Ming Ye

Education

- Ph.D. in Hydrology, 2002, University of Arizona, Tucson, AZ
- B.S. in Geology 1997, Nanjing University, China

Working Experience

- 2007-present, Assistant/Associate/Full Professor, Department of Earth, Ocean, and Atmospheric Science and Department of Scientific Computing, Florida State University (FSU), Tallahassee, FL
- 2004-2006, Assistant Research Professor, Desert Research Institute (DRI), Las Vegas, NV
- 2002-2004, Post-doctoral Research Associate, Pacific Northwest National Laboratory (PNNL), Portland, OR

Groundwater Sensitivity Analysis under Model and Scenario Uncertainty: Not Look for Keys under the Lamppost

Ming Ye (mye@fsu.edu)
Department of Earth, Ocean, and Atmospheric Science
Florida State University

University of Alabama
11/10/2017

NOT to Look for the Keys under the Lamppost

- A police officer sees a drunken man intently searching the ground near a lamppost and asks him the goal of his quest.
- The inebriate replies that he is looking for his car keys.
- The officer helps for a few minutes without success then he asks whether the man is certain that he dropped the keys near the lamppost.

"No," is the reply, "I do not know where I lost the keys." "Why look here?" asks the surprised and irritated officer. "The light is much better here," the intoxicated man responds with aplomb.

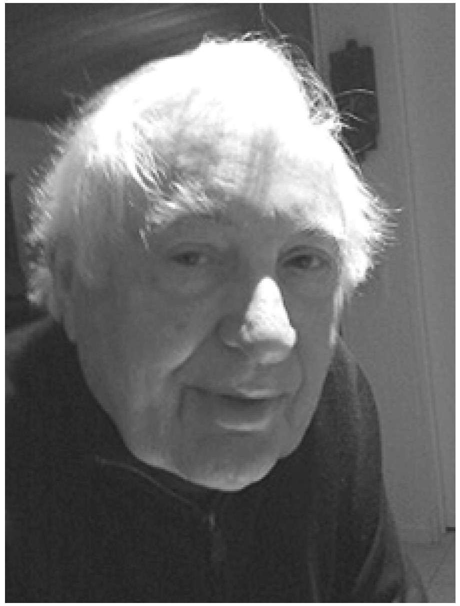
Do we do the same in groundwater modeling?

Lamppost in Groundwater Modeling

Bredehoeft (2010): Models and Model Analysis

How do we, as hydrogeologists, use the models today?

- "We collect all the available geologic data and create cross sections and isopach maps of various hydrostratigraphic layers.
- We decide that the geologic information is incomplete, and drill additional test wells. By this time, the project is 75% to 80% complete, and now we build 'the groundwater model'.
- We run the model, and we find it is sensitive to data that we failed to collect (e.g., the vertical hydraulic conductivity of a confining layer). Perhaps even our concept of the system is invalid.
- But there is neither time nor money left in the project budget – the model informs us that we need another project. As my colleague Ivan Barnes used to say: *Further work is indicated."*



Member of National Academy of Engineering, Elected in 1994

“There are a number of things **wrong** with this approach.

- Our data collection was based on an implicit (or explicit) conceptual model of the system.
- The last thing we did was test this conceptual model with the numerical model, and we found that we failed to collect important field data.
- In the end, we built a complex model that was state of the art numerically, calibrating it using optimization techniques.
- The calibration suggested changes, model revisions, different data, but we were out of time and budget. *Sound familiar?*”

We are looking for the keys under the lamppost!

“I may be like Don Quixote jousting with windmills; but to build bigger groundwater models just because it is feasible, or in vogue, seems to me meaningless – certainly it does not lead to increased understanding.”

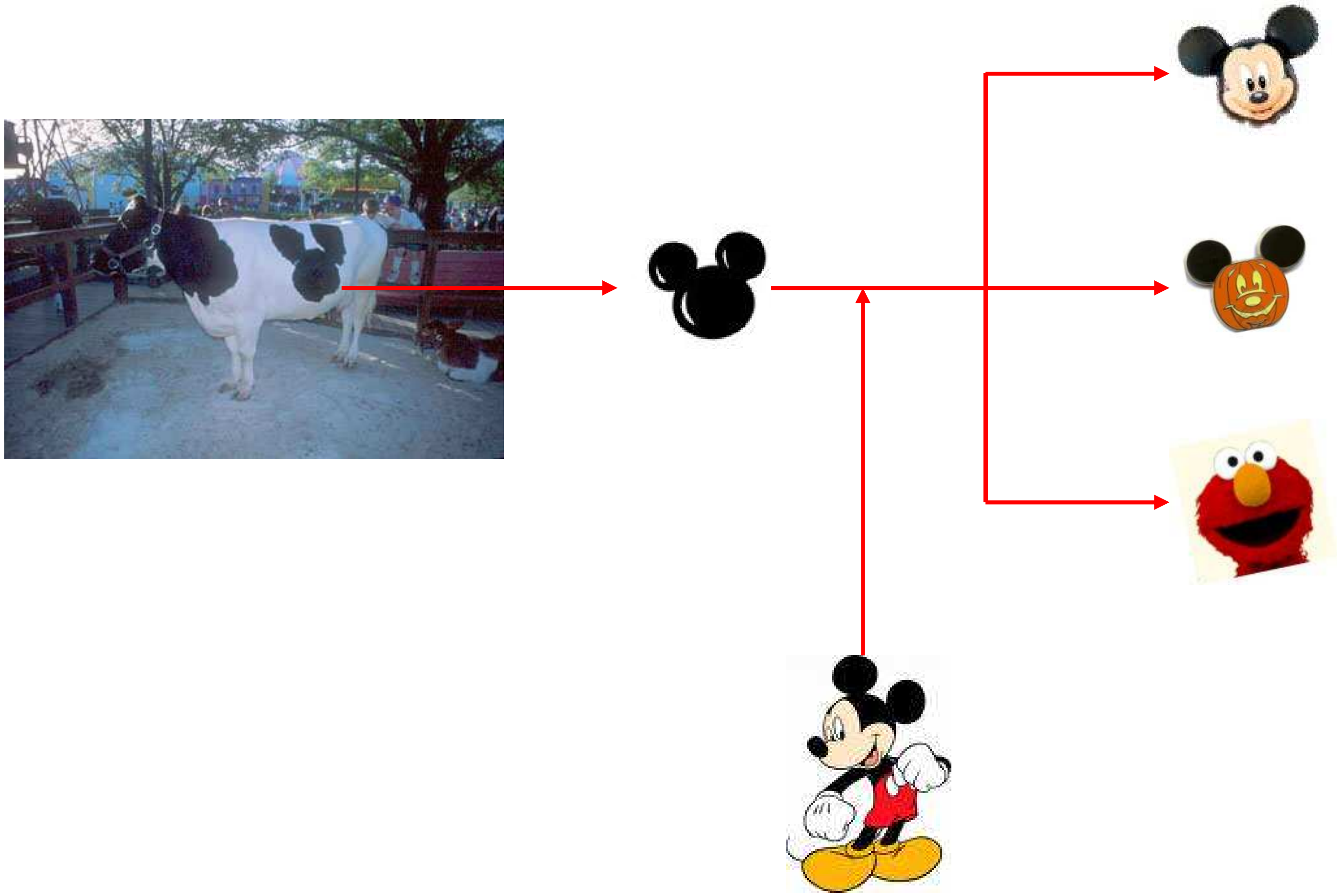
The Conceptual Model Problem: Surprise

- "Surprise is defined as the collection of new information that renders one's original conceptual model invalid."
- "Limited empirical data indicate that surprises occur in 20–30% of model analyses."

Prototype	Modeler	Model type	Surprise	Comments
Arkansas Valley	Konikow	Transport	No	Need longer period of calibration
Blue River	Emery	Flow	No	Need better parameters
Coachella Valley	Swain	Flow	Yes	Recharge events unanticipated
Houston	Jorgensen	Flow/ subsidence	?	Iterative modeling
HYDROCOIN	Konikow	Transport	Yes	Boundary condition modeled poorly
INEL	Robertson	Transport	No	Need better parameters
Los Alamos	Los Alamos	Unsaturated flow	?	Flow through unsaturated zone not understood
Los Angeles area	Bredehoeft	Flow	Yes	Flow vectors 90° off in model
Ontario U tailings	Flavelle	Transport	Yes	Need more than one distribution coefficient
Phoenix	Anderson	Flow	No	Need better parameters
Summitville	Bredehoeft	Flow	Yes	Seeps on mountain unaccounted for
Santa Barbara	Bredehoeft	Transport	?	Fault zone flow unaccounted for
WIPP	WIPP project	Flow	Yes	Salt had 1–3% interstitial brine
Yucca Mountain	YM project	Unsaturated flow	Yes	Chlorine 36 indicates fast flow path
Other models	15	Flow/ transport	No	Bredehoeft's consulting—no conceptual problems
Total	29		7 yes (3 ?)	

