Risk Analysis and Decision-Making Under Uncertainty: A Strategy and its Applications

Ming Ye (mye@fsu.edu)
Florida State University
Mary Hill (mchill@ku.edu)
University of Kansas

This research is supported by
DOE Early Career Award: DE-SC0008272

Acknowledge the efforts of ISCMEM Working Group 2–
Federal Scientists Working for Coordinated Uncertainty
Analysis and Parameter Estimation
Since 2002

Learning from the NAS workshop

- Anna Willett, director of the Interstate Technology and Regulatory Council (ITRC), identified that "Different agencies conduct risk assessments differently, and there should be better alignment among agencies and states on risk assessment and risk reduction."
- Many of the NRC committees stress that "A formalized decision-making process provides consistency and transparency in agency decisions."
- NRC (2005) enumerated a list of characteristics of a credible decision-making process, highlighting the importance of credible and believable results.

Our Observations

A. Tower of Babel
B. "Colossal" computational burdens

Our observation A: Tower of Babel

- Tower of Babel: The number of model/risk analysis methods
- Even methods that measure the same aspect of model/risk analysis are often presented in ways that make them difficult to compare.
- Impedes clear dialogue between scientists, decision-makers, and stakeholders.
- Obscure the very model comparisons and evaluations that need to be transparent

Hill et al. (2015, Ground Water)

A strategy for the "Tower of Babel"

1. Introduce typology to organize the diversity of model/risk analysis methods.

   Organize common concerns and applicable methods.

Questions addressed by model analysis

Model Adequacy

1 How can many data types with variable quality be included?

2 Is model misfit/overfit a problem? Are prior knowledge and data subsets inconsistent?

3 How nonlinear is the problem?

Sensitivity and Uncertainty

Observations (Obs) ⟷ Parameters (Pars)

4 What pars can and cannot be estimated with the obs?

5 Are any pars dominated by one obs and, thus, its error?

6 How certain are the par values?

7 Which obs are important and unimportant to pars?

Parameters (Pars) ⟷ Predictions (Preds)

8 Which pars are important and unimportant to preds?

9 How certain are the preds?

10 Which pars contribute most and least to pred uncertainty?

Observations (Obs) ⟷ Predictions (Preds)

11 Which existing and potential obs are important to preds?

12 For multi-model analysis, which models are likely to produce accurate preds?

Risk Assessment

13 What risk is associated with a given decision strategy?

Common questions	Frugal methods	Expensive methods
Model Adequacy		
1 How can many data types with variable quality be included?	Error-based weighting and SOO or MAP	MOO, Pareto curve
2 Is model misfit/overfit a problem? Are prior knowledge and data subsets inconsistent?	Compare fit to a priori error analysis using s_n^2, $s_{(n-p)}^2$, RMSE, Nash-Sutcliffe, graphs	MOO, Pareto curve
3 How nonlinear is the problem?	Intrinsic nonlinearity, DELSA	DELSA, Explore objective function,
Sensitivity and Uncertainty		
Observations (Obs) ↔ Parameters (Pars)		
4 What pars can and cannot be estimated with the obs?	Scaled local stats (CSS, ID, PCC, etc.), SVD, DoE, MoM (OAT)	DoE, MoM (OAT), eFAST, Sobol', RSA, CSE
5 Are any pars dominated by one obs and, thus, its error?	Scaled local stats (Leverage, DFBETAS)	Cross Validation
6 How certain are the par values?	Par uncertainty intervals	Par uncertainty intervals[3]
7 Which obs are important and unimportant to pars?	Scaled local stats (Leverage, Cook's D)	Cross Validation
Parameters (Pars) ↔ Predictions (Preds)		
8 Which pars are important and unimportant to preds?	Scaled local stats (PSS, etc.), DELSA	DELSA, FAST, Sobol'
9 How certain are the preds?	z/SD_z, Pred uncertainty intervals	Pred uncertainty intervals, MCMC, multi-model analysis
10 Which pars contribute most and least to pred uncertainty?	Scaled local stats (PPR)	FAST, Sobol'
Observations (Obs) ↔ Predictions (Preds)		
11 Which existing and potential obs are important to the preds?	Scaled local stats (OPR)	Cross Validation
12 For multi-model analysis, which models are likely to produce accurate preds?	Analyze model fit and estimated parameters, AIC, AICc, BIC, KIC	Cross Validation
Risk Assessment		
13 What risk is associated with a given decision strategy?	Combine uncertainty analysis and scenario simulation. Smooth cost function.	Combine uncertainty analysis and scenario simulation. Cost function need not be smooth.

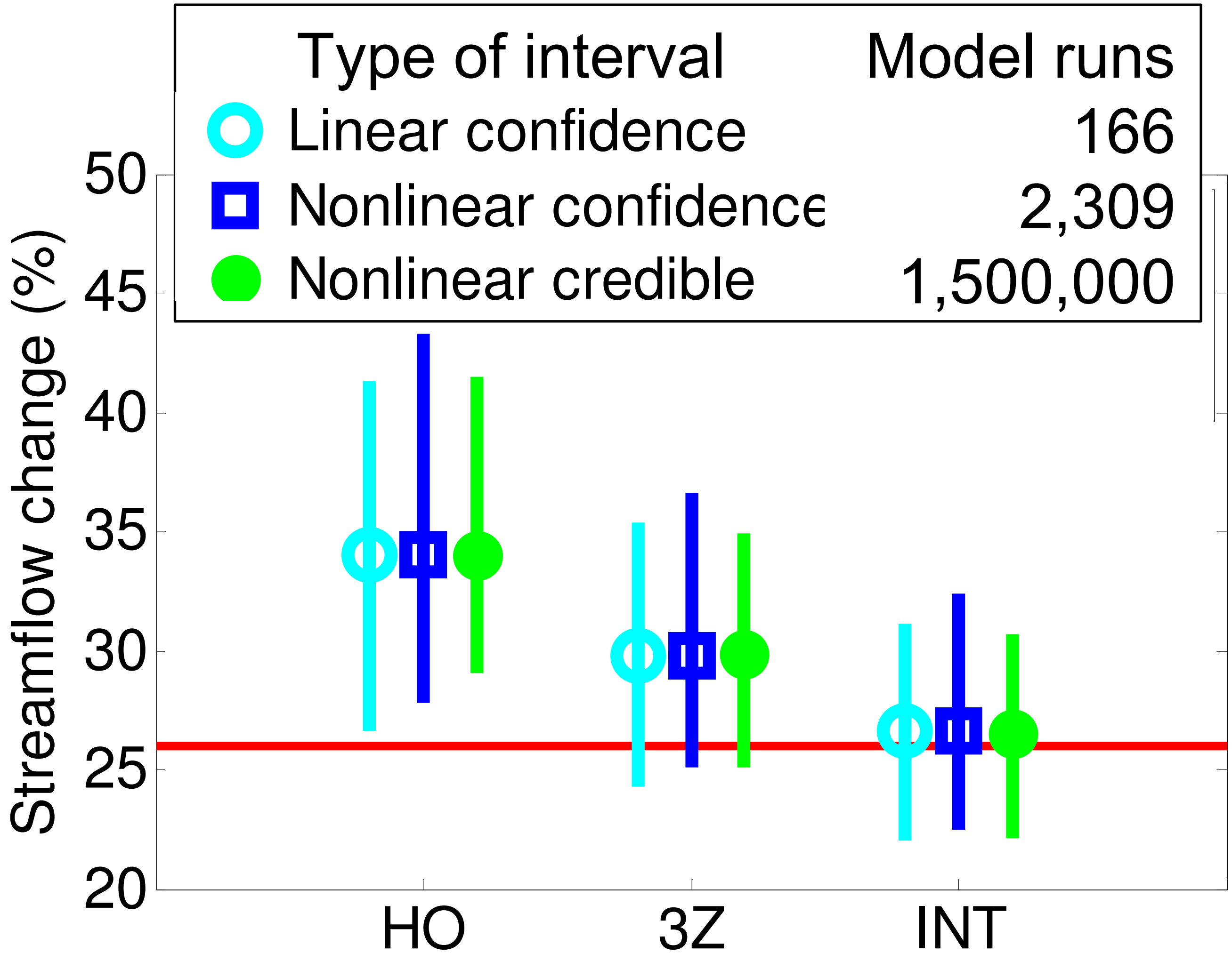
A strategy for the “Tower of Babel”

1. Introduce typology to organize the diversity of model/risk analysis methods.

   Organize common concerns and applicable methods.

2. Examine relation of methods through theoretical and empirical investigation

Example: Uncertainty Evaluation



Lu Ye Hill 2012 WRR

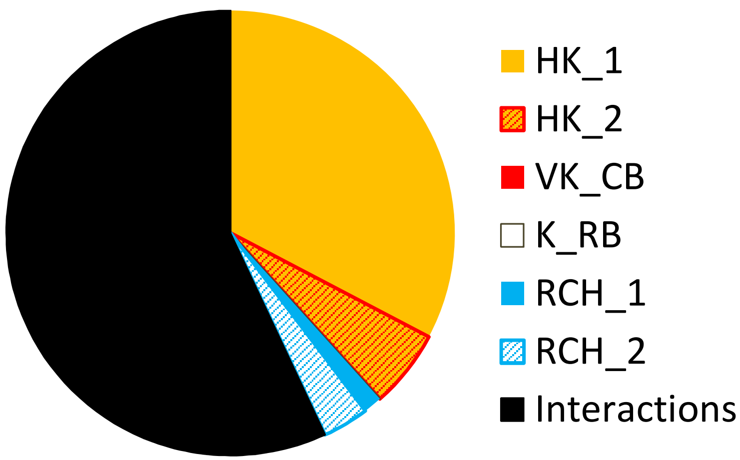
A strategy for the “Tower of Babel”

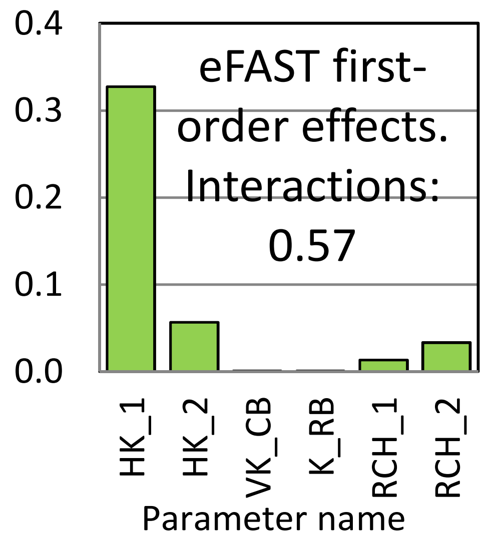
1. Introduce typology to organize the diversity of model/risk analysis methods.

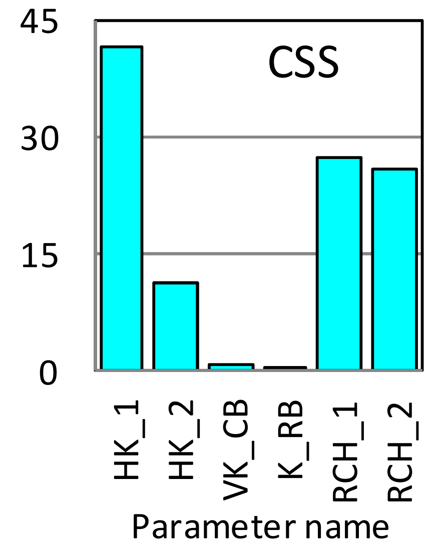
    Organize common concerns and applicable methods.

2. Examine relation of methods through theoretical and empirical investigation
3. Use comparable metrics and presentation of results even when different analysis methods are used

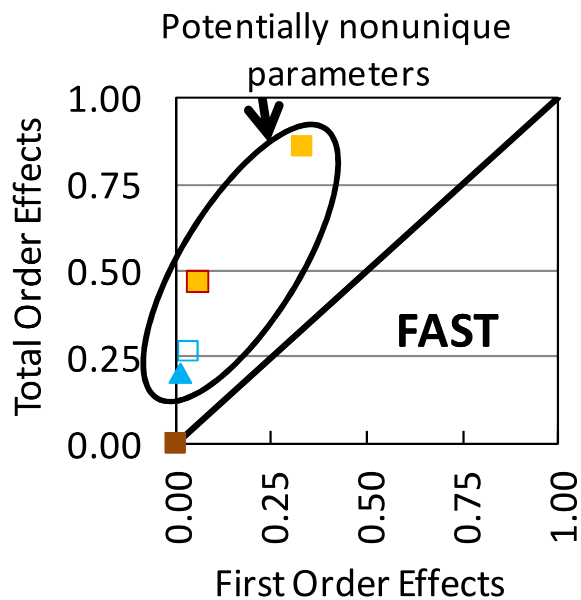
Identify important parameters

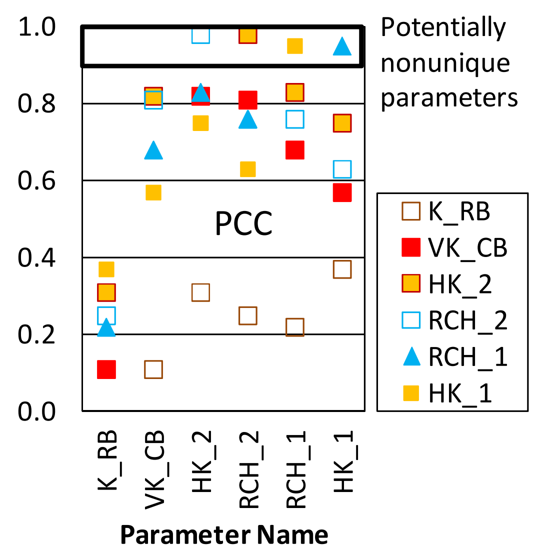






Not doing a good job yet measuring or communicating parameter interactions





Our observation 2: "colossal" computational burdens

- Goliath-size computational demand
  - Sobol' & MCMC often require 10,000++ model runs.
  - Impractical when runs take from seconds to months.
- Modeler choices
  - too few model runs to obtain numerically stable results, or
  - oversimplified models merely to obtain short execution times for model/risk analysis methods (Tail wags dog)
- Transparency suffers
  - Most computationally expensive methods based on sampling. No closed form analytical equation -- difficult to determine properties. Further, the computational demands impedes replication of results.

Our observation 2:
"colossal" computational burdens

Computer Burden		Number of model runs needed for analysis: Few	Number of model runs needed for analysis: Many
Length of one forward model run	Short	Small	Small to moderate
	Long	Moderate to Large	Colossal

Most methods parallelizable, so not used as factor.

Use relative terms so categories adapt to computer resources available.

Small ➔ one analysis < few hours.

Evaluate alternative models, refined discretization, etc

Common questions	Frugal methods	Expensive methods
Model Adequacy		
1 How can many data types with variable quality be included?	Error-based weighting and SOO or MAP	MOO, Pareto curve
2 Is model misfit/overfit a problem? Are prior knowledge and data subsets inconsistent?	Compare fit to a priori error analysis using s_n^2, $s_{(n-p)}^2$, RMSE, Nash-Sutcliffe, graphs	MOO, Pareto curve
3 How nonlinear is the problem?	Intrinsic nonlinearity, DELSA	DELSA, Explore objective function,
Sensitivity and Uncertainty		
Observations (Obs) ⟷ Parameters (Pars)		
4 What pars can and cannot be estimated with the obs?	Scaled local stats (CSS, ID, PCC, etc.), SVD, DoE, MoM (OAT)	DoE, MoM (OAT), eFAST, Sobol', RSA, CSE
5 Are any pars dominated by one obs and, thus, its error?	Scaled local stats (Leverage, DFBETAS)	Cross Validation
6 How certain are the par values?	Par uncertainty intervals	Par uncertainty intervals[3]
7 Which obs are important and unimportant to pars?	Scaled local stats (Leverage, Cook's D)	Cross Validation
Parameters (Pars) ⟷ Predictions (Preds)		
8 Which pars are important and unimportant to preds?	Scaled local stats (PSS, etc.), DELSA	DELSA, FAST, Sobol'
9 How certain are the preds?	z/SD_z, Pred uncertainty intervals	Pred uncertainty intervals, MCMC, multi-model analysis
10 Which pars contribute most and least to pred uncertainty?	Scaled local stats (PPR)	FAST, Sobol'
Observations (Obs) ⟷ Predictions (Preds)		
11 Which existing and potential obs are important to the preds?	Scaled local stats (OPR)	Cross Validation
12 For multi-model analysis, which models are likely to produce accurate preds?	Analyze model fit and estimated parameters, AIC, AICc, BIC, KIC	Cross Validation
Risk Assessment		
13 What risk is associated with a given decision strategy?	Combine uncertainty analysis and scenario simulation. Smooth cost function.	Combine uncertainty analysis and scenario simulation. Cost function need not be smooth.

