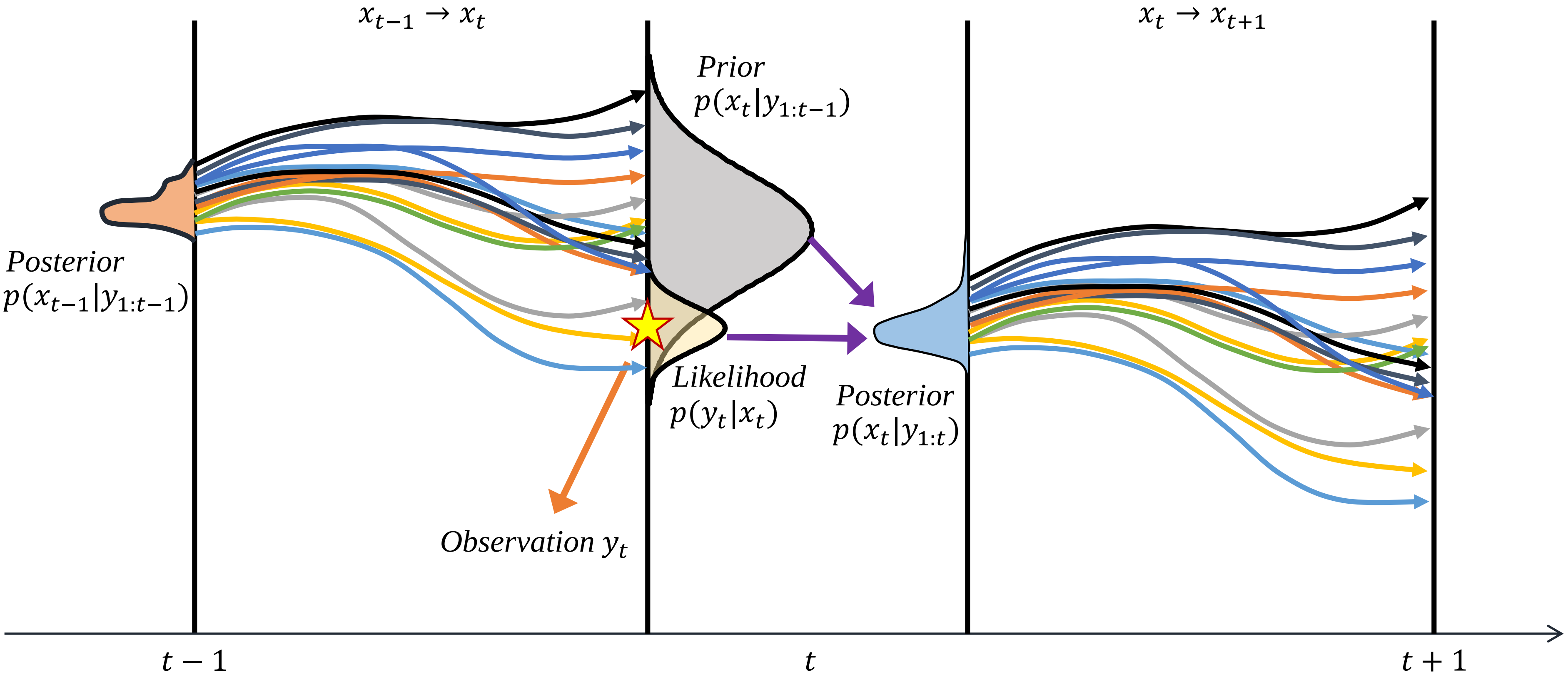


Data Assimilation: Particle Filter

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

$$p(x_{t+1} \mid Y_t) = \int_{x_t} p(x_{t+1} \mid x_t)\, p(x_t \mid Y_t)\, dx_t$$

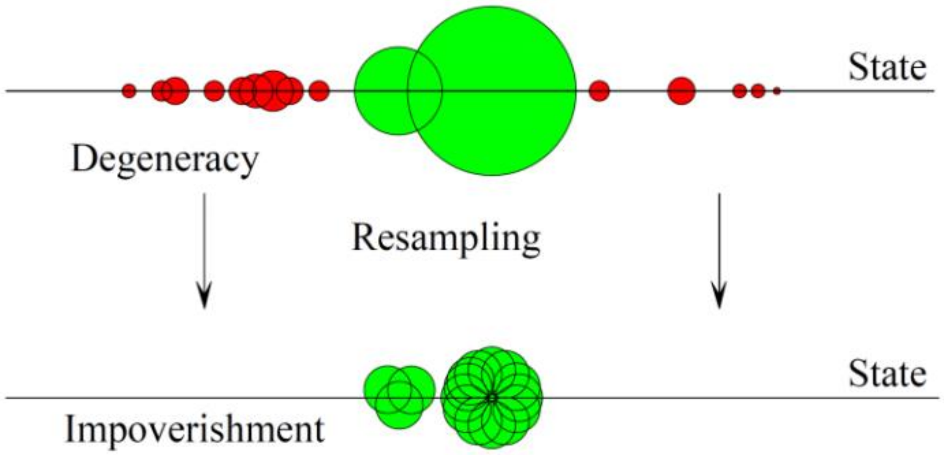
$$p(y_{t+1} \mid Y_t) = \int_{x_{t+1}} p(y_{t+1} \mid x_{t+1})\, p(x_{t+1} \mid Y_t)\, dx_{t+1}$$



Enhancement of DA by Evolutionary Particle Filter

Particle Filters (PFs) have received increasing attention from different disciplines as an effective tool to improve model predictions in nonlinear and non-Gaussian dynamical systems.

- Despite the success of the PF, one concern has been the particle degeneracy.
- To alleviate this problem, Sampling-Importance Resampling (SIR) is used to force particles to areas of high likelihood by multiplying high weighted particles while discarding low weighted particles.
- This, however, may cause another problem: sample impoverishment or loss of diversity in particles.



➢ How to Enhance Particle Filter DA?

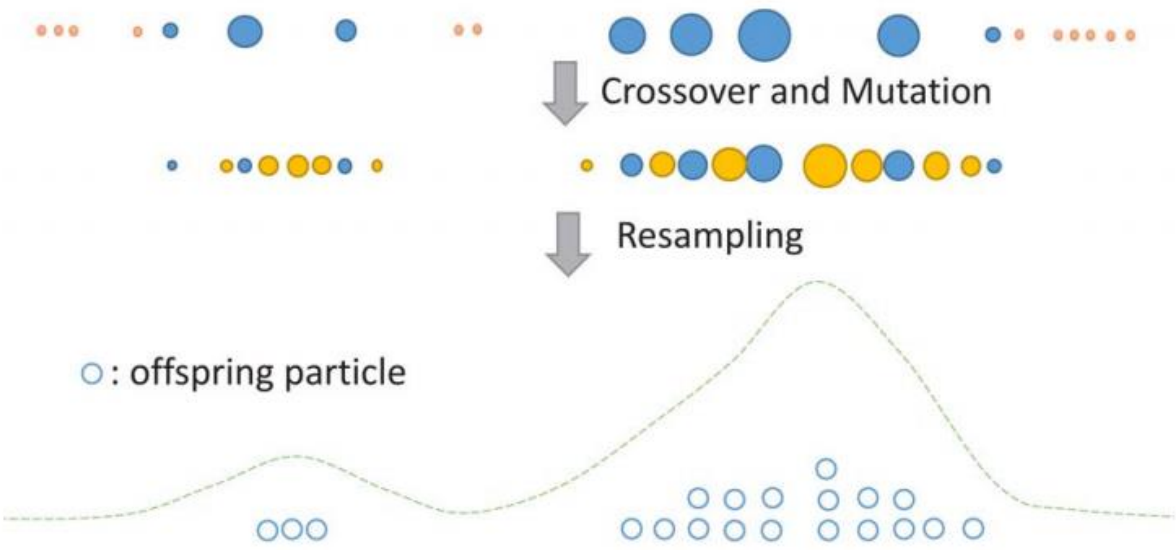
Markov Chain Monte Carlo (MCMC) with Particle Filter (***Moradkhani et al., 2012***)

Intelligent search and optimization methods categorized as Metaheuristic Algorithms (MAs) have also been used to mitigate the degeneracy problem.

