SAND2018-7036C

Accelerating Sequential Tempered MCMC for Fast Bayesian Inference and Uncertainty Quantification

Thomas A. Catanach

MCQMC 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
  - Posterior Reliability
  - Model Model Selection
- Sequential Tempered MCMC
- Accelerating ST-MCMC using ROMMA
- Water Distribution System Posterior Reliability Analysis
- Identifying context for a Biological Circuit
- Conclusion

Bayesian Methods

Motivation

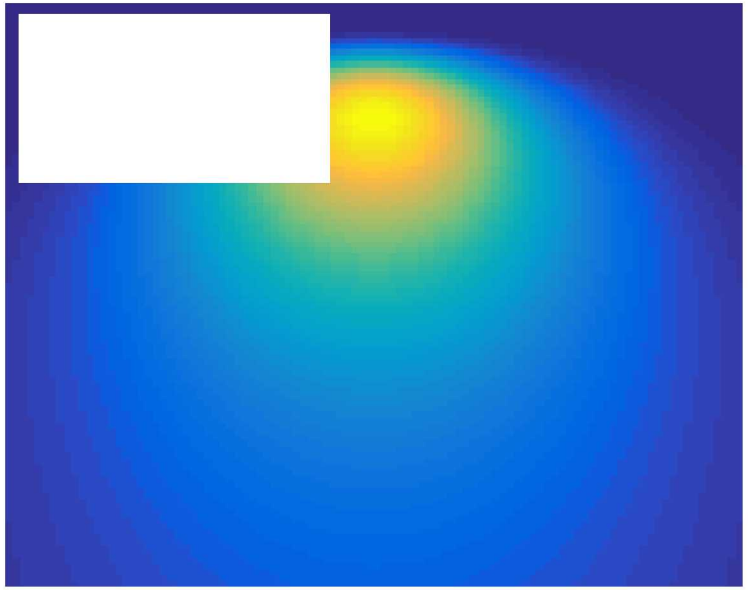
- The Bayesian Perspective:
  - Probability distributions quantify uncertainty due to insufficient information
- Bayesian methods for identification and estimation are critical to the robust system analysis

Goal:

Provide MCMC methods for computationally intensive Bayesian inference problems like model selection and posterior rare events analysis.

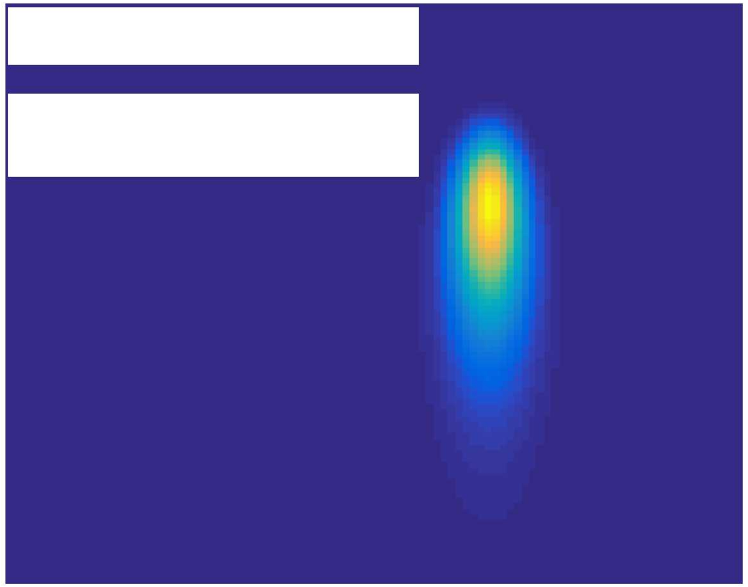
The Bayesian Inference Problem

Observations: $\mathcal{D}$



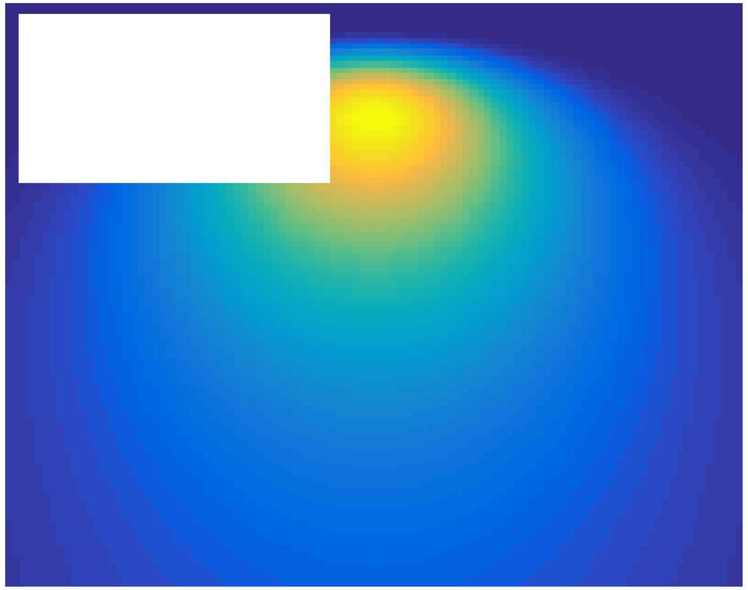
Bayes' Theorem

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



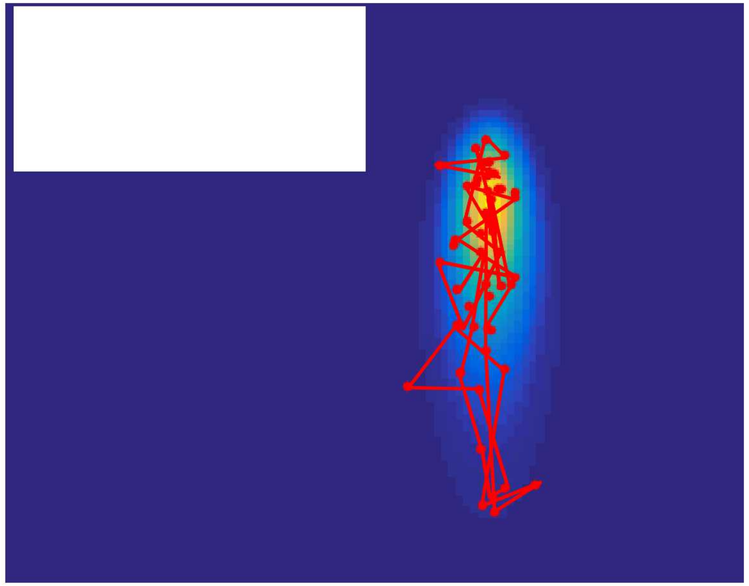
The Bayesian Inference Problem

Observations: $\mathcal{D}$



Bayes' Theorem

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



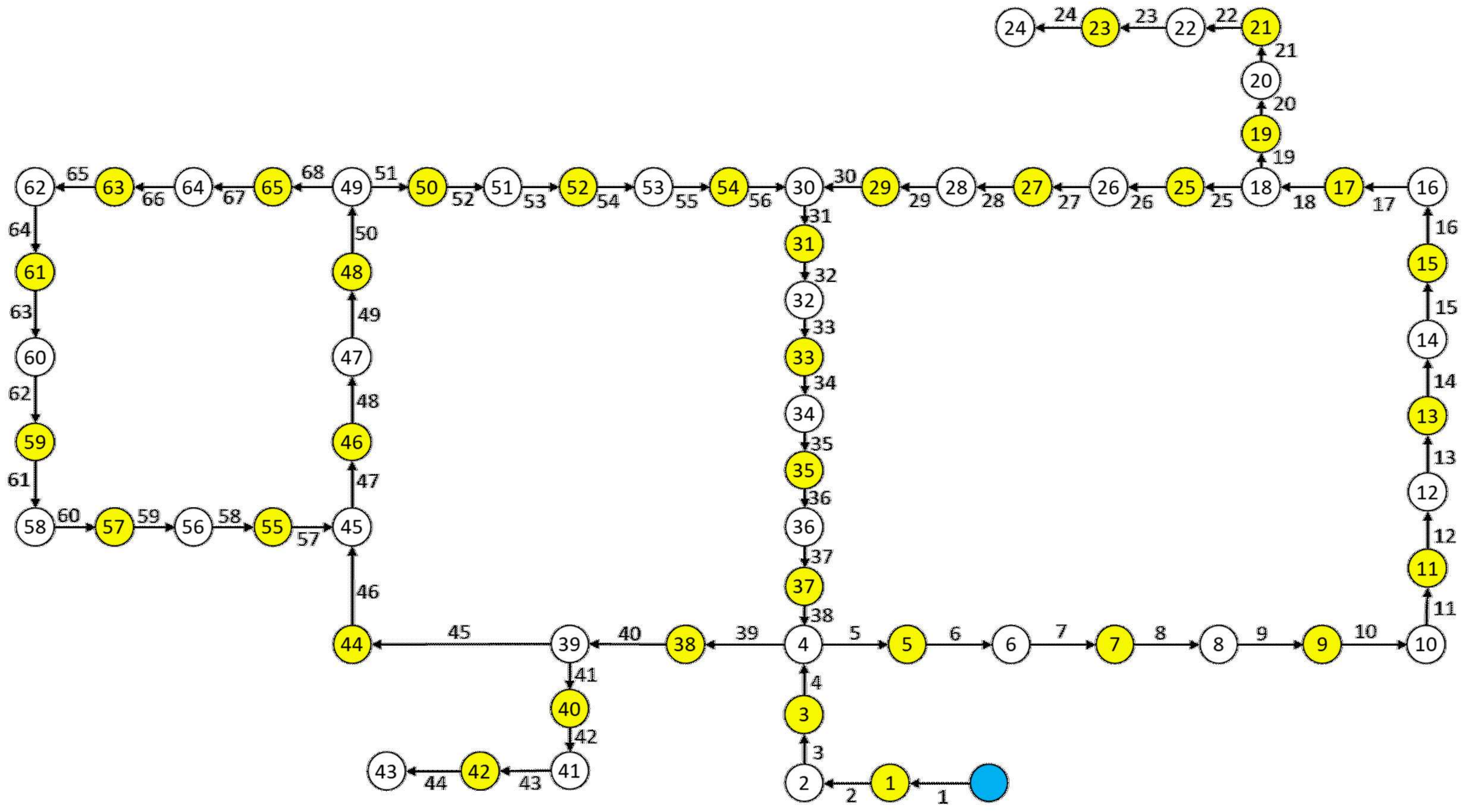
Posterior Estimation: $\mathbb{E}\left[g\left(\theta\right) \mid \mathcal{D}, \mathcal{M}\right] = \int g\left(\theta\right) p\left(\theta \mid \mathcal{D}, \mathcal{M}\right) d\theta \approx \frac{1}{N}\sum_{i=1}^{N} g\left(\theta_i\right)$

Model Evidence: $p\left(\mathcal{D} \mid \mathcal{M}\right) = \int p\left(\mathcal{D} \mid \theta, \mathcal{M}\right) p\left(\theta \mid \mathcal{M}\right) d\theta$

Target Application Problems

Example Inference Problem: Water Distribution[1]

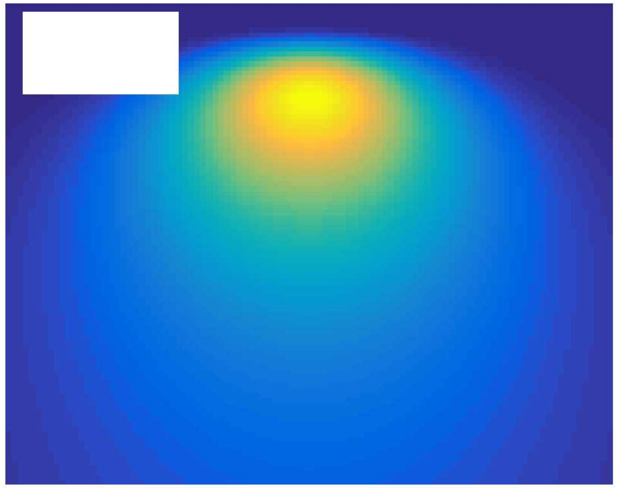
Leak Detection and Posterior Failure Probability Assessment