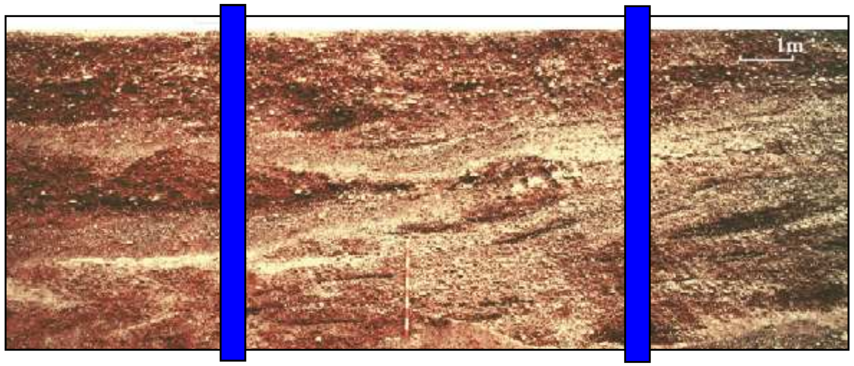
Bredehoeft (2005)

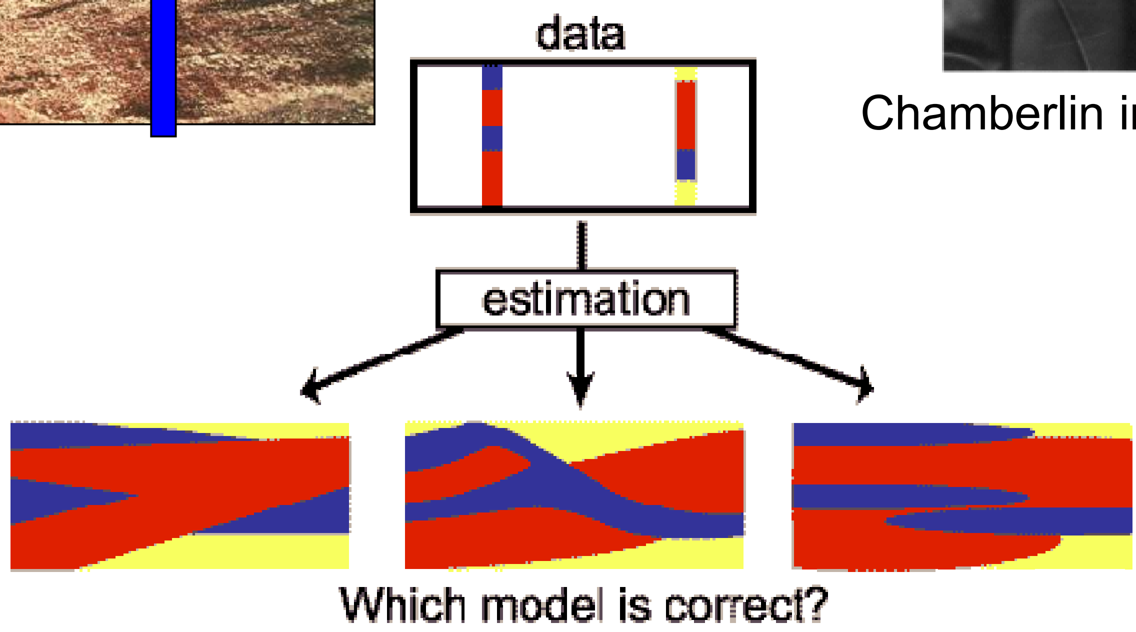
Surprise if Using a Wrong Conceptual Model



Chamberlin's Solution (1890)

- The method of the ruling theory
- The method of the working hypothesis
- The method of multiple working hypothesis





Chamberlin in the 1870s

Look for the Keys under **Multiple Lampposts**

- Model Selection

Select one out of multiple models (ZERO SUM)

- Model Averaging

Use multiple models for making predictions (AGREEING ON DISAGREE)

"All Models Are Wrong: How Do We Know Which Are Useful?“

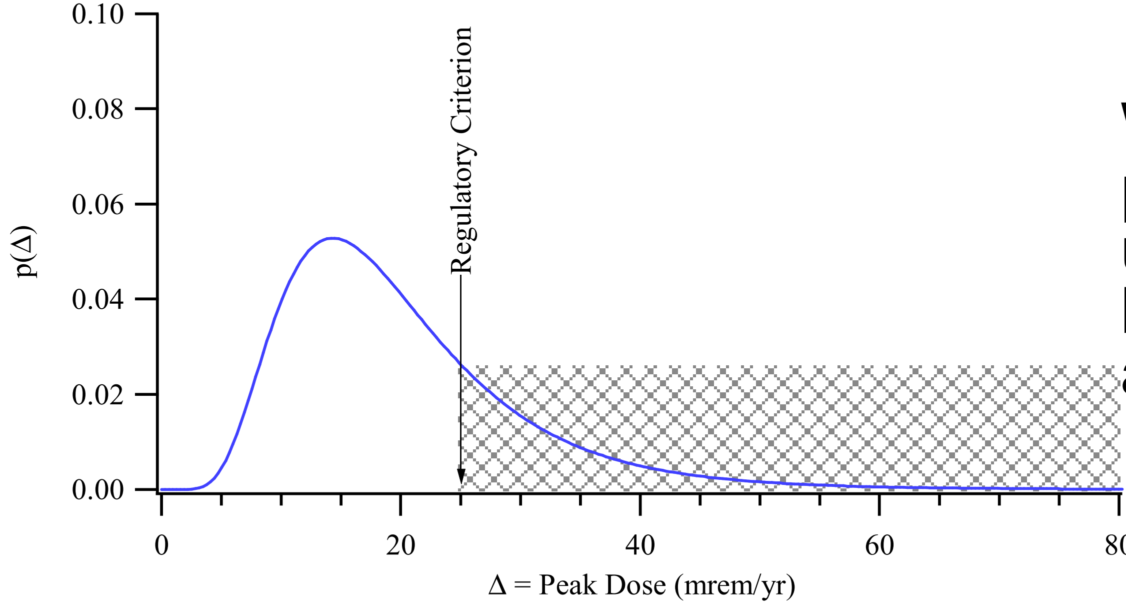
Eileen Poeter 2006 Darcy Lecturer

“Seeing Things Differently: Rethinking the Relationship Between Data, Models, and Decision-Making”

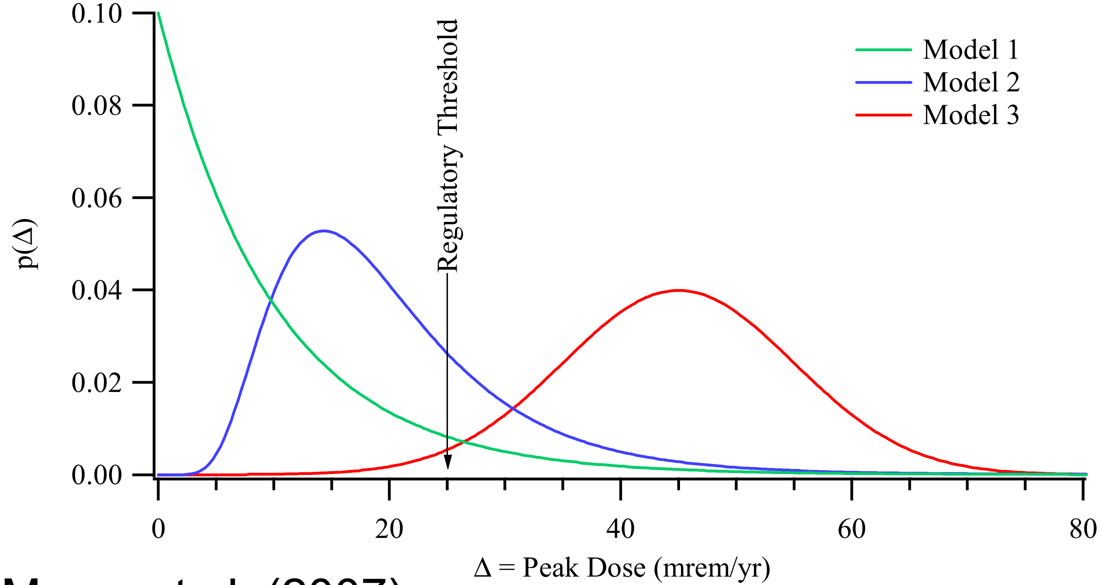
Ty Ferre 2016 Darcy Lecturer

I have been there!

Model Averaging



When we make model predictions, we used to use only one model based on our judgements and assumptions.

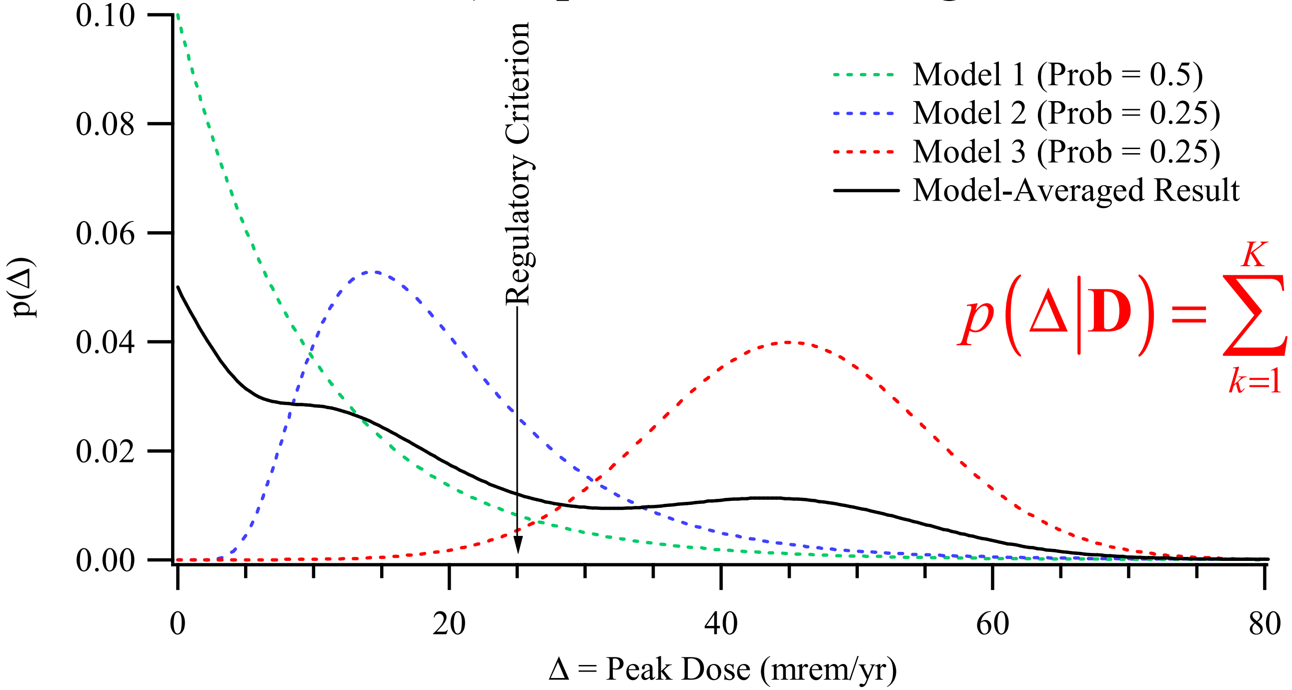


However, we are not always certain about the “right” judgments and assumptions to make.

Meyer et al. (2007)

Model Averaging

(*Draper*, 1995; *Hoeting et al.*, 1999; *Neuman*, 2002; *Ye et al.*, 2004)



$$p(\Delta|\mathbf{D}) = \sum_{k=1}^{K} \boxed{p(\Delta|M_k,\mathbf{D})}\,\boxed{p(M_k|\mathbf{D})}$$

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 Bayesian model averaging strategy of Neuman (2002) *encourages exploration of varied conceptual frameworks and assumptions at all stages of hydrogeologic model development*.