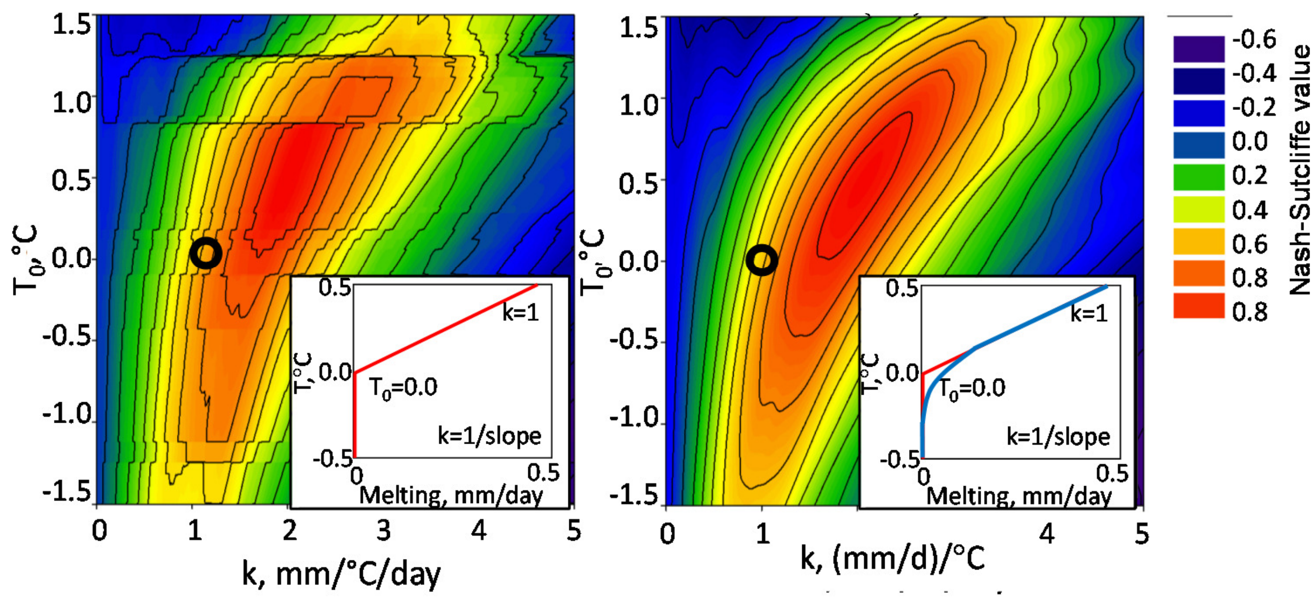
Our observation 2:
"colossal" computational burdens

Computer Burden		Number of model runs needed for analysis	
		Few	Many
Length of one forward model run	Short	Small	Small to moderate
	Long	Moderate to Large	Colossal

Most methods parallelizable, so not used as factor.
Use relative terms so categories adapt to computer resources available.
Small ➔ one analysis < few hours.
Evaluate alternative models, refined discretization, etc

What makes run times long and requires many model runs?

- Numerical daemons (Kavetski Clark 2010 WRR)
- Are some models more irregular than the real world?

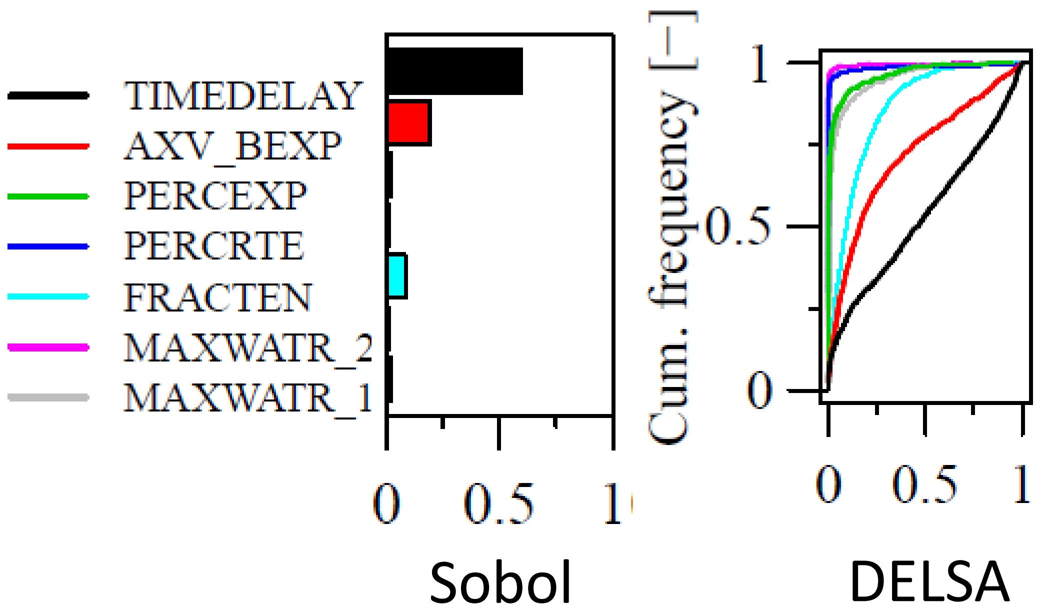


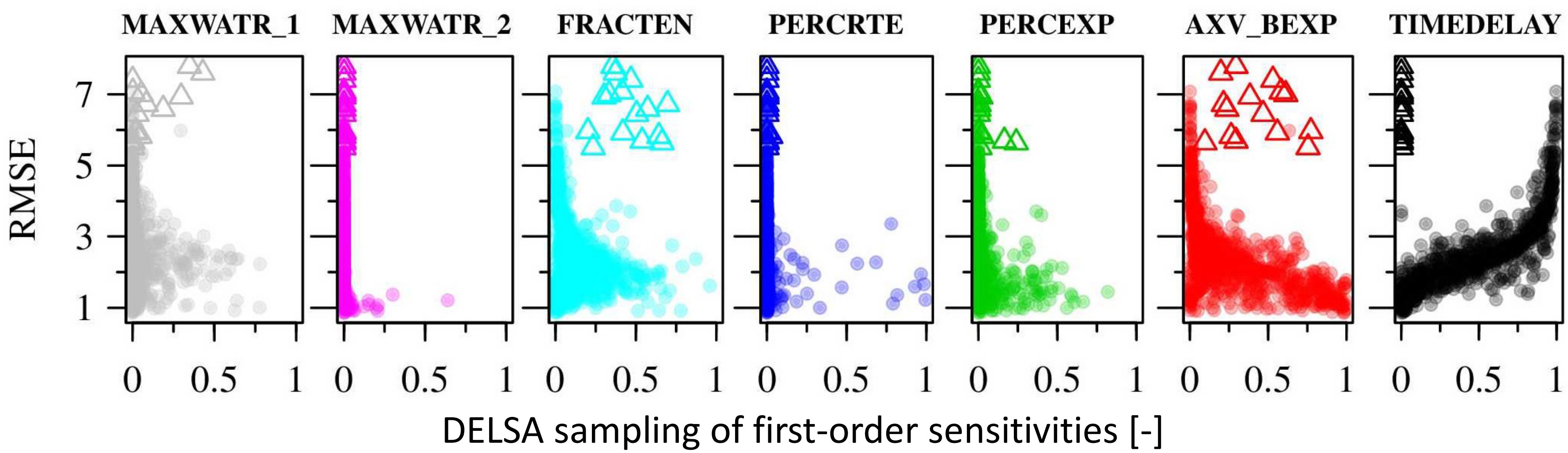
Kavetski and Kuczera 2007 WRR

How can the parameter space be explored?

- Sensitivity analysis – important and unimportant parameters
- Identify numerical

Dig deeper into DELSA for FUSE-016

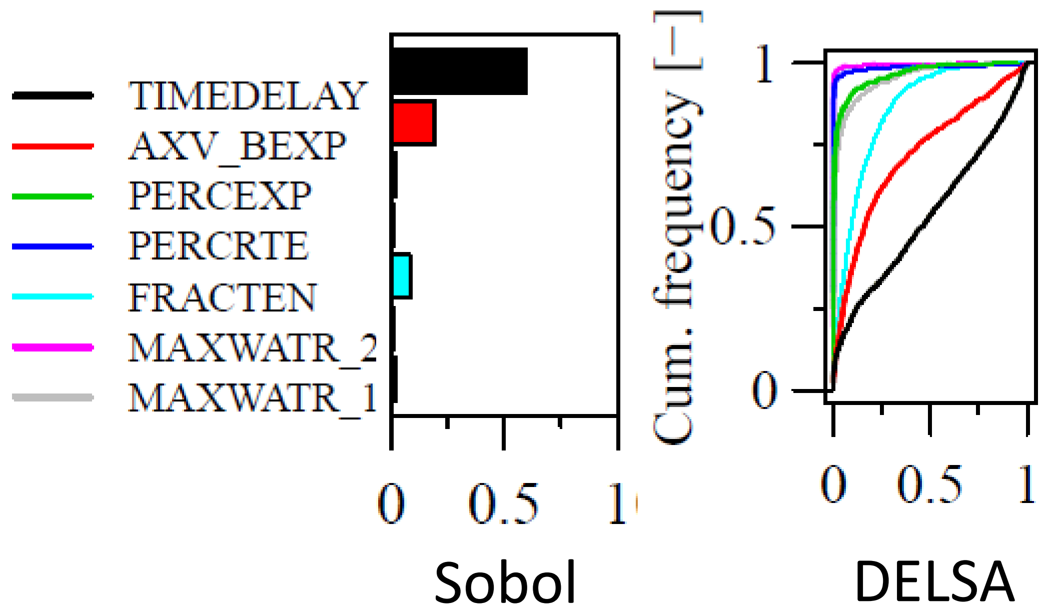




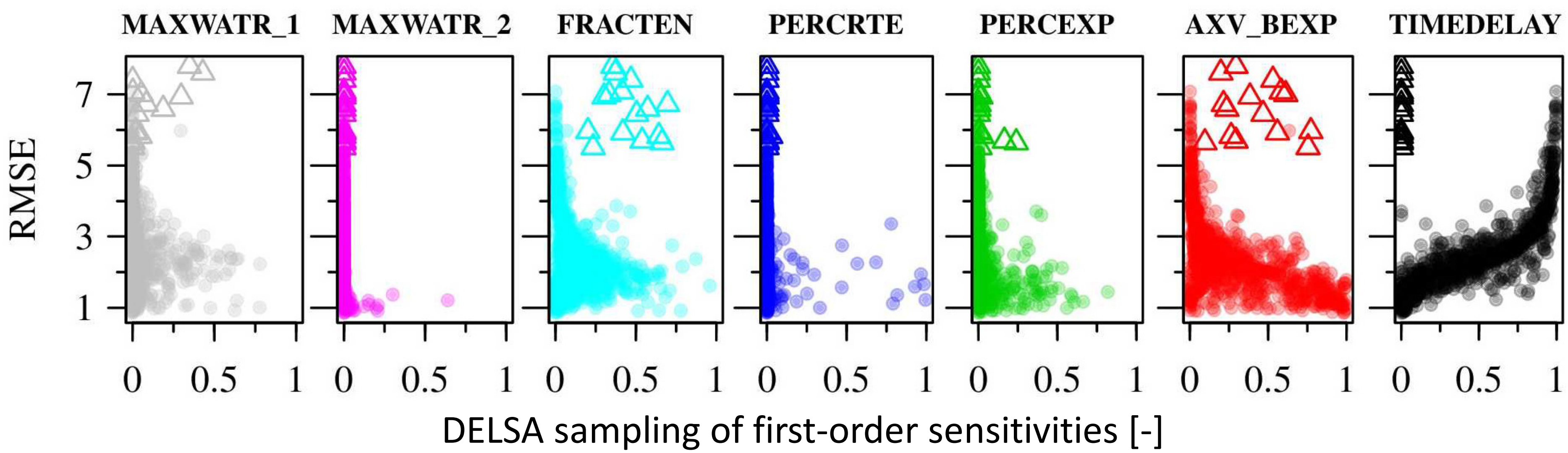
TIMEDELAY, the most important parameter based on Sobol', is only important for poorly fitting models.

This is where the article ends.

Dig deeper into DELSA for FUSE-016



Questions raised



Are inconvenient global methods useful for these problems?

Are very convenient local methods useful for these problems?

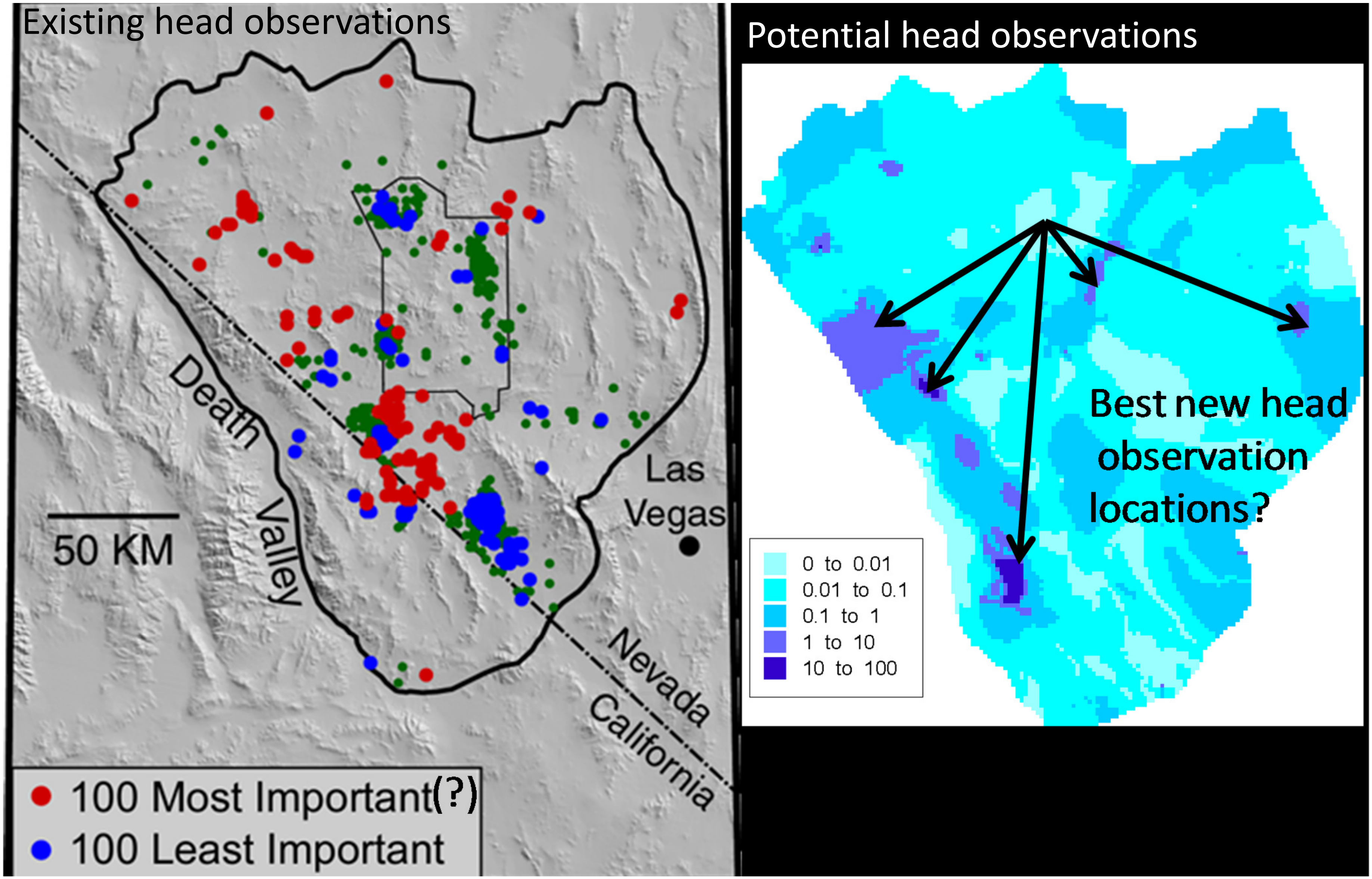
Use DELSA to explore the model and its simulated dynamics

Methods to alleviate "colossal" computational demands

- Computationally "frugal" analysis methods (few runs)
  - Frugal = small number of model executions
  - Rely on local methods, linearization of nonlinear models, and/or Gaussian assumptions
  - Example: Morris and DELSA for sensitivity analysis
- Computationally "efficient" model (short run time)
  - Eliminate "numerical daemons"
  - Build *surrogates* of computationally demanding models
    - Can afford many surrogate model executions
    - Need accurate surrogates
    - New method: Sparse grid collocation

Reduce effort for routine tasks so new things can be done

Here, identify observations that, through the model, dominate predictions



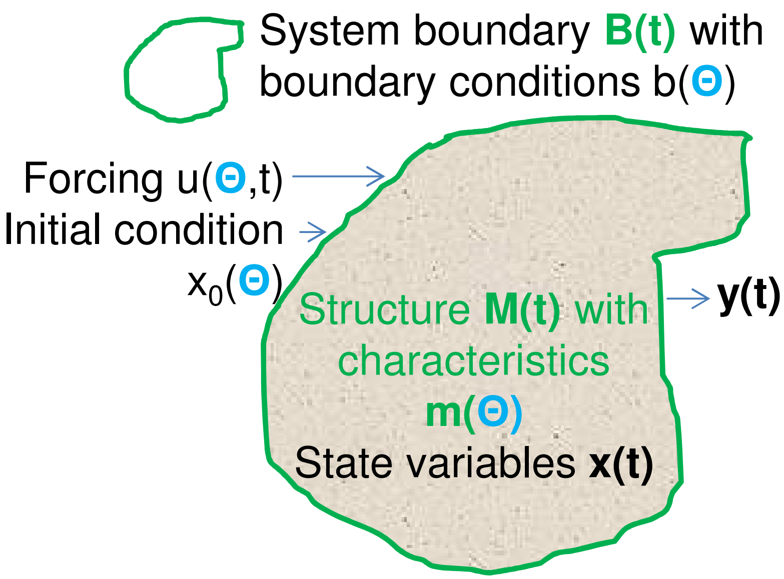
Our Observations

A. Tower of Babel
B. “Colossal” computational burdens

The strategy

- Unified strategy – possible?
  - DOD has approximately 38,000 sites and invests up to $2 billion a year in cleanup activities.
  - DOE has 90 out of 107 sites that have completed cleanup activities. Remediation of the remaining sites is expected to cost more than $300 billion and take 40 more years (DOE 2013).
- Paul Black: *Environmental remediation is site specific, but aspects of the risk assessment for a given site are often not site specific.*