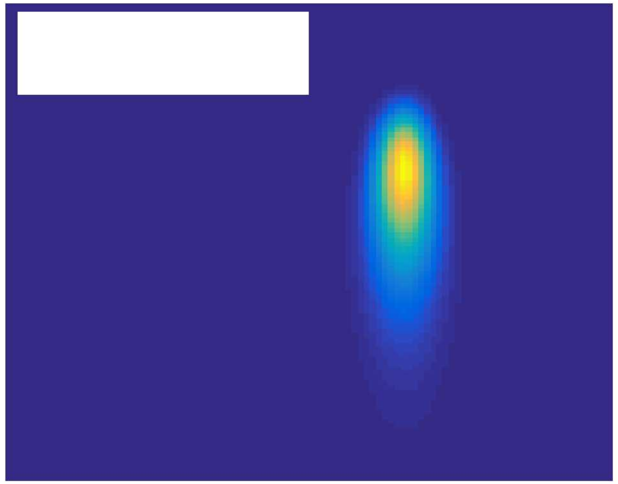
[1] Cunha and Sousa 1999

Example Inference Problem: System Identification

Prior distribution of the water system parameters



Posterior distribution of the water system parameters

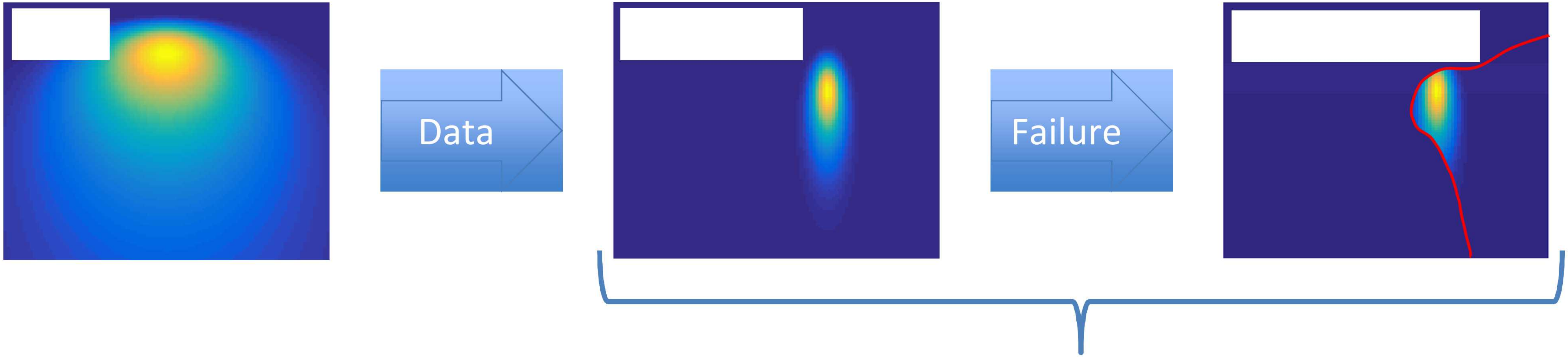


Example Inference Problem: Reliability Analysis

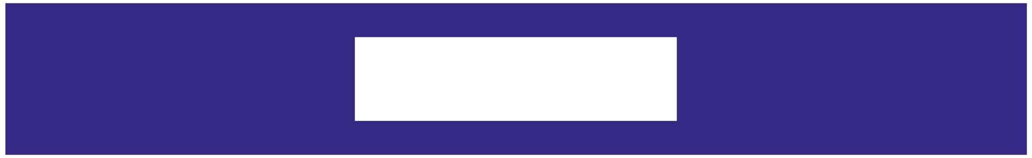
Prior distribution of the water system parameters

Posterior distribution of the water system parameters

Posterior distribution of failed water system parameters

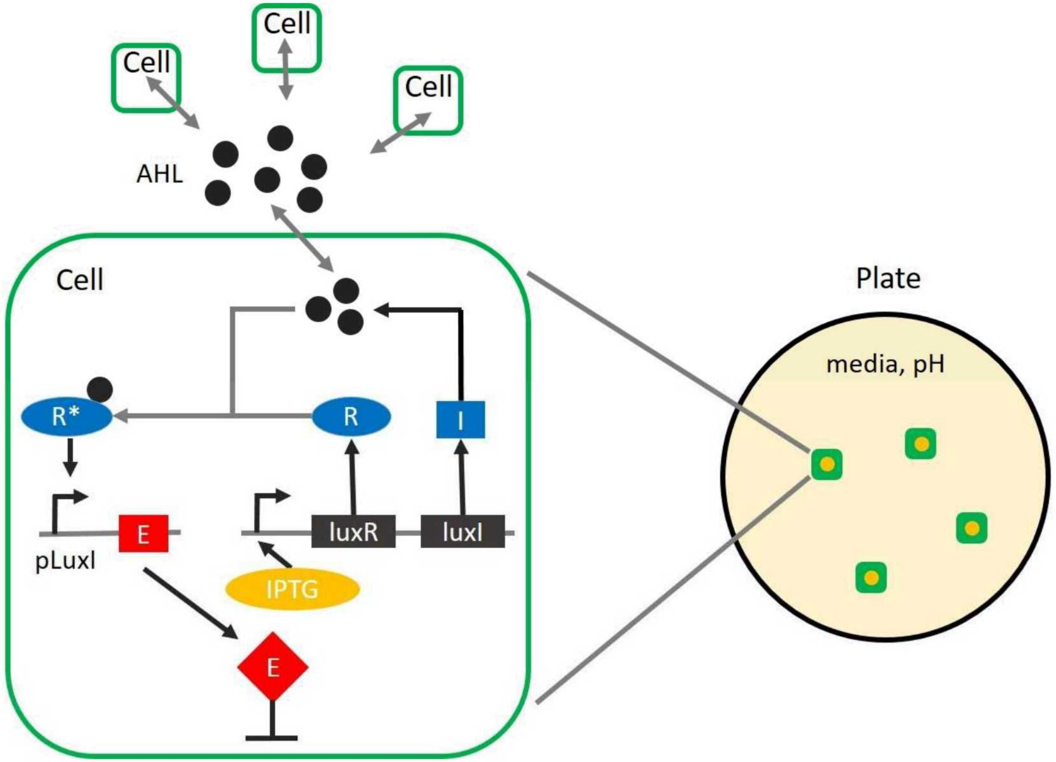


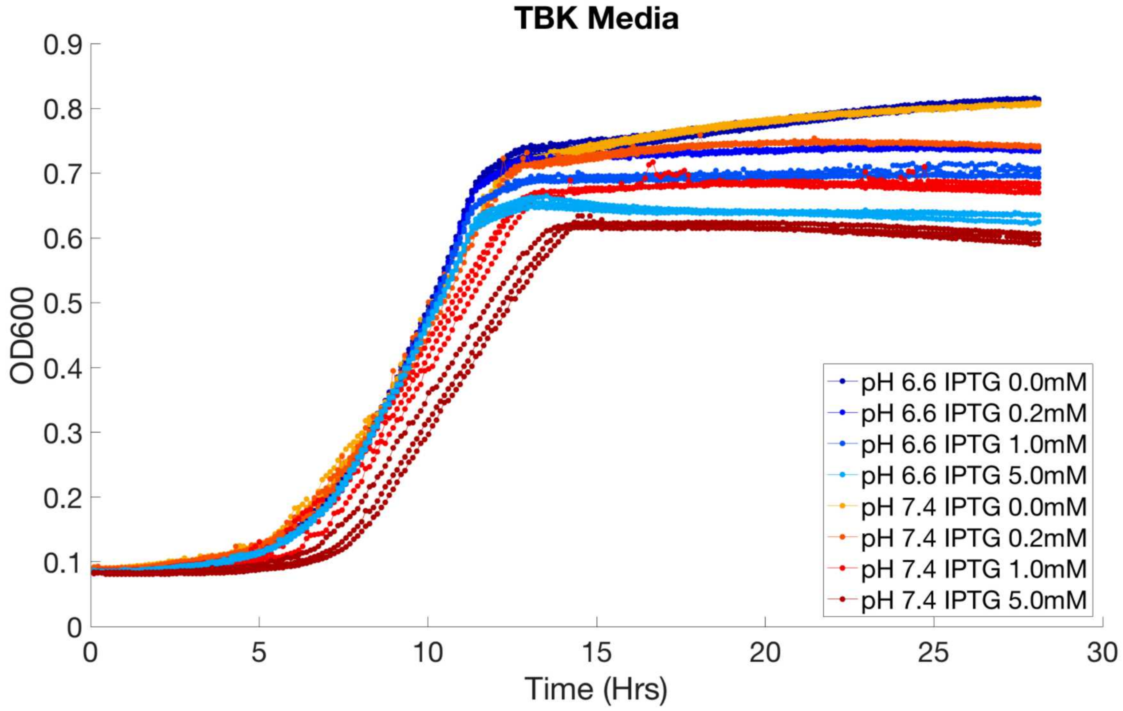
Posterior Estimate of Failure Probability



Example Inference Problem: Model Selection

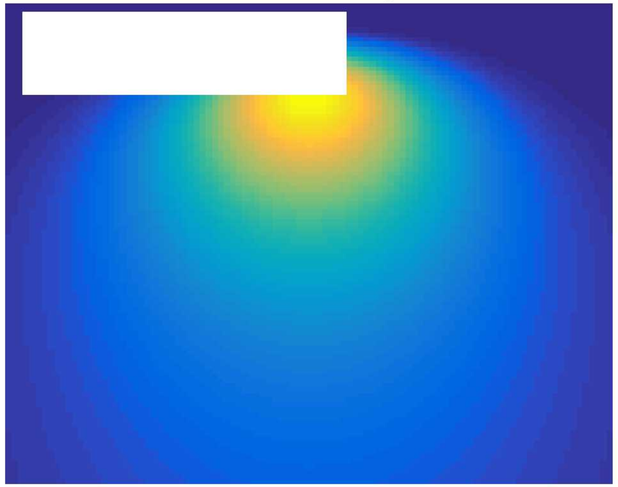
Inferring biological context for a growth control synthetic bio-circuit





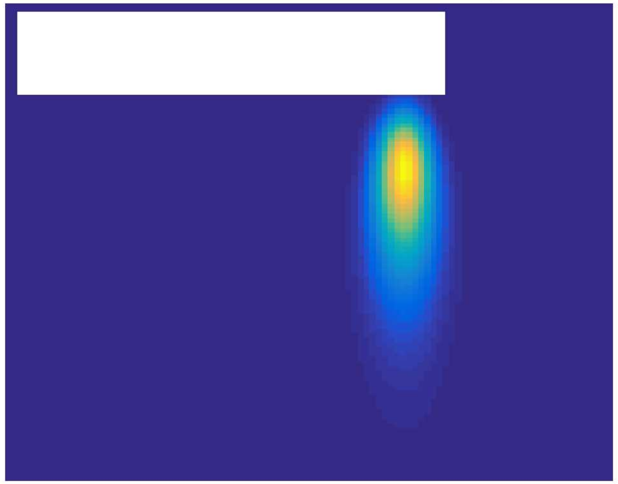
Example Inference Problem: Model Selection

Prior distribution of the bio-circuit model parameters



Data

Posterior distribution of the bio-circuit model parameters



Probability of a bio-circuit context model

Solving Nested Bayesian Inference Problems

Nested Sampling for Model Selection and Rare Events

Model Class Probability:

$$p(\mathcal{M} \mid z) = \frac{p(z \mid \mathcal{M})\, p(\mathcal{M})}{p(z)} \propto \left(\int p(z \mid \theta, \mathcal{M})\, p(\theta \mid \mathcal{M})\, d\theta \right) p(\mathcal{M})$$

Model Class Evidence:

$$\int p(z \mid \theta, \mathcal{M})\, p(\theta \mid \mathcal{M})\, d\theta = \prod_{i=1}^{l} c_i$$