- ❑ Genetic Algorithm (GA), *(Higuchi, 1997; Kwok et al., 2005; Park et al., 2009)*
- ❑ Evolution Strategy (ES), *(Uosaki et al., 2003; Uosaki et al., 2004)*
- ❑ Particle Swarm Optimization (PSO), *(Wang et al., 2006; Li et al., 2013)*
- ❑ Ant Colony Optimization (ACO), *(Xu et al., 2009; Park et al., 2010; Zhu et*
- ❑ Immune Genetic Algorithm (IGA), *(Han et al., 2011)*
- ❑ Inverse Weed Optimization, *(Ahmadi et al., 2012)*

➢ Earlier Version of GA- Particle Filter and limitations

An Illustration of GA-PF algorithm: *(Yin et al., 2015)*



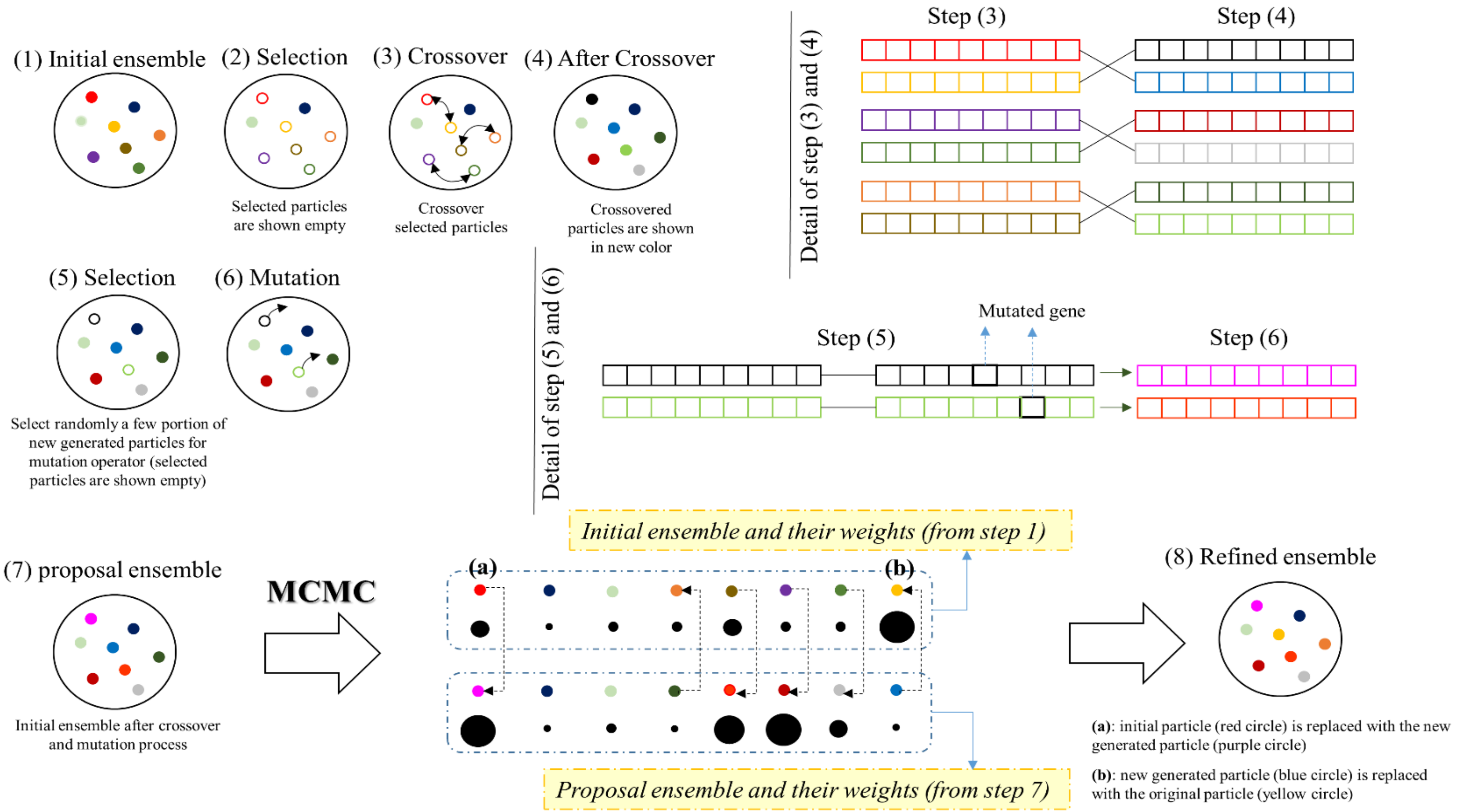
Limitation of using a single evolutionary algorithm with the PF:

- ❑ This approach reduces the weights of large-weight particles and may lead to sub-optimal performance.
- ❑ It is possible that the shuffled particles after the GA operation move outside the posterior distribution and lead to a degraded performance!

A new approach (GA-MCMC)

We modified the GA approach suitable for hydrogeoscience applications. In particular, we use a MCMC move inside the GA to guide the PF performance.

- This procedure is introduced as a GA-MCMC process.