From a modeling perspective



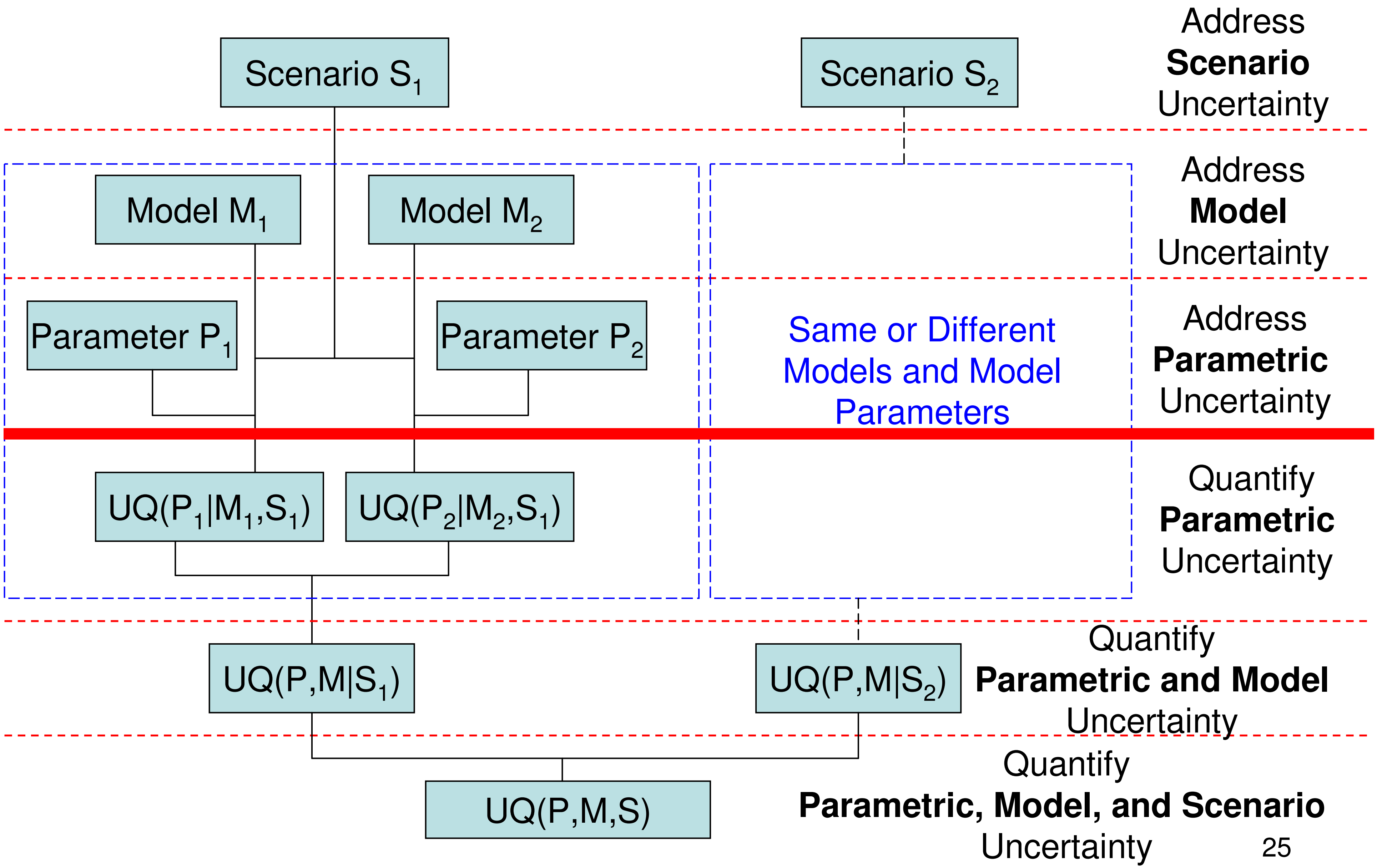
A model is composed of seven different components:

- System boundary (*B*),
- Forcing (*u*),
- Initial states (x_0),
- Structure (*M*),
- Parameters (ϑ),
- States (*x*), and
- Outputs (*y*).

The diagram can be used to

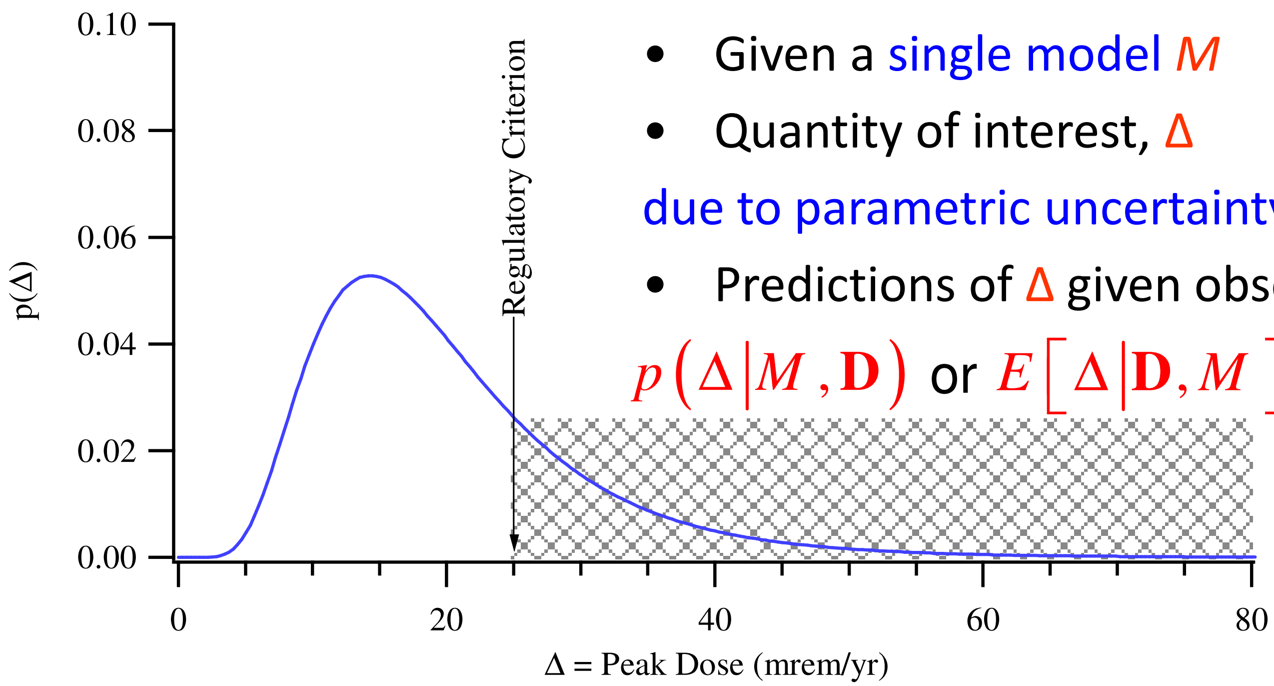
- Organize the current methods of model/risk analysis
- Develop new methods/frameworks of uncertainty quantification

A Comprehensive and Hierarchical Framework



Model uncertainty: why it matters and what to do about it (Draper)

- "In a typical application of the statistical paradigm, there's some quantity Δ about which I'm at least partially uncertain,
- and I wish to quantify my uncertainty about Δ, for the purpose of
  - sharing this information with other people (inference) or
  - helping myself or others to make a choice in the face of this uncertainty (decision-making).
- Uncertainty quantification is usually based on a probability model M, which relates Δ to known quantities (such as data values D)."



- Given a single model M
- Quantity of interest, Δ

due to parametric uncertainty of the model

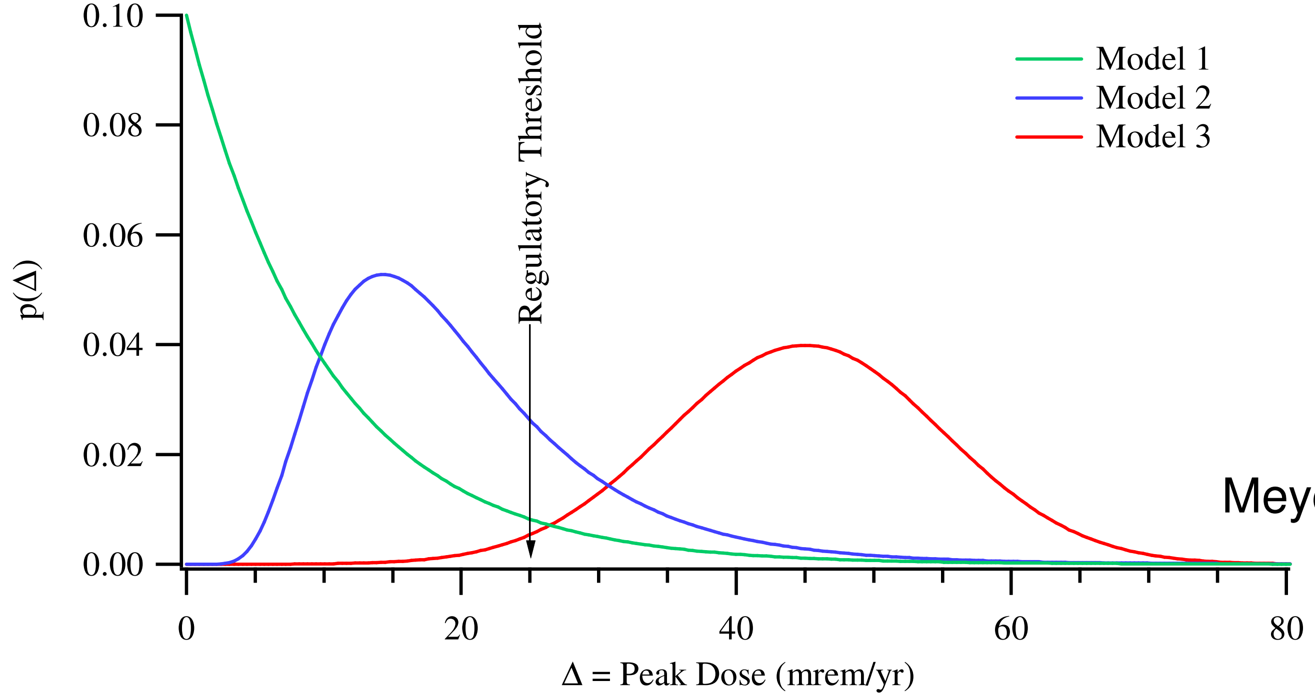
- Predictions of Δ given observation **D**

$p\left(\Delta \middle| M, \mathbf{D}\right)$ or $E\left[\Delta \middle| \mathbf{D}, M\right]$ and $Var\left[\Delta \middle| \mathbf{D}, M\right]$

Meyer et al. (2007)

Model Uncertainty

"*M* will in turn be based on assumptions and judgments on my part about how Δ and D are related, but I'm not always certain about the "right" assumptions and judgments to make."

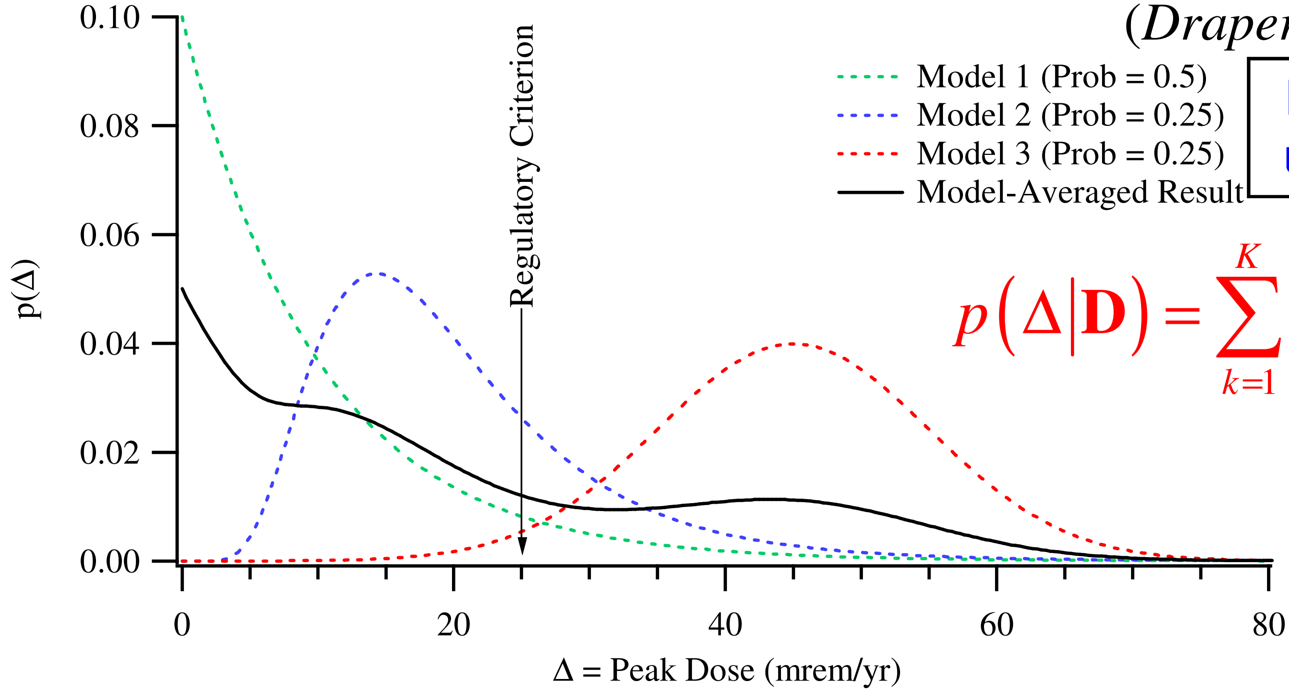


Meyer et al. (2007)

"To be completely honest, then, I have to acknowledge two sources of uncertainty: I'm uncertain about Δ, and I'm also uncertain about how to quantify my uncertainty about Δ. This second source is model uncertainty."

Bayesian Model Averaging (BMA)

(*Draper*, 1995; *Hoeting et al.* 1999)

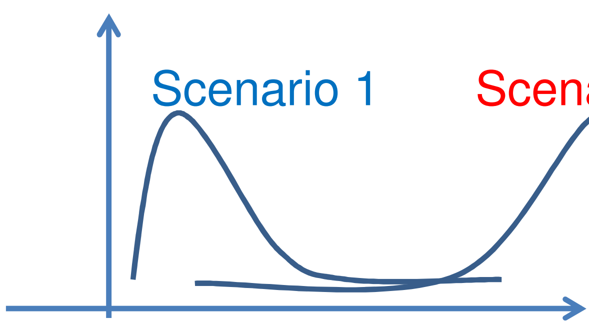


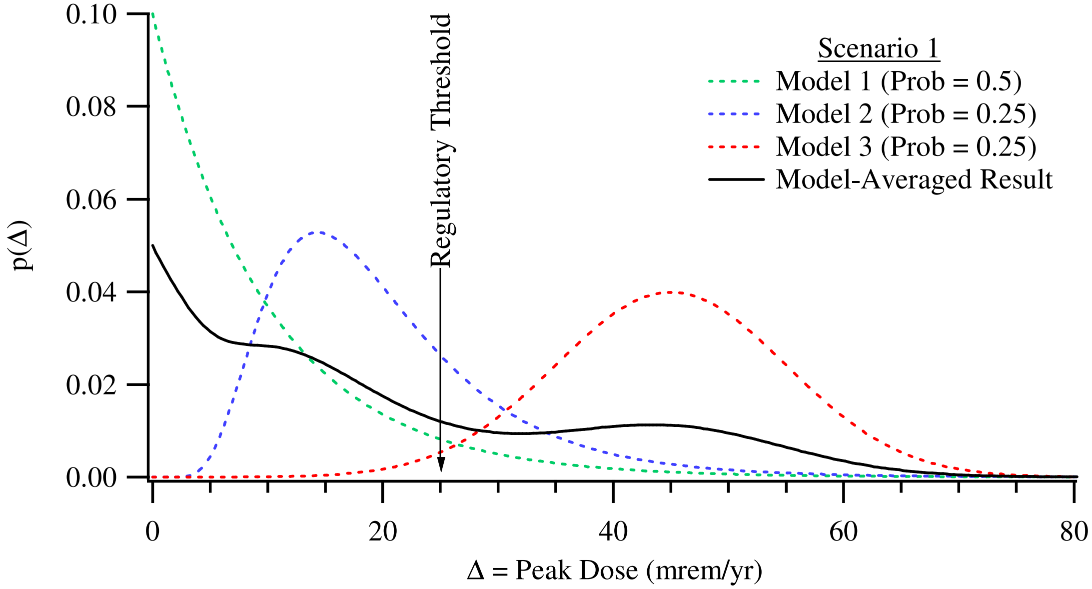
$$p(\Delta|\mathbf{D}) = \sum_{k=1}^{K} \overbrace{p(\Delta|M_k,\mathbf{D})}^{\text{Parametric uncertainty}}\, \underbrace{p(M_k|\mathbf{D})}_{\text{Posterior model probability}}$$

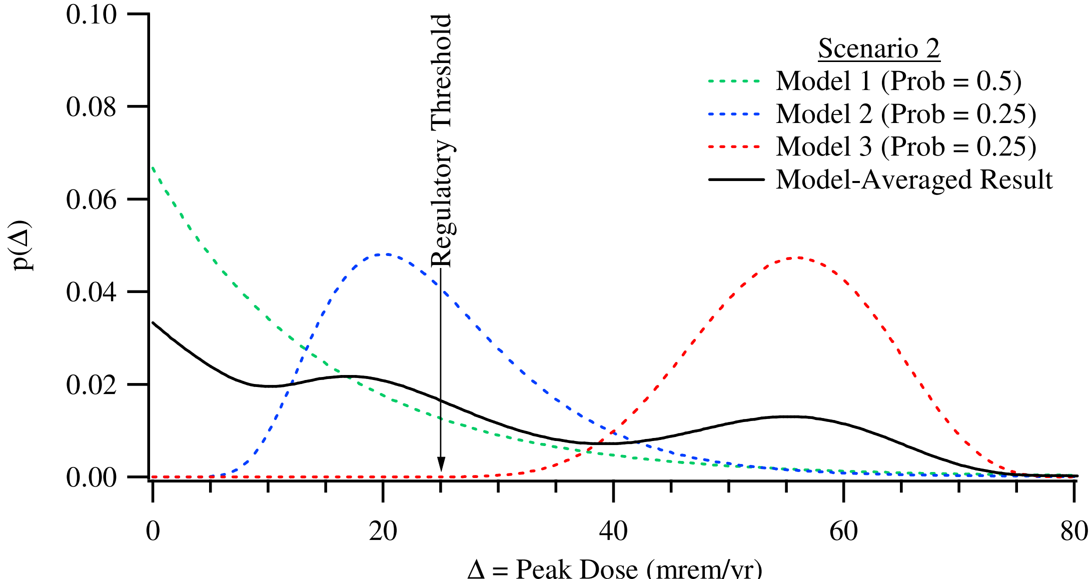
- Each model alternative has some merit in reproducing aspects of the physical system, this merit being quantified by each model’s probability.
- The model probability is interpreted as a relative measure with respect to the other model alternatives considered.
- The individual model results are presented along with the model-average results and the model probabilities, a fully informed decision can be made.