Decompose into evidence for intermediate distributions

ith Distribution Evidence Estimate:

$$c_i = \int p(z \mid \theta, \mathcal{M})^{\Delta\beta_i} \frac{p(z \mid \theta, \mathcal{M})^{\beta_{i-1}}\, p(\theta \mid \mathcal{M})}{\prod_{j=1}^{i-1} c_j} d\theta \approx \frac{1}{N} \sum_{k=1}^{N} p(z \mid \theta_{i-1,k}, \mathcal{M})^{\Delta\beta_i}$$

Level i Likelihood — Level i prior — Monte Carlo Estimate

Motivation

- ST-MCMC methods use parallel chains that interact with each other to speed up convergence
- ST-MCMC methods evolve to the posterior through a series of intermediate distributions which enable them to solve the model selection and failure probability estimation problems
- Advanced MCMC kernels could be used to enhance performance for problems with constraints and where prior information is important

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

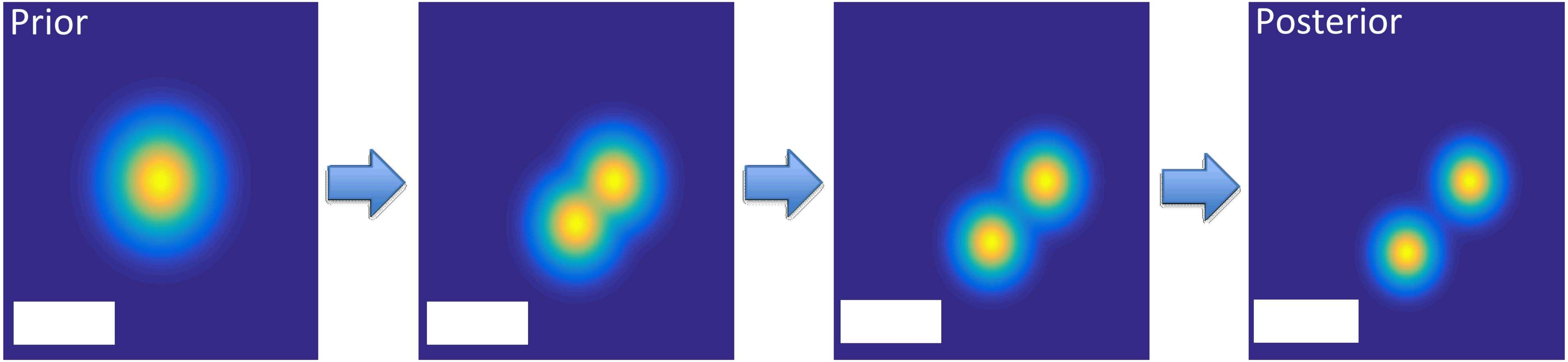
Annealing

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

Importance Resampling

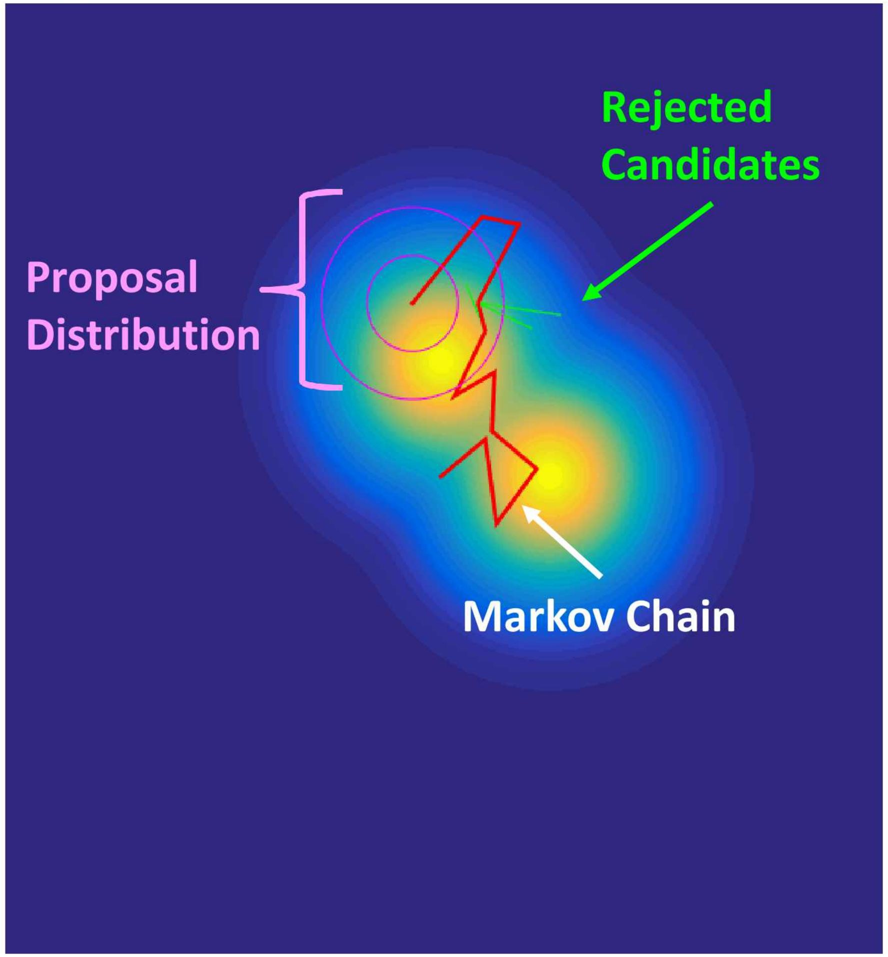
- Resampling the population rebalances the weights as the distribution changes. This discards unlikely samples and replicates likely samples
- Multinomial Resampling from level *i-1* to level *i*:

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

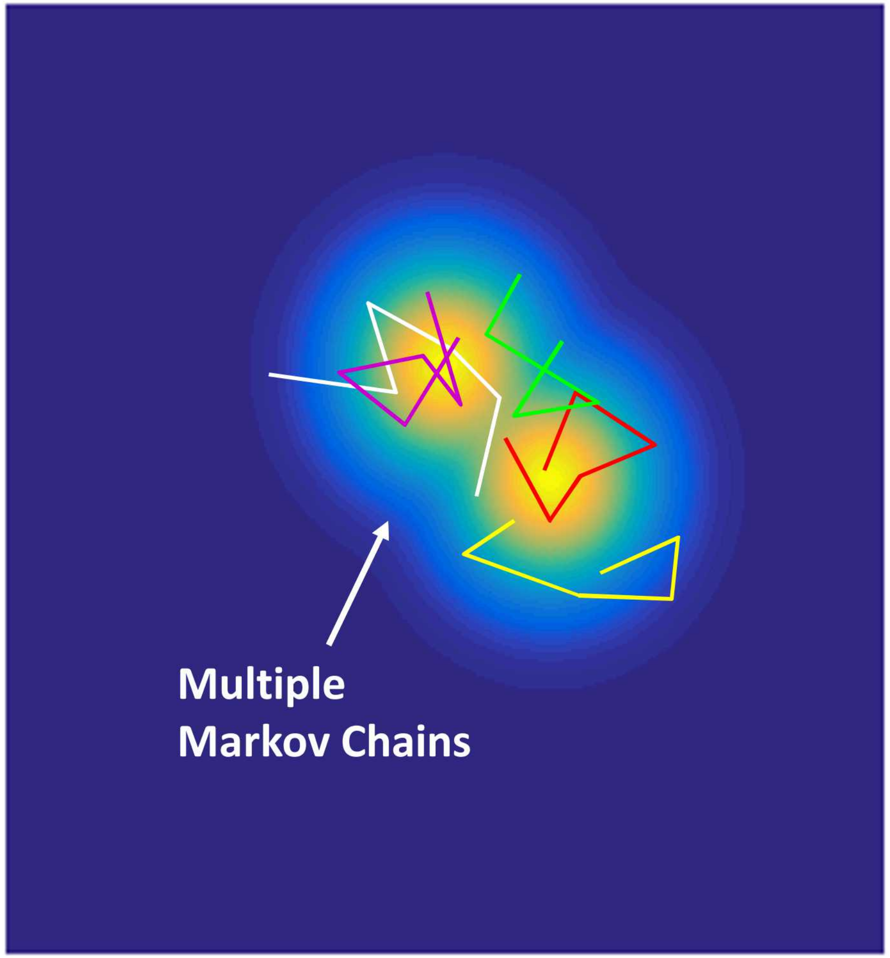
Sample weight: $w\left(\theta_{i-1,j}\right) \propto p\left(\mathcal{D} \mid \theta_{i-1,j}, \mathcal{M}\right)^{\Delta\beta_i}$

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

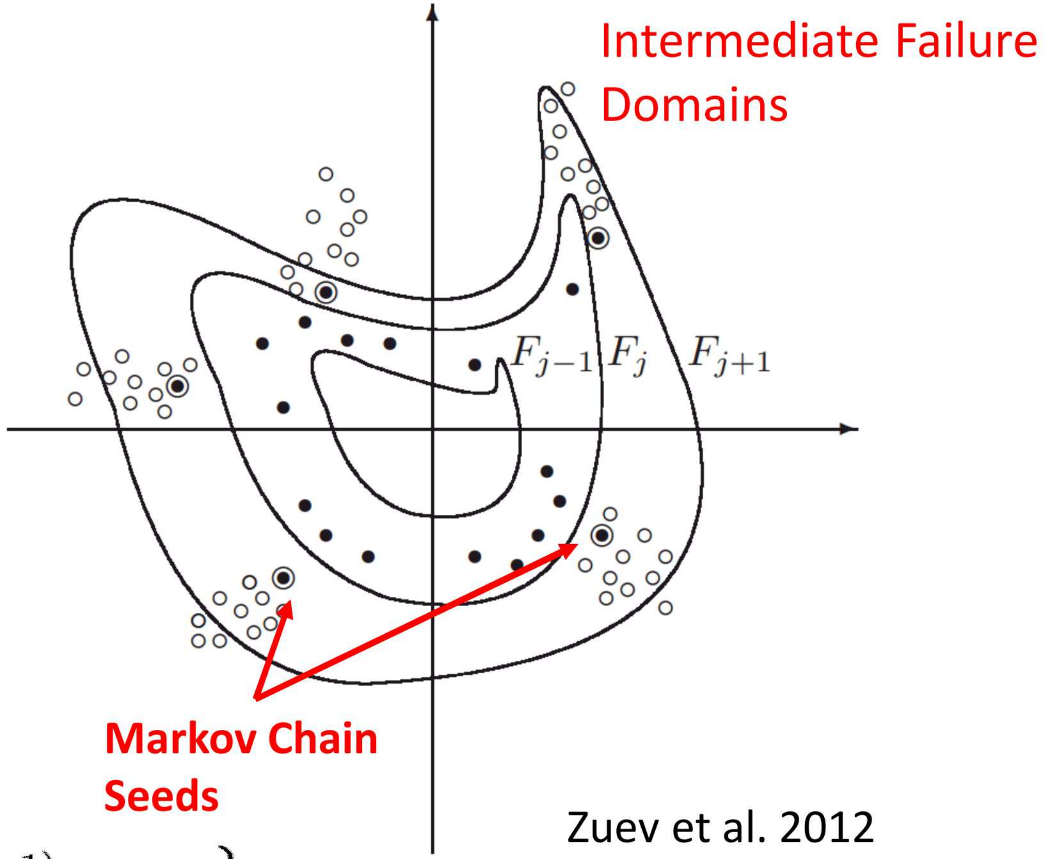
Relationship to Rare-Events algorithms like Subset Simulation [1]

- For estimating rare event probabilities, intermediate levels are defined using level sets of the failure function

$$P(\mathcal{F} \mid \mathcal{M}) = \int \mathbb{1}\{\theta \in \mathcal{F}\}\, p(\theta \mid \mathcal{M})\, d\theta$$

$$= \prod_{k=1}^{s} \frac{\int \mathbb{1}\{\theta \in \mathcal{F}_{\beta_k}\}\, p(\theta \mid \mathcal{M})\, d\theta}{\int \mathbb{1}\left\{\theta \in \mathcal{F}_{\beta_{k-1}}\right\} p(\theta \mid \mathcal{M})\, d\theta}$$

$$= \prod_{k=1}^{s} c_k$$



Zuev et al. 2012

$$c_k = \int \mathbb{1}\{\theta \in \mathcal{F}_{\beta_k}\}\, p\left(\theta \mid \mathcal{F}_{\beta_{k-1}}, \mathcal{M}\right) d\theta \approx \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\left\{\theta_i^{(k-1)} \in \mathcal{F}_{\beta_k}\right\}$$

[1] Au and Beck 2001

Rank-One Modified Metropolis Algorithm

Modified Metropolis Algorithm

- The Modified Metropolis Algorithm (MMA[1]) was developed to efficiently sample high dimensional distributions where prior information is important.
  - Nested problems where intermediate distributions are close to the prior
  - Unidentifiable inference problems
  - Priors that enforce constraints
- MMA builds up a candidate sample component-wise using prior information which speeds up sampling.
- MMA is a form of delayed acceptance algorithm where the proposal is optimized to explore the prior which is informative about the posterior.