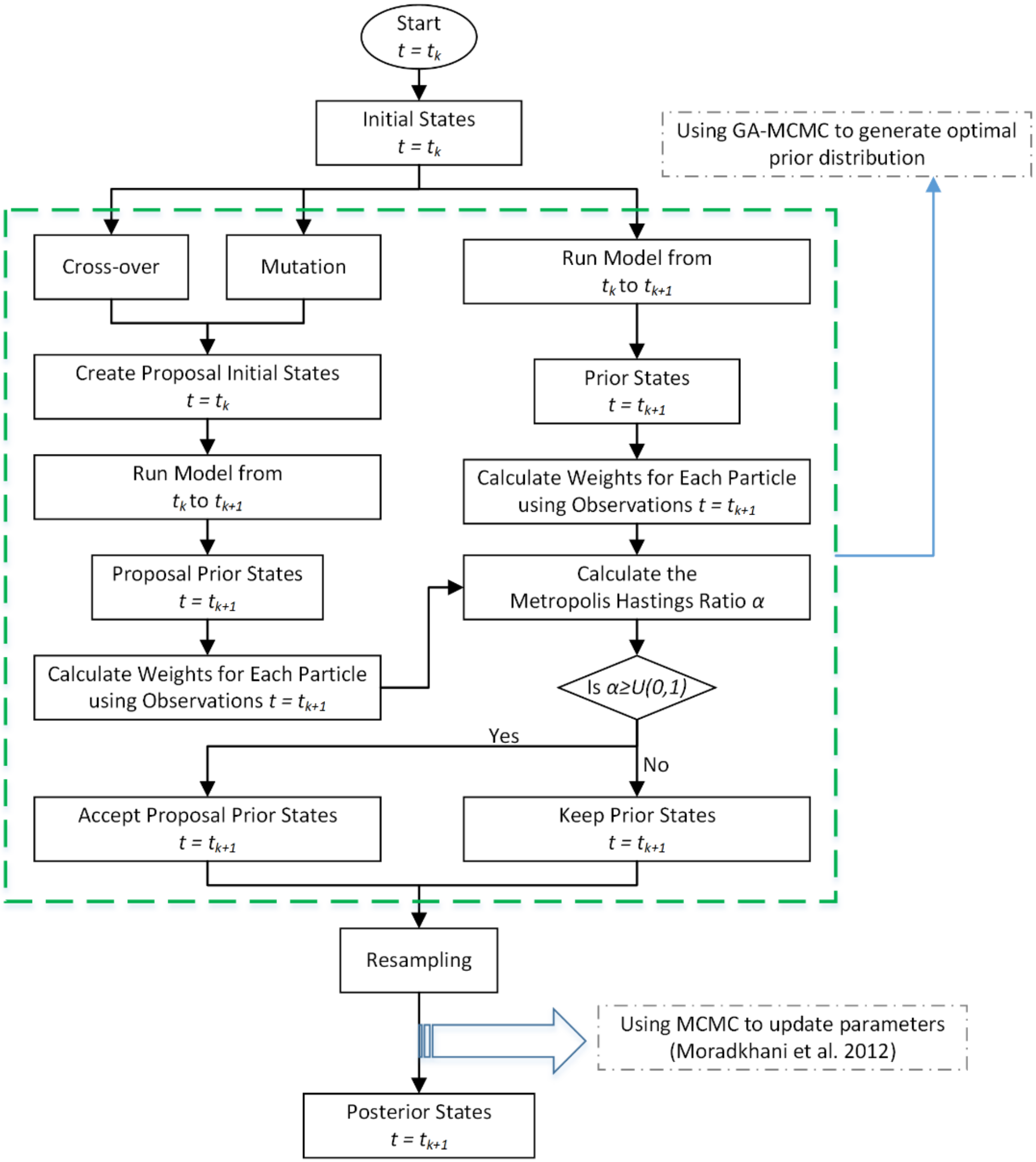


Evolutionary Particle Filter with MCMC

- The proposed approach takes our earlier developed PF-MCMC algorithm (Moradkhani et al., 2012) as a benchmark to further improve the assimilation results.

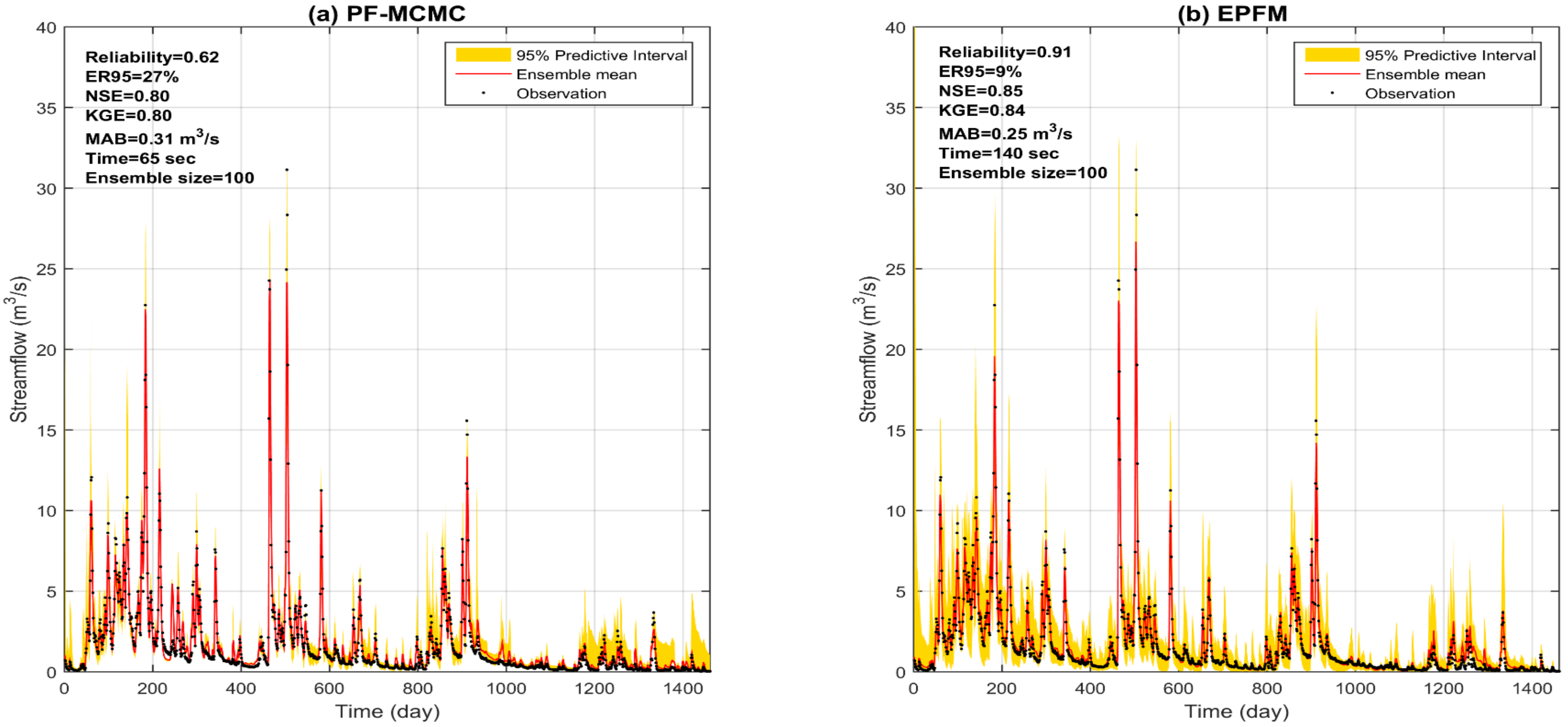
- The presented hybrid PF approach, the so-called Evolutionary Particle Filter with MCMC (EPFM), joins the strengths of GA-MCMC and PF-MCMC algorithms.

- MCMC is used twice: a) before resampling step in order to accept or reject the new generated state variable which leads to an optimal prior distribution, and b) after resampling step during parameter updating step.