Member of National Academy of Engineering, Elected in 1998

Bayesian Model Averaging: Posterior Model Probability

$$\boxed{\text{Posterior}} \xleftarrow{\text{DATA: }\mathbf{D}} \boxed{\text{Prior}}$$

Model Probability:

$$\boxed{p\left(M_k \mid \mathbf{D}\right)} = \frac{p\left(\mathbf{D} \mid M_k\right) \boxed{p\left(M_k\right)}}{\sum_{l=1}^{K} p\left(\mathbf{D} \mid M_l\right) p\left(M_l\right)}$$

Model Evidence

$$p\left(\mathbf{D} \mid M_k\right) = \int p\left(\mathbf{D} \mid \boldsymbol{\theta}_k, M_k\right) p\left(\boldsymbol{\theta}_k \mid M_k\right) d\boldsymbol{\theta}_k$$

- $\boldsymbol{\theta}_k$: multiple-dimensional parameters of model M_k
- Need to calculate the multivariate integration. Computationally expensive!!!

Computational Challenge

$$p(\mathbf{D}|M)=\int p(\mathbf{D}|\boldsymbol{\theta},M)\,p(\boldsymbol{\theta}|M)\,d\boxed{\boldsymbol{\theta}}$$

Arithmetic mean $\hat{p}_{AM}(\mathbf{D}|M)\approx\frac{1}{n}\sum_{i=1}^{n}p(\mathbf{D}\,|\,\boldsymbol{\theta}_i,M)$

Harmonic mean $\hat{p}_{HM}(\mathbf{D}|M)\approx n\Big/\sum_{i=1}^{n}\frac{1}{p(\mathbf{D}\,|\,\boldsymbol{\theta}_i,M)}$

Straightforward MC implementation, but do not work well.

Nested sampling $p(\mathbf{D}\,|\,M)=\int_0^1 L(X\,|\,\mathbf{D},M)\,d\boxed{X}$

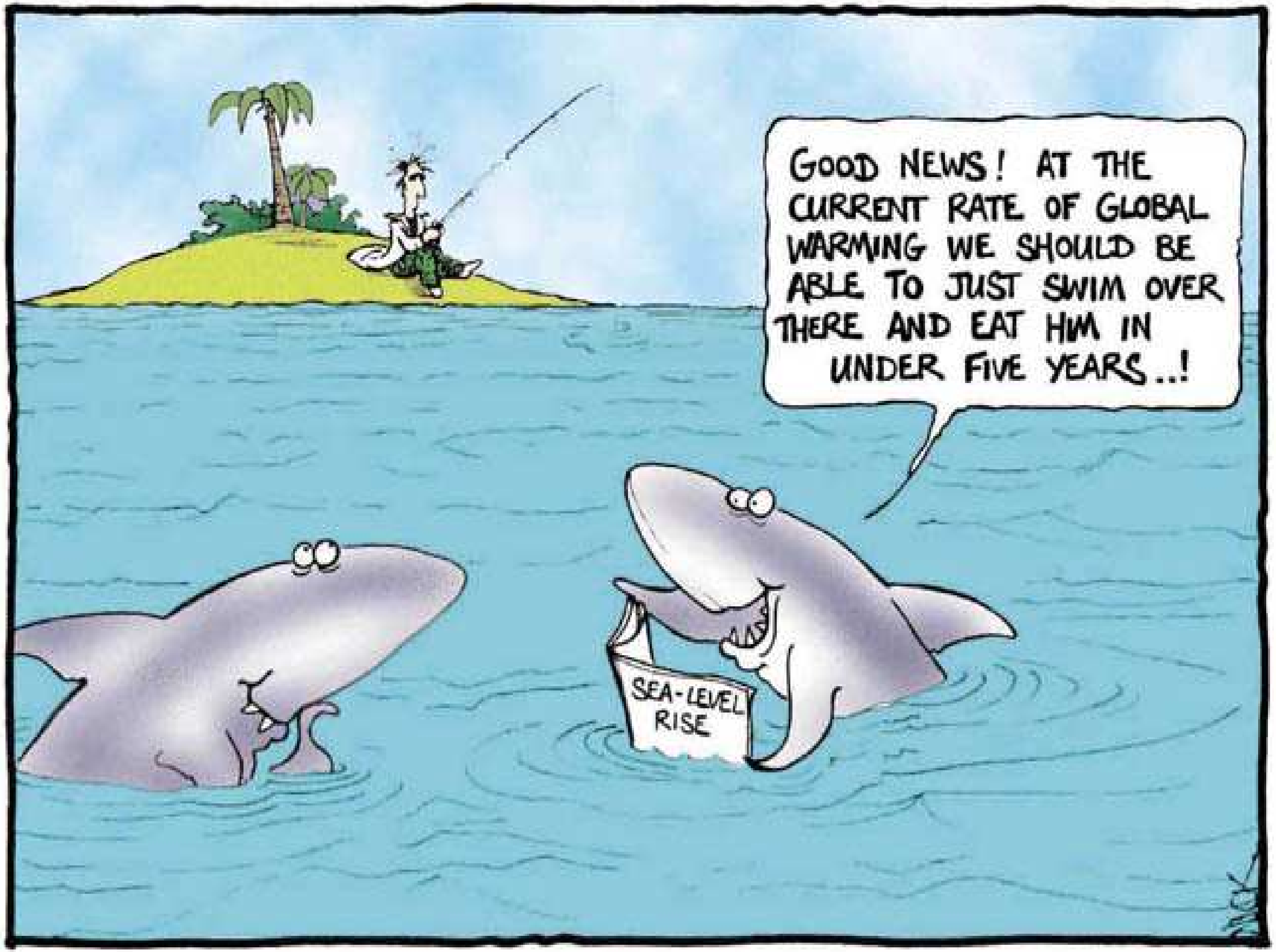
Thermodynamic Integration $p(\mathbf{D}\,|\,M)=\exp\int_0^1 E_{\boldsymbol{\theta}_\beta}[\ln p(\mathbf{D}|\boldsymbol{\theta},M)]\,d\boxed{\beta}$

Steppingstone Sampling $p(\mathbf{D}|M)=\frac{Z_1}{Z_0}=\left(\frac{Z_1}{Z_{\beta_{K-1}}}\right)\cdots\left(\frac{Z_{\beta_k}}{Z_{\beta_{k-1}}}\right)\cdots\left(\frac{Z_{\beta_1}}{Z_0}\right)$

Liu et al. (2016, WRR), Zeng et al. (2017, WRR), Elshall et al. (2017, WRR)

Groundwater Sensitivity Analysis under Model and Scenario Uncertainty: Not Look for Keys under the Lamppost