Scenario Uncertainty

- Two scenarios
- The three models are the same under each scenario.
- The three models have different predictions under different scenarios.







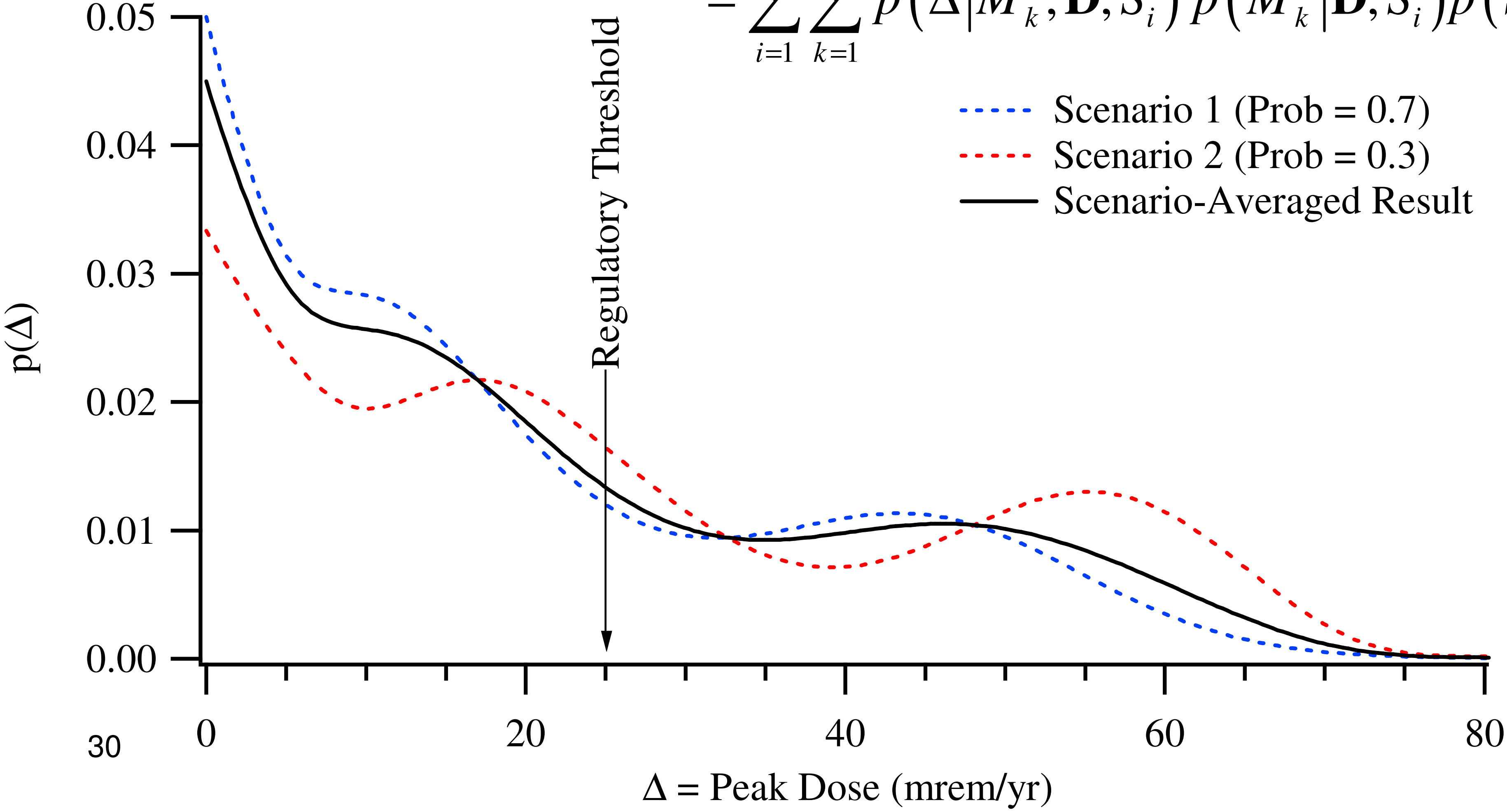
Scenario Averaging Approach

Model and parametric uncertainty

$$p\left(\Delta|\mathbf{D}\right)=\sum_{i=1}^{I}p\left(\Delta|\mathbf{D},S_i\right)p\left(S_i\right)$$

Scenario probability

$$=\sum_{i=1}^{I}\sum_{k=1}^{K}p\left(\Delta|M_k,\mathbf{D},S_i\right)p\left(M_k|\mathbf{D},S_i\right)p\left(S_i\right)$$



Uncertainty Decomposition

Data uncertainty in Δ

$$
\begin{aligned}
Var(\boldsymbol{\Delta}) = {} & E_S E_{M|S} E_{\boldsymbol{\theta}|M,S} \boxed{Var(\boldsymbol{\Delta} \mid \boldsymbol{\theta}, M, S)} \\
& + E_S E_{M|S} \boxed{Var_{\boldsymbol{\theta}|M,S}} E(\boldsymbol{\Delta} \mid \boldsymbol{\theta}, M, S) \\
& + E_S \boxed{Var_{M|S}} E_{\boldsymbol{\theta}|M,S} E(\boldsymbol{\Delta} \mid \boldsymbol{\theta}, M, S) \\
& + \boxed{Var_S} E_{M|S} E_{\boldsymbol{\theta}|M,S} E(\boldsymbol{\Delta} \mid \boldsymbol{\theta}, M, S)
\end{aligned}
$$

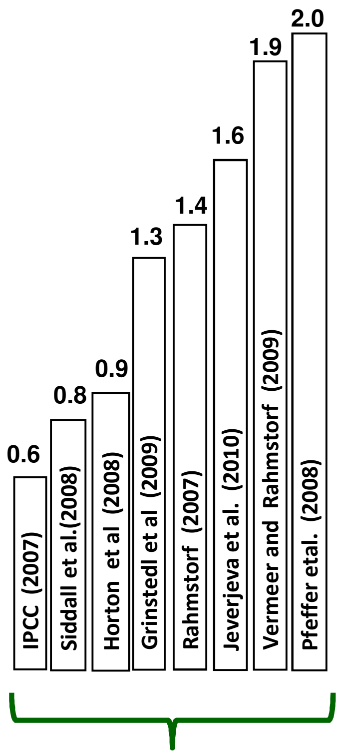
Scenario uncertainty in S

Parametric uncertainty in θ

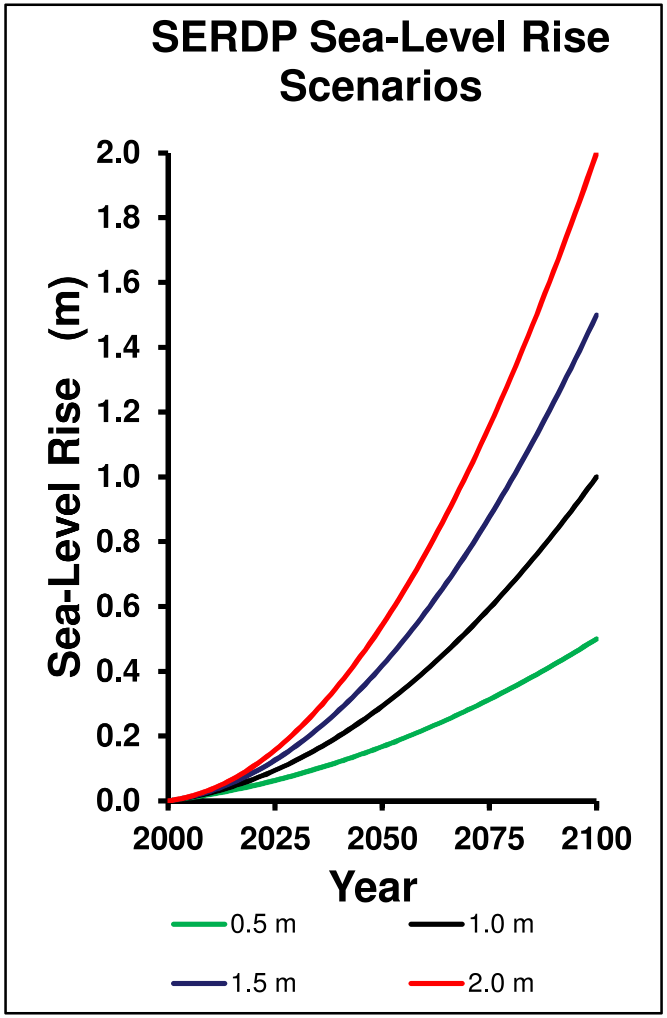
Model uncertainty in M

Sea-Level Rise (SLR) Scenarios

Projecting Future Sea-Level Rise



Recent Published Model Estimates of Year 2100 Sea-Level Rise (m)

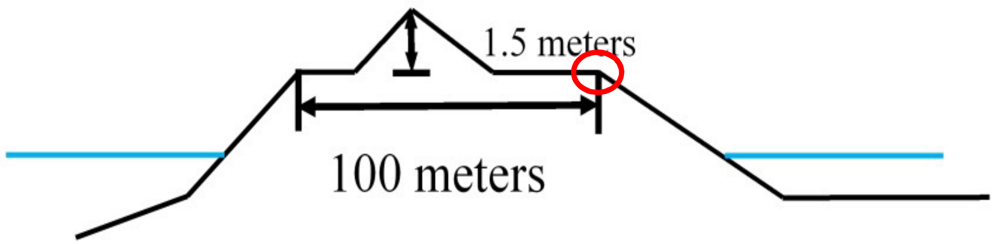


$$S - S_0 = a\,(Y - Y_0) + b\,(Y - Y_0)^2$$

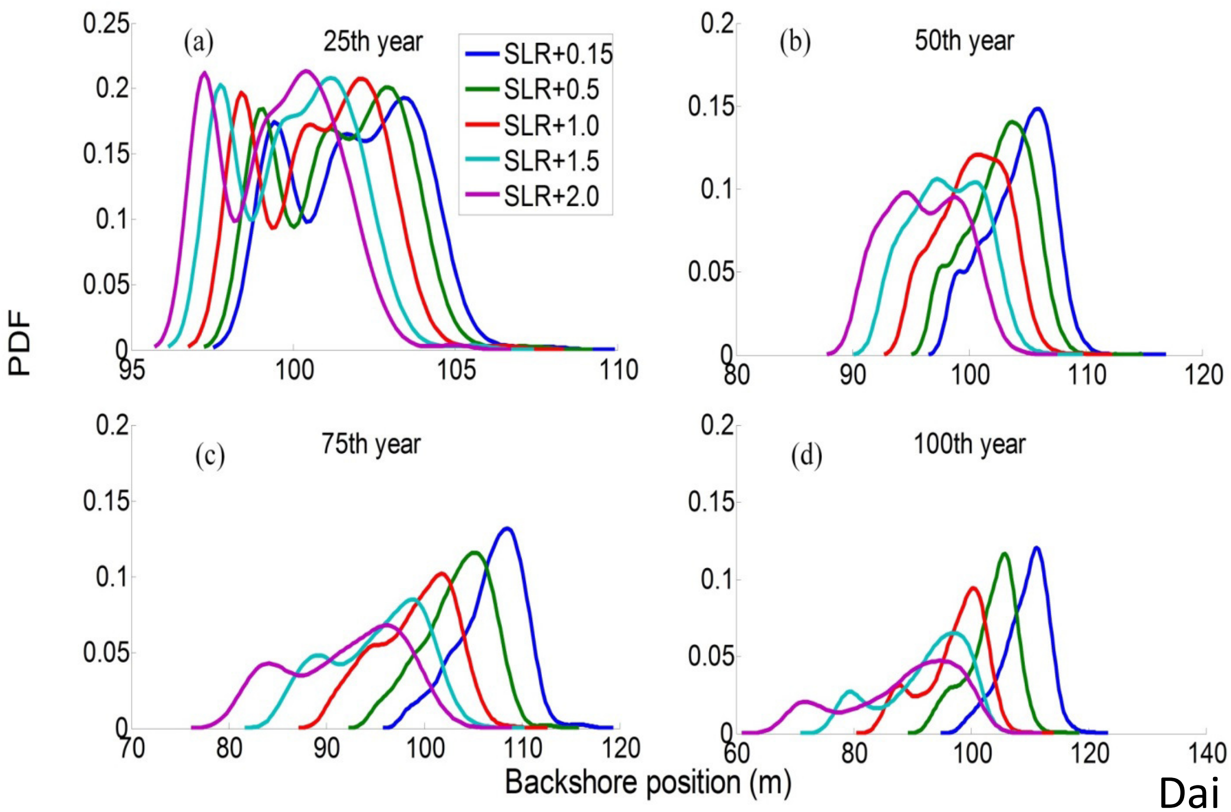
S = Sea level at year Y (taken to be 2100)
S_0 = Sea level at year Y_0 (taken to be 2000)
a = Initial rate of SLR
b = Rate of increase in rate of rise

- Empirical equations to predict sea-level rise given by SERDP
- Predicted sea-level rises agree with literature data.

Impacts of SLR and storms on backshore position



- A single coastal model
- Multiple scenarios of SLR
- Random storm parameters



The PDFs become dramatically different in shape and data range, when simulation time increases.

This manifests the impacts of SLR scenarios on predictive uncertainty.