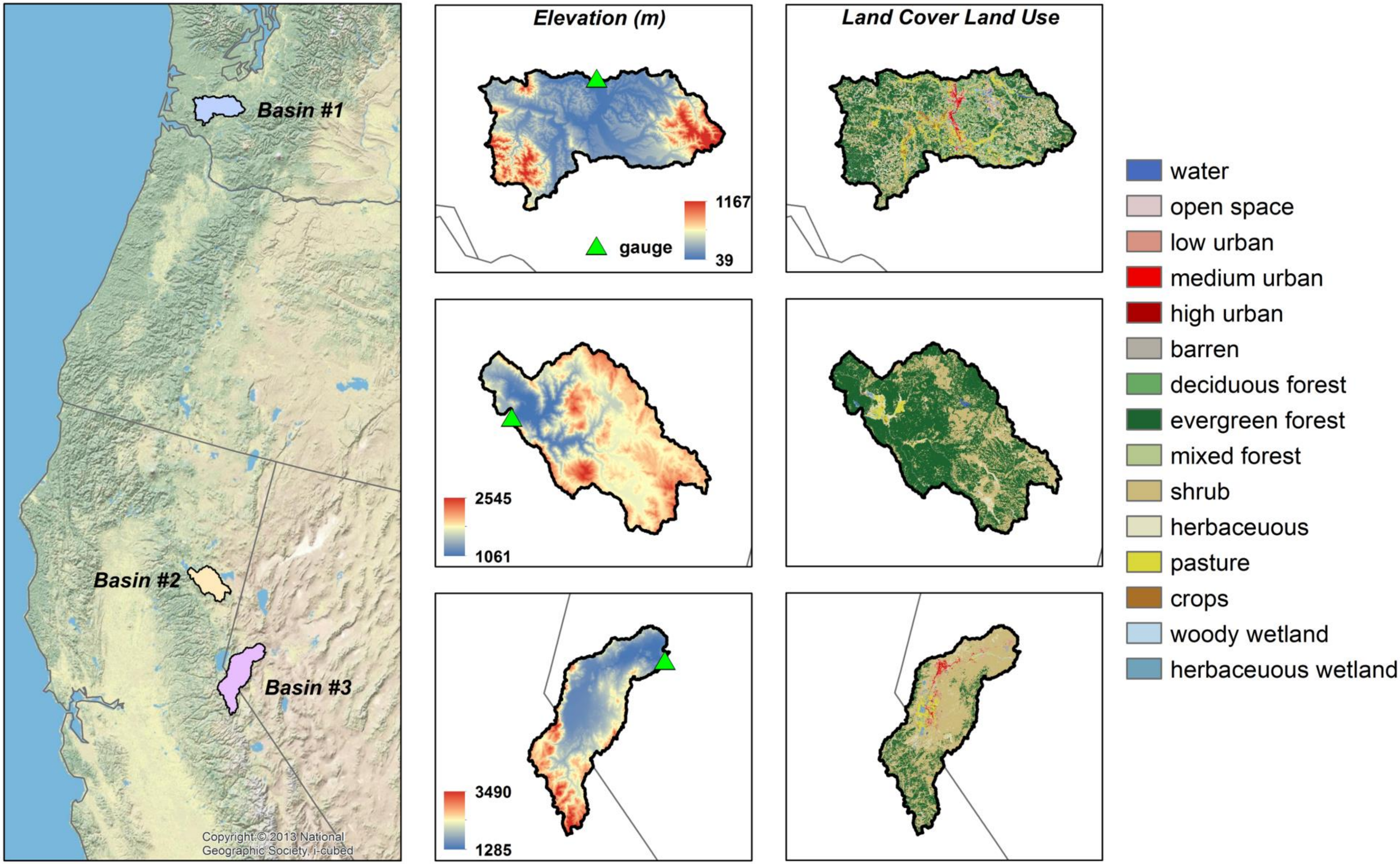


➢ A Synthetic Case

The Sacramento Soil Moisture Accounting Model (SAC-SMA) was used to simulate the streamflow in both a synthetic and three real data assimilation experiments.



➢ Three real case studies



➢ Why Evolutionary Particle Filter?