Sea-Level Rise Scenarios

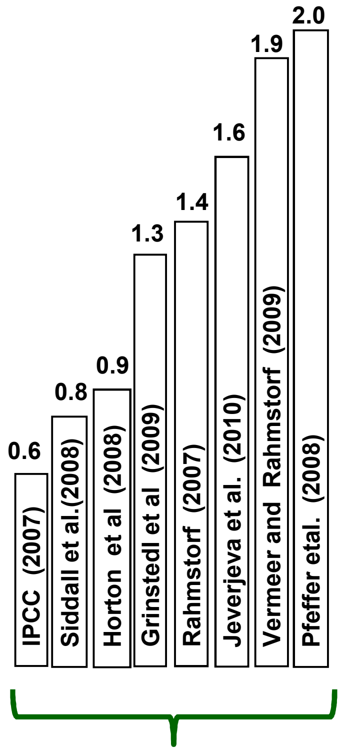


Sea-Level Rise Scenarios

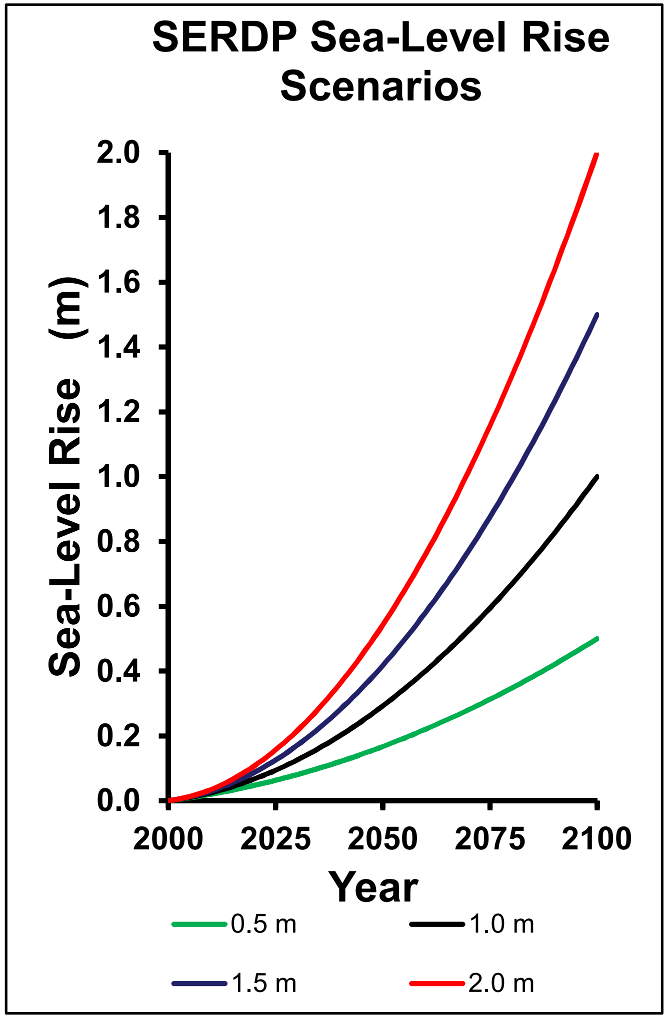
Projecting Future Sea-Level Rise

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

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



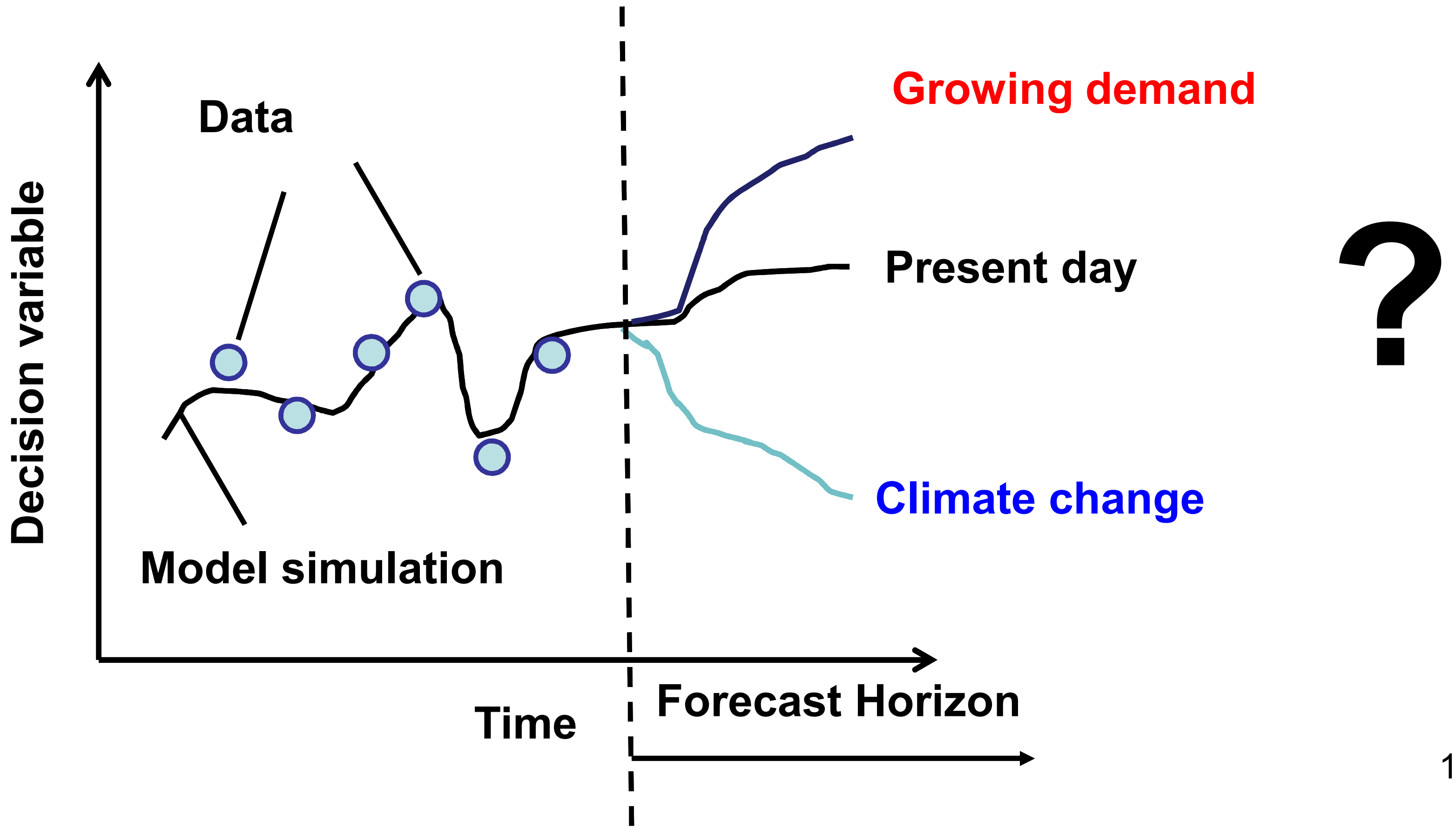
Recent Published Model Estimates of Year 2100 Sea-Level Rise (m)



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

Scenario Uncertainty

IPCC (2000, p.62): “scenarios are images of the future, or alternative futures. They are neither predictions nor forecasts. Rather, each scenario is one alternative image of how the future might unfold. A set of scenarios assists in the understanding of possible future developments of complex systems.”

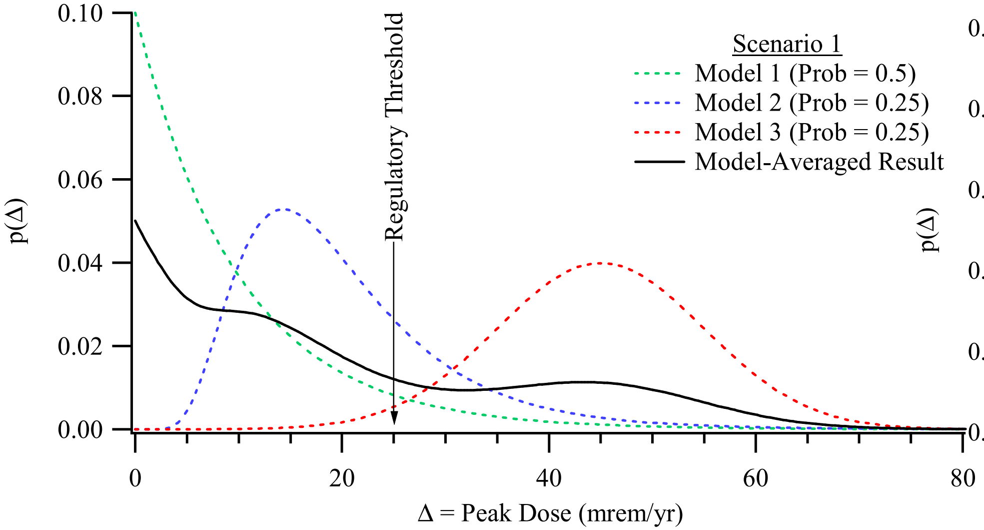


Scenario Averaging

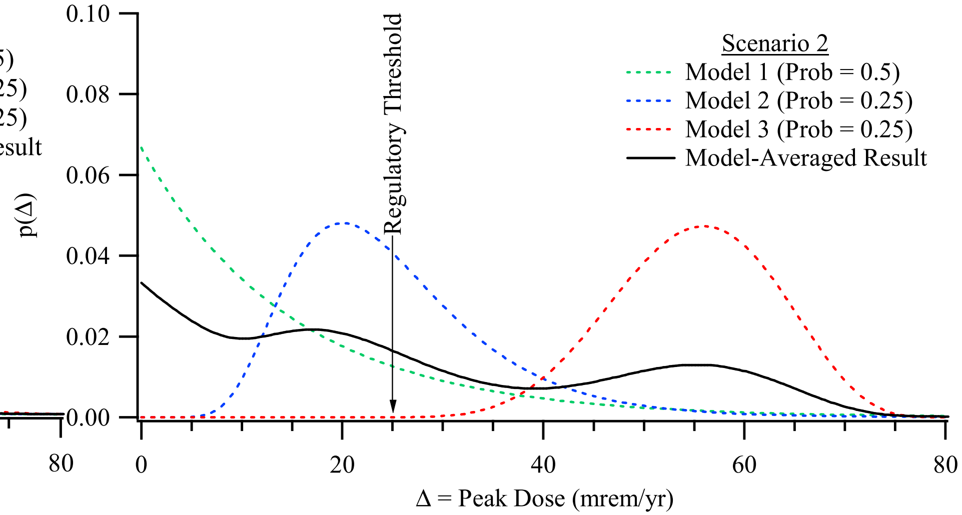
$$E\left(\Delta \middle| \mathbf{D}\right) = \sum_{i=1}^{I} E\left(\Delta \middle| \mathbf{D}, S_i\right) \boxed{p\left(S_i\right)}$$

Scenario probability

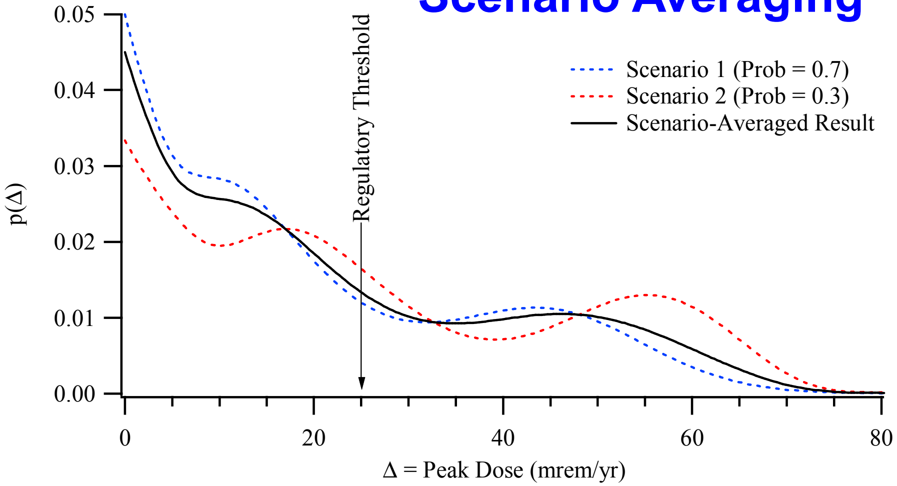
Scenario 1



Scenario 2



Scenario Averaging



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

Meyer et al. (2007)