Dai et al. (2014, J. Coastal Research)

Total effect sensitivity index

For a single model and a single scenario

$$Var(\boldsymbol{\Delta}) = \cancel{E_S E_{M|S} E_{\boldsymbol{\theta}|M,S} Var(\boldsymbol{\Delta} \mid \boldsymbol{\theta}, M, S)}$$ = constant for data uncertainty

$$+ E_S E_{M|S} Var_{\boldsymbol{\theta}|M,S} E(\boldsymbol{\Delta} \mid \boldsymbol{\theta}, M, S)$$

$$\cancel{+ E_S Var_{M|S} E_{\boldsymbol{\theta}|M,S} E(\boldsymbol{\Delta} \mid \boldsymbol{\theta}, M, S)}$$ = 0 No model uncertainty

$$\cancel{+ Var_S E_{M|S} E_{\boldsymbol{\theta}|M,S} E(\boldsymbol{\Delta} \mid \boldsymbol{\theta}, M, S)}$$ = 0 No scenario uncertainty

$$Var(\boldsymbol{\Delta}) = Var_{\boldsymbol{\theta}|M,S} E(\boldsymbol{\Delta} \mid \boldsymbol{\theta}, M, S)$$

$$= V_{\theta_i|M,S} E_{\boldsymbol{\theta}_{\sim i}|M,S} (E(\Delta \mid \boldsymbol{\theta}, M, S) \mid \theta_i)$$

$$+ E_{\theta_i|M,S} V_{\boldsymbol{\theta}_{\sim i}|M,S} (E(\Delta \mid \boldsymbol{\theta}, M, S) \mid \theta_i)$$

$$S_{T_i} = \frac{E_{\boldsymbol{\theta}_{\sim i}|M,S} V_{\theta_i|M,S} (E(\Delta \mid \boldsymbol{\theta}, M, S) \mid \boldsymbol{\theta}_{\sim i})}{V_{\boldsymbol{\theta}|M,S} E(\Delta \mid \boldsymbol{\theta}, M, S)}$$

Total effect sensitivity index

For multiple models but a single scenario

$$S_{T_i}^{\mathbf{M}} = \frac{E_{\mathbf{M}|S} E_{\boldsymbol{\theta}_{\sim i}|\mathbf{M},S} V_{\theta_i|\mathbf{M},S}(E(\Delta\,|\,\boldsymbol{\theta},\mathbf{M},S)\,|\,\boldsymbol{\theta}_{\sim i})}{E_{\mathbf{M}|S} V_{\boldsymbol{\theta}|\mathbf{M},S} E(\Delta\,|\,\boldsymbol{\theta},\mathbf{M},S)}$$

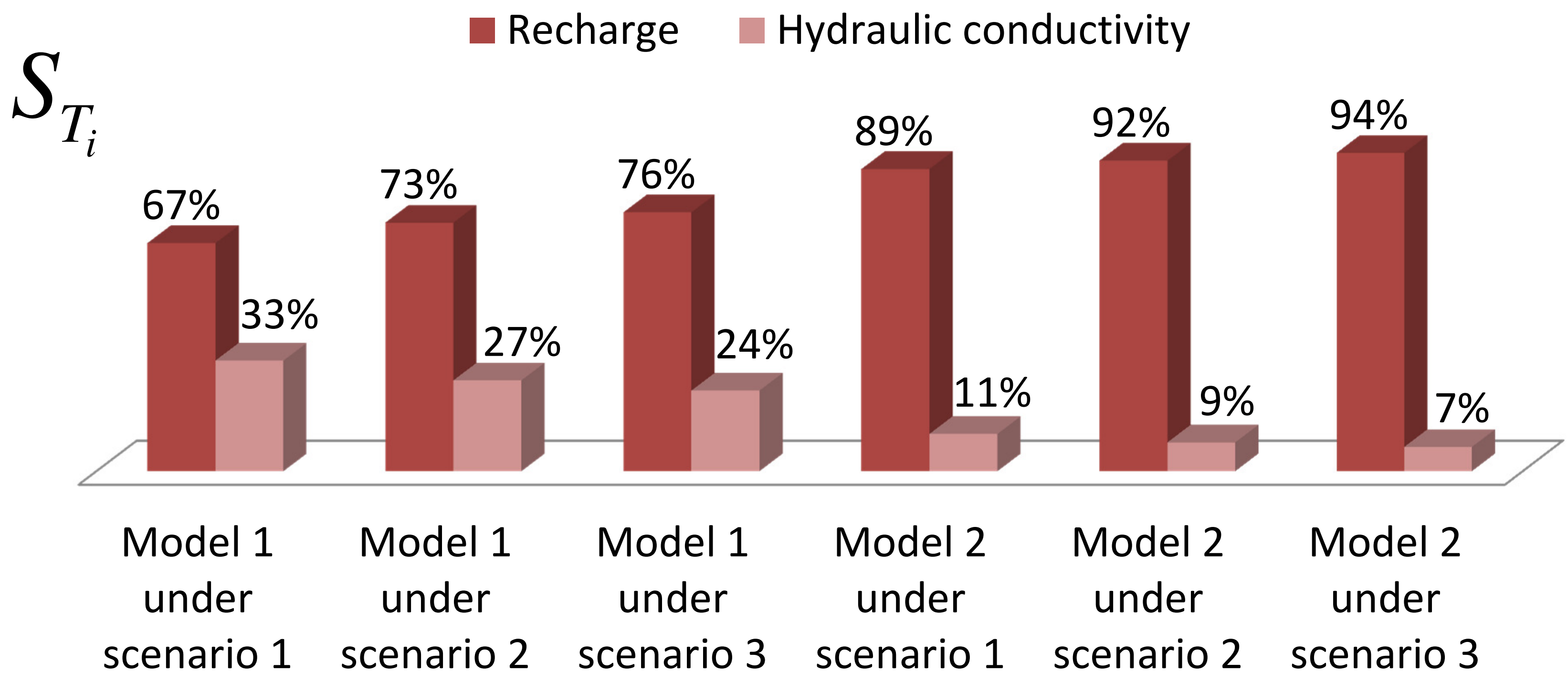
$$= \frac{\sum_{M} P(M\,|\,S) E_{\boldsymbol{\theta}_{\sim i}|M,S} V_{\theta_i|M,S}(E(\Delta\,|\,\boldsymbol{\theta},M,S)\,|\,\boldsymbol{\theta}_{\sim i})}{\sum_{M} P(M\,|\,S) V_{\boldsymbol{\theta}|M,S} E(\Delta\,|\,\boldsymbol{\theta},M,S)}$$

For multiple models and multiple scenarios

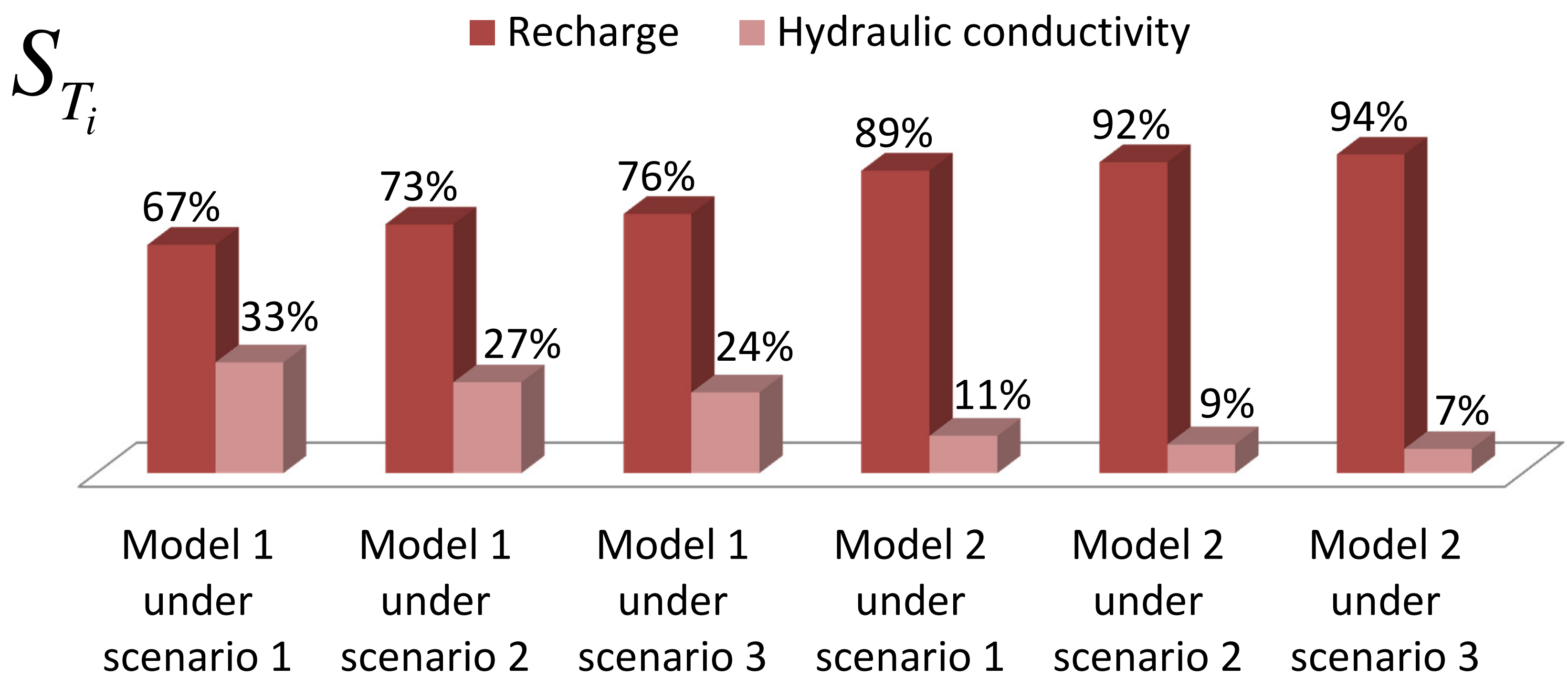
$$S_{T_i}^{\mathbf{MS}} = \frac{E_{\mathbf{s}} E_{\mathbf{M}|\mathbf{S}} V_{\boldsymbol{\theta}_{\sim i}|\mathbf{M},\mathbf{S}} E_{\theta_i|\mathbf{M},\mathbf{S}}(E(\Delta\,|\,\boldsymbol{\theta},\mathbf{M},\mathbf{S})\,|\,\boldsymbol{\theta}_{\sim i})}{E_{\mathbf{s}} E_{\mathbf{M}|\mathbf{S}} V_{\boldsymbol{\theta}|\mathbf{M},\mathbf{S}} E(\Delta\,|\,\boldsymbol{\theta},\mathbf{M},\mathbf{S})}$$

$$= \frac{\sum_{S}\sum_{M} P(S) P(M\,|\,S) E_{\boldsymbol{\theta}_{\sim i}|M,S} V_{\theta_i|M,S}(E(\Delta\,|\,\boldsymbol{\theta},M,S)\,|\,\boldsymbol{\theta}_{\sim i})}{\sum_{S}\sum_{M} P(S) P(M\,|\,S) V_{\boldsymbol{\theta}|M,S} E(\Delta\,|\,\boldsymbol{\theta},M,S)}$$

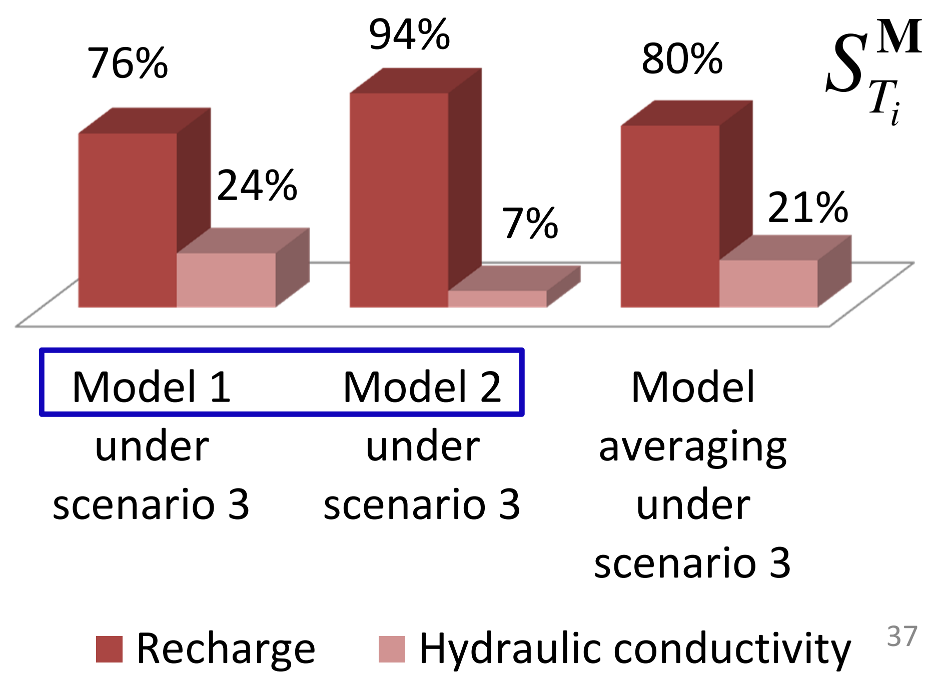
Head Sensitivity Analysis:
Individual models and individual scenarios



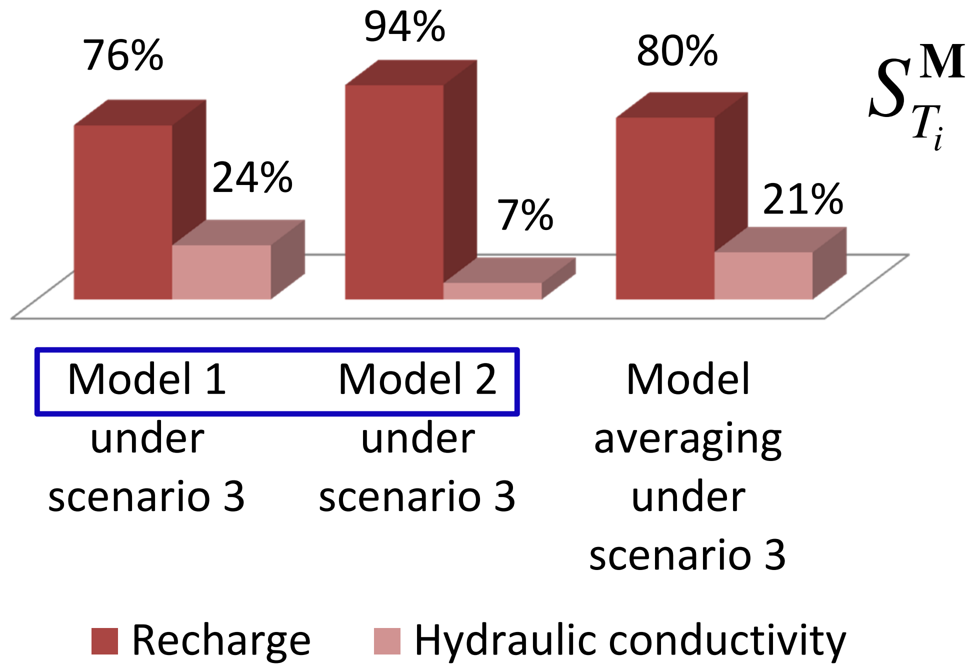
- Parameter sensitivity is different for different models under different scenarios.
- Using a single model and a single scenario may lead to biased identification of important parameters.

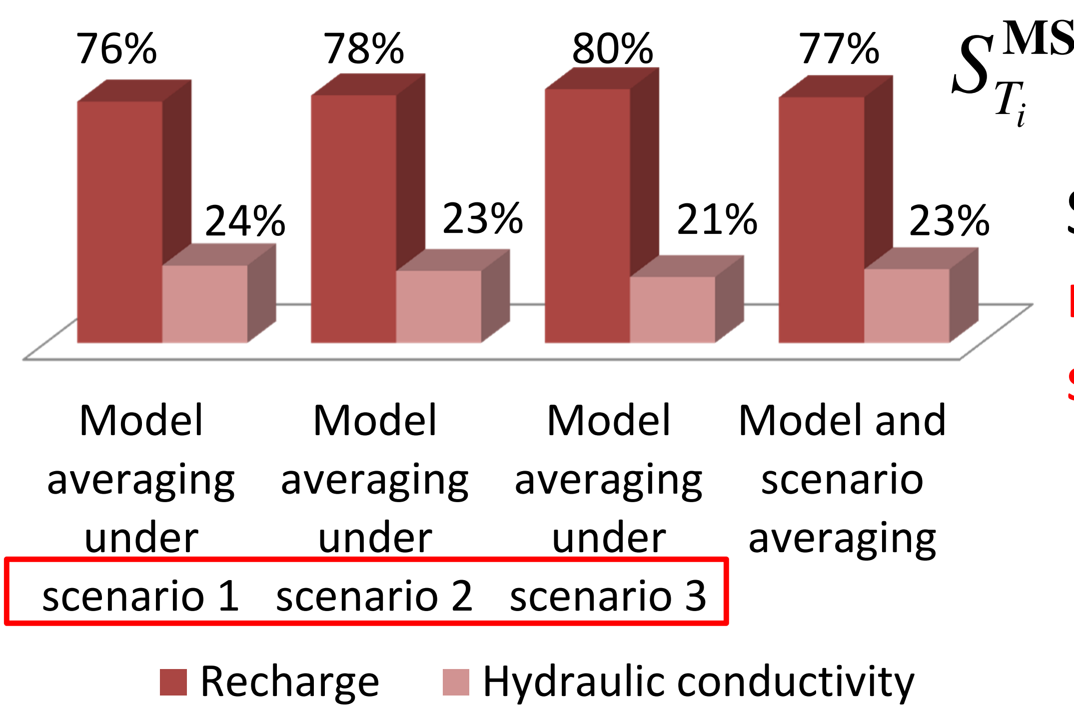


- Sensitivity for multiple models under a single scenario.



Sensitivity for multiple models under a single scenario





Sensitivity for multiple models under multiple scenarios

How methods relate?

Common questions	Frugal methods	Expensive methods
Model Adequacy		
[1] How can many data types with variable quality be included?	Error-based weighting and SOO or MAP	MOO, Pareto curve
[2] Is model misfit/overfit a problem? Are prior knowledge and data subsets inconsistent?	Compare fit to a priori error analysis using s_n^2, $s_{(n-p)}^2$, RMSE, Nash-Sutcliffe, graphs	MOO, Pareto curve
[3] How nonlinear is the problem?	Intrinsic nonlinearity, DELSA	DELSA, Explore objective function,
Sensitivity and Uncertainty		
Observations (Obs) ⟷ Parameters (Pars)		
[4] What pars can and cannot be estimated with the obs?	Scaled local stats (CSS, ID, PCC, etc.), SVD, DoE, MoM (OAT)	DoE, MoM (OAT), eFAST, Sobol', RSA, CSE
[5] Are any pars dominated by one obs and, thus, its error?	Scaled local stats (Leverage, DFBETAS)	Cross Validation
[6] How certain are the par values?	Par uncertainty intervals	Par uncertainty intervals[3]
[7] Which obs are important and unimportant to pars?	Scaled local stats (Leverage, Cook's D)	Cross Validation
Parameters (Pars) ⟷ Predictions (Preds)		
[8] Which pars are important and unimportant to preds?	Scaled local stats (PSS, etc.), DELSA	DELSA, FAST, Sobol'
[9] How certain are the preds?	z/SD_z, Pred uncertainty intervals	Pred uncertainty intervals, MCMC, multi-model analysis
[10] Which pars contribute most and least to pred uncertainty?	Scaled local stats (PPR)	FAST, Sobol'
Observations (Obs) ⟷ Predictions (Preds)		
[11] Which existing and potential obs are important to the preds?	Scaled local stats (OPR)	Cross Validation
[12] For multi-model analysis, which models are likely to produce accurate preds?	Analyze model fit and estimated parameters, AIC, AICc, BIC, KIC	Cross Validation
Risk Assessment		
[13] What risk is associated with a given decision strategy?	Combine uncertainty analysis and scenario simulation. Smooth cost function.	Combine uncertainty analysis and scenario simulation. Cost function need not be smooth.