[1] Au and Beck 2001

MMA Description

Step k:
for $i = 1$ **to** N_{steps} **do**
  Draw $\xi \sim \mathcal{N}(0, I_{N_d})$
  Set $\hat{\theta} = \theta^i$
  for $j = 1$ **to** N_d **do**
    Perform a component-wise update
    Accept or Reject component-wise update according to prior
  end
  Accept or Reject full update according to the data
end

Assumes indpendent prior: $\pi(\theta) = \prod_{j=1}^{N_d} \pi_j(\theta_j)$
σ_j is the proposal standard deviation of the j^{th} component

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

Rank-one Modified Metropolis Algorithm

- The Rank-one Modified Metropolis Algorithm (ROMMA) extends MMA to handle general priors and correlated proposal distributions.
- Instead of component-wise updates, ROMMA makes a series of rank-one updates according to chose basis.
- By being able to handle correlations ROMMA performs well on both prior and posterior problems i.e. posterior rare-events and we see significant performance gains over MMA or RWM.

ROMMA Description

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

Randomly choose forward or reverse ordering of components

Draw $\xi \sim \mathcal{N}(0, I_{N_d})$

Compute the transformed components

Set $\tilde{\theta} = \theta^i$
for $j = 1$ **to** N_d **do**

Perform rank one update

Accept or Reject rank one update according to prior

end

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

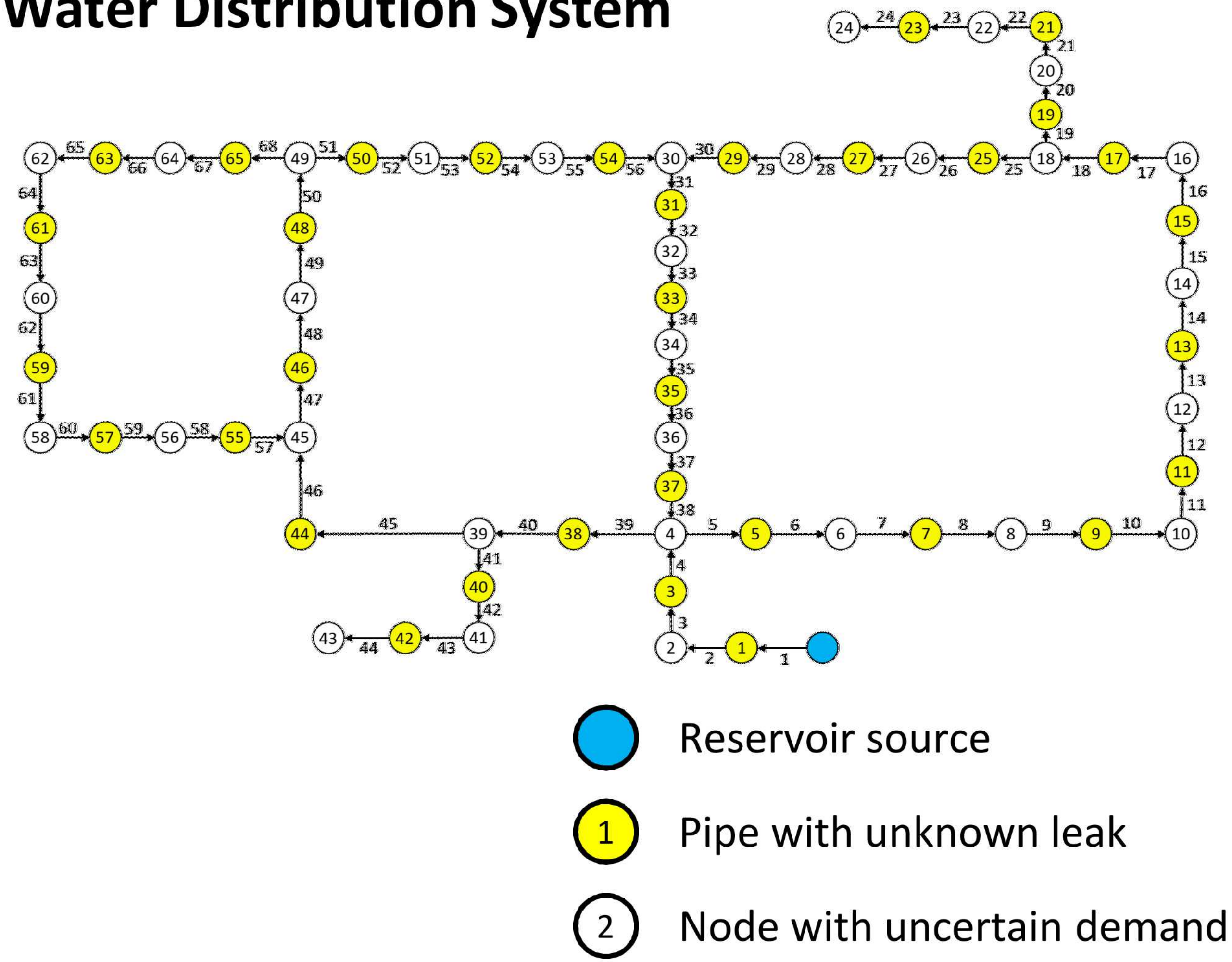
Water Distribution System Reliability: Finding Rare Events

Water Distribution System Reliability

Problem Formulation:

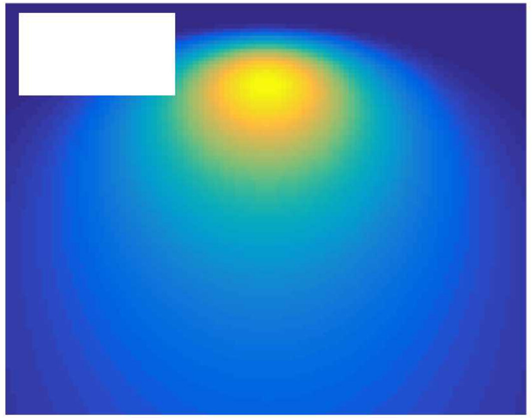
- Estimate the probably of not meeting minimum pressure requirements
- Uncertain demands, leak positions, and leak sizes
- Data is available giving the node pressures under different loading conditions

Water Distribution System



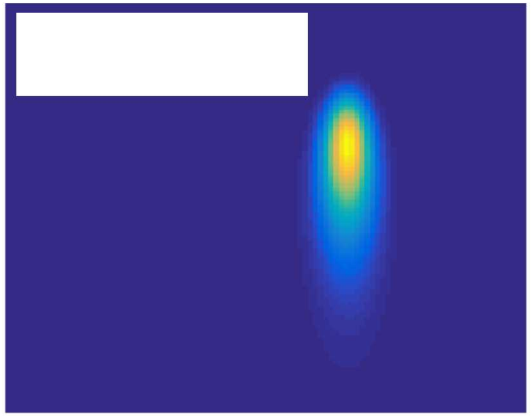
Water System Reliability Analysis

Prior distribution of the water system parameters



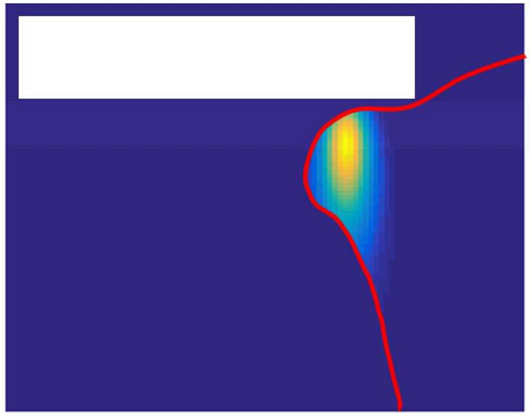
Data

Posterior distribution of the water system parameters

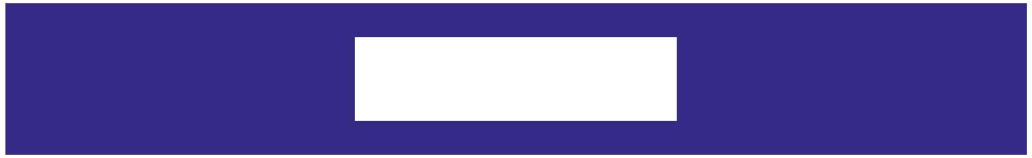


Failure

Posterior distribution of failed water system parameters

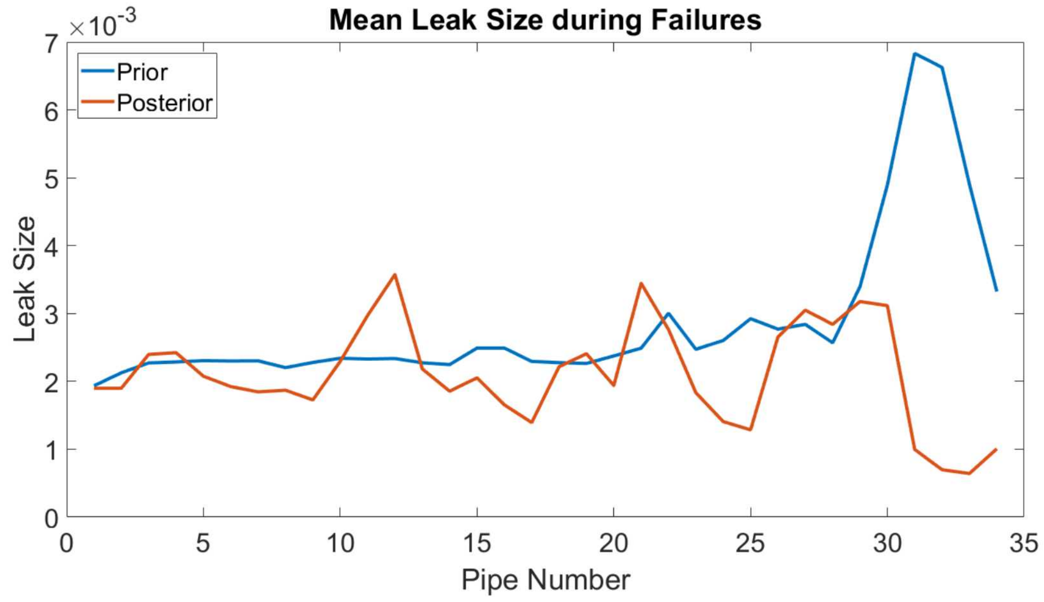


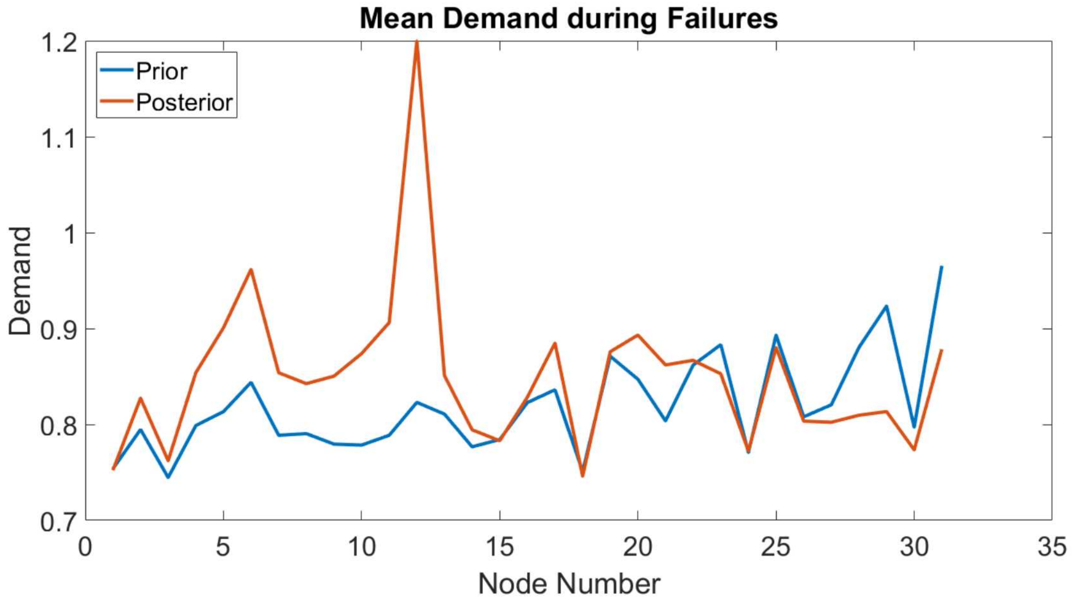
Posterior Estimate of Failure Probability