My Experience

- 2002-2004 (PNNL): Hanford Site, WA

Nuclear Regulatory Commission

- 2004-2006 (DRI): Death Valley Regional Flow System, NV

Department of Energy

- 2009-2012 (FSU): Death Valley Regional Flow System, NV

National Science Foundation

- 2009-2012 (FSU): Naturita Site, CO

Department of Energy

- 2009-2012 (FSU): Eglin Air Force Base, FL

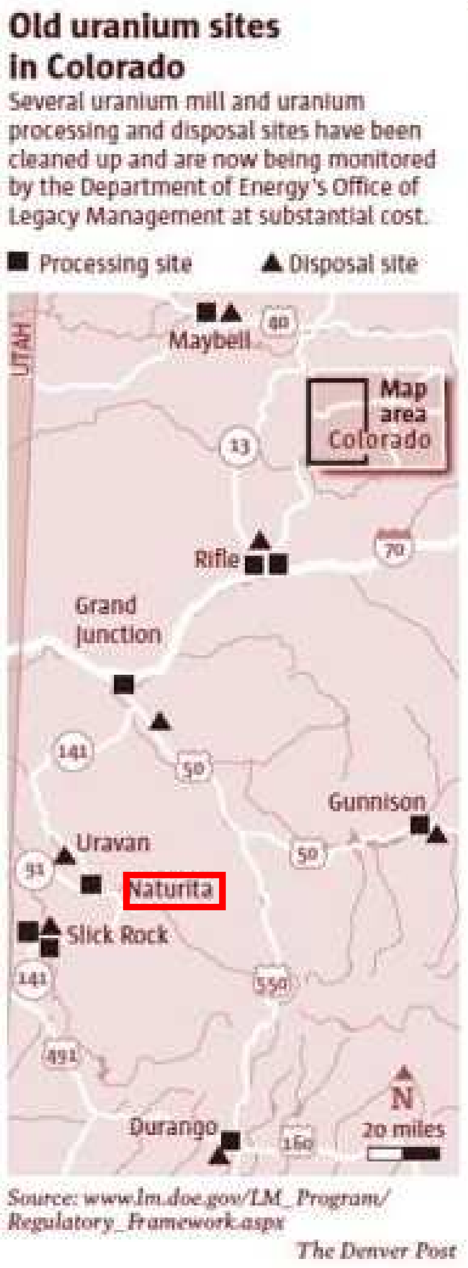
Department of Defense

- 2012 – 2018 (FSU): Naturita Site, CO, and Oak Ridge Site, TN

Department of Energy

Early Career Award: $750,000

Uranium Mill Tailing at Naturita, CO



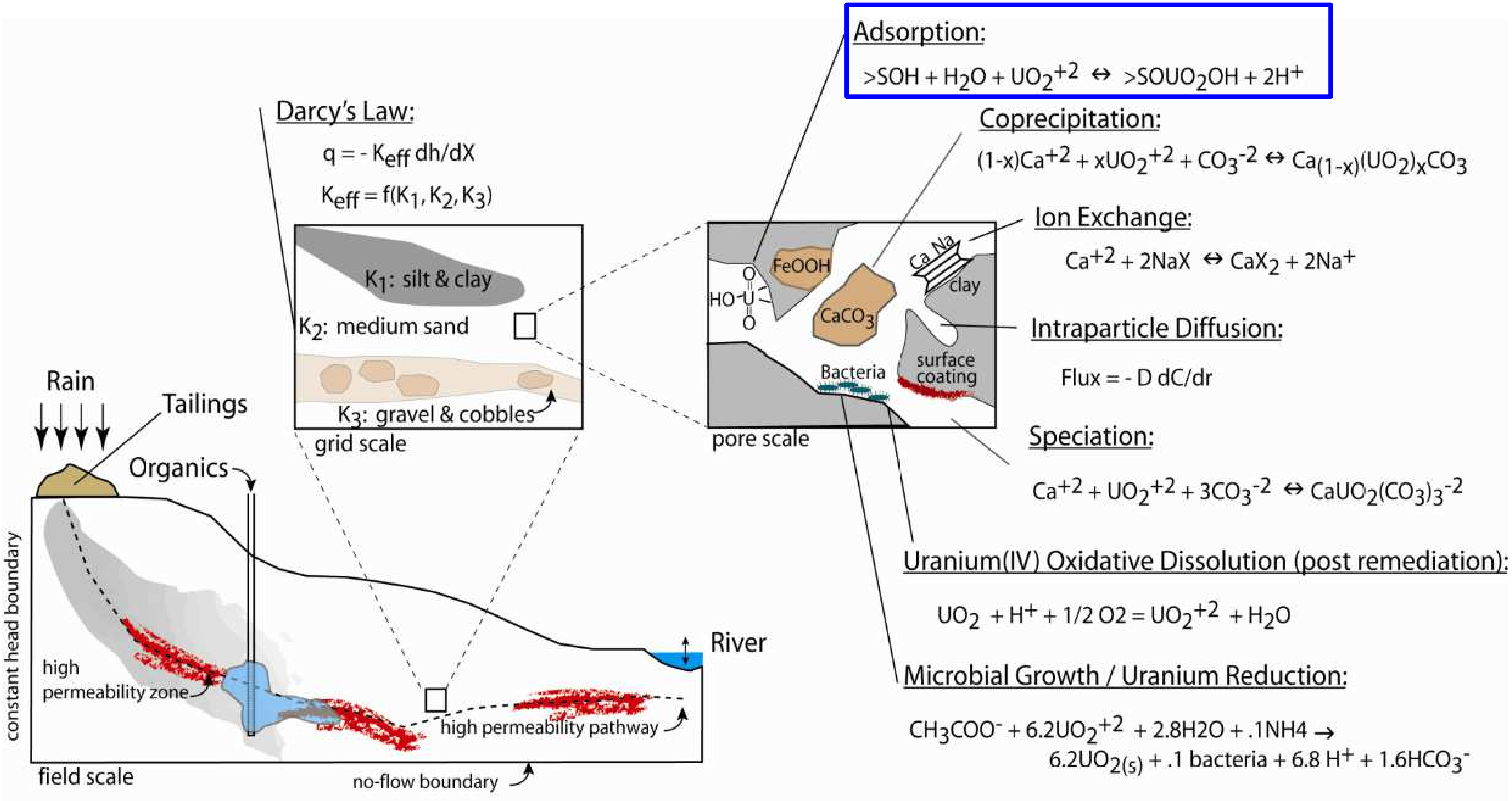
Toxic legacy of uranium haunts proposed Colorado mill, By Nancy Lofholm, *The Denver Post, 9/5/2010*

- **NATURITA MILL AND DISPOSAL SITE**
  $86.3 million Cost of cleanup The Vanadium Corp. of America began operating the mill in 1939. The mill processed 704,000 tons of uranium ore for the Manhattan Project from 1942 to 1958. In the late 1970s, a private corporation bought the tailings pile and moved it to another site called Hecla/Durita to extract additional uranium and vanadium.
- **Left behind:** At and around the original mill, 138 acres were contaminated. Groundwater beneath the site was contaminated.
- **The fix:** From 1993 to 1997, DOE removed 800,000 yards of contaminated material and put it in a disposal site near Uravan. Contamination was left in place on 22 acres. More than one acre was left because the radiation levels were so high that workers would have been at risk.

http://www.denverpost.com/news/ci_15996355

Conceptual Model Challenges in Groundwater Reactive Transport Modeling

Uncertainty in model parameters, structures, and Scenarios



Surface Complexation Models

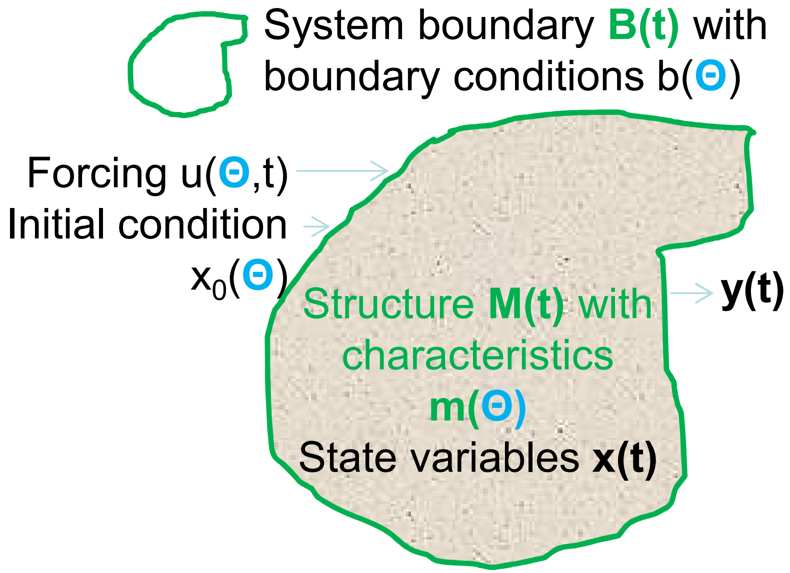
Kohler et al. (1996, WRR)

- Seven models (C1-C7) of surface complexation with different level of complexity (numbers of functional groups).
- Question: Which model to use?

Gary: I do not know. You tell me, Ming!

Model	Reactions
C1	$\boxed{S_1OH} + UO_2^{+2} + H_2O = S_1OUO_2OH + 2H^+$
C2	$S_1OH + UO_2^{+2} + H_2O = S_1OUO_2OH + 2H^+$ $S_2OH + UO_2^{+2} + H_2O = S_2OUO_2OH + 2H^+$
C3	$S_1OH + UO_2^{+2} + H_2O = S_1OUO_2OH + 2H^+$ $S_2OH + UO_2^{+2} = S_2OUO_2^{+2} + H^+$
C4	$\boxed{S_1OH} + UO_2^{+2} + H_2O = S_1OUO_2OH + 2H^+$ $\boxed{S_2OH} + UO_2^{+2} + H_2O = S_2OUO_2OH + 2H^+$ $S_2OH + UO_2^{+2} = S_2OUO_2^{+2} + H^+$
C5	$\boxed{S_1OH} + UO_2^{+2} + H_2O = S_1OUO_2OH + 2H^+$ $\boxed{S_2OH} + UO_2^{+2} + H_2O = S_2OUO_2OH + 2H^+$ $S_2OH + UO_2^{+2} = S_2OUO_2^{+2} + H^+$ $\boxed{S_3OH} + UO_2^{+2} + H_2O = S_3OUO_2OH + 2H^+$
C6	$S_1OH + UO_2^{+2} + H_2O = S_1OUO_2OH + 2H^+$ $S_2OH + UO_2^{+2} + H_2O = S_2OUO_2OH + 2H^+$ $S_2OH + UO_2^{+2} = S_2OUO_2^{+2} + H^+$ $S_3OH + UO_2^{+2} = S_3OUO_2^{+2} + H^+$
C7	$S_1OH + UO_2^{+2} + H_2O = S_1OUO_2OH + 2H^+$ $S_2OH + UO_2^{+2} + H_2O = S_2OUO_2OH + 2H^+$ $S_3OH + UO_2^{+2} + H_2O = S_3OUO_2OH + 2H^+$

