SAND2018-3858C

Speeding Up Sequential Tempered MCMC for Fast Bayesian Inference and Uncertainty Quantification

Thomas A. Catanach

SIAM UQ 2018

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND NO. 2017-13509 C

Overview

- Bayesian Inference and Uncertainty Quantification Problems
- Introduction to MCMC
- Sequential Tempered MCMC
- Posterior Reliability Analysis using ST-MCMC
- Conclusion

Bayesian Methods

Motivation

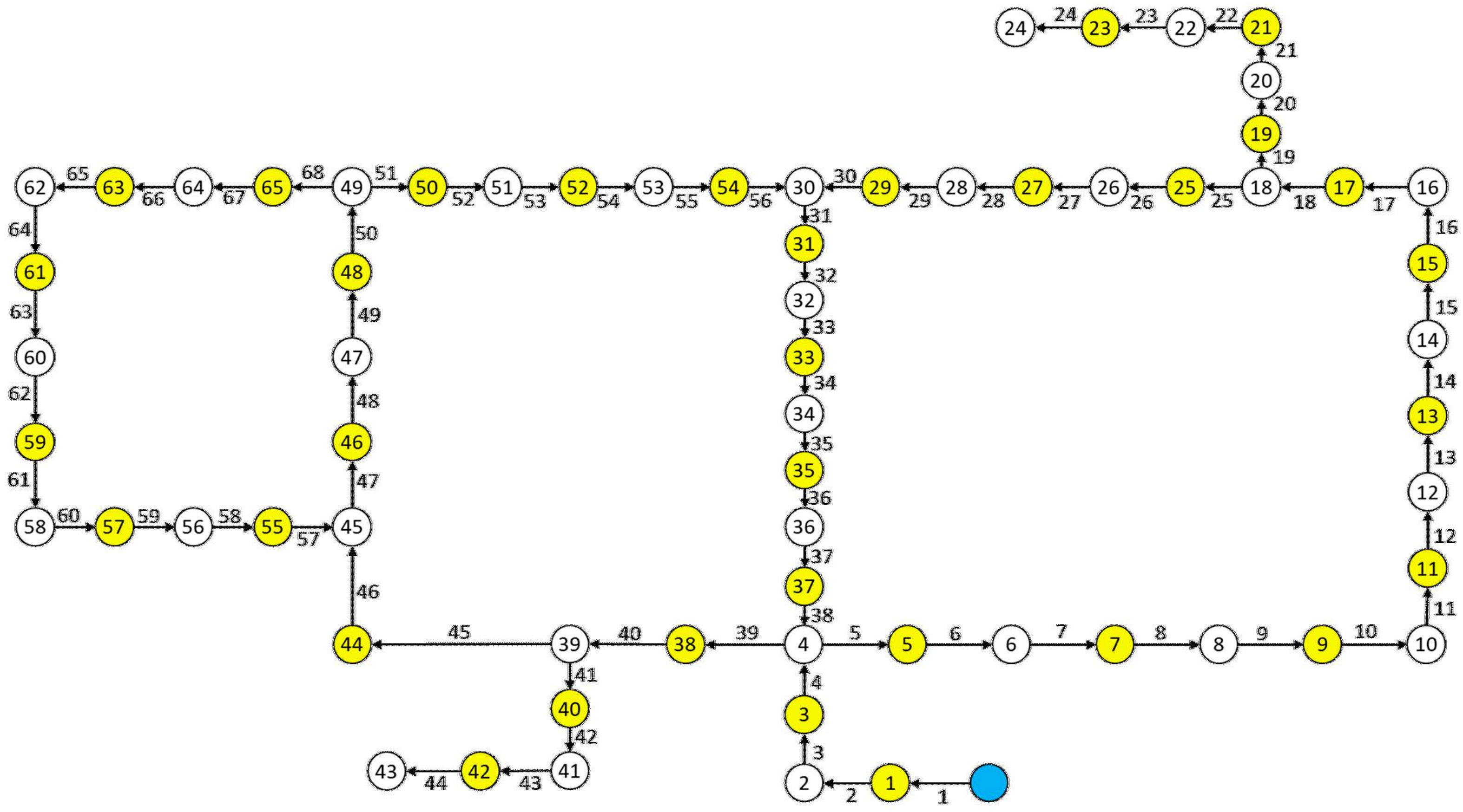
- The Bayesian Perspective:
  - Probability distributions quantify uncertainty due to insufficient information
- Bayesian methods for identification and estimation are critical to the robust system analysis

Goal:

Provide MCMC methods for computationally intensive Bayesian inference problems in complex systems

Example Inference Problem: Water Distribution[1]

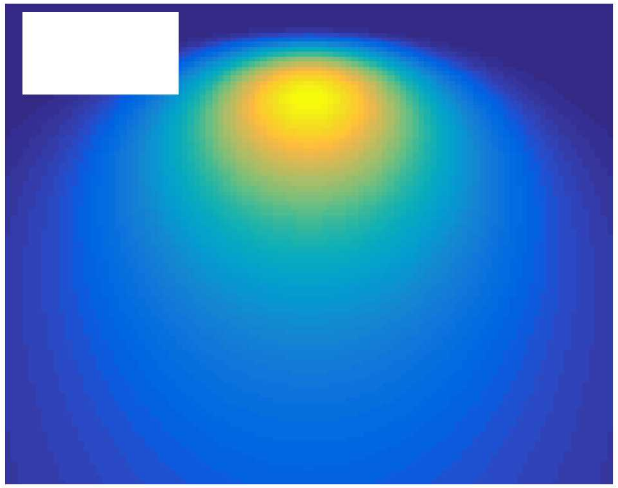
Leak Detection and Posterior Failure Probability Assessment



[1] Cunha and Sousa 1999

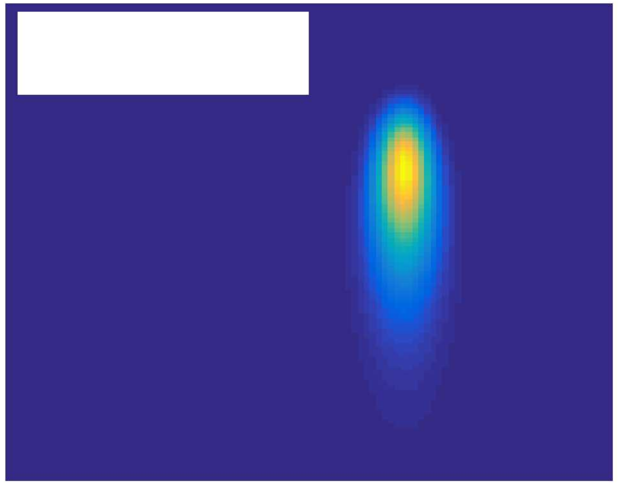
Example Inference Problem: System Identification

Prior distribution of the water system parameters



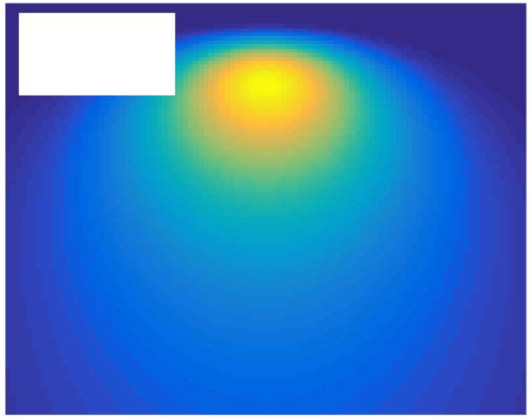
Data

Posterior distribution of the water system parameters



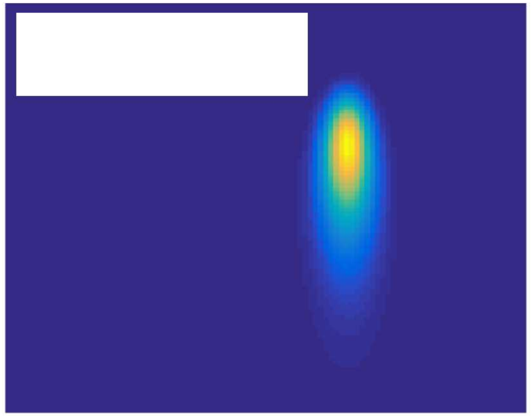
Example Inference Problem: Reliability Analysis