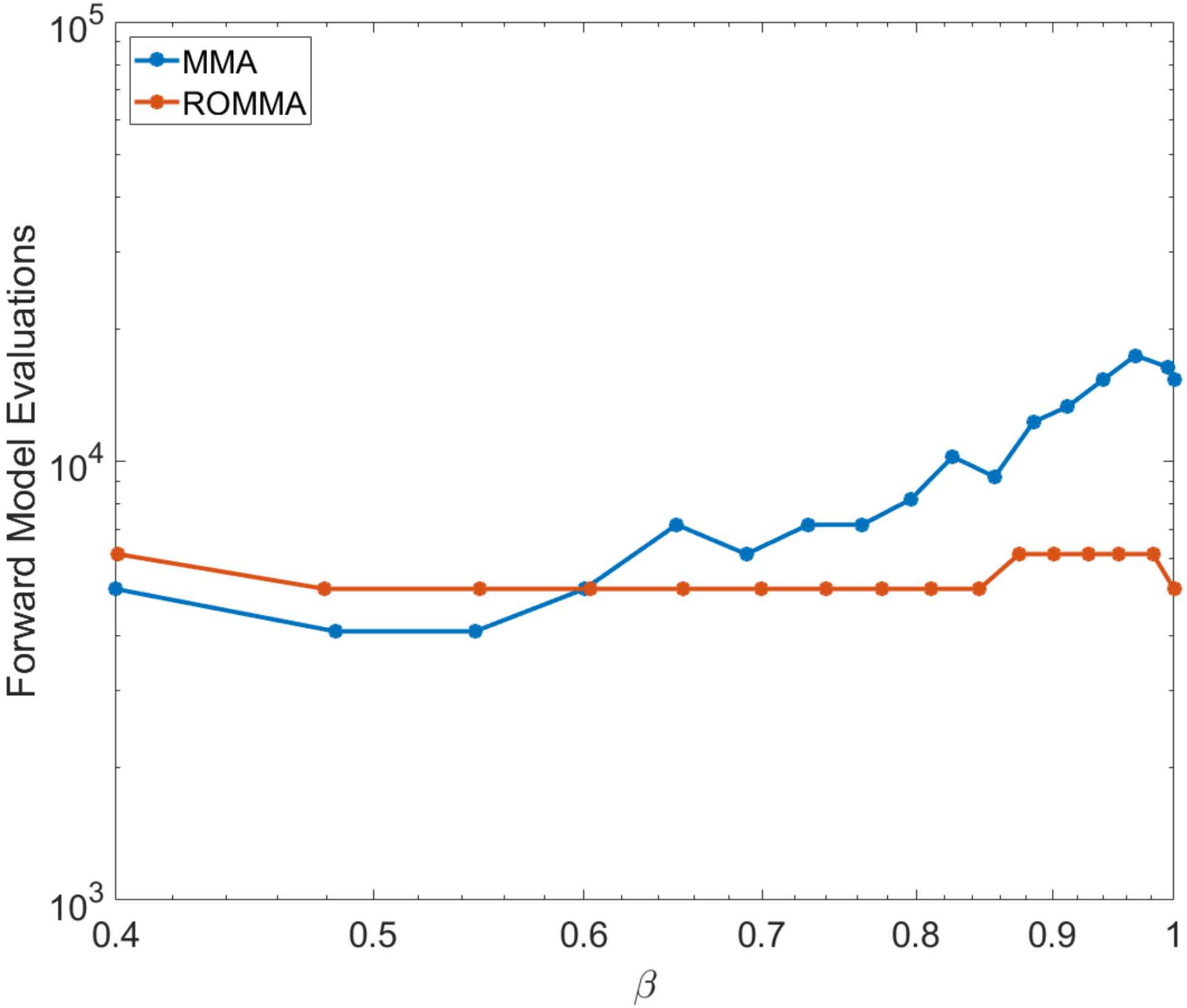
Water System Reliability Results

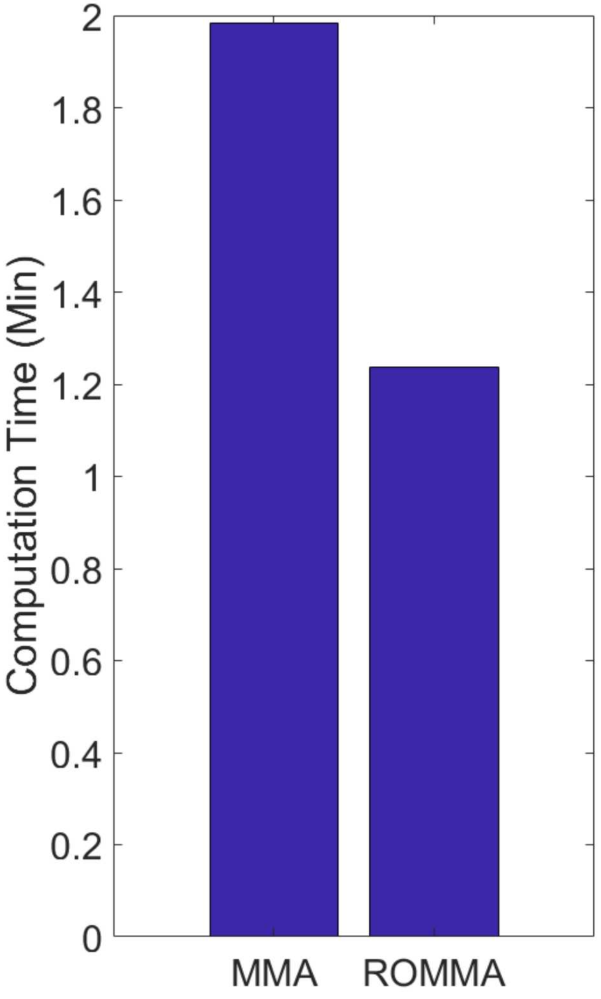
	MMA/RWM ST-MCMC Computational Time (min)	ROMMA ST-MCMC Computation Time (min)
Prior Reliability (1.5 x 10^{-5})	2.0	1.2
Posterior Inference	605.5	20.3
Posterior Reliability (3.0 x 10^{-7})	206.0	36.4