Bayesian UQ Framework



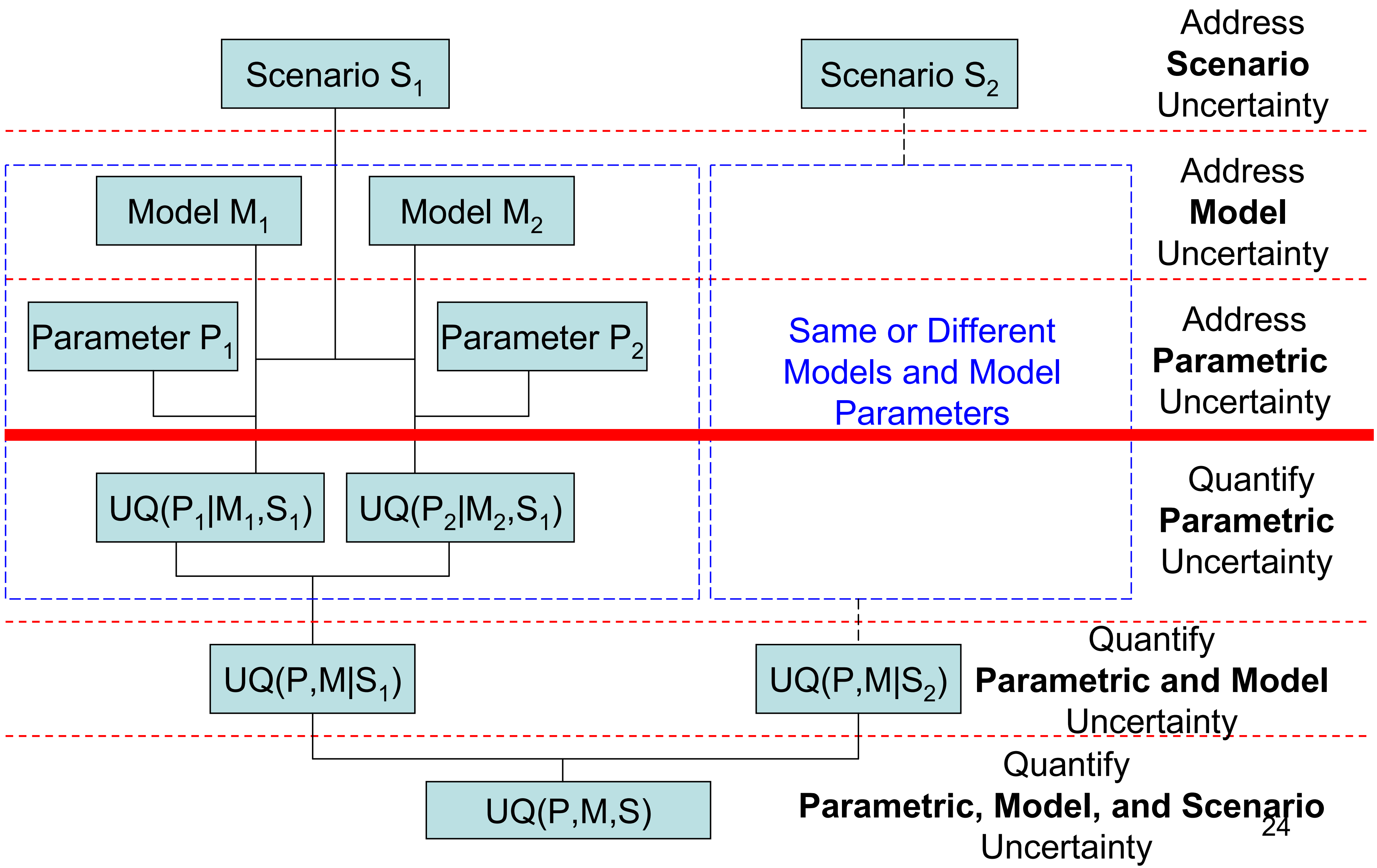
A model is composed of seven different components:

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

The sources of the predictive uncertainty, from a system point of view, can be categorized into

(1) Scenario uncertainty in system input (*u*),
(2) Model uncertainty in model structure (*M*),
(3) Parameter uncertainty in model parameters (θ), and
(4) Data uncertainty in observations.

A Comprehensive and Hierarchical Framework



Bayesian UQ Framework: Scenario, Model, and Parametric Uncertainty

- Scenario uncertainty

$$Var(\Delta) = E_s \boxed{Var(\Delta \mid S)} + Var_s E(\Delta \mid S)$$

Within-scenario Between-scenario

- Model uncertainty of a given scenario

$$Var(\Delta \mid S) = E_{M|S} \boxed{Var(\Delta \mid M, S)} + Var_{M|S} E(\Delta \mid M, S)$$

Within-model Between-model

- Parametric uncertainty of a given scenario and a given model

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

Data Parametric

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

Groundwater Sensitivity Analysis under Model and Scenario Uncertainty: Not Look for Keys under the Lamppost

Sensitivity Analysis

Razavi and Gupta (2015, WRR)

- Sensitivity analysis describes different “intuitive” understandings of the sensitivity of one or more model responses to different factors such as model parameters or forcings.
- Objectives of sensitivity analysis:
  - Assessment of Similarity
  - Factor Importance and Function
  - Regions of Sensitivity
  - Factor Interdependence
  - Factor and Model Reduction
  - Uncertainty Apportionment: Quantitative attribution of the uncertainty in model response to different factors (sources of uncertainty), with the goal of identifying where best to focus efforts at improved factor characterization so as to achieve reductions in total uncertainty

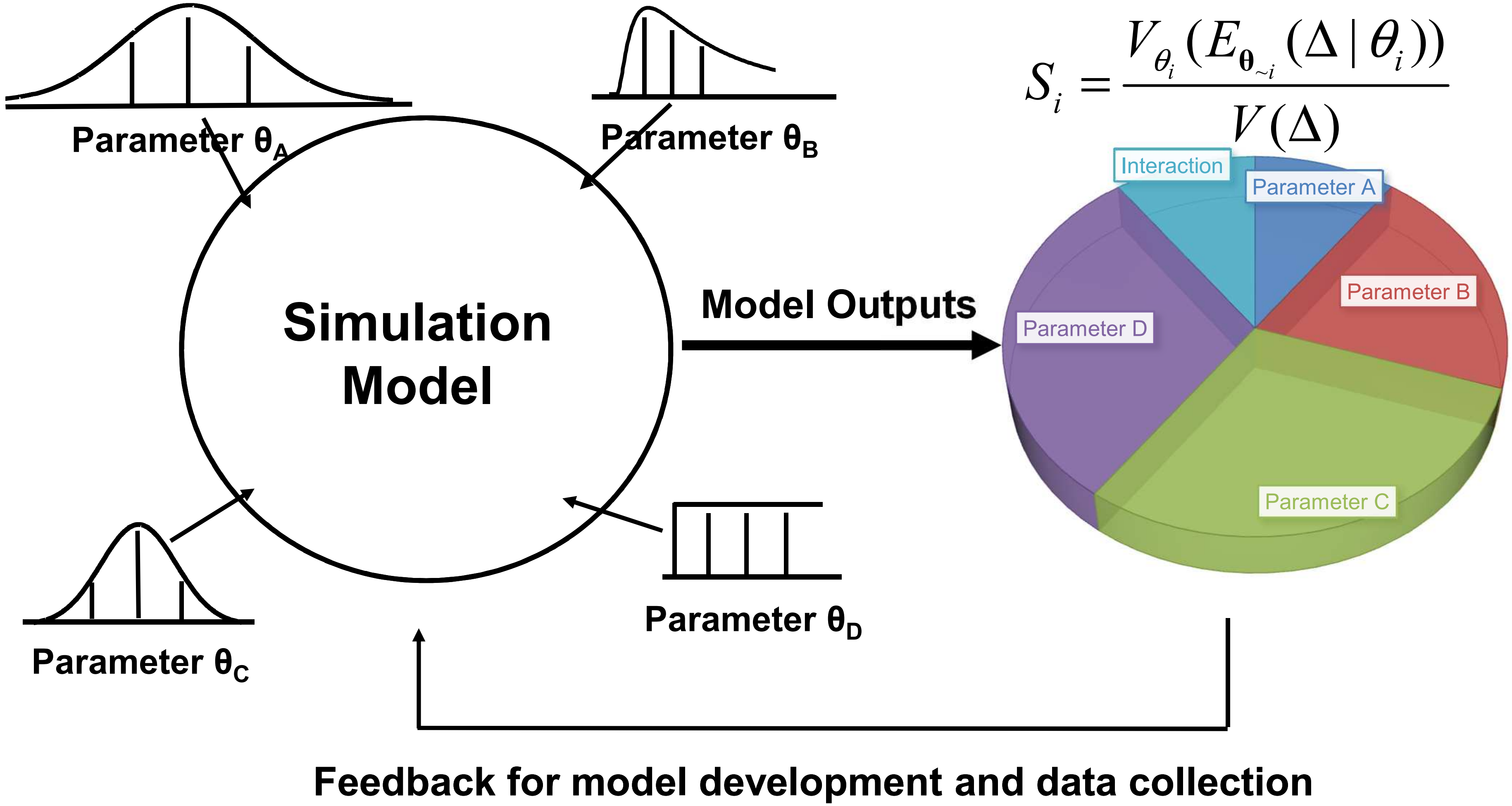
Global Sensitivity Analysis for a Single Model

Question: If we are not certain about what model parameter values to use, can we identify the important parameters?

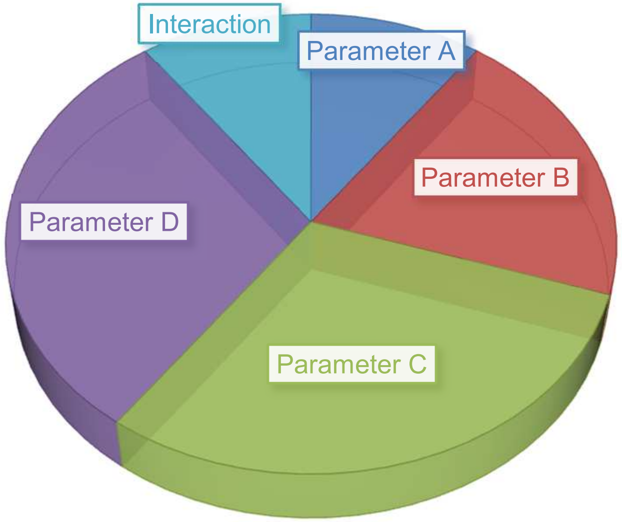
$$\Delta = f(\theta_A, \theta_B, \theta_C, \theta_D)$$

Sobol' first-order effect sensitivity index

$$S_i = \frac{V_{\theta_i}(E_{\boldsymbol{\theta}_{\sim i}}(\Delta \mid \theta_i))}{V(\Delta)}$$



Challenges of Global Sensitivity Analysis under Model Uncertainty



$$\Delta = f(\boldsymbol{\theta}_A, \boldsymbol{\theta}_B, \boldsymbol{\theta}_C, \boldsymbol{\theta}_D)$$

$$S_i = \frac{V_{\theta_i}(E_{\boldsymbol{\theta}_{\sim i}}(\Delta \mid \theta_i))}{V(\Delta)}$$

S_A	S_B	S_C	S_D
10%	20%	20%	**50%**

What if there is another plausible model/scenario?