Prior distribution of the water system parameters



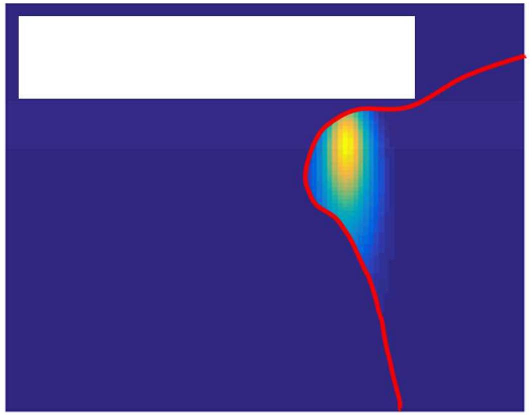
Data

Posterior distribution of the water system parameters

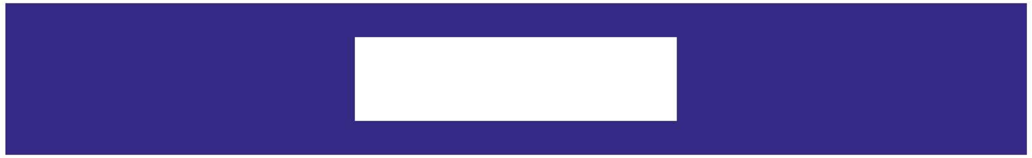


Failure

Posterior distribution of failed water system parameters



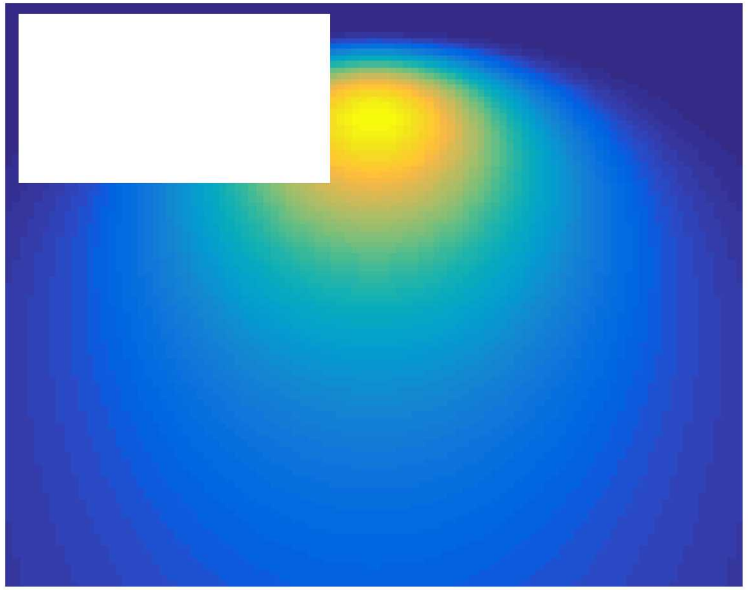
Posterior Estimate of Failure Probability



Bayesian Inference and MCMC

The Bayesian Inference Problem

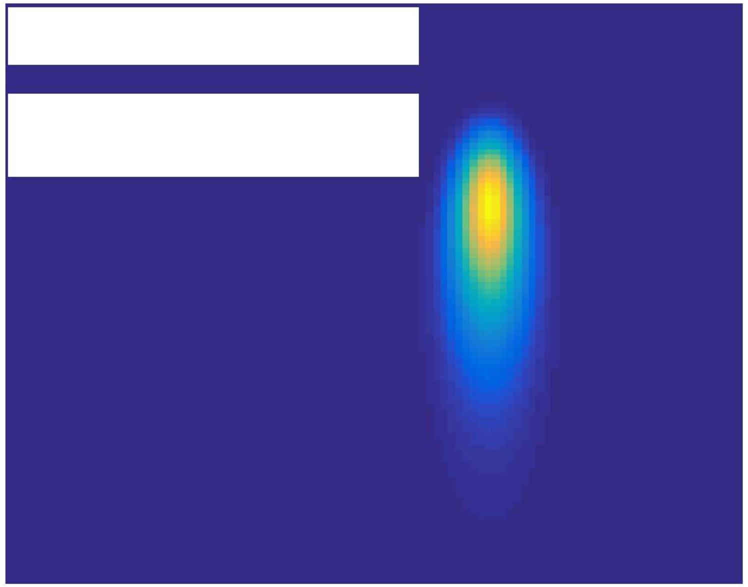
Sandia National Laboratories



Observations: $\mathcal{D}$

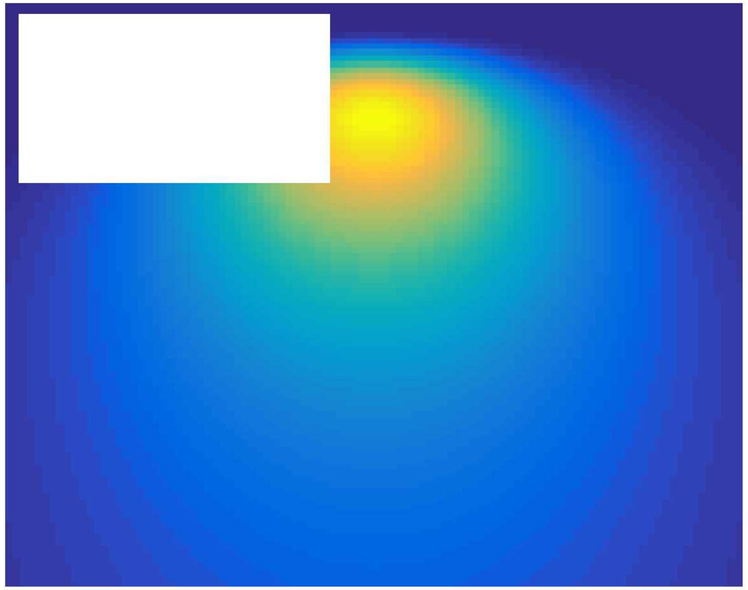
Bayes' Theorem

$$p(\theta \mid \mathcal{D}, \mathcal{M}) = \frac{p(\mathcal{D}|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})}$$



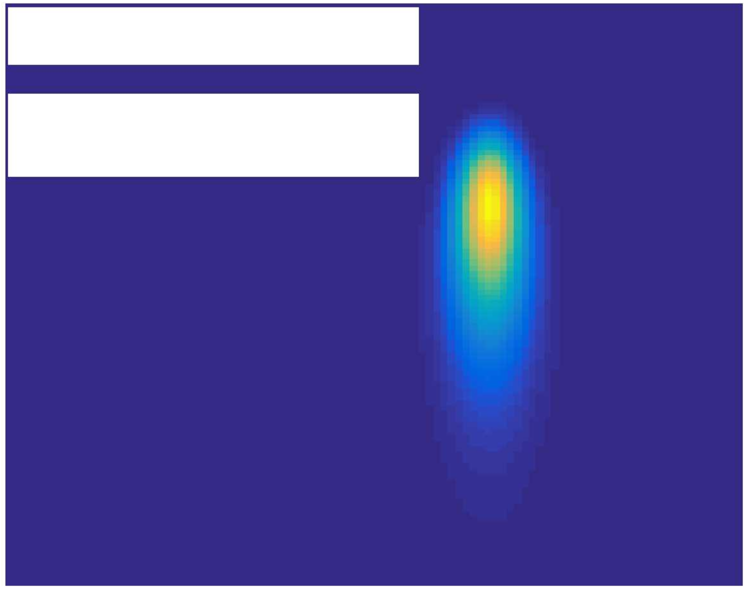
The Bayesian Inference Problem

Observations: $\mathcal{D}$



Bayes' Theorem

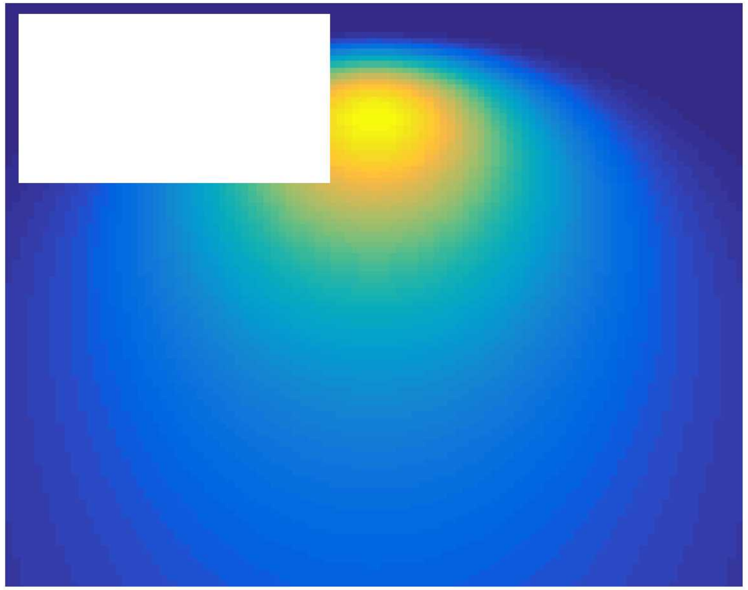
$$p\left(\theta \mid \mathcal{D}, \mathcal{M}\right) = \frac{p(\mathcal{D}|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})}$$



$$p\left(\mathcal{D} \mid \mathcal{M}\right) = \underbrace{\int p\left(\mathcal{D} \mid \theta, \mathcal{M}\right) p\left(\theta \mid \mathcal{M}\right) d\theta}_{\text{Intractable}}$$