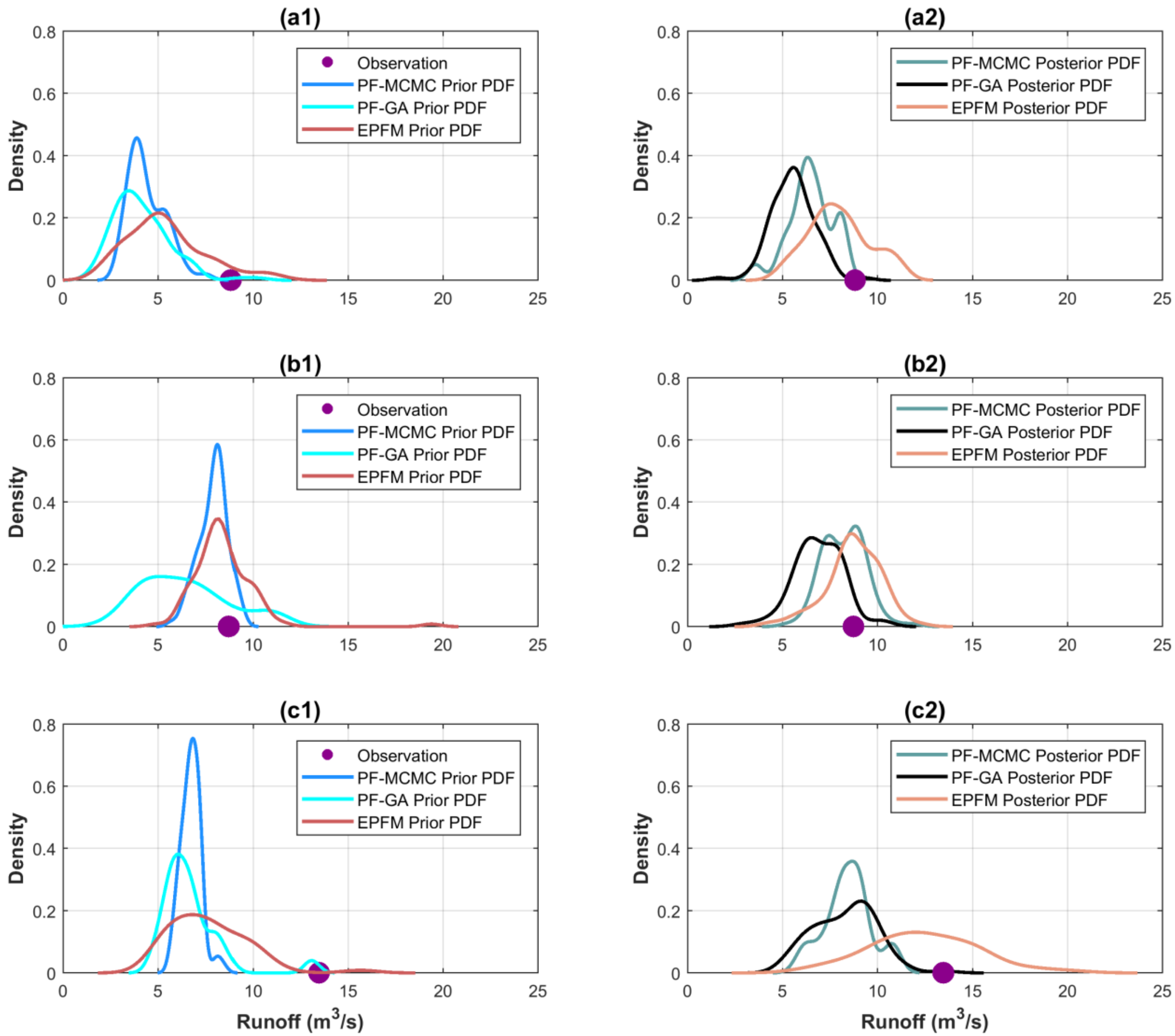
Prior Density

Posterior Density

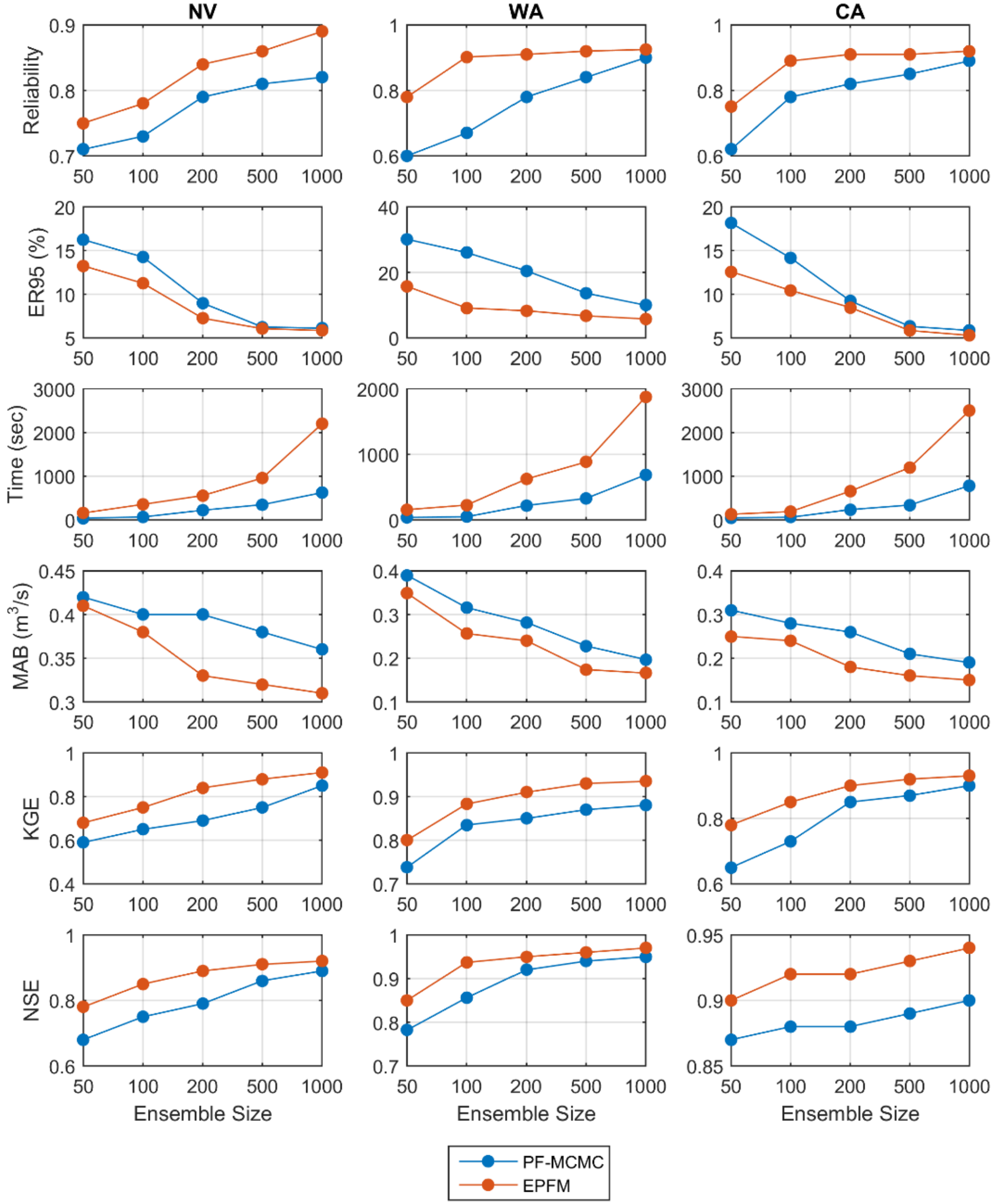
day = 50

day = 431

day = 761

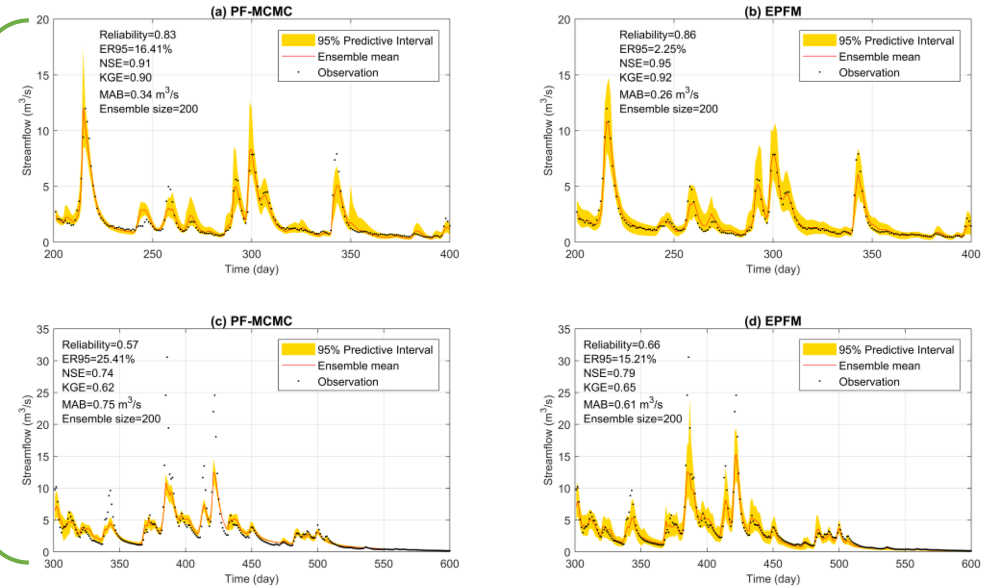


➢ Performance Assessment- real case studies

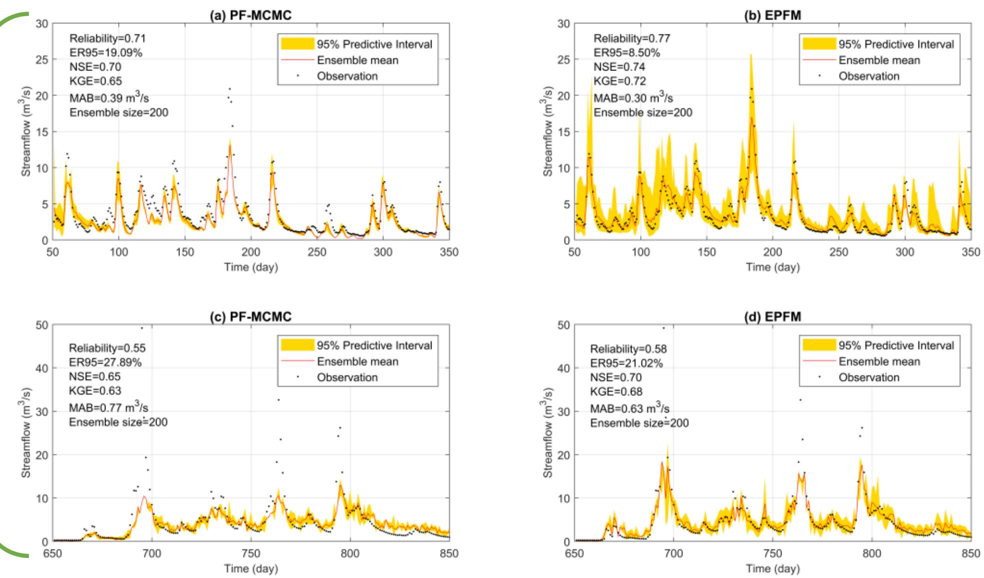


➢ Forecasting skill

❑ The comparison of the PF-MCMC and EPFM skills in one-day ahead streamflow forecast for the synthetic (a and b) and a real case study (c and d), Chehalis River Basin in WA, during the flood season.



❑ The comparison of the PF-MCMC and EPFM skills in five-day ahead streamflow forecast for the synthetic (a and b) and a real case study (c and d), Chehalis River Basin in WA, during the flood season.