Are the parameters important to one model/scenario also important to another model/scenario?

Challenges of Global Sensitivity Analysis under Model Uncertainty

$$\Delta = M_1(\boldsymbol{\theta}_A, \boldsymbol{\theta}_B, \boldsymbol{\theta}_C, \boldsymbol{\theta}_D)$$

$$\Delta = M_2(\boldsymbol{\theta}_A, \boldsymbol{\theta}_B, \boldsymbol{\theta}_C)$$

$$S_i = \frac{V_{\theta_i}(E_{\boldsymbol{\theta}_{\sim i}}(\Delta \mid \theta_i))}{V_1(\Delta)}$$

$$S_i = \frac{V_{\theta_i}(E_{\boldsymbol{\theta}_{\sim i}}(\Delta \mid \theta_i))}{V_2(\Delta)}$$

	S_A	S_B	S_C	S_D
Model M_1	10%	20%	20%	**50%**
Model M_2	**50%**	20%	10%	N/A

- Different models may have different parameters
- Sensitivity index is not comparable cross models.

S_B = 20% for models M_1 and M_2.

M_1: 20% of the variance of 100, $V_1(\Delta)$.

M_2: 20% of the variance of 10,000, $V_1(\Delta)$.

Dai and Ye (2015, Journal of Hydrology)

Parametric Uncertainty Under Model and Scenario Uncertainty

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

Basic Idea:

Important parameters are identified

- Not for a single model/scenario
- but for multiple models/scenarios on the average sense.

This new sensitivity index copes with model/scenario uncertainty, and can avoid wrong identification of important parameters given the uncertainty.

New Sensitivity Index

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

For a single model and a single scenario

(no model and scenario uncertainty)

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

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

New Sensitivity Index

For multiple models but a single scenario

(model uncertainty and no scenario uncertainty)

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

$$S_{Ti} = \frac{\boxed{E_{M|S}} E_{\theta_{\sim i}|M,S} Var_{\theta_i|M,S} (E(\Delta \,|\, \theta, M, S) \,|\, \theta_{\sim i})}{\boxed{E_{M|S}} Var_{\theta|M,S} E(\Delta \,|\, \theta, M, S)}$$

$$= \frac{\boxed{\sum_M P(M \,|\, S)} E_{\theta_{\sim i}|M,S} Var_{\theta_i|M,S} (E(\Delta \,|\, \theta, M, S) \,|\, \theta_{\sim i})}{\boxed{\sum_M P(M \,|\, S)} Var_{\theta|M,S} E(\Delta \,|\, \theta, M, S)}$$

New Sensitivity Index

For multiple models and multiple scenarios

(model uncertainty and scenario uncertainty)

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

$$S_{Ti} = \frac{\boxed{E_s E_{M|S}} E_{\theta_{\sim i}|M,S} Var_{\theta_i|M,S}(E(\Delta \,|\, \theta, M, S) \,|\, \theta_{\sim i})}{\boxed{E_s E_{M|S}} Var_{\theta|M,S} E(\Delta \,|\, \theta, M, S)}$$

$$= \frac{\boxed{\sum_S \sum_M P(S) P(M \,|\, S)} E_{\theta_{\sim i}|M,S} Var_{\theta_i|M,S}(E(\Delta \,|\, \theta, M, S) \,|\, \theta_{\sim i})}{\boxed{\sum_S \sum_M P(S) P(M \,|\, S)} Var_{\theta|M,S} E(\Delta \,|\, \theta, M, S)}$$

Global Sensitivity Analysis for Identifying Important Parameters of Nitrogen Nitrification and Denitrification Under Model and Scenario Uncertainty

Ming Ye (mye@fsu.edu)

Department of Scientific Computing, Florida State University

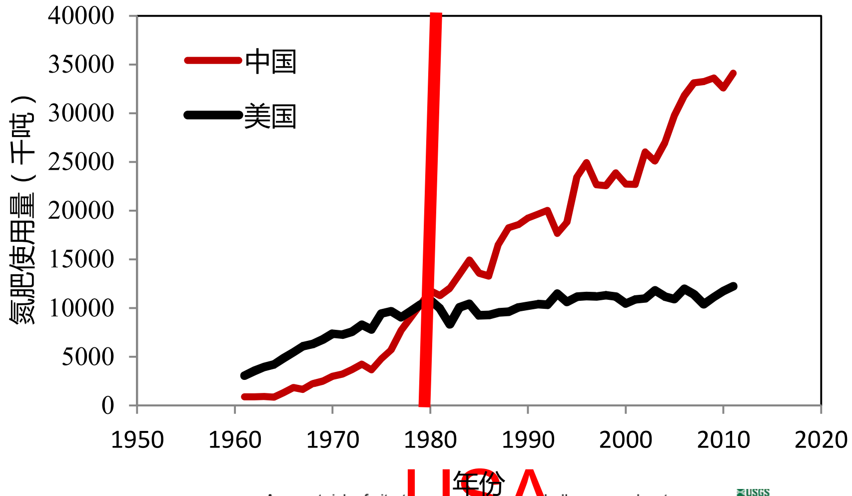
Zhuowei Chen, Liangsheng Shi,

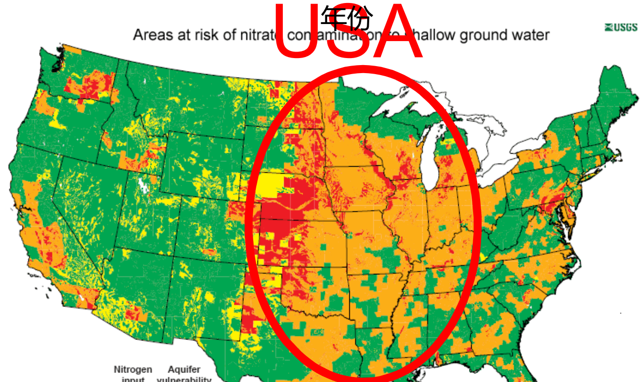
Yan Zhu, and Jinzhong Yang

State Key Laboratory of Water Resources and Hydropower Engineering Sciences, Wuhan University

Chen et al. (2017, WRR, Under Revision, RRR)

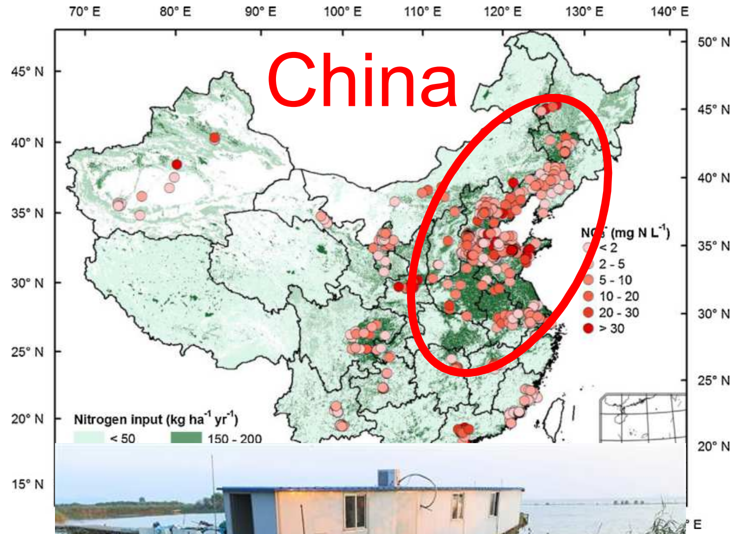
Nitrogen Contamination due to Fertilizer Use