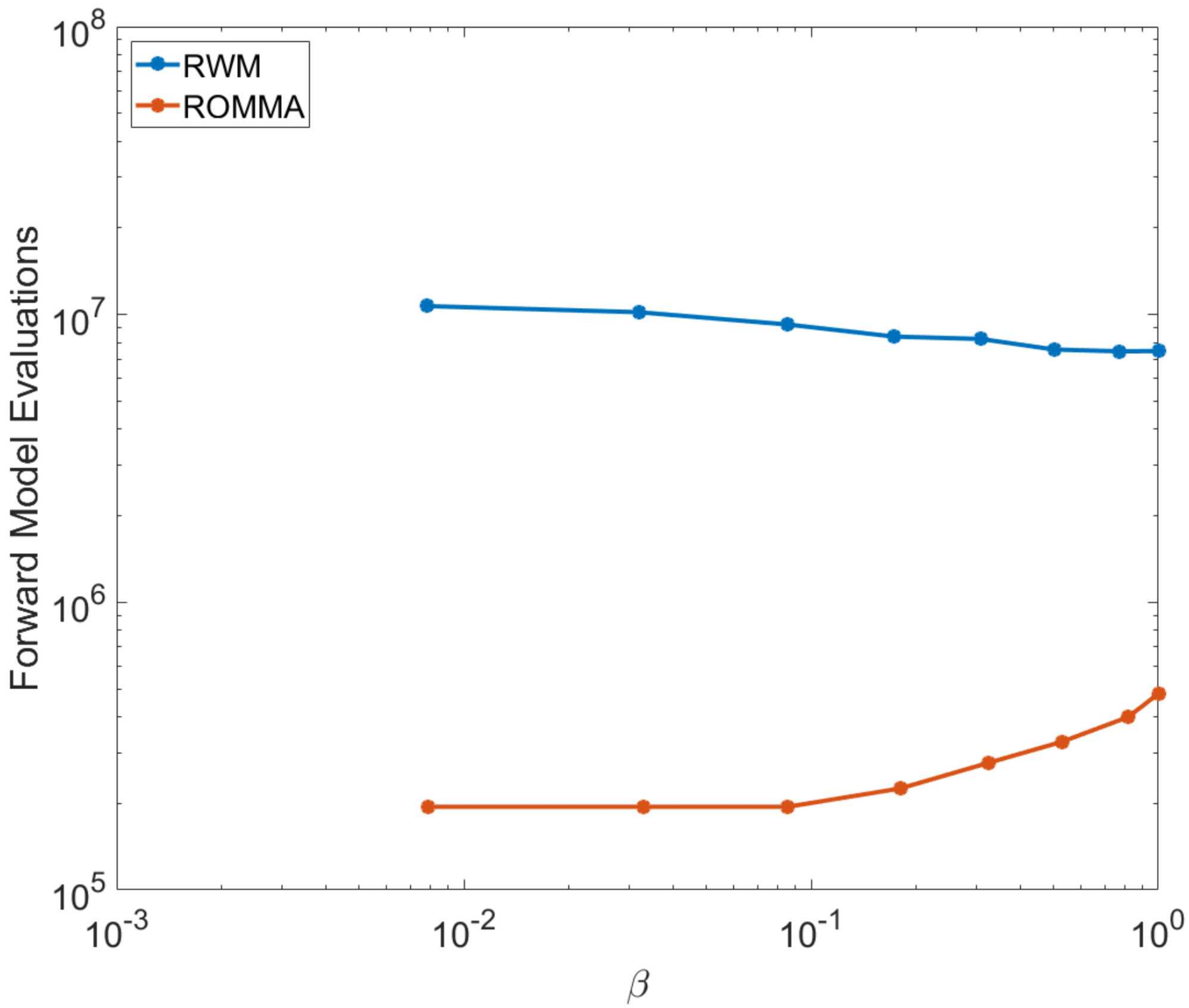


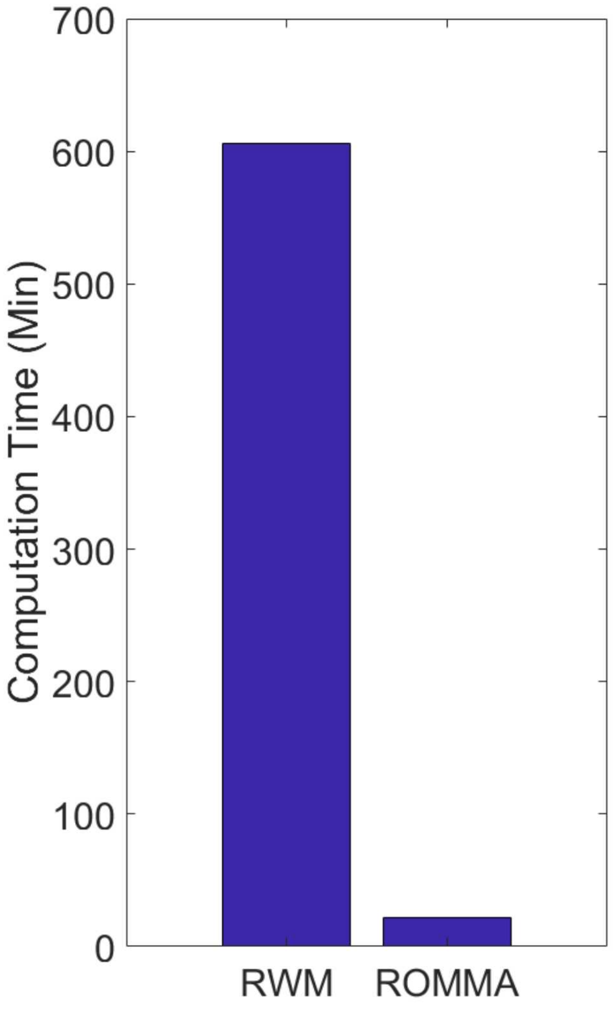
Prior Reliability Comparison



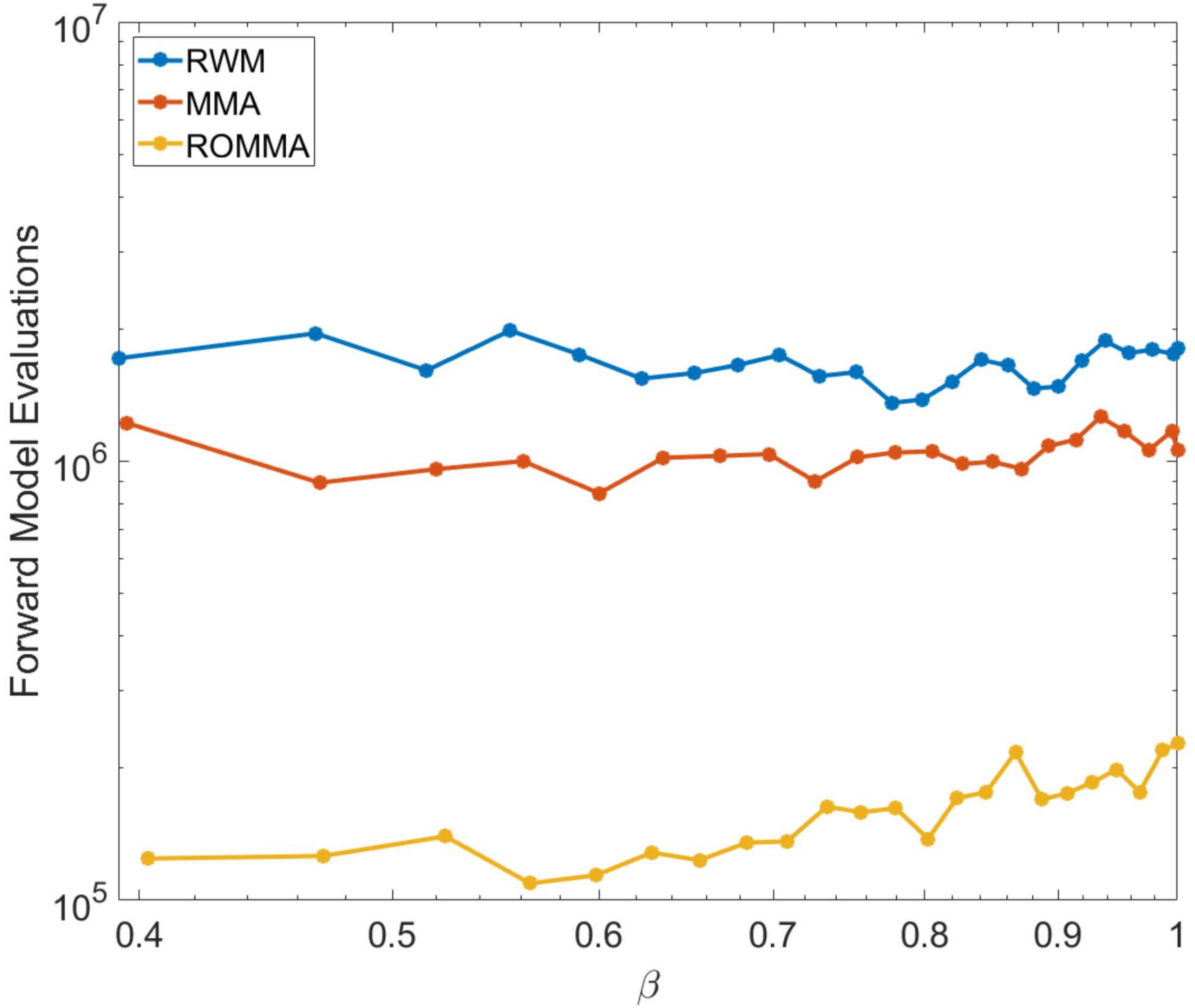


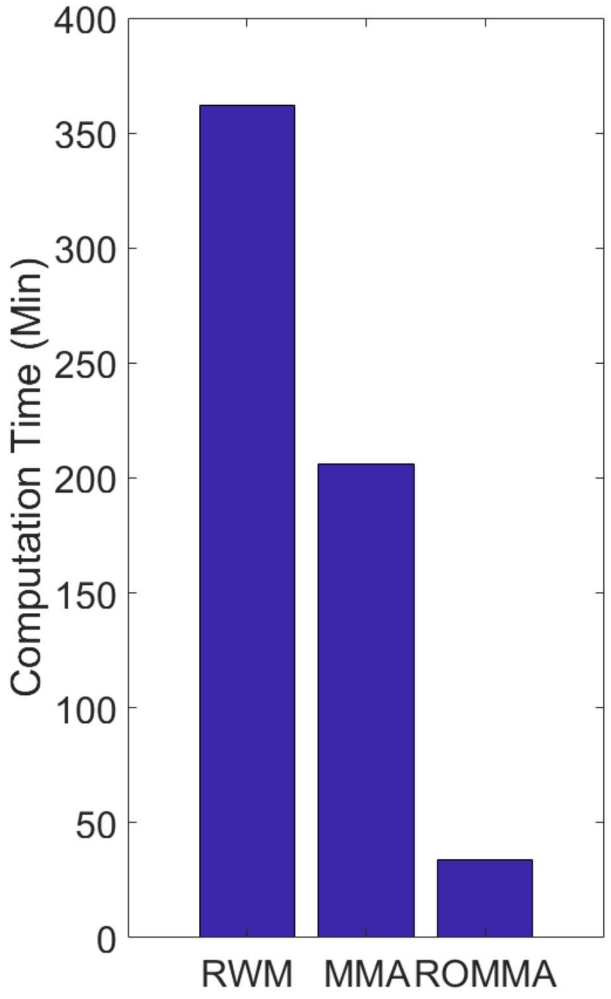
Posterior Sampling Comparison





Posterior Reliability Comparison



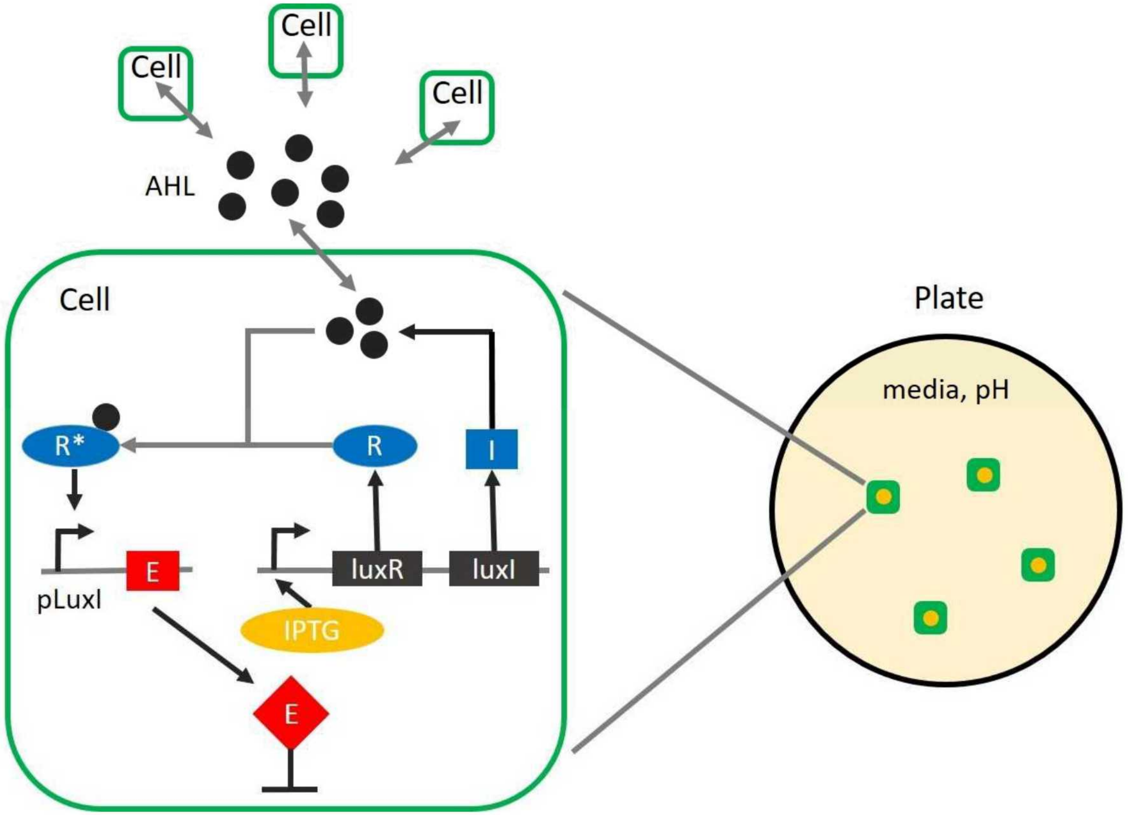


Identifying Biological Context:
Model Selection

Investigating Context Dependence

- Context dependence in synthetic biological circuits causes parts and modules to behave unpredictably in the cell under different experimental conditions.
- We investigate context dependence by examining the cell growth regulation circuit[1].
- We identify context models using Bayesian model selections and identify biological parameters in mathematical circuit model.