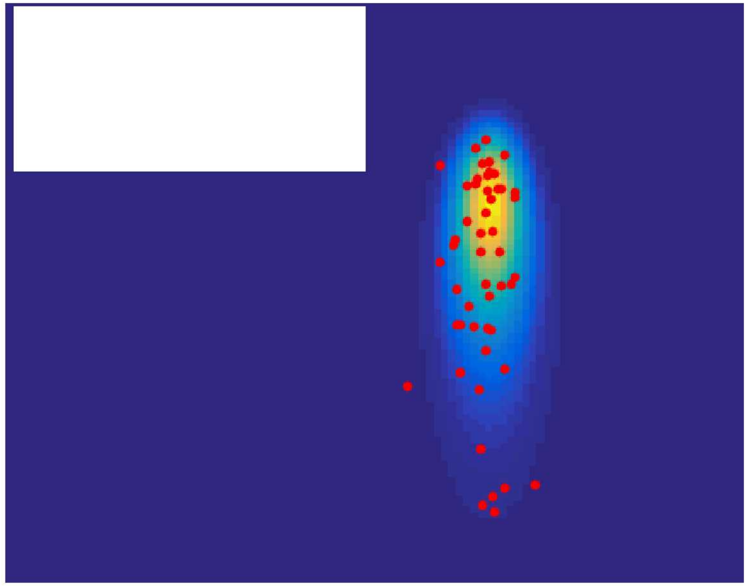
The Bayesian Inference Problem

Observations: $\mathcal{D}$



Bayes' Theorem

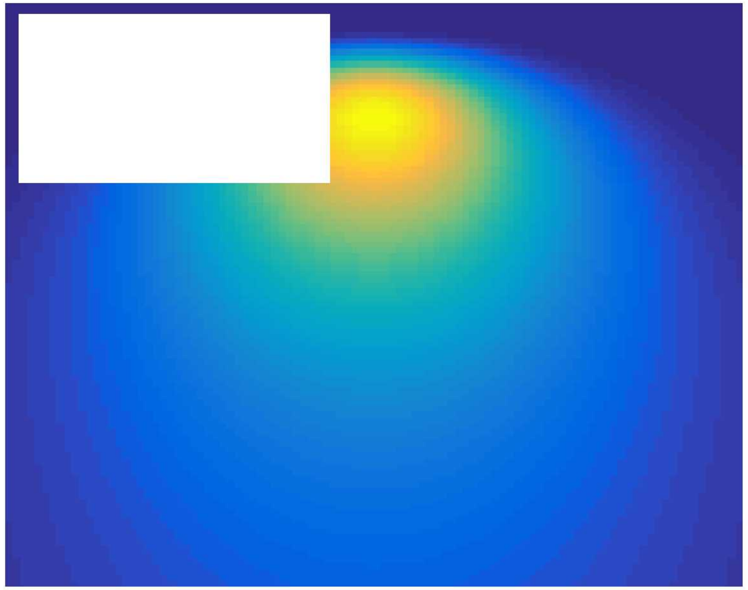
$$p\left(\theta \mid \mathcal{D}, \mathcal{M}\right) = \frac{p(\mathcal{D}|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})}$$



Posterior Estimation: $\mathbb{E}\left[g\left(\theta\right) \mid \mathcal{D}, \mathcal{M}\right] = \int g\left(\theta\right) p\left(\theta \mid \mathcal{D}, \mathcal{M}\right) d\theta \approx \frac{1}{N} \sum_{i=1}^{N} g\left(\theta_i\right)$

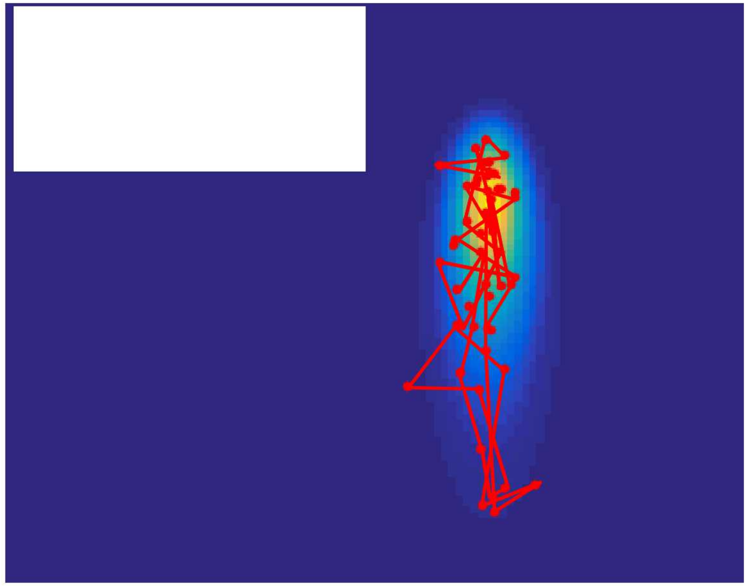
The Bayesian Inference Problem

$$\text{Observations: } \mathcal{D}$$



Bayes' Theorem

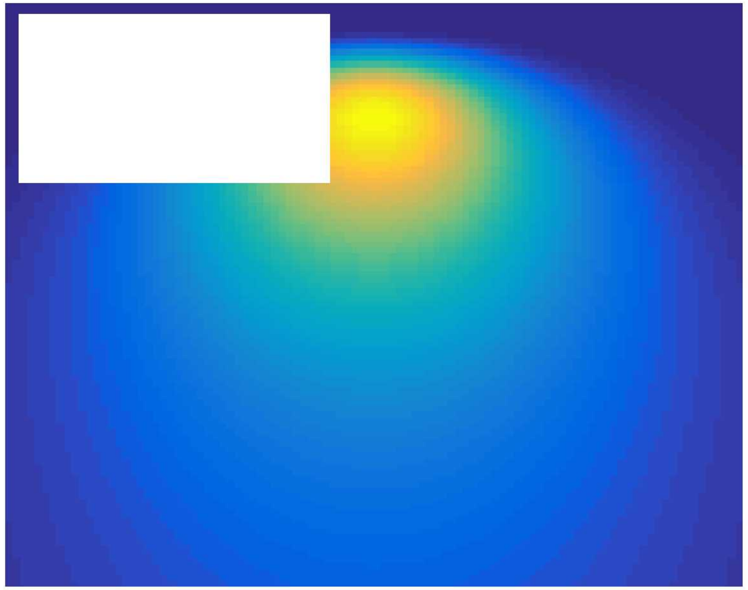
$$p(\theta \mid \mathcal{D}, \mathcal{M}) = \frac{p(\mathcal{D}|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})}$$



Exploration of the space by proposal distribution + Accept/Reject correction = Metropolis-Hastings MCMC

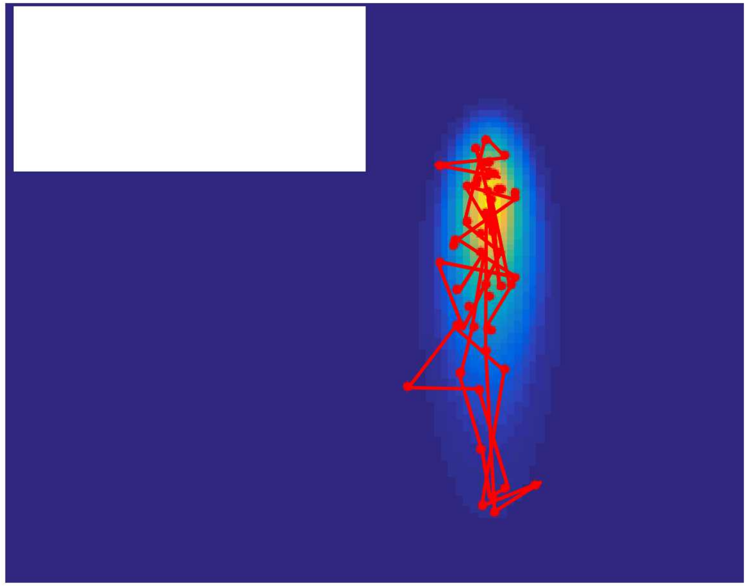
The Bayesian Inference Problem

$$\text{Observations: } \mathcal{D}$$



Bayes' Theorem

$$p(\theta \mid \mathcal{D}, \mathcal{M}) = \frac{p(\mathcal{D}|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})}$$



Posterior Estimation: $\mathbb{E}\left[g(\theta) \mid \mathcal{D}, \mathcal{M}\right] = \int g(\theta)\, p(\theta \mid \mathcal{D}, \mathcal{M})\, d\theta \approx \frac{1}{N}\sum_{i=1}^{N} g(\theta_i)$

Effective Number of Samples: $ESS\left[g(\theta_{1:N})\right] = \dfrac{var\left[g(\theta)\right]}{var\left[\frac{1}{N}\sum_{i=1}^{N} g(\theta_i)\right]}$

Metropolis-Hastings Algorithm

1. Initialize the state θ_1 randomly, usually according to the prior, set $n = 1$

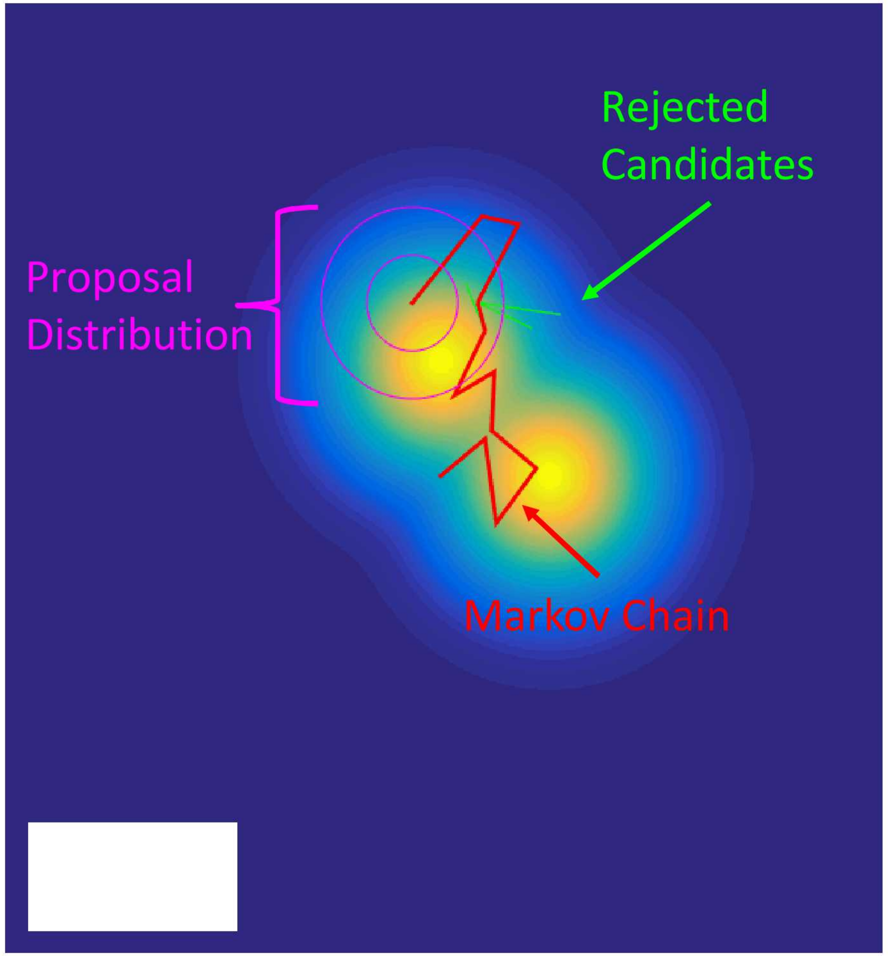
2. Pick a candidate state θ'_{n+1} according to the proposal $Q\left(\theta'_{n+1} \mid \theta_n\right)$