Advances in Water Resources 111 (2018) 192–204

ELSEVIER

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Enhancing hydrologic data assimilation by evolutionary Particle Filter and Markov Chain Monte Carlo

Peyman Abbaszadeh, Hamid Moradkhani*, Hongxiang Yan

Water Resources and Remote Sensing Lab, Department of Civil and Environmental Engineering, Portland State University, 1930 SW 4th Ave., suite 200, Portland, OR 97201, United States

A R T I C L E I N F O

Keywords:
Particle Filters
Markov Chain Monte Carlo
Genetic algorithm
Hydrologic prediction

A B S T R A C T

Particle Filters (PFs) have received increasing attention by researchers from different disciplines including the hydro-geosciences, as an effective tool to improve model predictions in nonlinear and non-Gaussian dynamical systems. The implication of dual state and parameter estimation using the PFs in hydrology has evolved since 2005 from the PF-SIR (sampling importance resampling) to PF-MCMC (Markov Chain Monte Carlo), and now to the most effective and robust framework through evolutionary PF approach based on Genetic Algorithm (GA) and MCMC, the so-called EPFM. In this framework, the prior distribution undergoes an evolutionary process based on the designed mutation and crossover operators of GA. The merit of this approach is that the particles move to an appropriate position by using the GA optimization and then the number of effective particles is increased by means of MCMC, whereby the particle degeneracy is avoided and the particle diversity is improved. In this study, the usefulness and effectiveness of the proposed EPFM is investigated by applying the technique on a conceptual and highly nonlinear hydrologic model over four river basins located in different climate and geographical regions of the United States. Both synthetic and real case studies demonstrate that the EPFM improves both the state and parameter estimation more effectively and reliably as compared with the PF-MCMC.