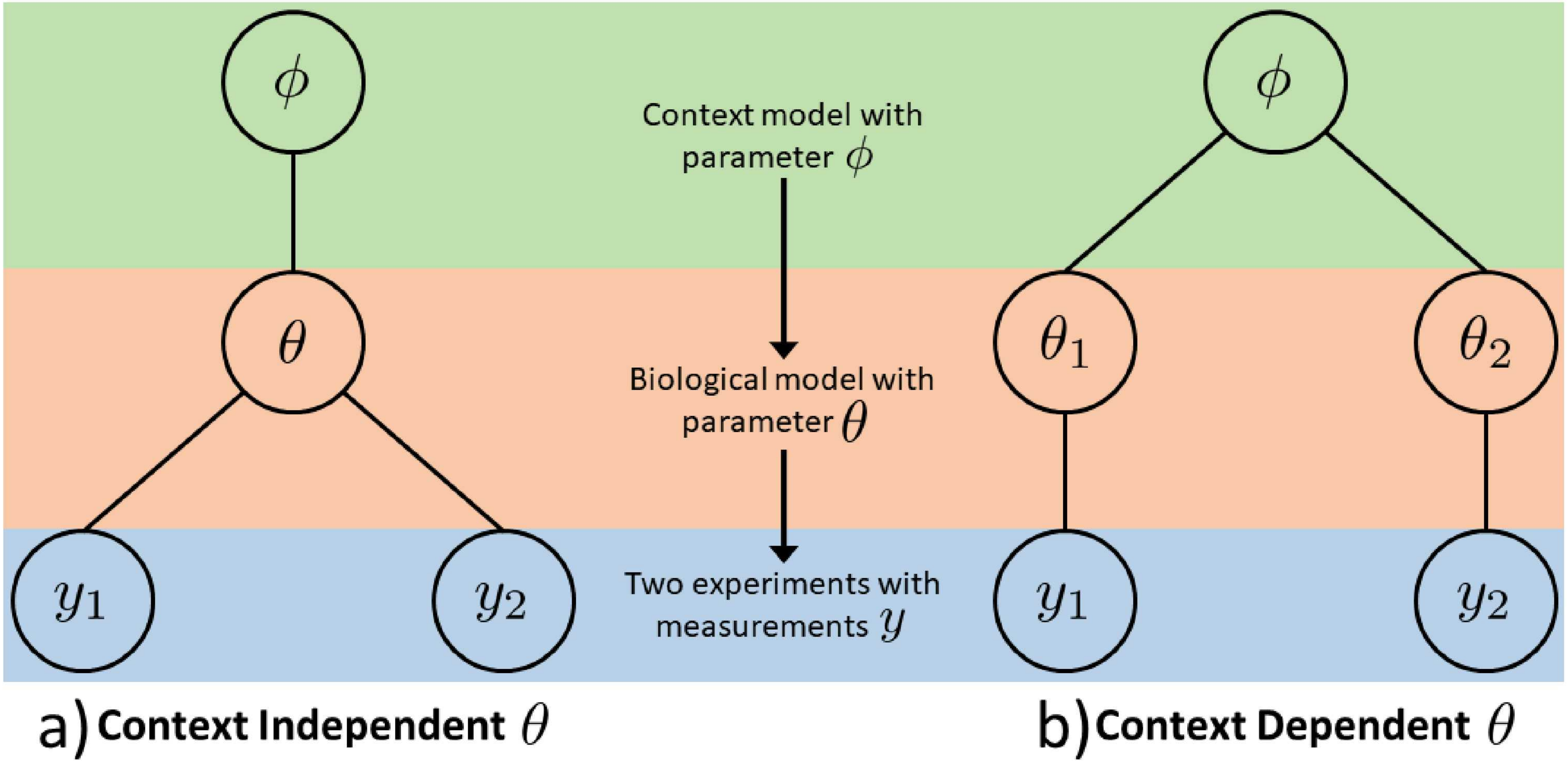
Synthetic Biological Growth Control Circuit



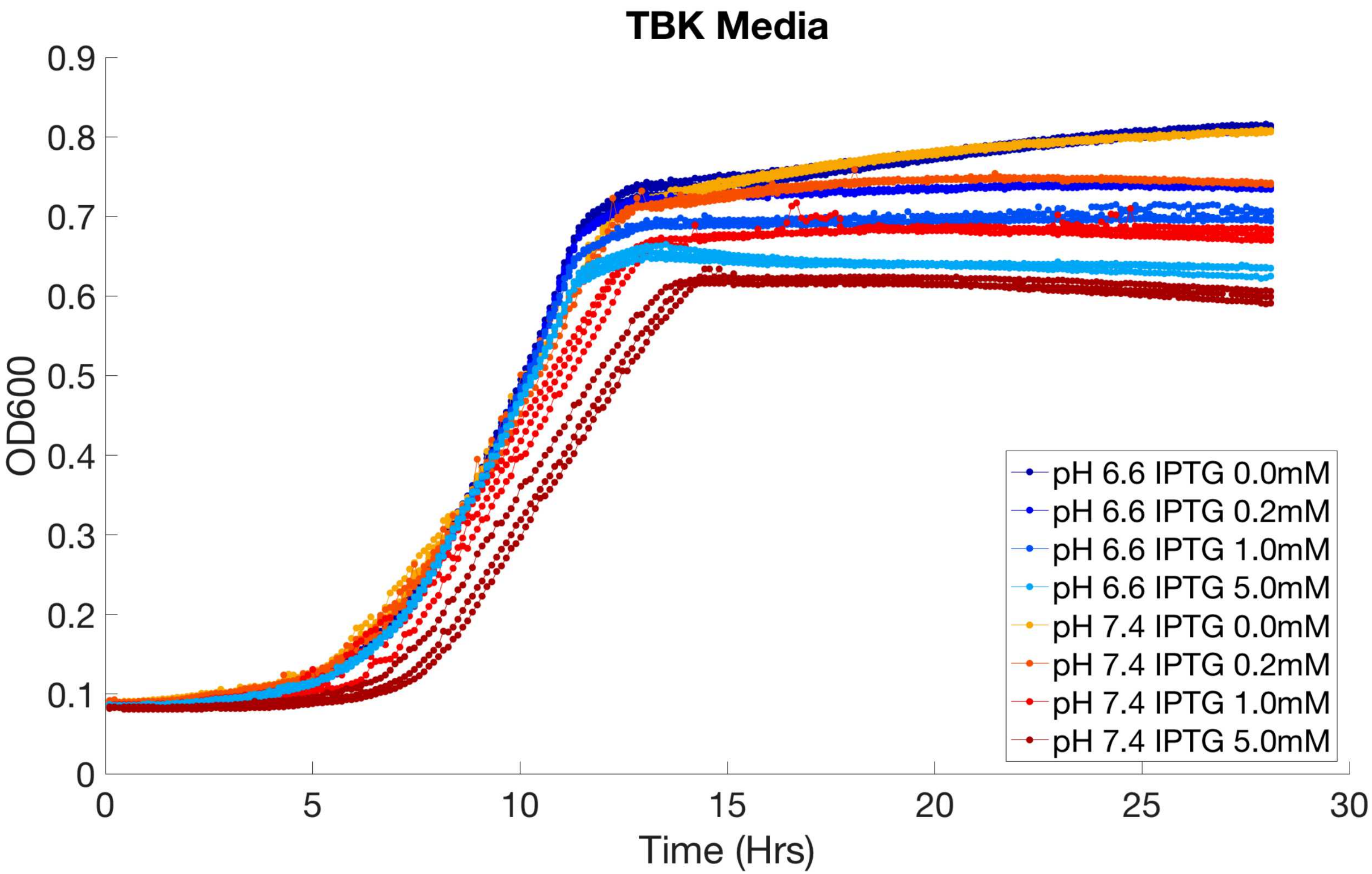
[1] You et al. 2004

Problem Formulation

Context dependence is the relationship of a biological system to conditions not explicitly described in our mathematical model because they are unknown or too complicated. Instead these relationships can be described using a stochastic model.



pH and Inducer concentration Dependence

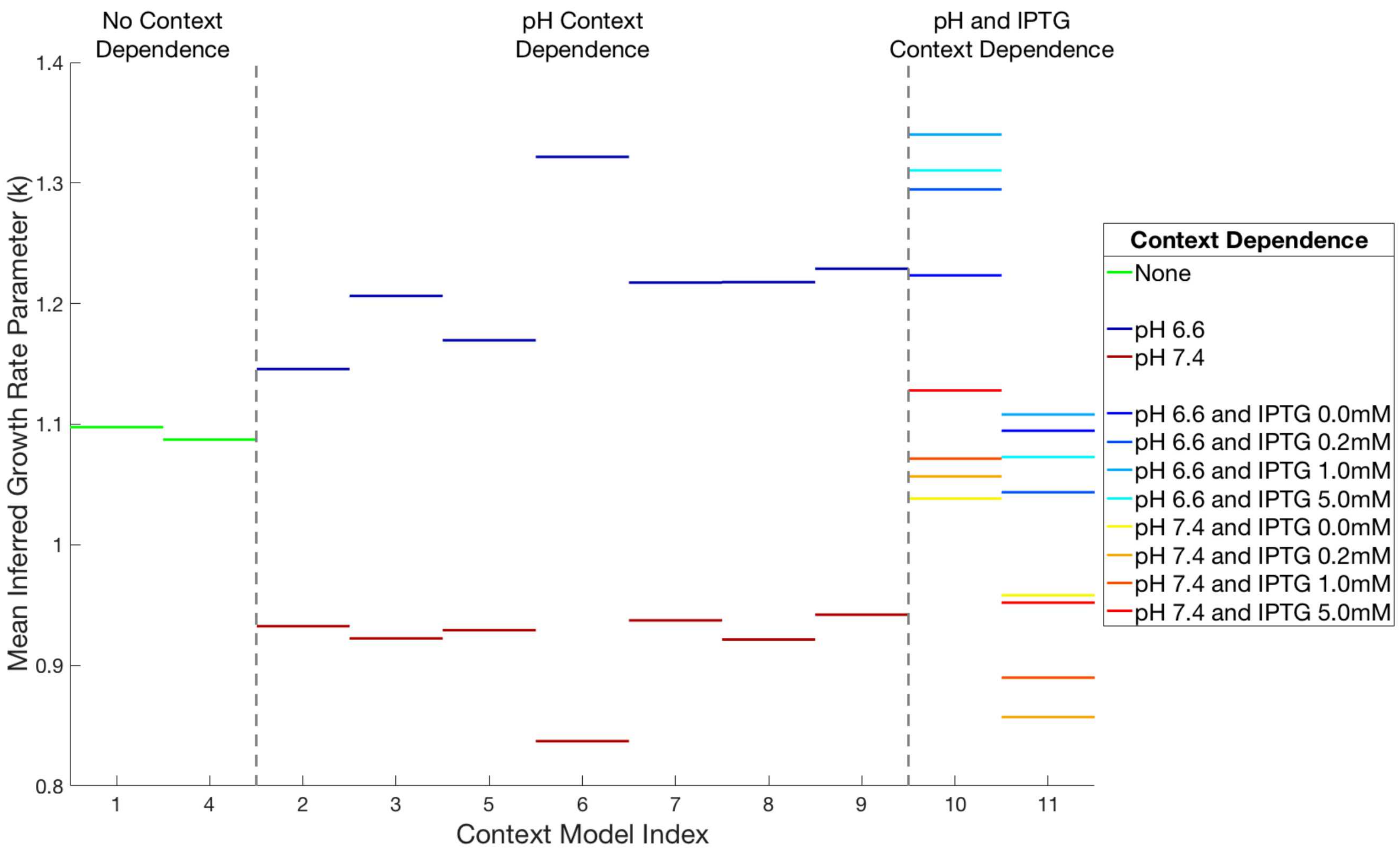


pH and Inducer concentration Dependence

Context Models and Probabilities

Model	k	N_m	v_A	d_A	t_0	A_0	**Prob.**
Model 1	Shared	Shared	Shared	pH	Shared	Shared	10^{-241}
Model 2	pH	Shared	Shared	pH	Shared	Shared	10^{-239}
Model 3	pH	pH	Shared	pH	Shared	Shared	10^{-230}
Model 4	Shared	Shared	pH IPTG	pH	Shared	Shared	10^{-207}
Model 5	pH	Shared	pH IPTG	pH	Shared	Shared	10^{-204}
Model 6	pH	pH	pH IPTG	Shared	Shared	Shared	10^{-202}
Model 7	pH	pH	pH IPTG	pH	Shared	Shared	10^{-195}
Model 8	pH	pH	IPTG	pH	Shared	Shared	10^{-204}
Model 9	pH	pH	pH IPTG	pH	pH	pH	10^{-196}
Model 10	pH IPTG	pH	pH IPTG	pH	Shared	Shared	10^{-58}
Model 11	pH IPTG	pH IPTG	pH IPTG	pH	Shared	Shared	≈ 1.0

The assumed context model significantly influences the inferred growth rate



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 proposals
- Combining Sequential Tempering with Multilevel-Multifidelity Hierarchies to reduce computational cost
- Better metrics for assessing correlation e.g. Canonical Correlation Analysis (CCA)

Conclusion

- Sequential Tempered MCMC methods are able to solving Bayesian system Identification, model selection posterior reliability problems
- MCMC proposals like ROMMA that incorporate knowledge about the prior or posterior can significantly speed up ST-MCMC algorithms
- Applications like water distribution system reliability and synthetic biological systems context dependence were enabled using these new techniques