3. Accept or reject the candidate according to a sampled uniform variable ζ on $[0, 1]$:

$$\theta_{n+1} = \begin{cases} \theta'_{n+1} & \zeta \leq \alpha\left(\theta'_{n+1} \mid \theta_n\right) \\ \theta_n & \zeta > \alpha\left(\theta'_{n+1} \mid \theta_n\right) \end{cases}$$

$$\alpha\left(\theta' \mid \theta\right) = \min\left(1, \frac{\pi\left(\theta'\right) Q\left(\theta \mid \theta'\right)}{\pi\left(\theta\right) Q\left(\theta' \mid \theta\right)}\right)$$

4. Increment n and go to step 2



Metropolis-Hastings Algorithm

1. Initialize the state θ_1 randomly, usually according to the prior, set $n = 1$
2. Pick a candidate state θ'_{n+1} according to the proposal $Q\left(\theta'_{n+1} \mid \theta_n\right)$
3. Accept or reject the candidate according to a sampled uniform variable ζ on $[0,1]$:

$$\theta_{n+1} = \begin{cases} \theta'_{n+1} & \zeta \leq \alpha\left(\theta'_{n+1} \mid \theta_n\right) \\ \theta_n & \zeta > \alpha\left(\theta'_{n+1} \mid \theta_n\right) \end{cases}$$

$$\alpha\left(\theta' \mid \theta\right) = \min\left(1, \frac{\pi\left(\theta'\right) Q\left(\theta \mid \theta'\right)}{\pi\left(\theta\right) Q\left(\theta' \mid \theta\right)}\right)$$

4. Increment n and go to step 2

Markov chain transition kernel $K\left(\theta' \mid \theta\right)$ (steps 2–3)

Designing the Markov Chain Monte Carlo Kernel

- Sufficient requirements to guarantee $\pi(\theta)$ is the stationary distribution of the Markov chain are Reversibility and Ergodicity

- Design objectives for choosing the Kernel $K(\theta' \mid \theta)$:
  - Minimizes the convergence time (burn-in) to the stationary distribution
  - Minimizes the correlation when sampling the stationary distribution

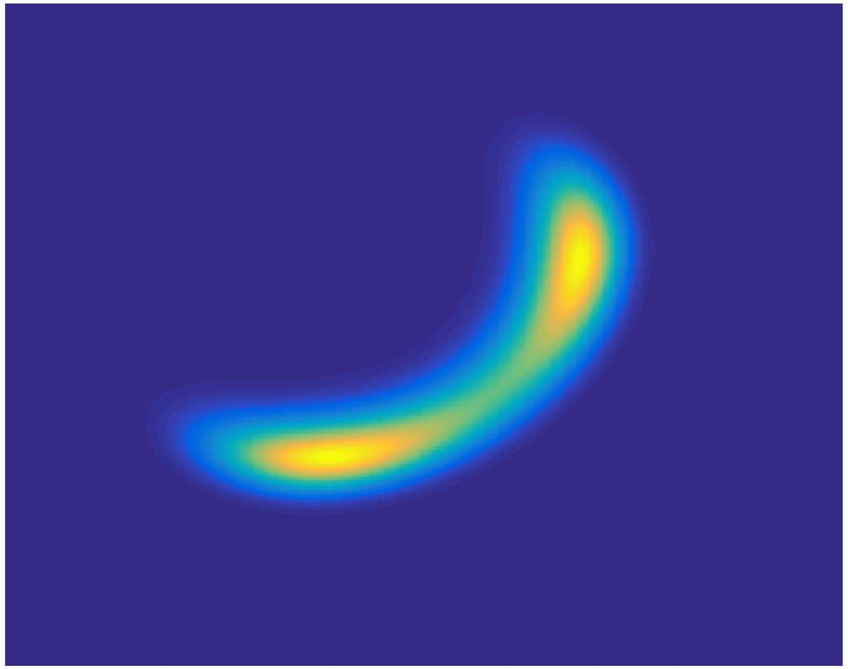
Limitations of Classic MH MCMC

- Challenging to explore complicated geometries distributions when the proposal distribution does not adapt

- Many model evaluations are necessary in high dimensions because the MH chain mixes slowly

- Parallelization and HPC are difficult because evolving the MH chain is sequential

Locally Identifiable Posterior Distribution



Parallel MCMC with Sequential Tempered MCMC

Motivation

- ST-MCMC methods use parallel chains that interact with each other to speed up convergence
- ST-MCMC methods also enable us to solve the model selection and failure probability estimation problems
- However, theoretical tools are still needed to aid in selecting algorithm parameters
- Advanced MCMC kernels could be used to enhance performance

Sequential Tempered MCMC

- ST-MCMC methods combine:
  1) Annealing: Introduce intermediate distributions
  2) MCMC: Explore the intermediate distributions
  3) Importance Resampling: Discard unlikely chains and multiply likely chains while maintaining the distribution
- Examples: SMC[1], Subset Simulation[2], TMCMC[3], AlTar/Catmip[4], AIMS[5], and AMSSA[6]

[1] Del Moral et al 2006
[2] S.K. Au and J.L. Beck 2001
[3] J. Ching and Y. C. Chen 2007
[4] J.L. Beck and K.M. Zuev 2013
[5] S.E Minson, M. Simons, J.L. Beck 2013
[6] E. Prudecio and S.H. Cheung 2012

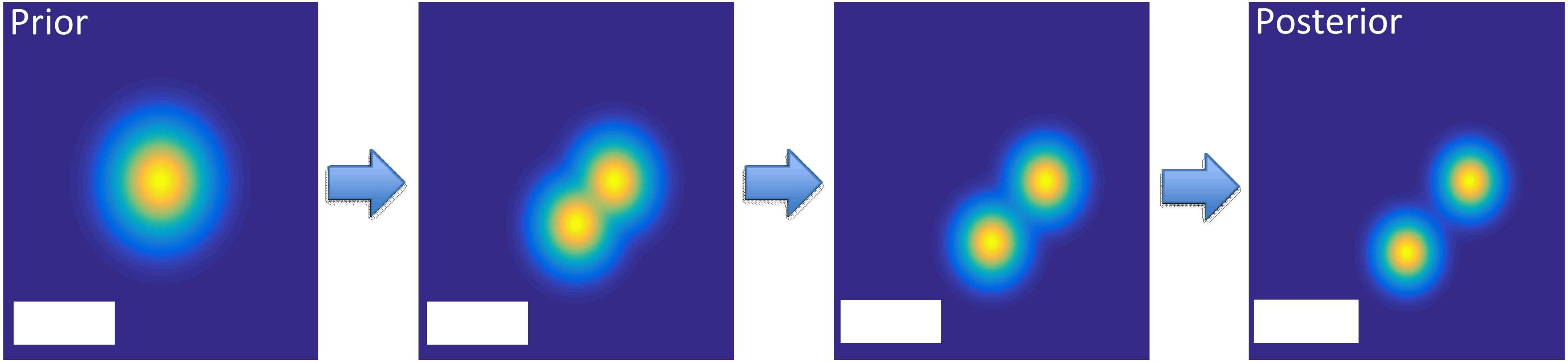
Annealing

β defines how much the data updates the intermediate distribution:

$$\pi_i(\theta) \propto p(\mathcal{D} \mid \theta, \mathcal{M})^{\beta_i}\, p(\theta \mid \mathcal{M}) \quad \beta_i \in [0,1]$$

Intermediate distributions at different β levels

Level 0: $\beta_0 = 0$ | Level 1: $\beta_1 = \beta_0 + \Delta\beta_1$ | Level 2: $\beta_2 = \beta_1 + \Delta\beta_2$ | Level n: $\beta_n = 1$



Annealing: Finding Δβ

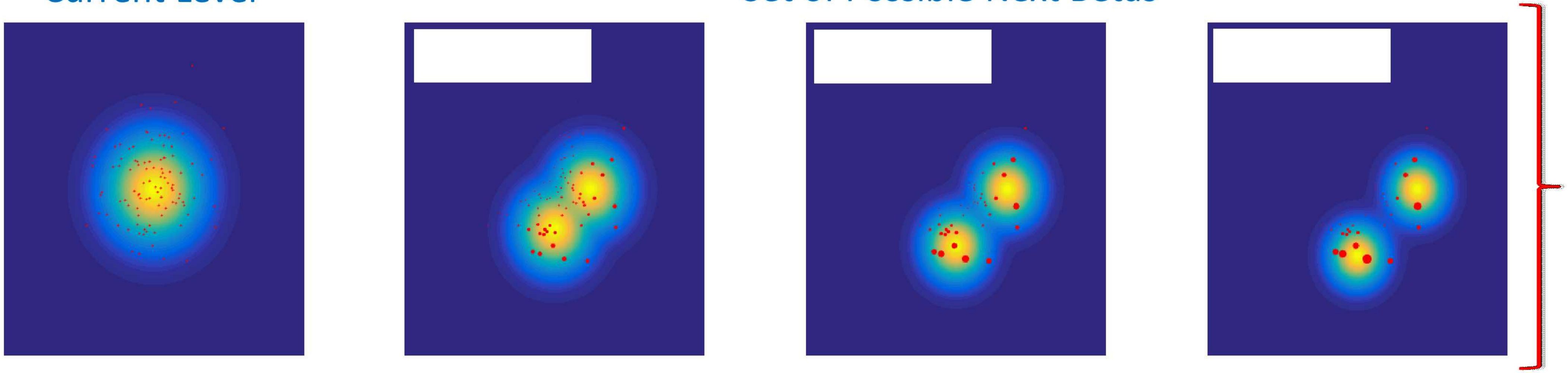
Find Δβ such that the coefficient of variation (κ) of the sample weights is 1