- Different methods give similar results.
- Select the methods that are most computationally efficient.

Bayesian vs. frequentist methods

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad\qquad \boldsymbol{\varepsilon} : N_n(\mathbf{0}, \mathbf{C}_\varepsilon)$$

- Regression methods

$$\hat{\mathbf{b}} : N_p(\boldsymbol{\beta}, (\mathbf{X}^T\mathbf{C}_\varepsilon^{-1}\mathbf{X})^{-1}) \propto \exp\left[-\frac{1}{2}\boxed{\left(\hat{\mathbf{b}}-\boldsymbol{\beta}\right)}^T \mathbf{X}^T\mathbf{C}_\varepsilon^{-1}\mathbf{X}\left(\hat{\mathbf{b}}-\boldsymbol{\beta}\right)\right]$$

$$g(\hat{\mathbf{b}}) : N(g(\boldsymbol{\beta}), \mathbf{Z}(\mathbf{X}^T\mathbf{C}_\varepsilon^{-1}\mathbf{X})^{-1}\mathbf{Z}^T)$$

$$\propto \exp\left[-\frac{1}{2}\boxed{\left(g(\hat{\mathbf{b}})-g(\boldsymbol{\beta})\right)}^T \left(\mathbf{Z}(\mathbf{X}^T\mathbf{C}_\varepsilon^{-1}\mathbf{X})^{-1}\mathbf{Z}^T\right)^{-1}\left(g(\hat{\mathbf{b}})-g(\boldsymbol{\beta})\right)\right]$$

- Regression methods

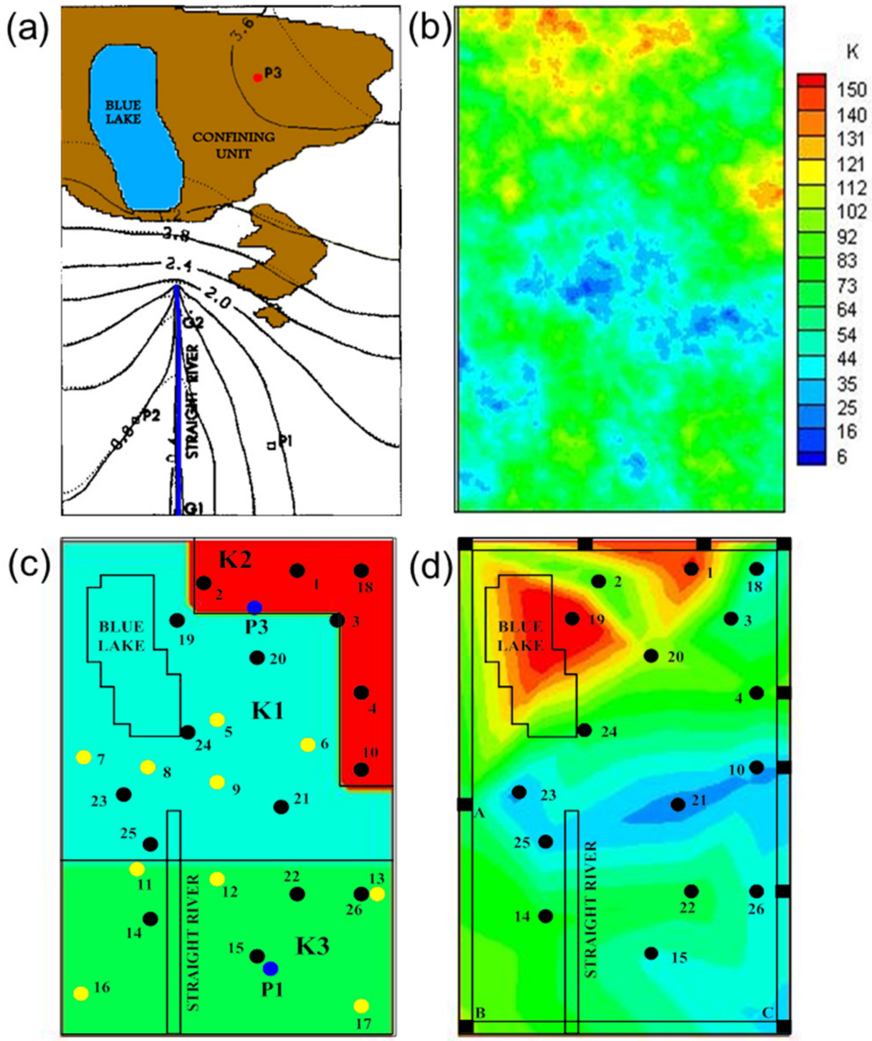
$$\boldsymbol{\beta} : N_p(\hat{\mathbf{b}}, (\mathbf{X}^T\mathbf{C}_\varepsilon^{-1}\mathbf{X})^{-1}) \propto \exp\left[-\frac{1}{2}\boxed{\left(\boldsymbol{\beta}-\hat{\mathbf{b}}\right)}^T \mathbf{X}^T\mathbf{C}_\varepsilon^{-1}\mathbf{X}\left(\boldsymbol{\beta}-\hat{\mathbf{b}}\right)\right]$$

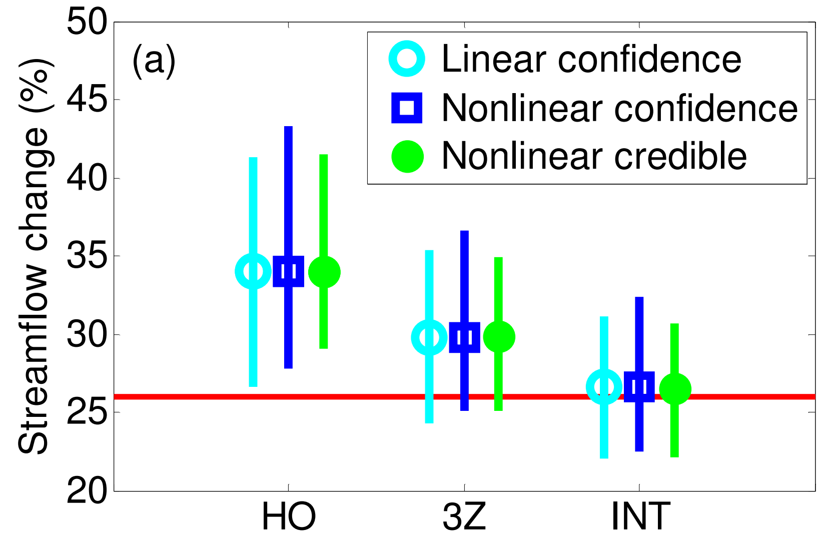
$$g(\boldsymbol{\beta}) : N_p(g(\hat{\mathbf{b}}), \mathbf{Z}(\mathbf{X}^T\mathbf{C}_\varepsilon^{-1}\mathbf{X})^{-1}\mathbf{Z}^T)$$

$$\propto \exp\left[-\frac{1}{2}\boxed{\left(g(\boldsymbol{\beta})-g(\hat{\mathbf{b}})\right)}^T \left(\mathbf{Z}(\mathbf{X}^T\mathbf{C}_\varepsilon^{-1}\mathbf{X})^{-1}\mathbf{Z}^T\right)^{-1}\left(g(\boldsymbol{\beta})-g(\hat{\mathbf{b}})\right)\right]$$

Lu et al. (2012, Water Resources Research)

Credible vs. confidence intervals





- The credible and confidence intervals are similar.
- Model runs for all three models (HO, 3Z, INT) are 166, 2,309, and 1.5 million, respectively.

Lu et al. (2012, Water Resources Research)

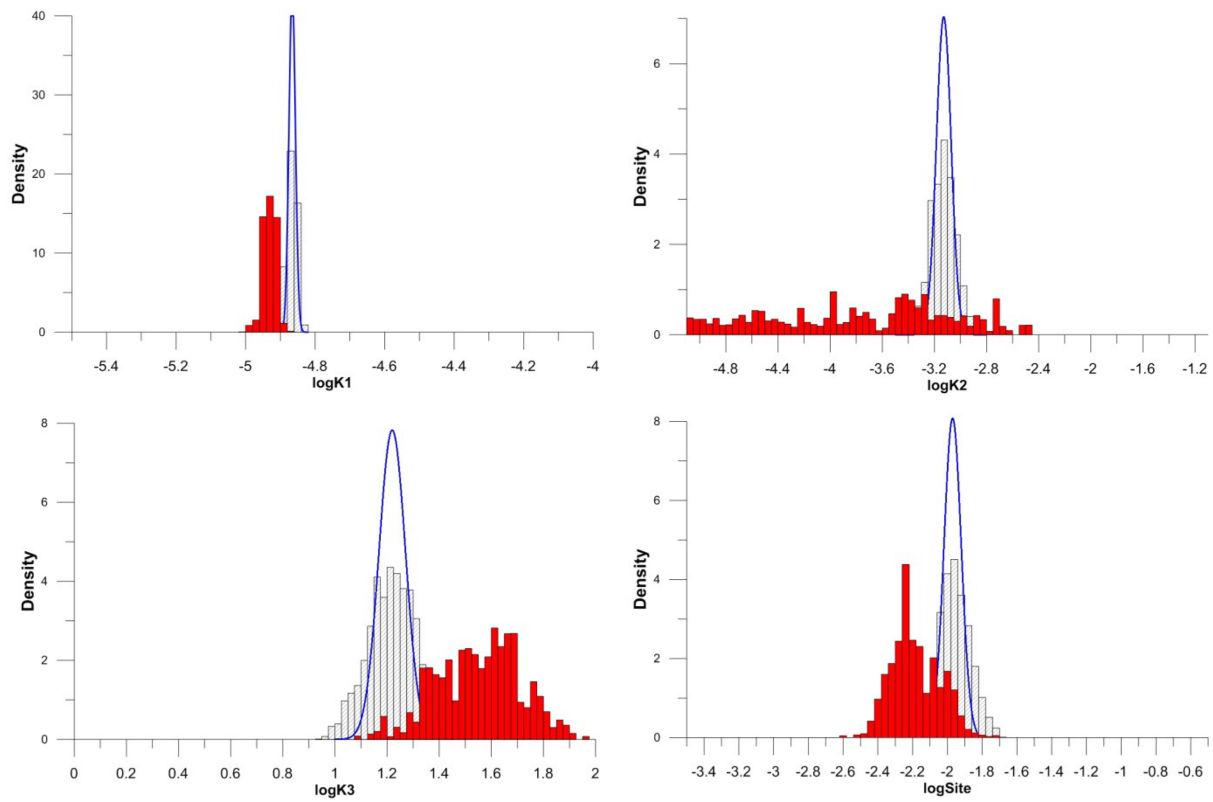
When the results are different

Surface complexation model

$$\boxed{S_1OH}+UO_2^{2+}+H_2O=S_1OUO_2OH+2H^+ \quad \text{logK1}$$

$$S_2OH+UO_2^{2+}+H_2O=S_2OUO_2OH+2H^+ \quad \text{logK2}$$

$$S_2OH+UO_2^{2+}=S_2OUO_2^{+}+H^+ \quad \text{logK3}$$



Density functions of the four model parameters:

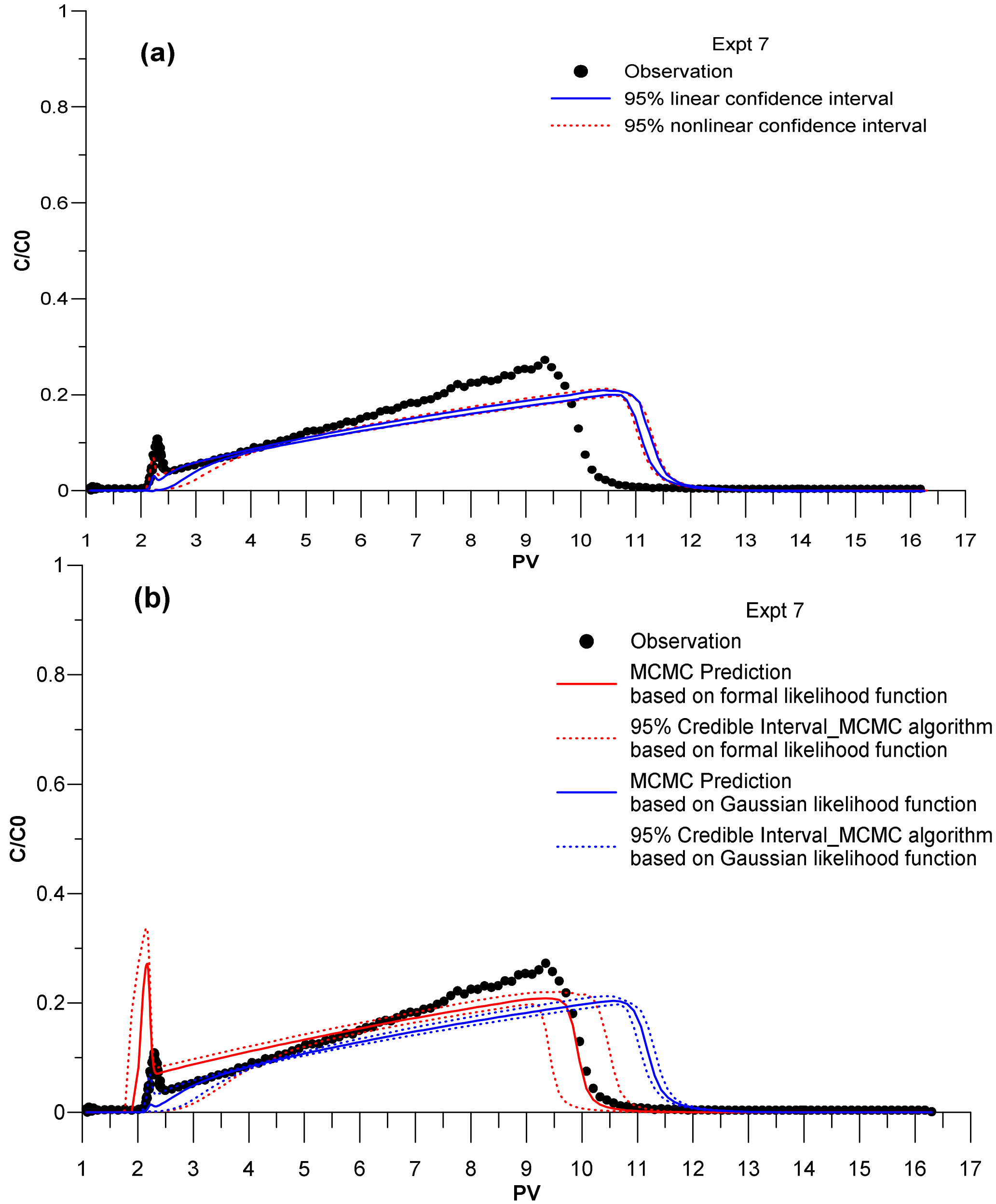
Blue: results of a regression method with **Gaussian** likelihood function

Grey: results of a Bayesian method with **Gaussian** likelihood function

Red: results of Bayesian method with **non-Gaussian** likelihood function

Shi et al. (2014, Water Resources Research)

Credible vs. confidence intervals



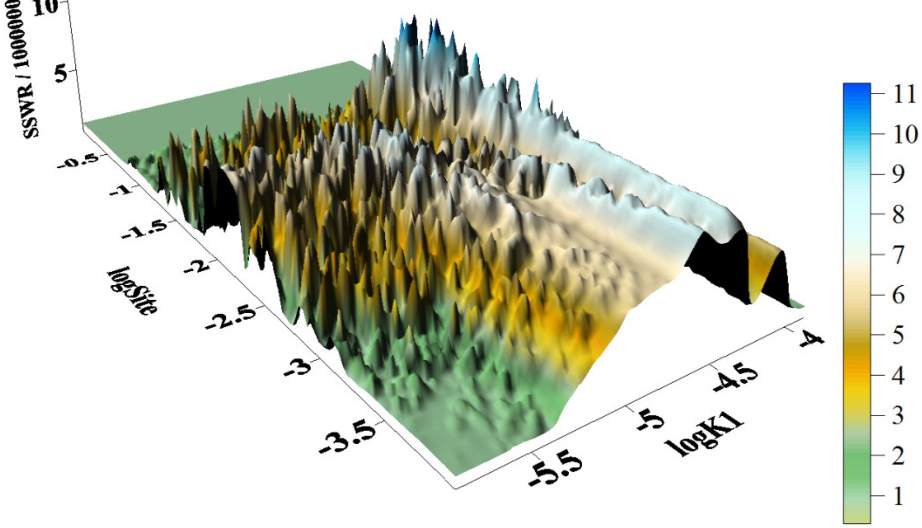
- Credible interval and confidence intervals are similar when Gaussian likelihood is used.
- The credible interval with non-Gaussian likelihood is superior.