Data Driven Model Uncertainty Estimation in

Hydrologic Data Assimilation

Water Resources Research

S. Pathiraja[1,2], H.Moradkhani[3], L. Marshall[2], A. Sharma[2], G. Geenens[4]

[1]Institut für Mathematik

Universität Potsdam

Potsdam

GERMANY

Email: pathiraja@uni-potsdam.de

[2]Water Research Centre

School of Civil and Environmental Engineering

University of New South Wales

Sydney, NSW

AUSTRALIA

[3]Department of Civil and Environmental Engineering

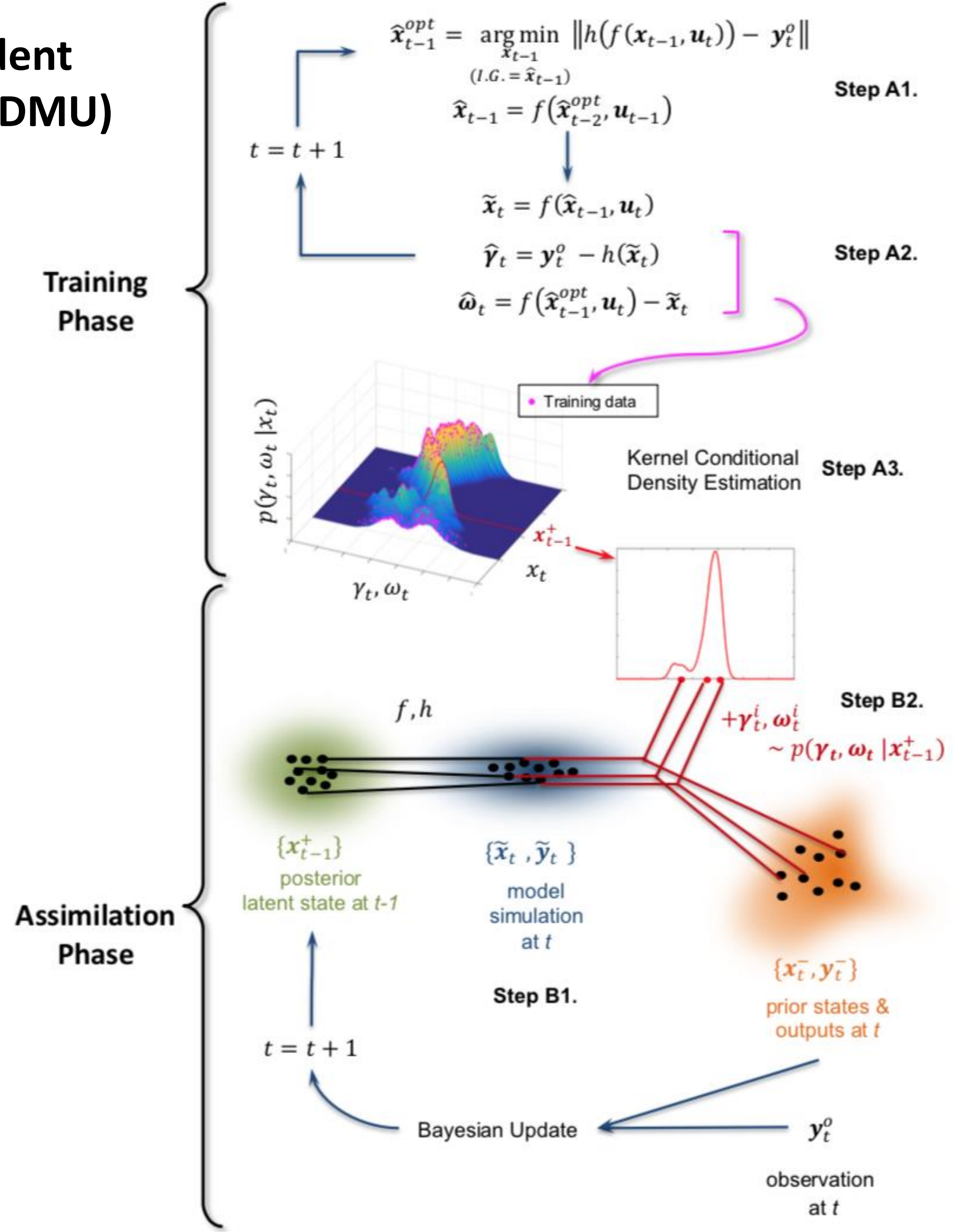
Portland State University

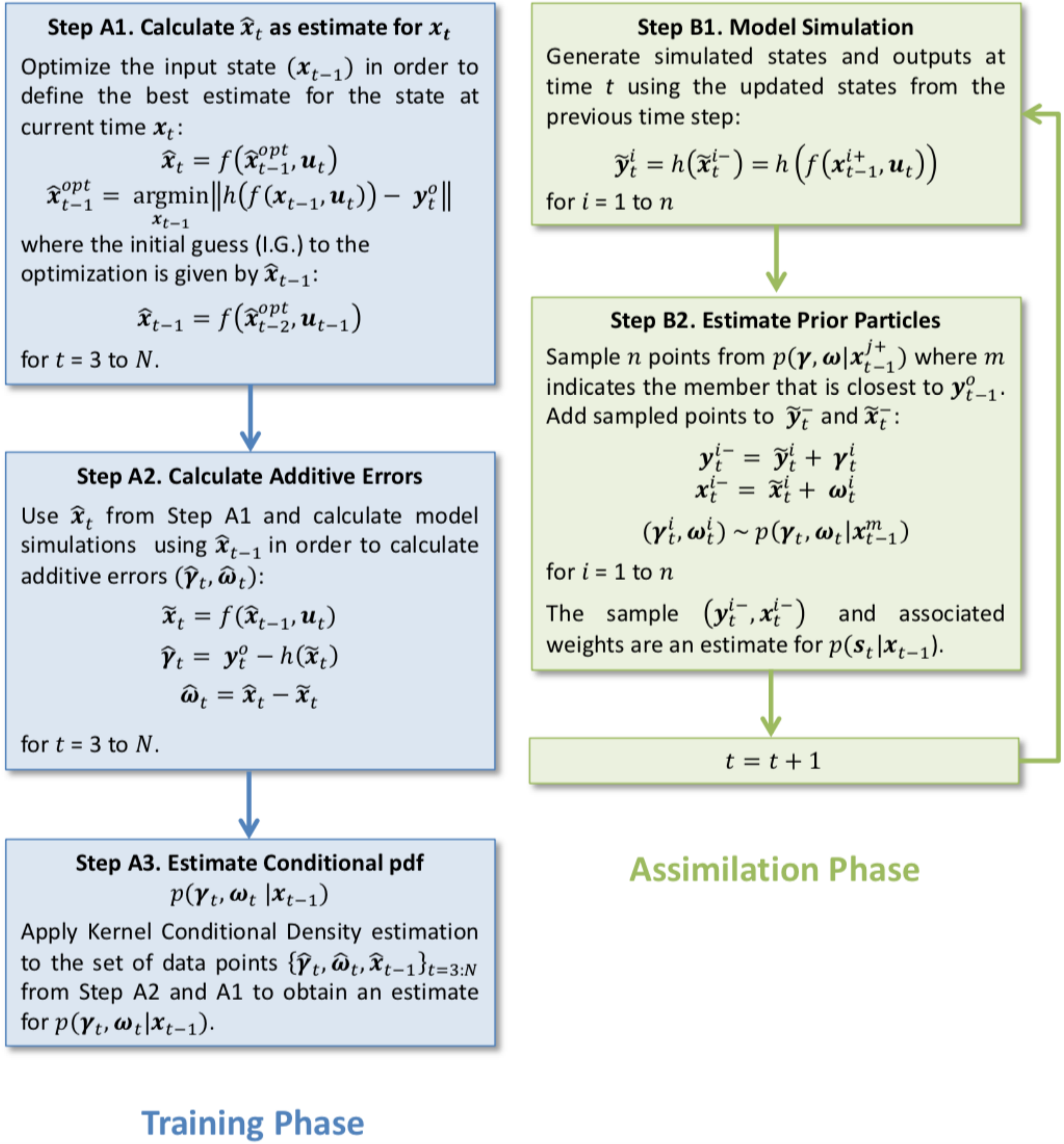
Portland, Oregon

USA

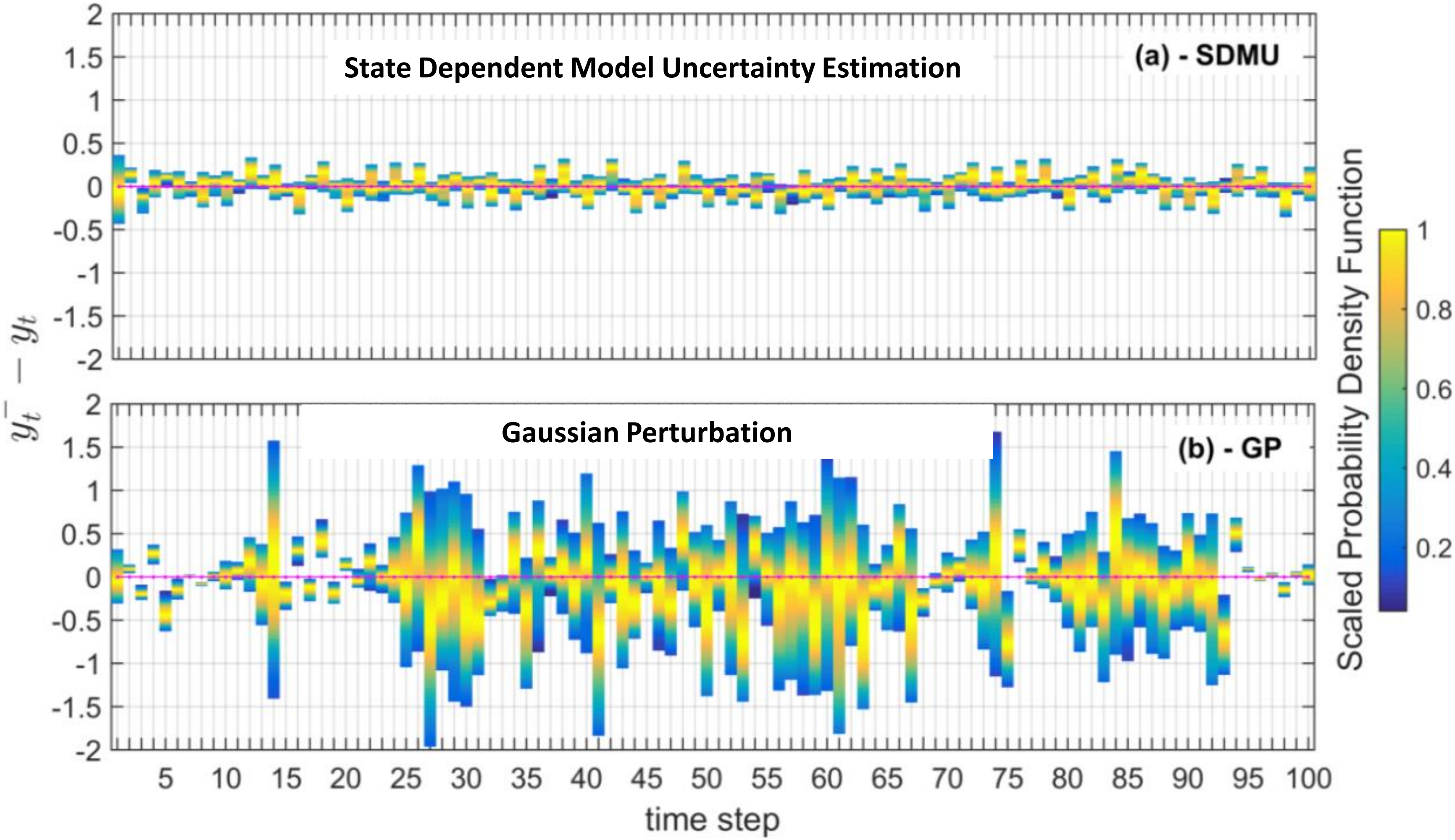
Schematic of the State-Dependent Model Uncertainty Estimation (SDMU) Method