Sample weight: $w\left(\theta_j\right) \propto p\left(\mathcal{D} \mid \theta_j, \mathcal{M}\right)^{\Delta\beta_i}$

Coefficient of variation: $\kappa\left(w\right) = \frac{\sigma(w)}{\bar{w}}$

Current Level

Set of Possible Next Betas



Weighted Sample Populations

Importance Resampling

- Resampling the population rebalances the weights as the distribution changes. This discards unlikely samples and replicates likely samples

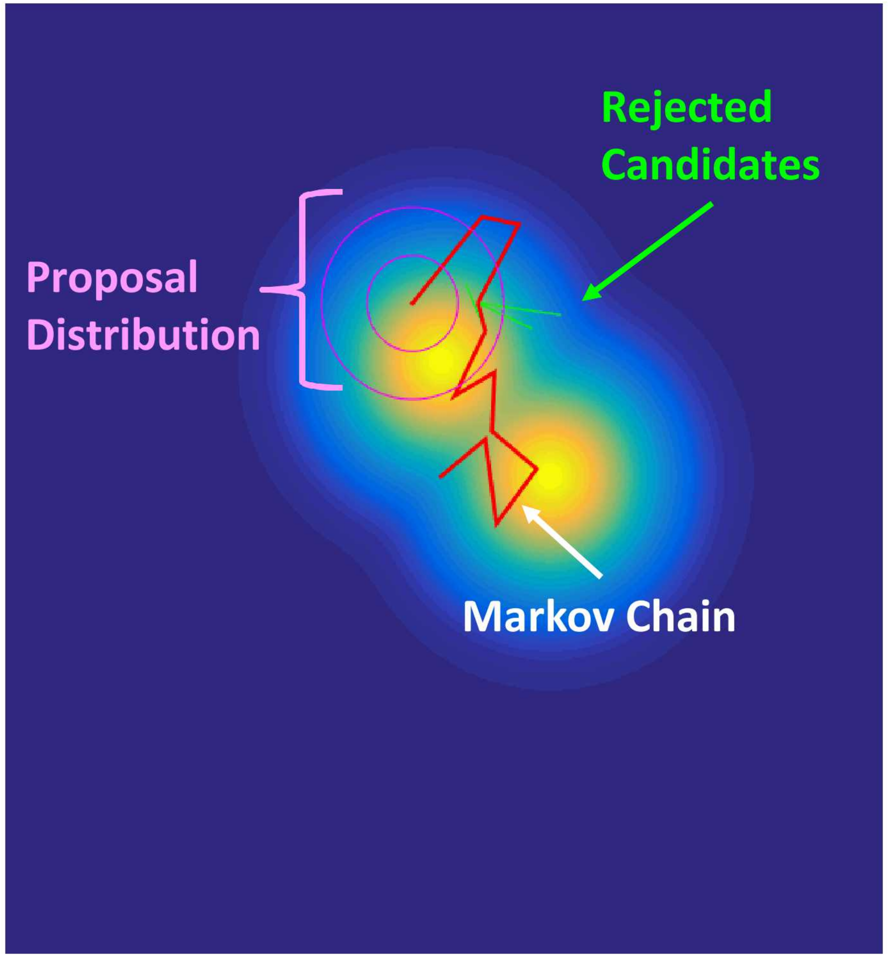
- Multinomial Resampling from level *i-1* to level *i*:

Probability of selecting sample k: $P\left(\theta_{i,j} = \theta_{i-1,k}\right) = w\left(\theta_{i-1,k}\right)$

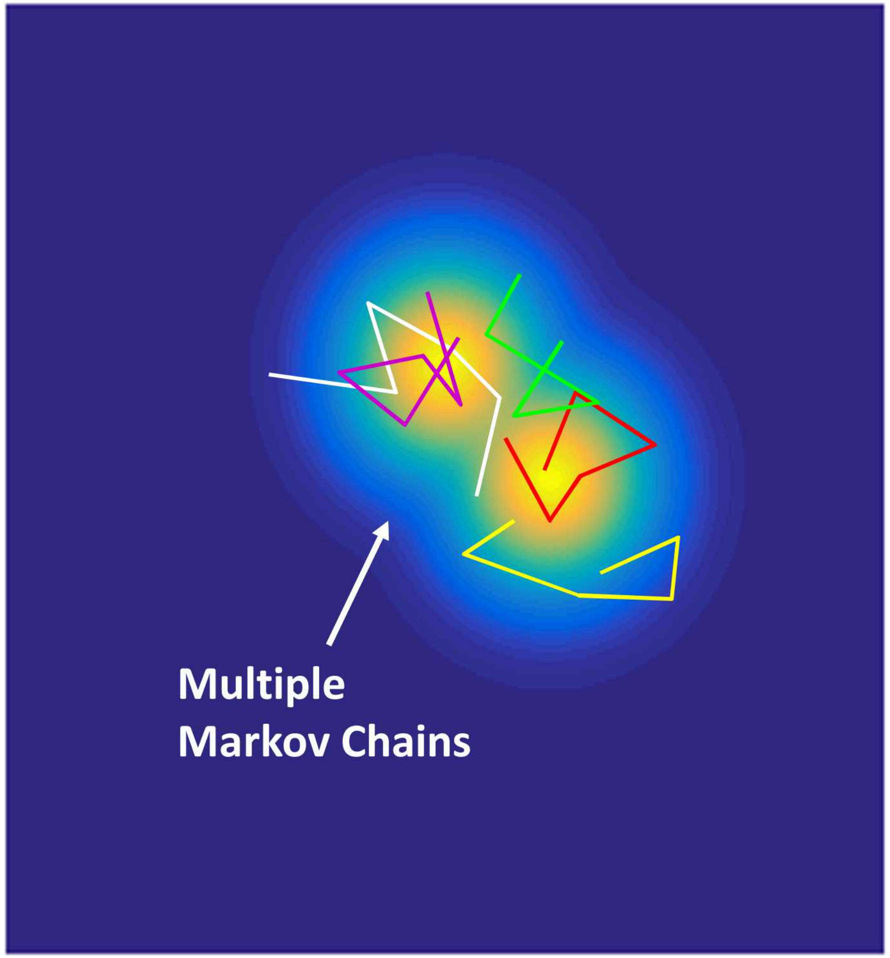
Sample weight: $w\left(\theta_{i-1,j}\right) \propto p\left(\mathcal{D} \mid \theta_{i-1,j}, \mathcal{M}\right)^{\Delta\beta_i}$

Metropolis Hastings MCMC with Parallel Chains

Single MH Markov Chain



Parallel MH Markov Chain



Designing the ST-MCMC Algorithm

- Algorithm Parameters
  - Number of parallel Markov Chains
  - Chain Length or target correlation
  - Annealing/convergence rate i.e. coefficient of variation target
- MCMC Algorithm
  - Freedom to choose the proposal distribution and its properties
  - Design of the Markov Chain kernel
- Resampling scheme for importance sampling

Contribution 1:

Theoretical results to estimate the ESS of the sample population and to choose algorithm parameters

Theoretical Study of Effective Sample Size in ST-MCMC

- We can approximate the evolution of the sample population ESS (n_k) using three MCMC parameters:

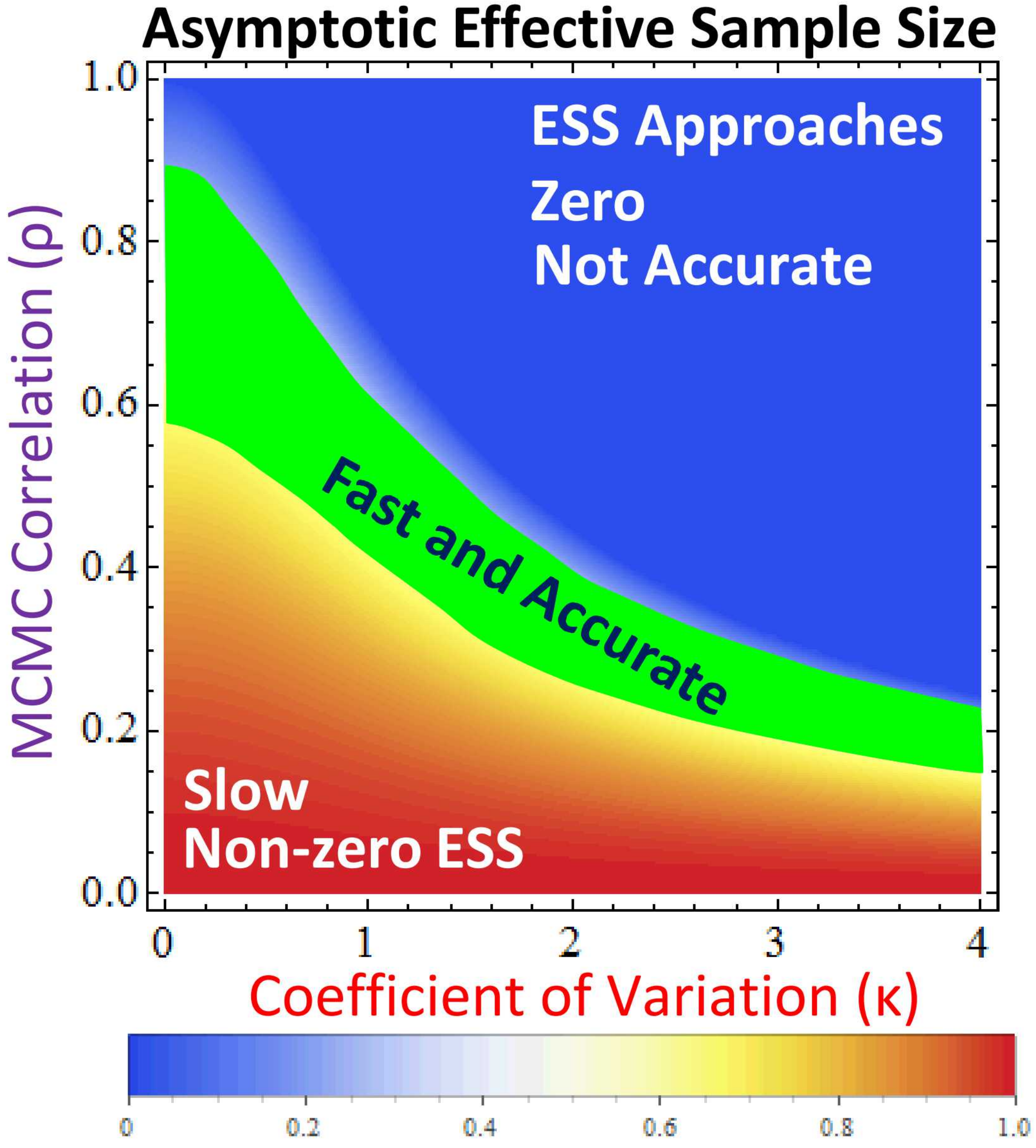
$$n_{k+1} = n_k \frac{N}{(N-1)\left(1+\kappa^2\right)\rho^2 + n_k}$$

Number of chains (N)

Coefficient of Variation (κ)

MCMC Correlation (ρ)

- Parameter estimation is possible when n_k does not asymptotically approach zero



Contribution 2:

Generalize the Modified Metropolis Algorithm (MMA[1]) to efficiently sample high dimensional distributions with constraints

[1] Au and Beck 2001

Rank One Modified Metropolis Algorithm (ROMMA)

- Sampling distributions with significant prior structure, like inequality constraints, can slow down Metropolis type algorithms in high dimensions
- Explicitly integrating prior constraint information into the MCMC proposal can rapidly improve mixing

ROMMA Description

Step k:

for $i = 1$ **to** N_{steps} **do**

- **Draw** $P = P_+$ or $P = P_-$ — Randomly choose forward or reverse ordering of components
- **Draw** $\xi \sim \mathcal{N}(0, I_{N_d})$
- **Set** $R = PSP^T$ — Compute the transformed components
- **Set** $\hat{\theta} = \theta^i$
- **for** $j = 1$ **to** N_d **do**
  - $\tilde{\theta} = \hat{\theta} + PR_j\xi_j$ — Perform rank one update
  - **Accept** $\hat{\theta} = \tilde{\theta}$ with prob. $\min\left[\frac{\pi(\tilde{\theta})}{\pi(\hat{\theta})}, 1\right]$ — Accept or Reject rank one update according to prior
- **end**
- **Accept** $\theta^{i+1} = \hat{\theta}$ with prob. $\min\left[\frac{p(\mathcal{D}|\hat{\theta})}{p(\mathcal{D}|\theta^i)}, 1\right]$ — Accept or Reject full update according to the data

end

S is $\sqrt{\Sigma}$ where Σ is the covariance

P_+ and P_- choose the ordering of the components

N_d is the number of components

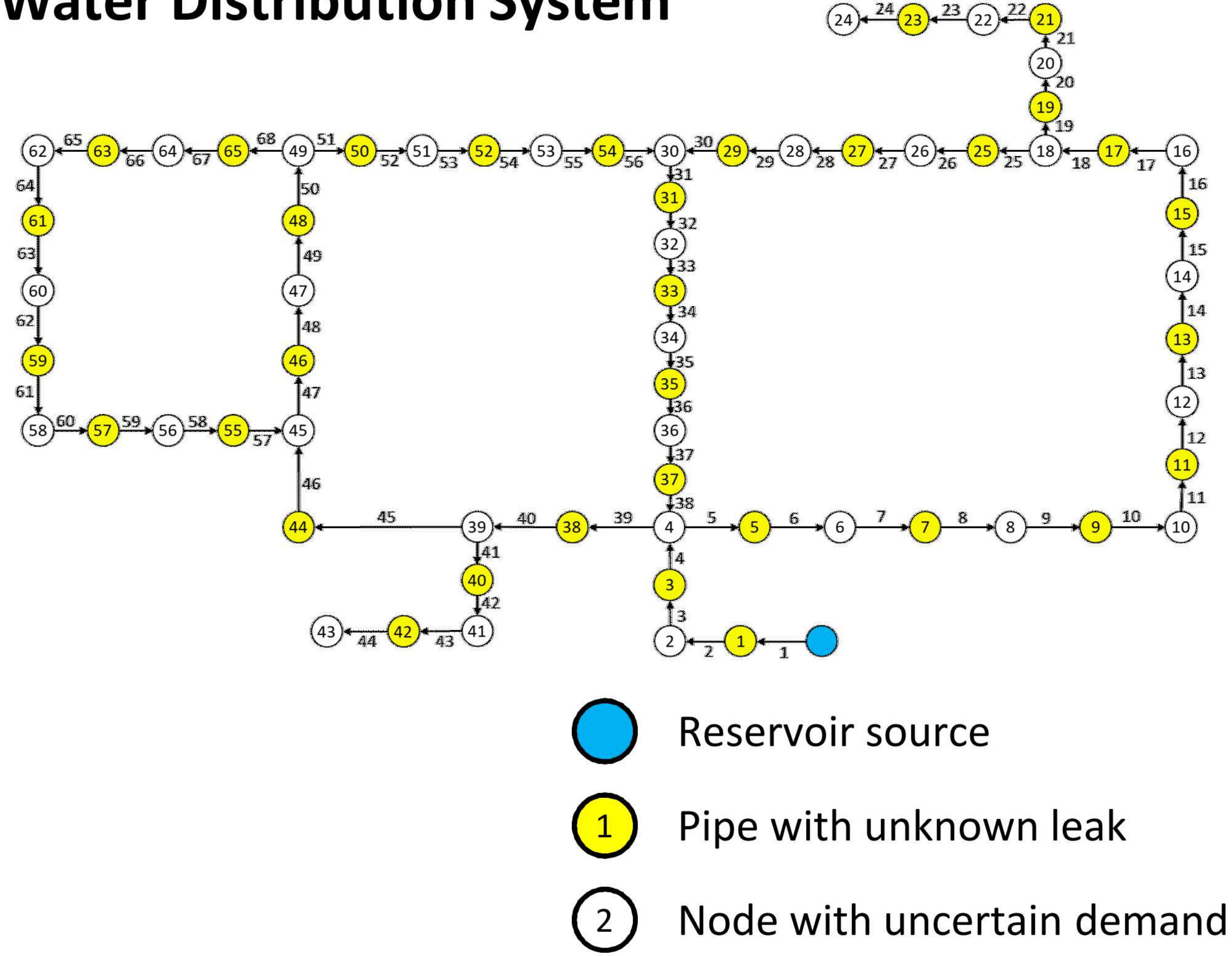
N_{steps} is the number of steps in the Markov chain

Water Distribution System Reliability

Problem Formulation:

- Estimate the probably of not meeting minimum pressure requirements
- Uncertain demands, leak positions, and leak sizes
- Data is available giving the node pressures under different loading conditions