Backup Slides

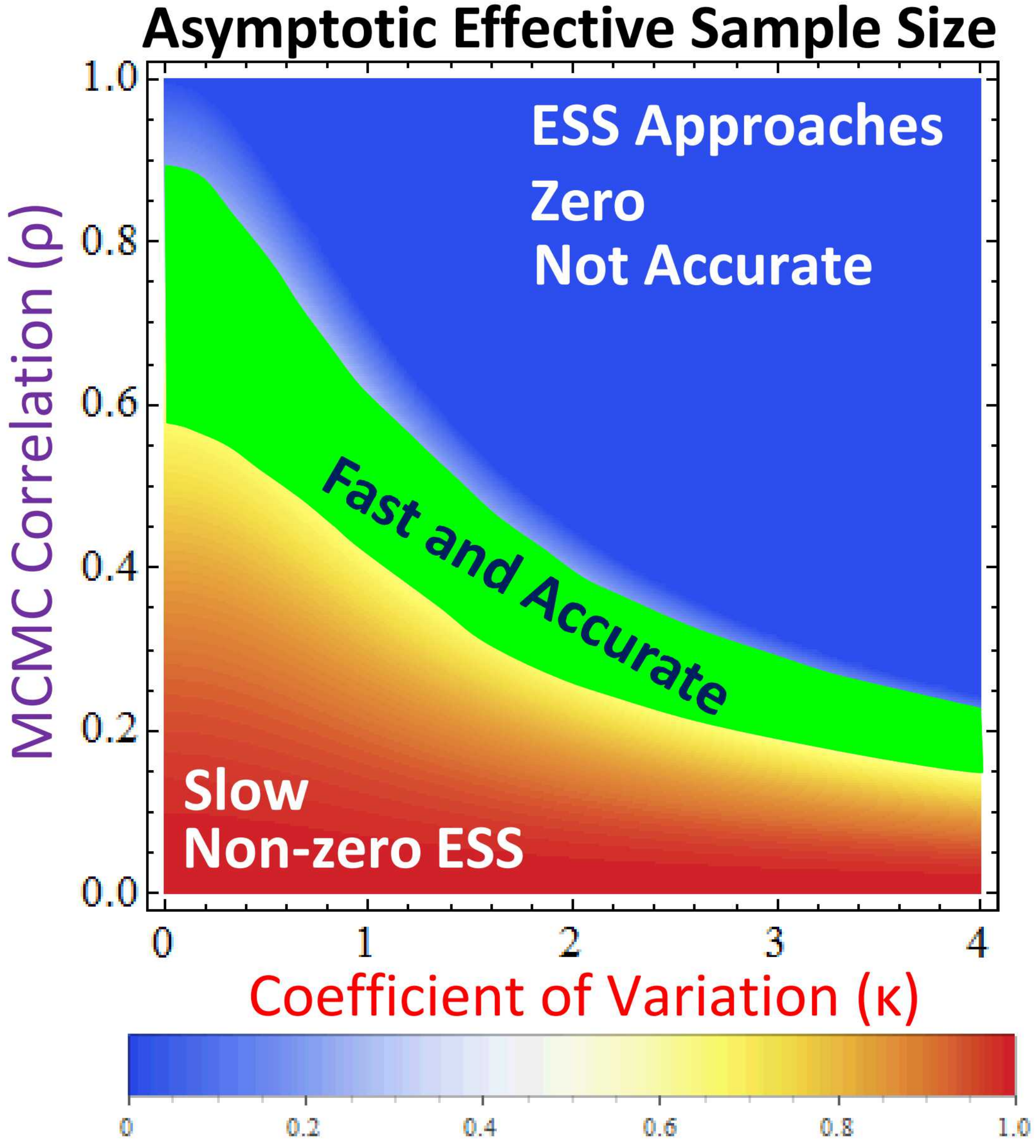
Theoretical Study of Effective Sample Size in ST-MCMC

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

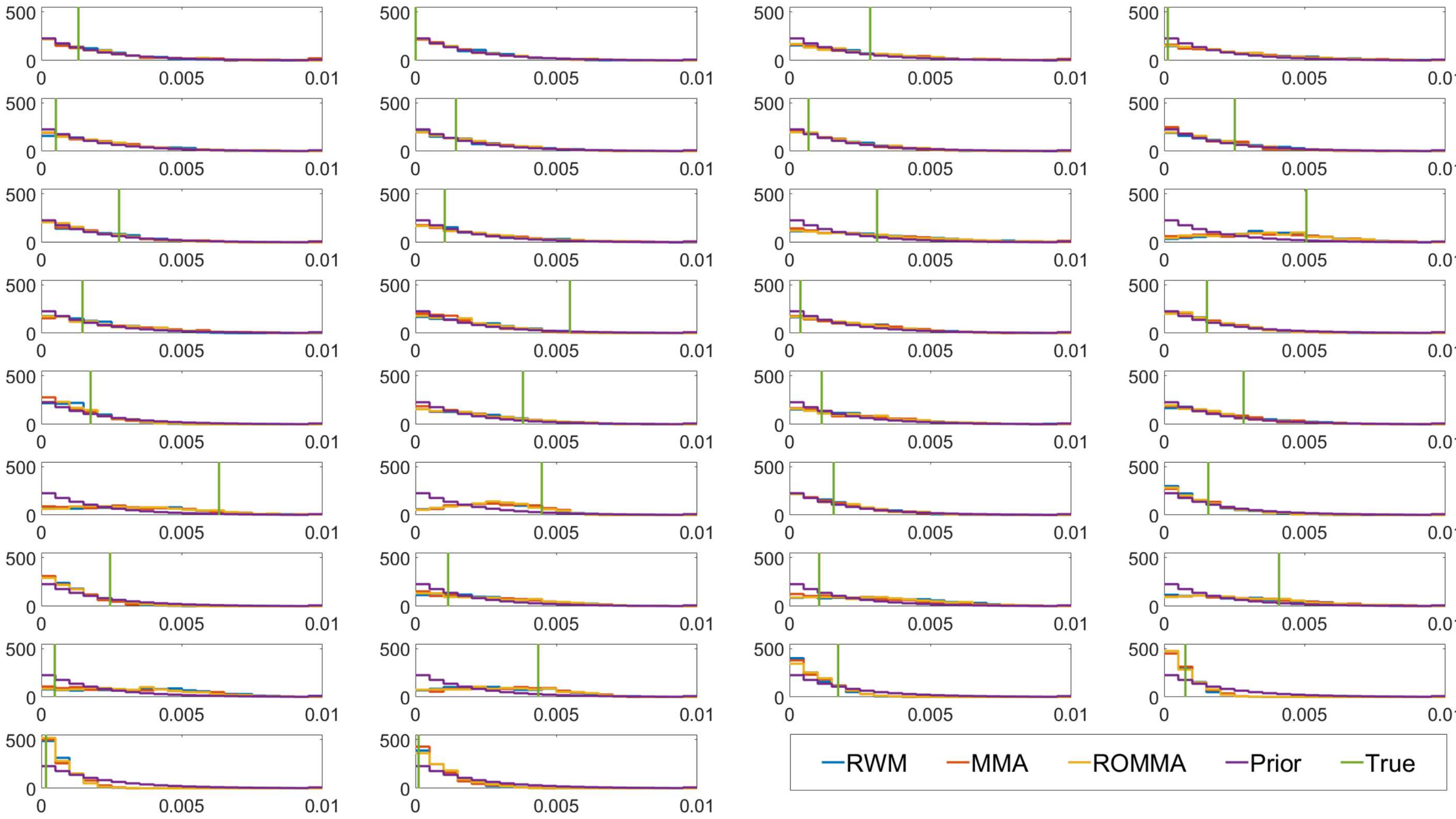
$$n_{k+1} = n_k \frac{N}{(N-1)\left(1+\kappa^2\right)\rho^2 + n_k}$$

Number of chains (N); Coefficient of Variation (κ); MCMC Correlation (ρ)

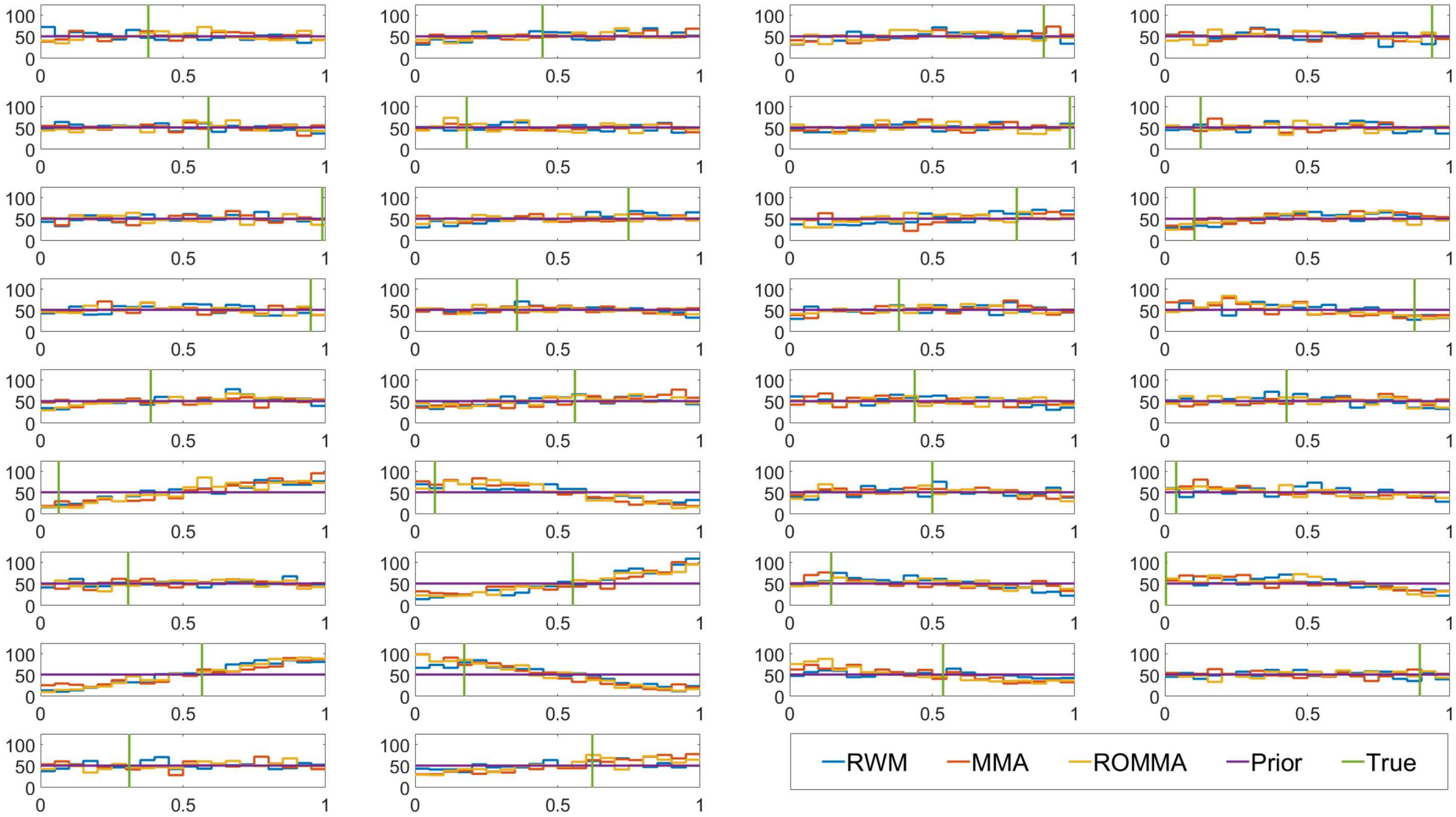
- Parameter estimation is possible when n_k does not asymptotically approach zero



Posterior Failure Leak Size



Posterior Failure Leak Position



Posterior Failure Demand