Shi et al. (2014, Water Resources Research)

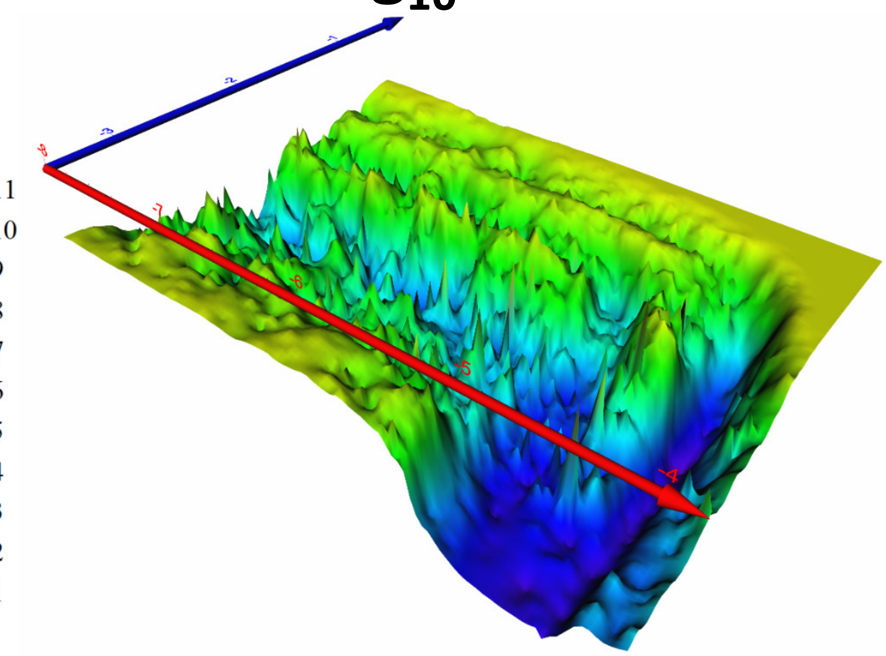
Why are the results different?

- The reaction model is highly nonlinear
- The response surface of SSWR (sum of squared weighted residuals) is extremely irregular.

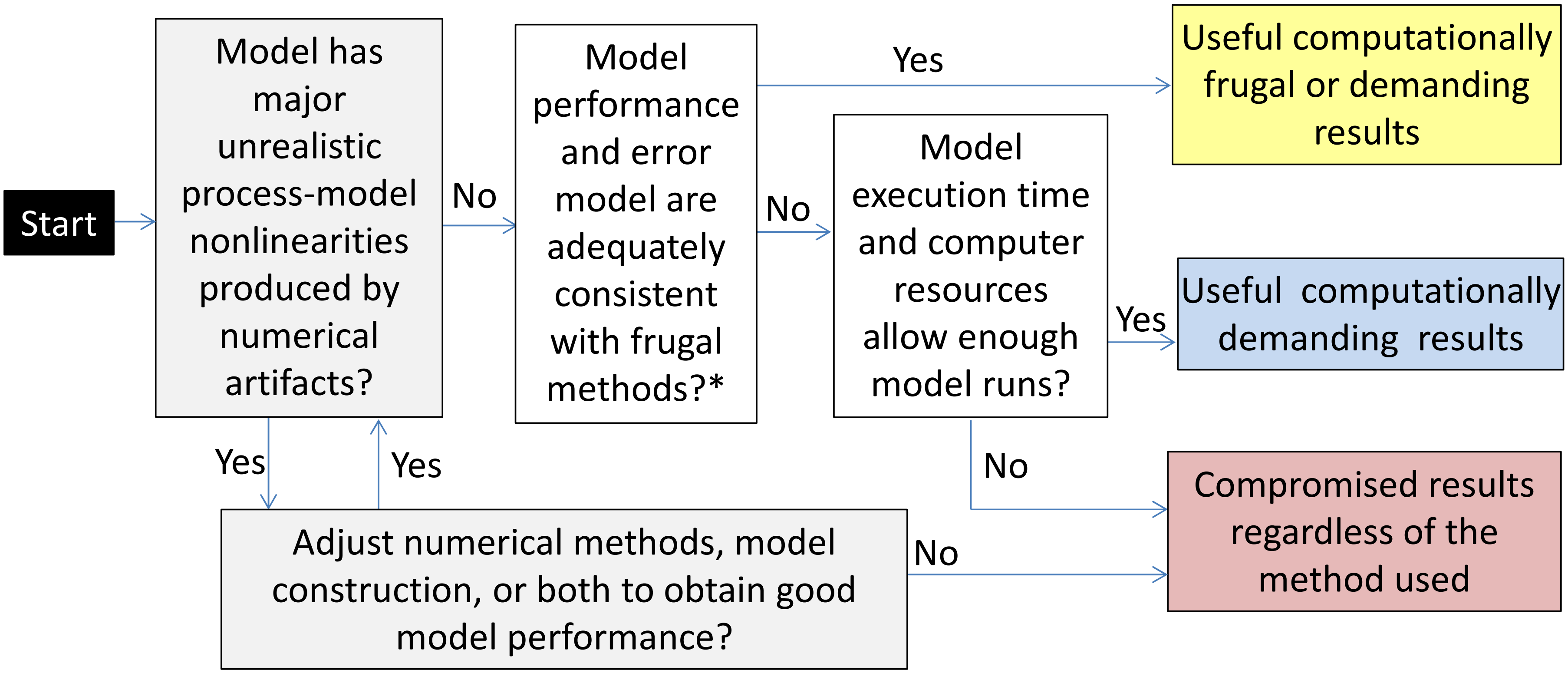
SSWR



$-\log_{10}$SSWR



Ad-hoc comparable metrics



*Yes: Simulated results vary smoothly as parameter values change for parameters of interest. If multiple local minima exist one can be identified as the realistic solution. The error model is unimodal. These characteristics are determined based on knowledge of model theory and numerical methods, and easily conducted tests.

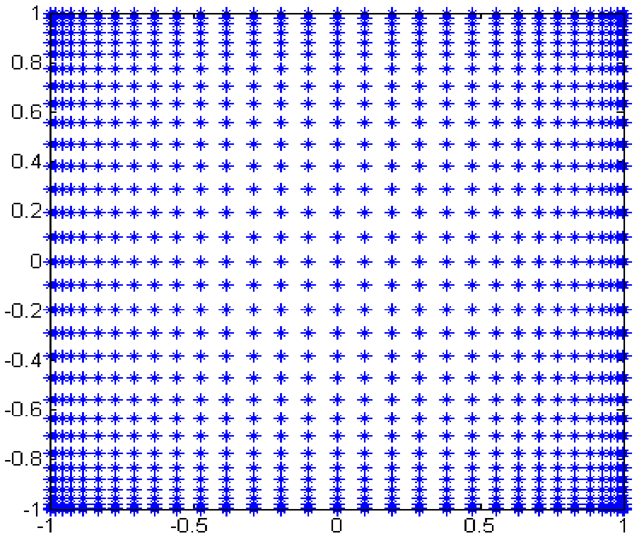
Computational Efficient Methods

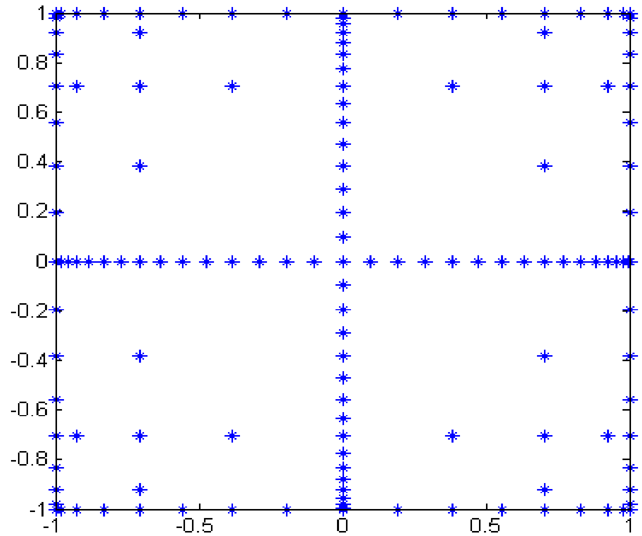
$$S_{T_i}^{\mathbf{M}} = \frac{E_{\mathbf{M}|S} E_{\boldsymbol{\theta}_{\sim i}|\mathbf{M},S} V_{\theta_i|\mathbf{M},S} (E(\Delta \mid \boldsymbol{\theta}, \mathbf{M}, S) \mid \boldsymbol{\theta}_{\sim i})}{E_{\mathbf{M}|S} V_{\boldsymbol{\theta}|\mathbf{M},S} E(\Delta \mid \boldsymbol{\theta}, \mathbf{M}, S)}$$

What if tens or even hundreds of thousands of model runs are needed?

- Build a surrogate model whose execution time is negligible
  - Polynomial chaos expansion
  - Sparse grid collocation
  - Reduced order modeling
- Run the surrogate model at least as possible

Sparse Grid Collocation (SGC)





Sparse grid interpolation

polynomials

$$\eta(\theta_1,\ldots,\theta_{N_\theta}) \approx \mathcal{I}^{L,N_\theta}(\eta) = \sum_{|l|\le L} (-1)^{L-|l|} \binom{N_\theta - 1}{L-|l|} \sum_{j_1=1}^{N_{l_1}} \ldots \sum_{j_{N_\theta}=1}^{N_{l_{N_\theta}}} \eta(\theta_{j_1},\ldots,\theta_{j_{N_\theta}}) \prod_{n=1}^{N_\theta} \boxed{\phi_{j_n}(\theta_n)}$$

Sparse grid integration

$$\int_\Gamma \eta(\boldsymbol{\theta})\, d\boldsymbol{\theta} \approx \int_\Gamma \mathcal{I}^{L,N_\theta}(\eta)\, d\boldsymbol{\theta} = \sum_{i=1}^{N_s} \omega_i \eta(\boldsymbol{\theta}_i)$$

MCMC with Use SGC Surrogate

Bayes' theorem for MCMC simulation

$$p\left(\boldsymbol{\theta}_k \middle| \mathbf{D}, M_k\right) = \frac{p\left(\mathbf{D} \middle| \boldsymbol{\theta}_k, M_k\right) p\left(\boldsymbol{\theta}_k \middle| M_k\right)}{p\left(\mathbf{D} \middle| M_k\right)}$$

Likelihood of standard MCMC

$$p\left(\mathbf{D} \middle| \boldsymbol{\theta}, M\right) = \frac{1}{(2\pi)^{N_D/2} \left|\Sigma_{\mathbf{D}}\right|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{D} - \boxed{\mathbf{f}(\boldsymbol{\theta})})^T \Sigma_D^{-1} (\mathbf{D} - \mathbf{f}(\boldsymbol{\theta}))\right]$$

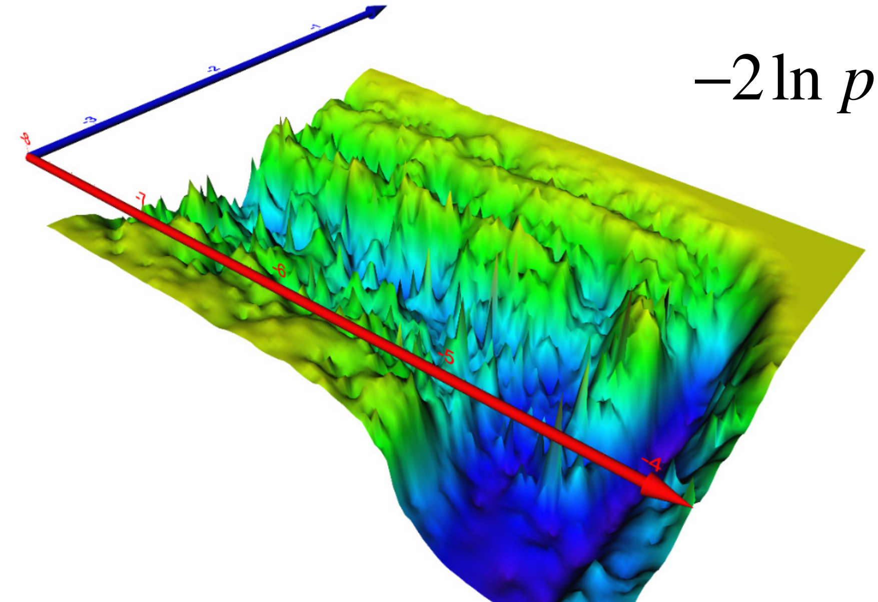
Sparse-grid surrogate of likelihood

$$\hat{p}\left(\mathbf{D} \middle| \boldsymbol{\theta}, M\right) = \frac{1}{(2\pi)^{N_D/2} \left|\Sigma_{\mathbf{D}}\right|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{D} - \boxed{\mathcal{I}^{L,N_\theta}(\boldsymbol{\theta})})^T \Sigma_D^{-1} (\mathbf{D} - \mathcal{I}^{L,N_\theta}(\boldsymbol{\theta}))\right]$$

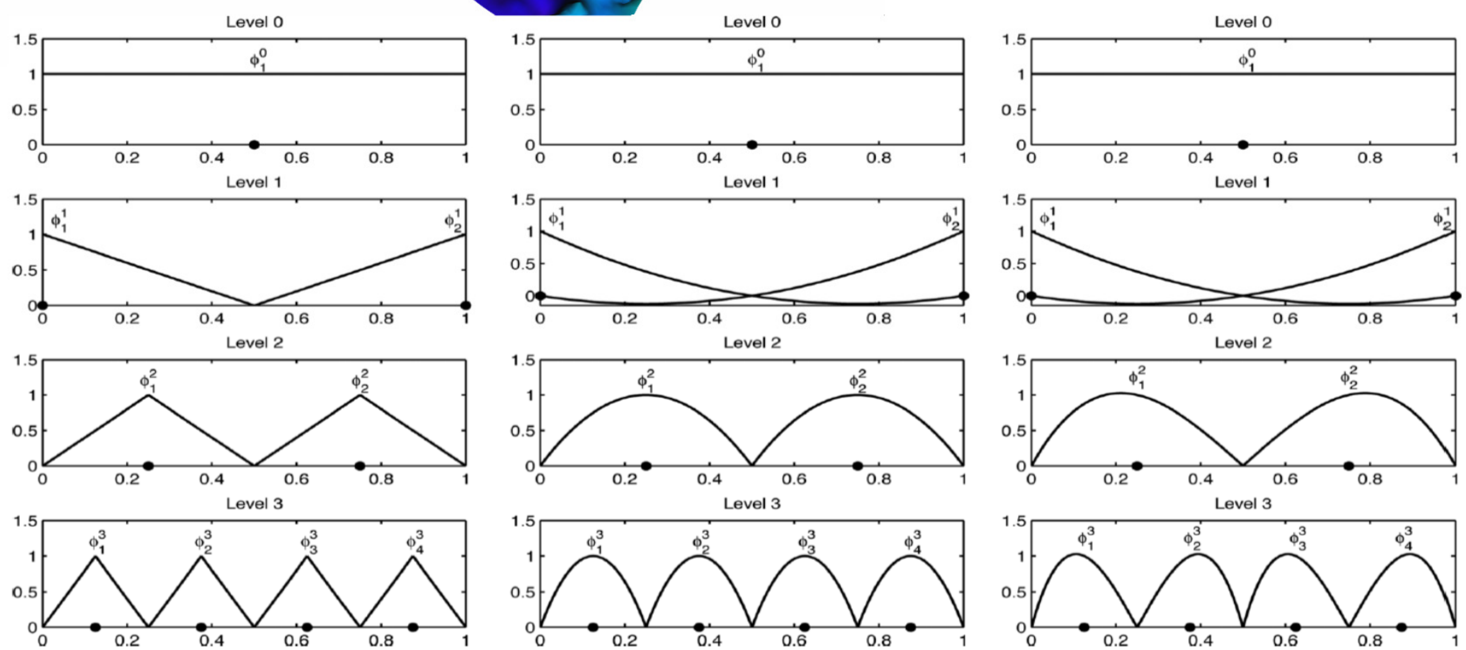
The idea is simple, but the implementation is NOT.

Response Surface of log Likelihood

$$-2\ln p\left(\mathbf{D}\middle|\boldsymbol{\theta},M\right)\propto(\mathbf{D}-\mathbf{f}(\boldsymbol{\theta}))^T\Sigma_D^{-1}(\mathbf{D}-\mathbf{f}(\boldsymbol{\theta}))$$

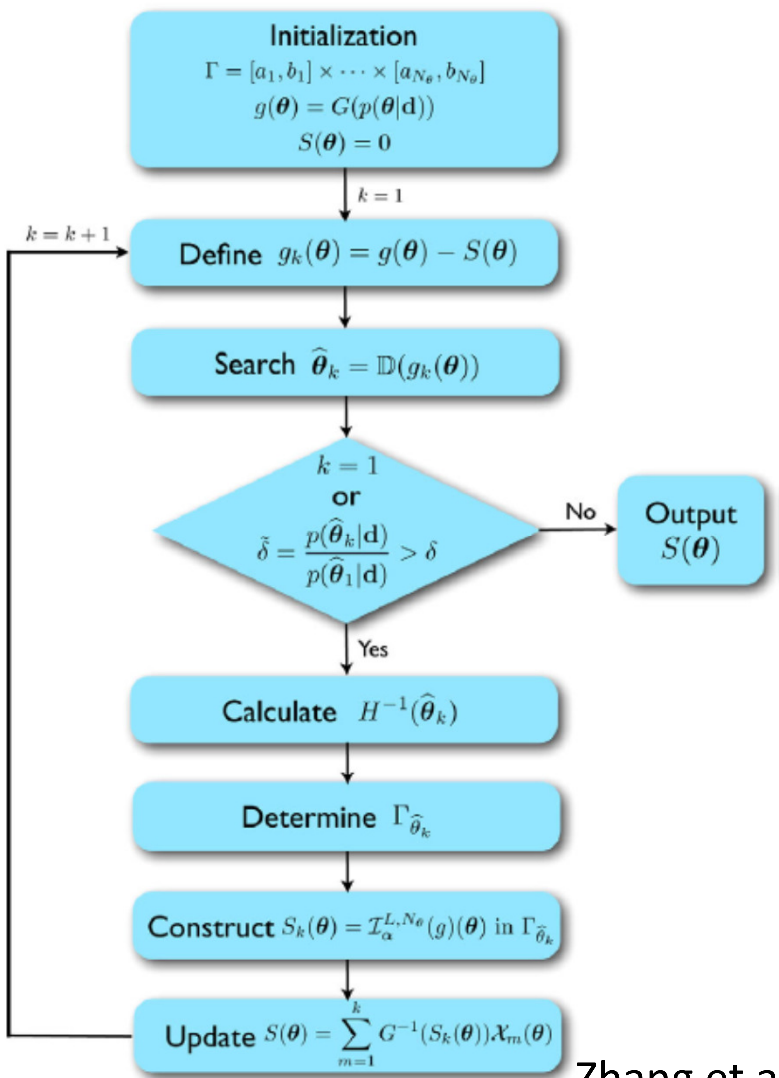


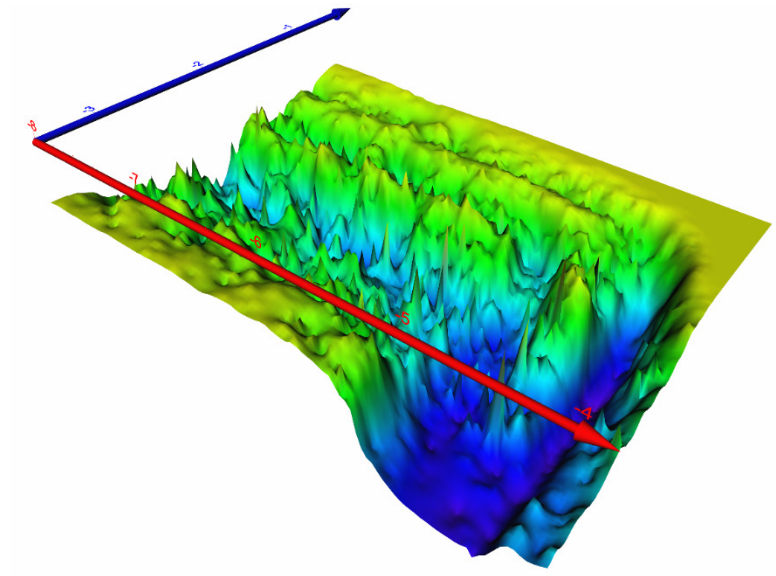
- Complicated response surface with multiple peaks and valleys.
- It is very difficult to build a surrogate for such a surface.