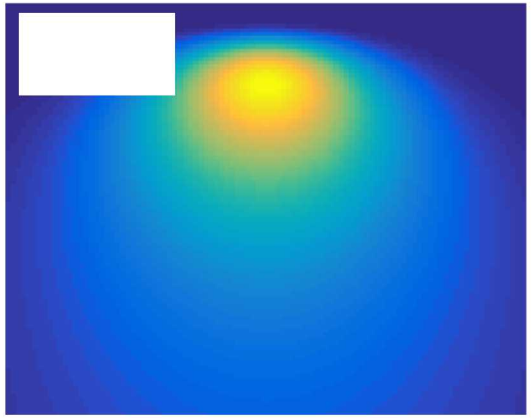
Water Distribution System



Water System Reliability Analysis

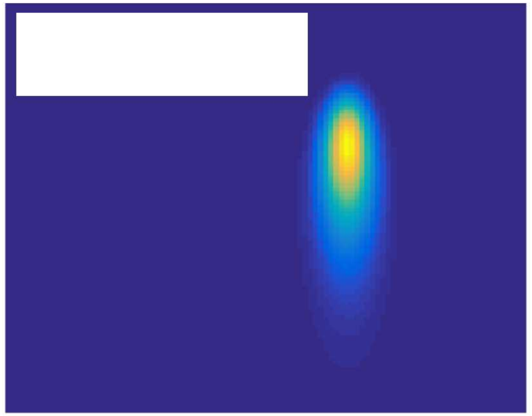
Sandia National Laboratories

Prior distribution of the water system parameters



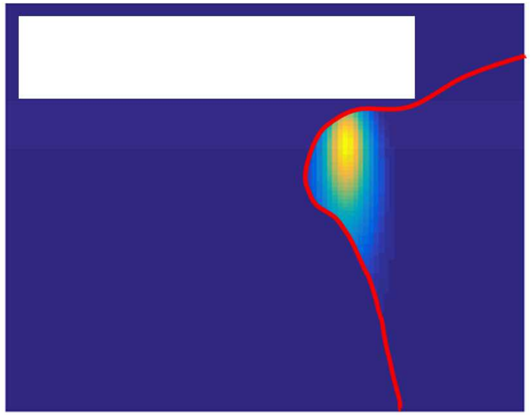
Data

Posterior distribution of the water system parameters



Failure

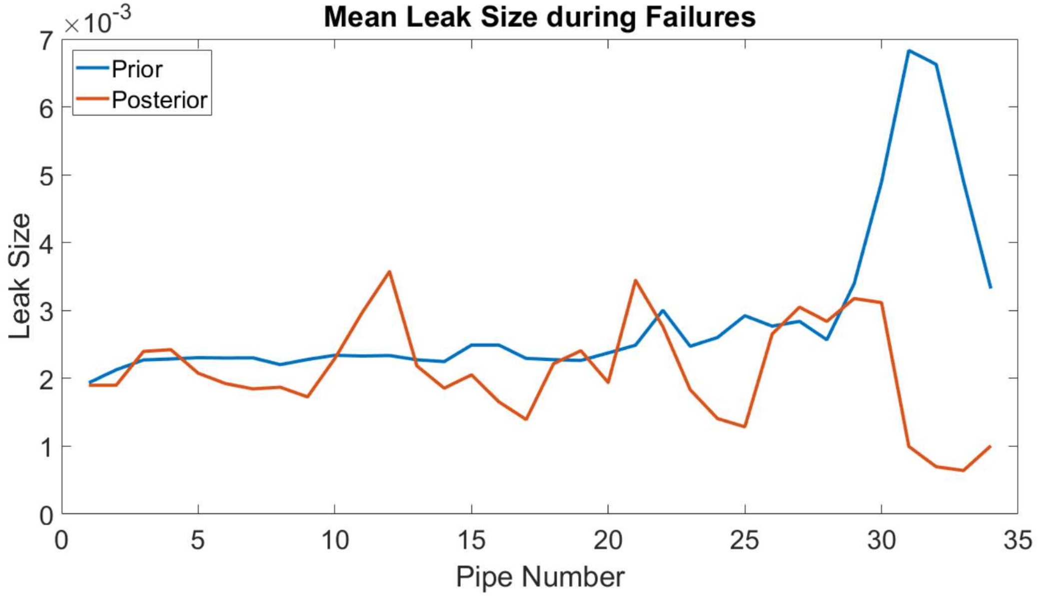
Posterior distribution of failed water system parameters

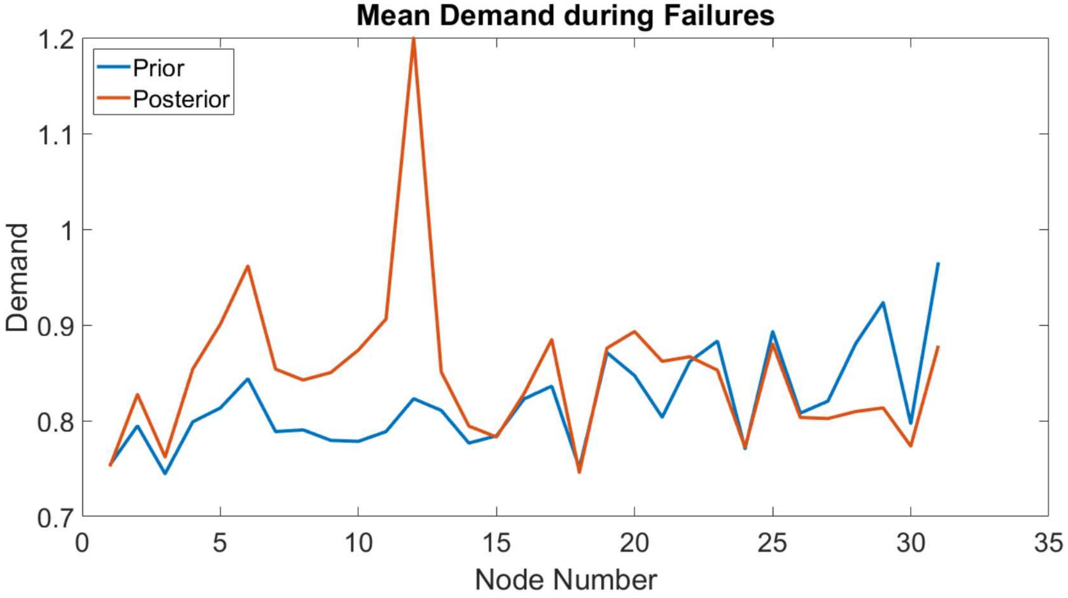


Posterior Estimate of Failure Probability

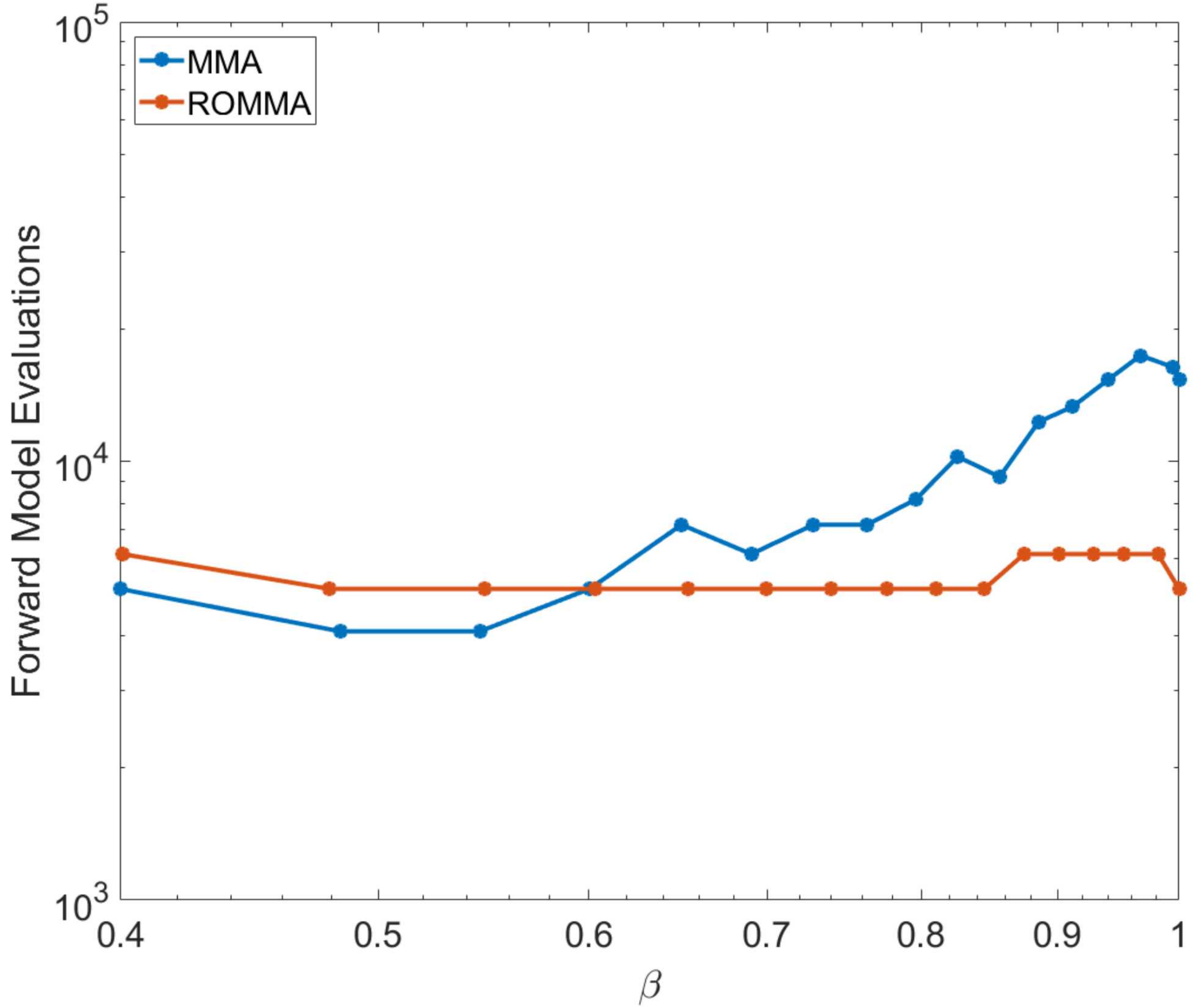
Water System Reliability Results

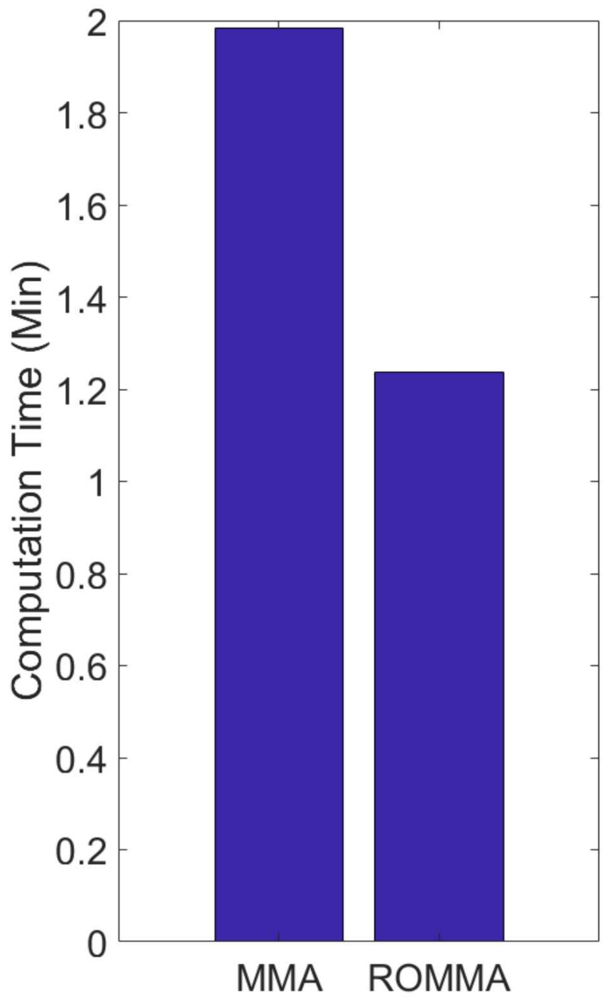
	MMA/RWM ST-MCMC Computational Time (min)	ROMMA ST-MCMC Computation Time (min)
Prior Reliability (1.5×10^{-5})	2.0	1.2
Posterior Inference	605.5	20.3
Posterior Reliability (3.0×10^{-7})	206.0	36.4