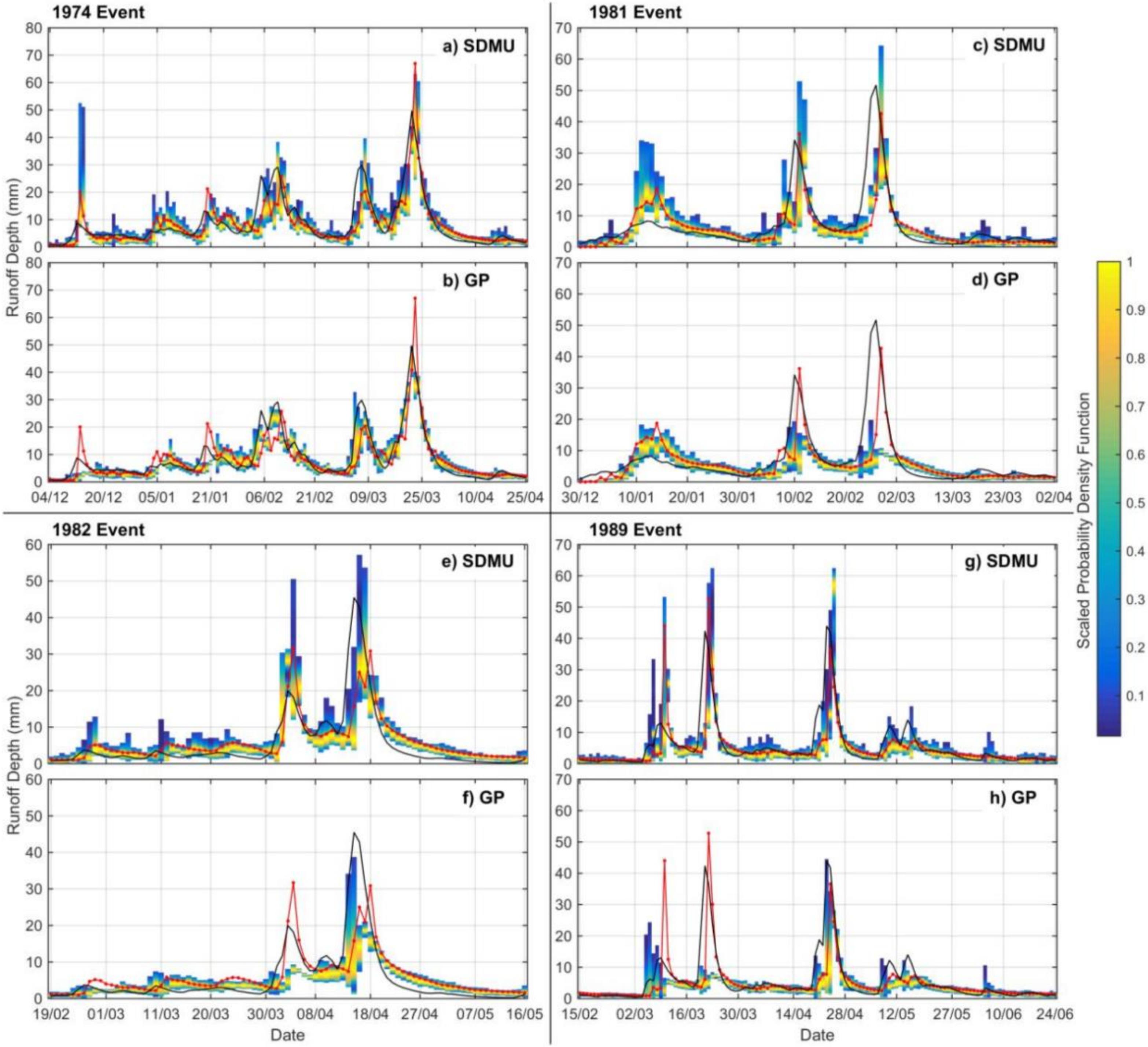
- The system states are partially observed
- Minimal prior knowledge of the model error processes is available, except that the errors display state dependence.
- It includes an approach for estimating the uncertainty in hidden model states, with the end goal of improving predictions of observed variables.
- A training period of model simulations and observations is used to develop the PDF of additive errors in observed variables conditioned on model states and inputs.




Step A1. Calculate $\widehat{\boldsymbol{x}}_t$ as estimate for $\boldsymbol{x}_t$
Optimize the input state $(\boldsymbol{x}_{t-1})$ in order to define the best estimate for the state at current time $\boldsymbol{x}_t$:
$\widehat{\boldsymbol{x}}_t = f(\widehat{\boldsymbol{x}}_{t-1}^{opt}, \boldsymbol{u}_t)$
$\widehat{\boldsymbol{x}}_{t-1}^{opt} = \underset{\boldsymbol{x}_{t-1}}{\operatorname{argmin}} \| h(f(\boldsymbol{x}_{t-1}, \boldsymbol{u}_t)) - \boldsymbol{y}_t^o \|$
where the initial guess (I.G.) to the optimization is given by $\widehat{\boldsymbol{x}}_{t-1}$:
$\widehat{\boldsymbol{x}}_{t-1} = f(\widehat{\boldsymbol{x}}_{t-2}^{opt}, \boldsymbol{u}_{t-1})$
for t = 3 to N.
Step A2. Calculate Additive Errors
Use $\widehat{\boldsymbol{x}}_t$ from Step A1 and calculate model simulations using $\widehat{\boldsymbol{x}}_{t-1}$ in order to calculate additive errors $(\widehat{\boldsymbol{\gamma}}_t, \widehat{\boldsymbol{\omega}}_t)$:
$\widetilde{\boldsymbol{x}}_t = f(\widehat{\boldsymbol{x}}_{t-1}, \boldsymbol{u}_t)$
$\widehat{\boldsymbol{\gamma}}_t = \boldsymbol{y}_t^o - h(\widetilde{\boldsymbol{x}}_t)$
$\widehat{\boldsymbol{\omega}}_t = \widehat{\boldsymbol{x}}_t - \widetilde{\boldsymbol{x}}_t$
for t = 3 to N.
Step A3. Estimate Conditional pdf
$p(\boldsymbol{\gamma}_t, \boldsymbol{\omega}_t \,|\boldsymbol{x}_{t-1})$
Apply Kernel Conditional Density estimation to the set of data points $\{\widehat{\boldsymbol{\gamma}}_t, \widehat{\boldsymbol{\omega}}_t, \widehat{\boldsymbol{x}}_{t-1}\}_{t=3:N}$ from Step A2 and A1 to obtain an estimate for $p(\boldsymbol{\gamma}_t, \boldsymbol{\omega}_t|\boldsymbol{x}_{t-1})$.
Training Phase
Step B1. Model Simulation
Generate simulated states and outputs at time t using the updated states from the previous time step:
$\widetilde{\boldsymbol{y}}_t^i = h(\widetilde{\boldsymbol{x}}_t^{i-}) = h\left(f(\boldsymbol{x}_{t-1}^{i+}, \boldsymbol{u}_t)\right)$
for i = 1 to n
Step B2. Estimate Prior Particles
Sample n points from $p(\boldsymbol{\gamma}, \boldsymbol{\omega}|\boldsymbol{x}_{t-1}^{j+})$ where m indicates the member that is closest to $\boldsymbol{y}_{t-1}^o$. Add sampled points to $\widetilde{\boldsymbol{y}}_t^-$ and $\widetilde{\boldsymbol{x}}_t^-$:
$\boldsymbol{y}_t^{i-} = \widetilde{\boldsymbol{y}}_t^i + \boldsymbol{\gamma}_t^i$
$\boldsymbol{x}_t^{i-} = \widetilde{\boldsymbol{x}}_t^i + \boldsymbol{\omega}_t^i$
$(\boldsymbol{\gamma}_t^i, \boldsymbol{\omega}_t^i) \sim p(\boldsymbol{\gamma}_t, \boldsymbol{\omega}_t|\boldsymbol{x}_{t-1}^m)$
for i = 1 to n
The sample $(\boldsymbol{y}_t^{i-}, \boldsymbol{x}_t^{i-})$ and associated weights are an estimate for $p(\boldsymbol{s}_t|\boldsymbol{x}_{t-1})$.
$t = t + 1$
Assimilation Phase

Scaled Probability Density Function of difference in predicted y and true y




1974 Event
1981 Event
1982 Event
1989 Event
a) SDMU
b) GP
c) SDMU
d) GP
e) SDMU
f) GP
g) SDMU
h) GP
Runoff Depth (mm)
Date
Scaled Probability Density Function
04/12
20/12
05/01
21/01
06/02
21/02
09/03
25/03
10/04
25/04
30/12
10/01
20/01
30/01
10/02
20/02
02/03
13/03
23/03
02/04
19/02
01/03
11/03
20/03
30/03
08/04
18/04
27/04
07/05
16/05
15/02
02/03
16/03
30/03
14/04
28/04
12/05
27/05
10/06
24/06