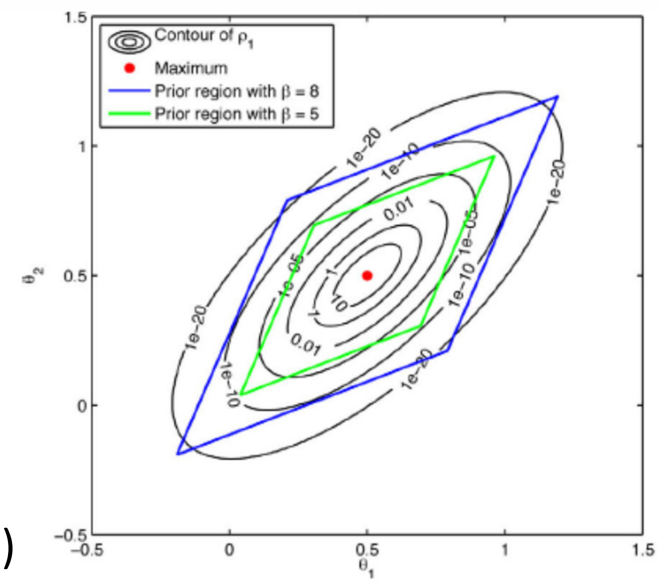


Zhang et al. (2013)

Adaptive Sparse-Grid
High-order Stochastic Collocation

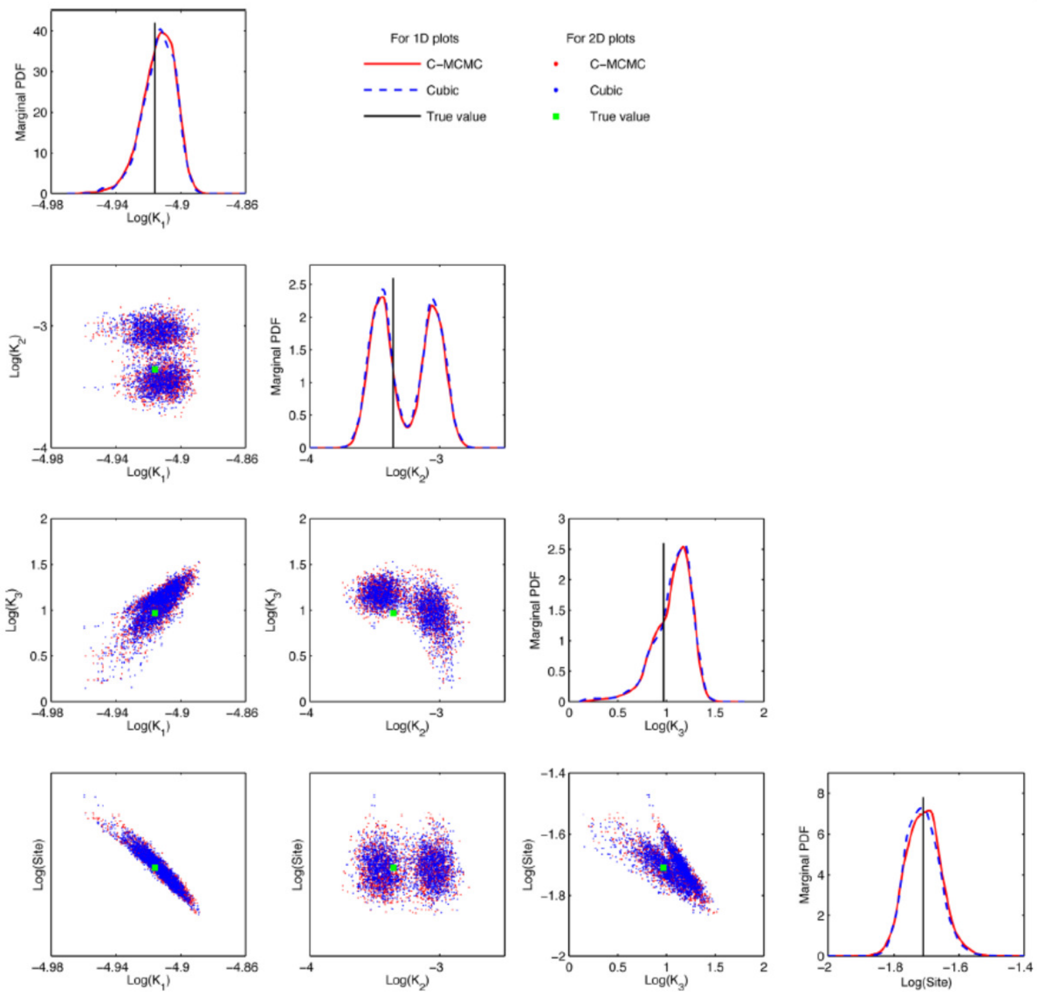






Zhang et al. (2013)

Evaluation of Accuracy and Efficiency



Accuracy

The probability densities obtained using SGC are almost identical to those of the original model.

- Efficiency
- 60,000 model runs needed for the original MCMC
- 9,647 models runs needed for building the surrogate

Zhang et al. (2013,)

Computational Efficient Methods

$$S_{T_i}^{\mathbf{M}} = \frac{E_{\mathbf{M}|S} E_{\boldsymbol{\theta}_{\sim i}|\mathbf{M},S} V_{\theta_i|\mathbf{M},S}\left(E(\Delta \mid \boldsymbol{\theta}, \mathbf{M}, S) \mid \boldsymbol{\theta}_{\sim i}\right)}{E_{\mathbf{M}|S} V_{\boldsymbol{\theta}|\mathbf{M},S} E(\Delta \mid \boldsymbol{\theta}, \mathbf{M}, S)}$$

Run the surrogate model at least as possible

$$V_{\theta_i}(E_{\boldsymbol{\theta}_{\sim i}}(\Delta \mid \theta_i)) = \boxed{E_{\theta_i}\left(\boxed{\left(E_{\boldsymbol{\theta}_{\sim i}}(\Delta \mid \theta_i)\right)}^2\right)} - \boxed{\left(E_{\theta_i}\boxed{\left(E_{\boldsymbol{\theta}_{\sim i}}(\Delta \mid \theta_i)\right)}\right)^2}$$

$$E_{\boldsymbol{\theta}_{\sim i}}(\Delta \mid \theta_i) \approx \mathcal{I}_1(l_1, d-1)[f(\boldsymbol{\theta}_{\sim i})]$$

$$E_{\theta_i}\left(\left(E_{\boldsymbol{\theta}_{\sim i}}(\Delta \mid \theta_i)\right)^2\right) \approx \mathcal{I}_2(l_2, 1)\left[\left(\mathcal{I}_1(l_1, d-1)[f(\boldsymbol{\theta}_{\sim i})]\right)^2\right]$$

$$E_{\theta_i}\left(E_{\boldsymbol{\theta}_{\sim i}}(\Delta \mid \theta_i)\right) \approx \mathcal{I}_3(l_3, 1)\left[\mathcal{I}_1(l_1, d-1)[f(\boldsymbol{\theta}_{\sim i})]\right]$$

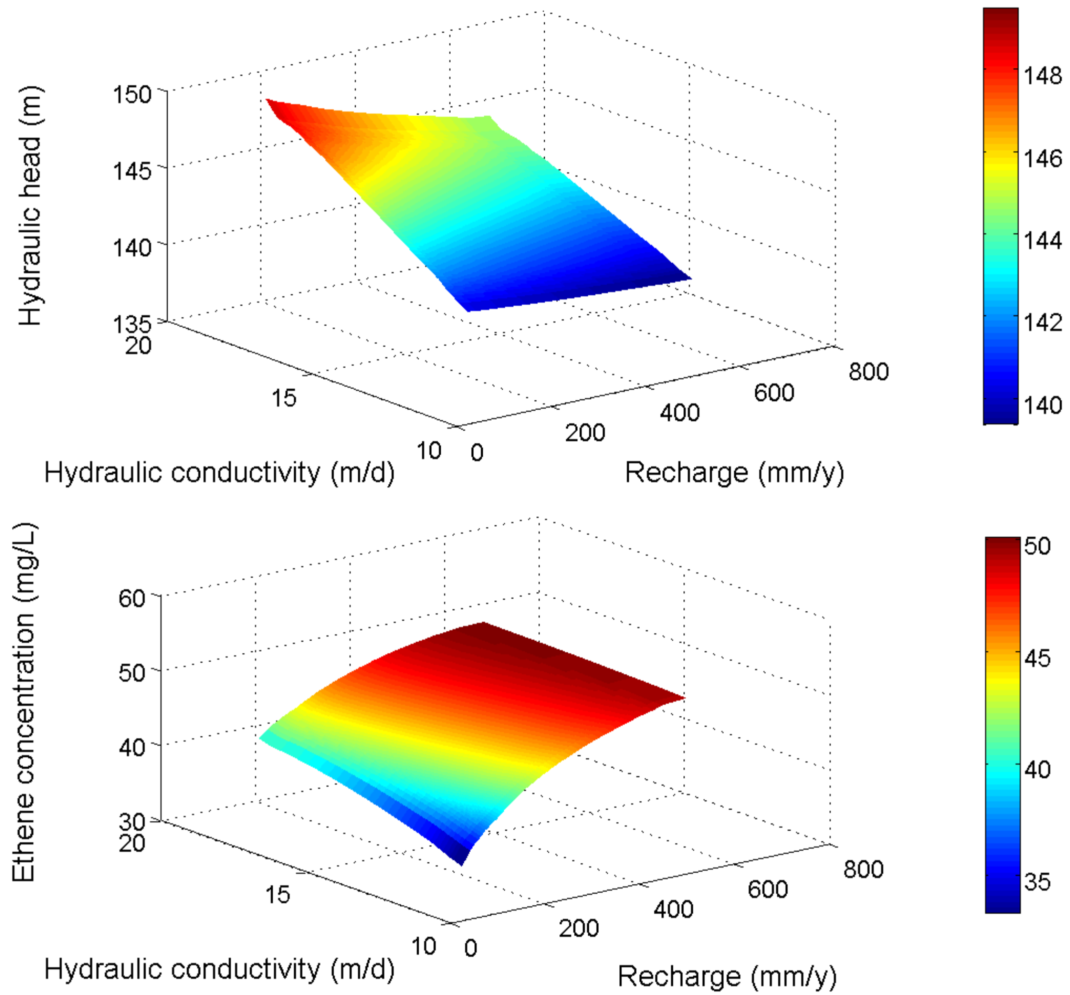
Computational Efficiency

$$S_{T_i}^{\mathbf{M}} = \frac{E_{\mathbf{M}|S} E_{\boldsymbol{\theta}_{\sim i}|\mathbf{M},S} V_{\theta_i|\mathbf{M},S}(E(\Delta|\boldsymbol{\theta},\mathbf{M},S)|\boldsymbol{\theta}_{\sim i})}{E_{\mathbf{M}|S} V_{\boldsymbol{\theta}|\mathbf{M},S} E(\Delta|\boldsymbol{\theta},\mathbf{M},S)}$$

NMCS	NME	S_{T_K}	S_{T_R}	NSGP	Level	NME	S_{T_K}	S_{T_R}
For hydraulic head at x = 6,000 m								
50	200	0.160	0.640	9	1	18	0.325	0.791
100	400	0.240	0.878	25	2	50	0.268	0.735
1,000	4,000	0.265	0.767	49	3	98	0.268	0.735
10,000	40,000	0.251	0.732					
100,000	400,000	0.269	0.735					
1,000,000	4,000,000	0.268	0.735					
For ethene concentration at x = 6,000 m and on 1,000 days								
50	300	0.700	0.052	21	1	84	0.900	-0.047
100	600	1.105	0.043	145	2	590	0.947	0.035
1,000	6,000	0.849	0.039	651	3	2,898	0.951	0.040
10,000	60,000	0.959	0.043	2,277	4	11,880	0.952	0.042
100,000	600,000	0.960	0.043					
1,000,000	6,000,000	0.952	0.042					

Reasons for the efficiency

Head and concentrations are smooth functions of model parameters.



Conclusion

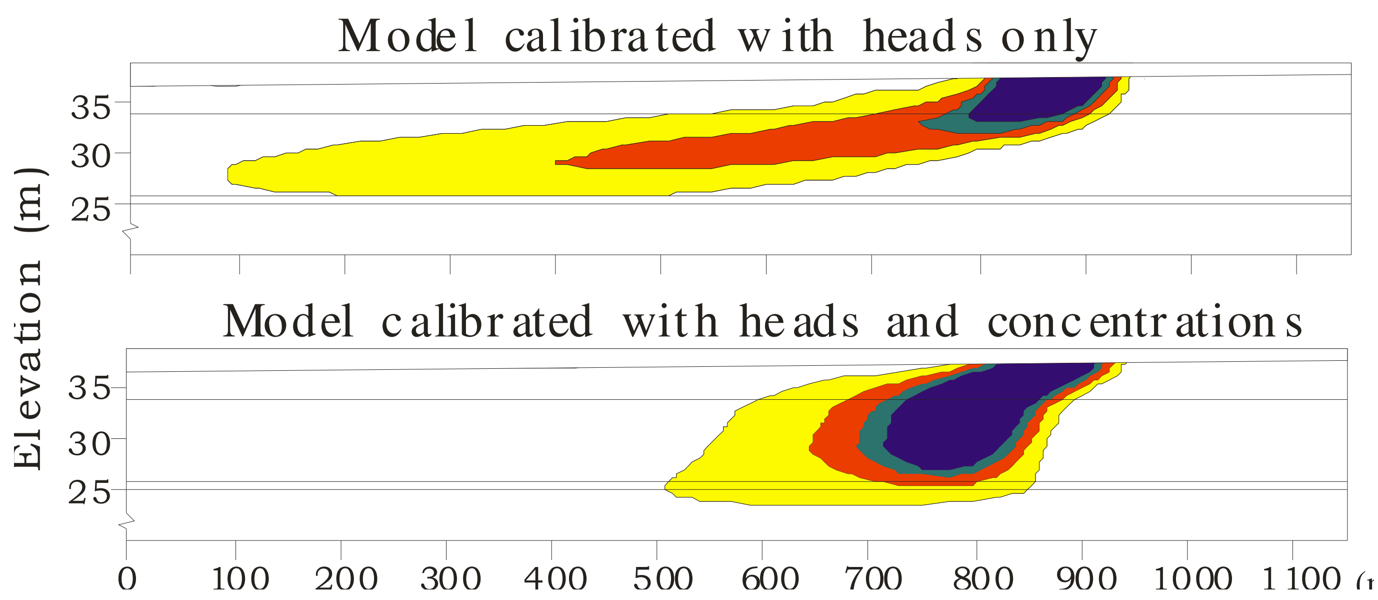
- We observed two problems in model/risk analysis: (1) the tower of Babel and (2) “colossal” computational burdens.
- We proposed a strategy that uses **a three-pronged line of attack** to
  - Organize the diverse methods from a modeling perspective
  - Analyze similarity and dissimilarity between the methods
  - Evaluate computationally fugal and efficient methods
- Using the methods of model and scenario averaging, we developed a comprehensive and hierarchical Bayesian framework to consider the uncertainties in model scenarios, structures, and parameters.
- Using the sparse grid collocation methods greatly improves computational efficiency for uncertainty quantification and risk analysis.

Test of Transport Prediction

Heads only

Head and concentrations

About a factor of 2 difference in transport distance



1990's: The hydrogeologic work will yield much better predictive skill than this