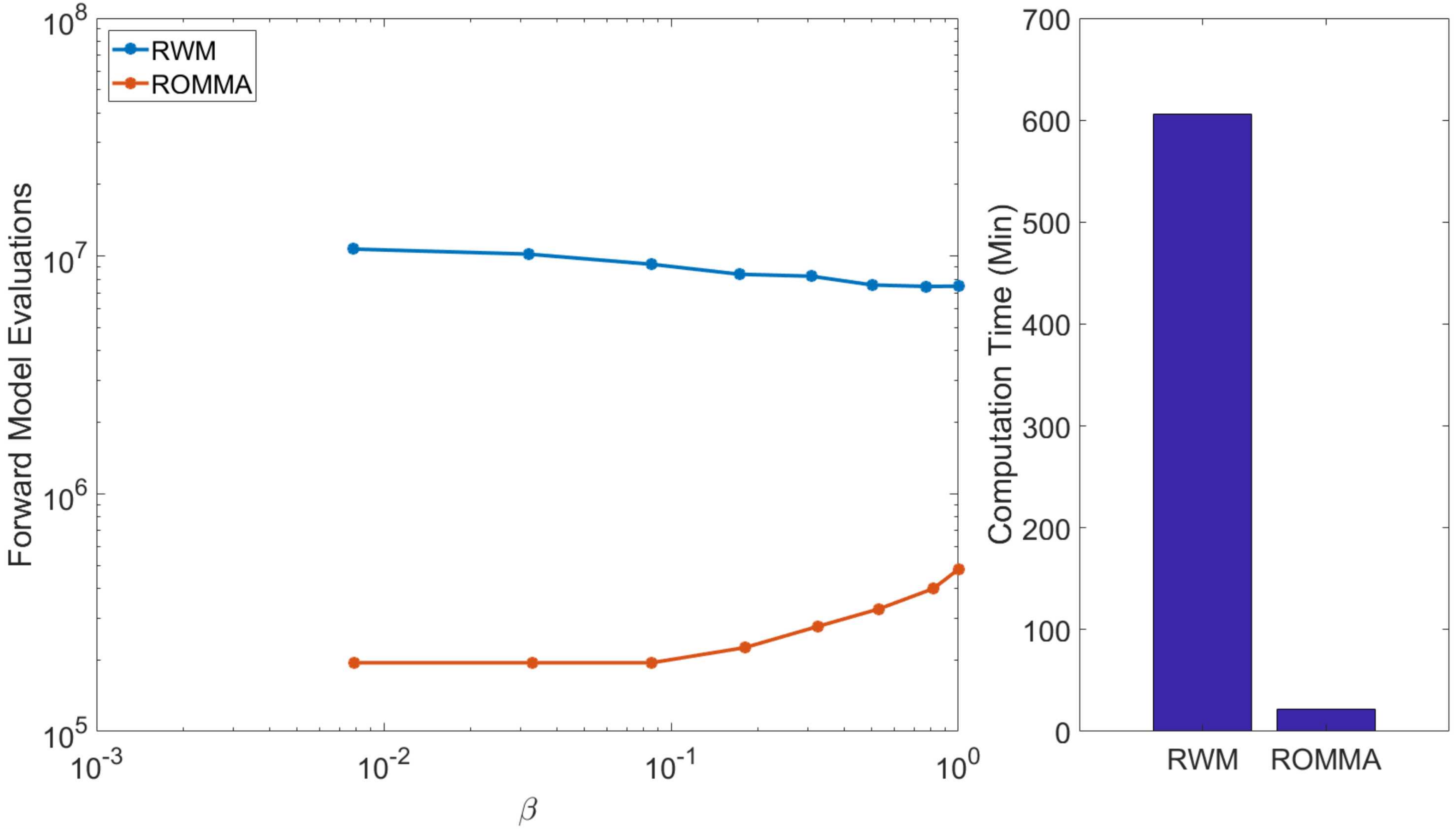


Prior Reliability Comparison

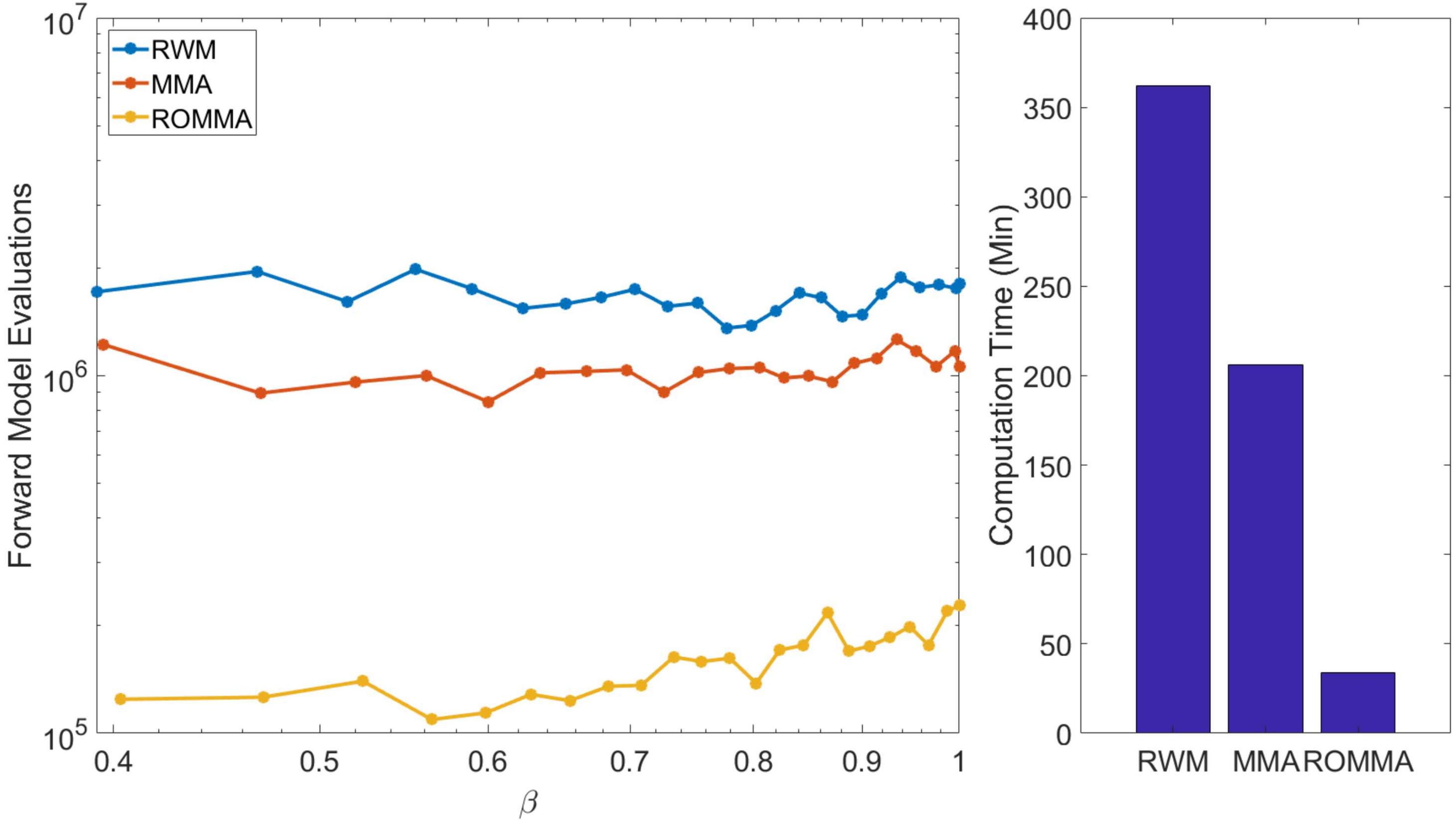




Posterior Sampling Comparison



Posterior Reliability Comparison



Future Directions for ST-MCMC

- Using the sample population to build a better estimate of the global properties of the posterior distribution to learn a more efficient MCMC proposal
- Combining Sequential Tempering with Multilevel-Multifidelity Hierarchies to reduce computational cost
- Better metrics for assessing correlation e.g. Canonical Correlation Analysis (CCA)

Conclusion

- Bayesian inference naturally expresses problems in system identification and uncertainty quantification

- Sequential Tempered MCMC methods improve efficiency and parallelism when solving System Identification and Posterior Reliability Problems

- MCMC proposals that incorporate knowledge about the prior or posterior can significantly help ST-MCMC algorithms scale