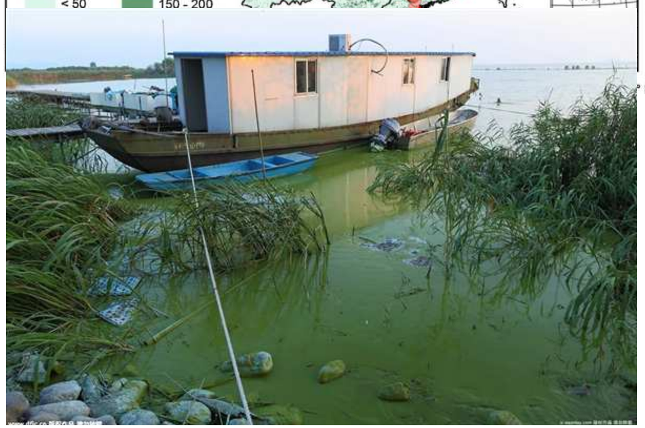




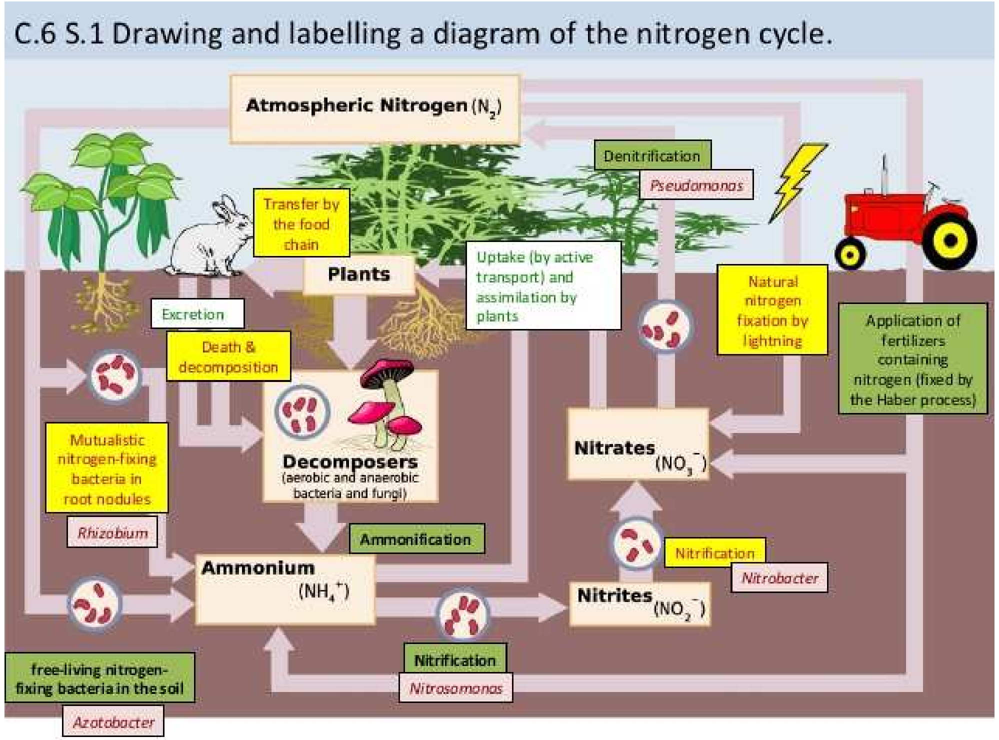
Nitrogen from fertilizer use

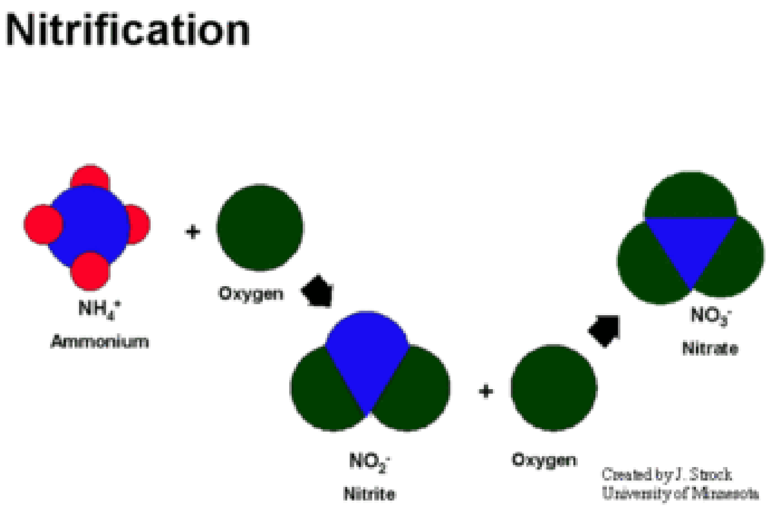
National Science Foundation, China
Overseas Collaborative Research
2,000,000 RMB ($300,000)

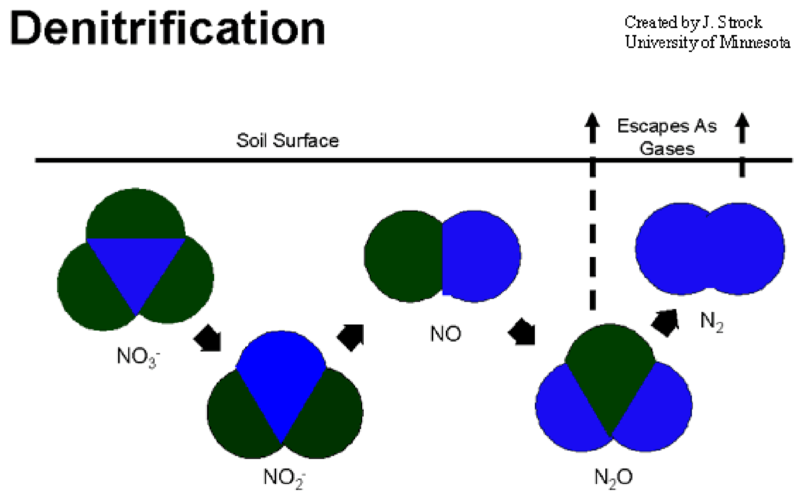




Nitrogen Cycle

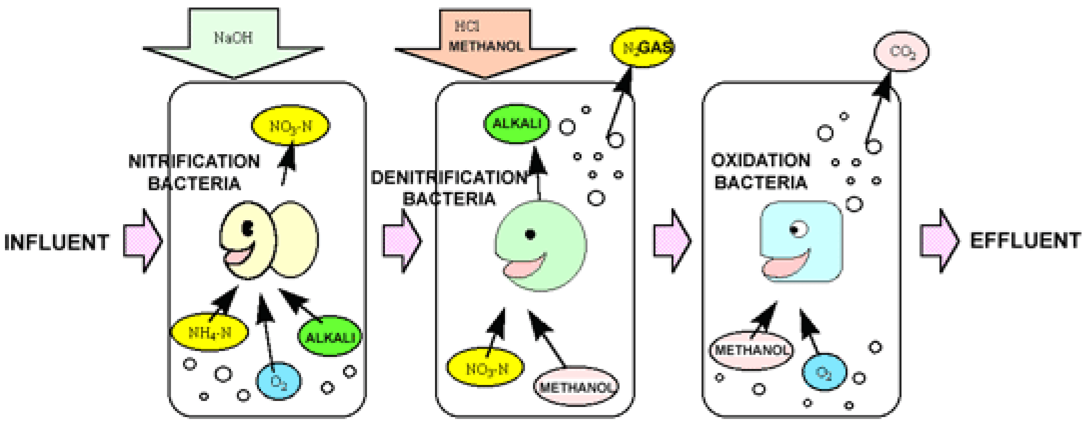






Nitrogen Reactive Transport Modeling

CONCEPT OF NITROGEN REMOVAL



Tallahassee Wastewater Treatment Plant ($270M)

ADE (for transport)

+ODE (for reactions)

$$\frac{\partial}{\partial x_i}\left(\theta D_{ij}\frac{\partial C_{NH4}}{\partial x_j}\right) - q_i\frac{\partial C_{NH4}}{\partial x_i} - (\theta + K_d\rho)\frac{\partial C_{NH4}}{\partial t} = -R_{nit}$$

$$R_{nit} = \frac{\partial \theta N}{\partial t} = -\boxed{K_{nit}}(\theta + \rho K_d)\boxed{f_m f_T}N$$

K_{nit} Optimal nitrification rate

f_m Reduction factors of soil moisture (m)

f_T Reduction factors of soil temperature (T)

Model Uncertainty in Literature

- *Hansen et al.* [1995] compared five models (ANIMO, SOILN, OMNI, NLEAP, and DAISY).
- *Frolking et al.* [1998] described four models (CENTURY, DNDC, Expert-N, and NASA-CASA) for simulating nitrous oxide (N_2O) emissions.
- *Wu and McGechan.* [1998] compared four models (ANIMO, SUNDIAL, SOILN, and DAISY).
- *Heinen* [2006] reviewed more than **fifty models**.

Alternative Models for Reduction Factors

T1 and M1 from SOILN [*Johnsson et al.*, 1987]

T2 and M2 from DAISY [*Hansen et al.*, 1991]

	Nitrification	Denitrification
T1	$f_T = Q_{10}^{(T-T_r)/10}$	$f_T = Q_{10}^{(T-T_r)/10}$
T2	$f_T = \begin{cases} 0, (T \leq 2°\text{C}) \\ 0.15(T-2), (2°\text{C} < T \leq 6°\text{C}) \\ 0.1T, (6°\text{C} < T \leq 20°\text{C}) \\ e^{\left(0.47-0.027T+0.00193T^2\right)}, \left(20°\text{C} < T \leq 40\right) \end{cases}$	$f_T = 1$
M1	$f_m = \begin{cases} \left(\frac{\theta-\theta_w}{\theta_{lo}-\theta_w}\right)^m, \theta_w \leq \theta < \theta_{lo} \\ 1, \theta_{lo} \leq \theta < \theta_{hi} \\ e_s + (1-e_s)\left(\frac{\theta_s-\theta}{\theta_s-\theta_{hi}}\right)^m, \theta_{hi} \leq \theta < \theta_s \end{cases}$	$f_m = \begin{cases} 0, \theta \leq \theta_d \\ \left(\frac{\theta-\theta_d}{\theta_s-\theta_d}\right)^{d_1}, \theta > \theta_d \end{cases}$
M2	$f_m = \begin{cases} \frac{\text{pF}}{1.5}, \text{pF} < 1.5 \\ 1, 1.5 \leq \text{pF} < 2.5 \\ 2 - \frac{2\text{pF}}{5}, 2.5 \leq \text{pF} < 5 \\ 0, \text{pF} > 5 \end{cases}$	$f_m = \begin{cases} 0, x_w < 0.8 \\ 2(x_w-0.8), 0.8 < x_w \leq 0.9 \\ 0.2 + 8(x_w-0.9), 0.9 < x_w \leq 1 \end{cases}$

Impacts of T/P Scenarios

- Temperature scenarios: S15, S5, and S25 for three yearly-average temperatures, affecting f_T directly
- Precipitation Scenarios: Present-day (13mm) and Wet (26mm), affecting moisture content (θ) directly and f_m indirectly.

	Nitrification
T1	$f_T = Q_{10}^{(T-T_r)/10}$
T2	$f_T = \begin{cases} 0, (T \le 2°\mathrm{C}) \\ 0.15(T-2), (2°\mathrm{C} < T \le 6°\mathrm{C}) \\ 0.1T, (6°\mathrm{C} < T \le 20°\mathrm{C}) \\ e^{(0.47-0.027T+0.00193T^2)}, (20°\mathrm{C} < T \le 40) \end{cases}$
M1	$f_m = \begin{cases} \left(\frac{\theta-\theta_w}{\theta_{lo}-\theta_w}\right)^m, \theta_w \le \theta < \theta_{lo} \\ 1, \theta_{lo} \le \theta < \theta_{hi} \\ e_s + (1-e_s)\left(\frac{\theta_s-\theta}{\theta_s-\theta_{hi}}\right)^m, \theta_{hi} \le \theta < \theta_s \end{cases}$
M2	$f_m = \begin{cases} \frac{\mathrm{pF}}{1.5}, \mathrm{pF} < 1.5 \\ 1, 1.5 \le \mathrm{pF} < 2.5 \\ 2 - \frac{2\mathrm{pF}}{5}, 2.5 \le \mathrm{pF} < 5 \\ 0, \mathrm{pF} > 5 \end{cases}$

Model Parameters and Parameter Importance

Parameters:

K_{nit}, K_{den}, T_r, Q_{10}, and m

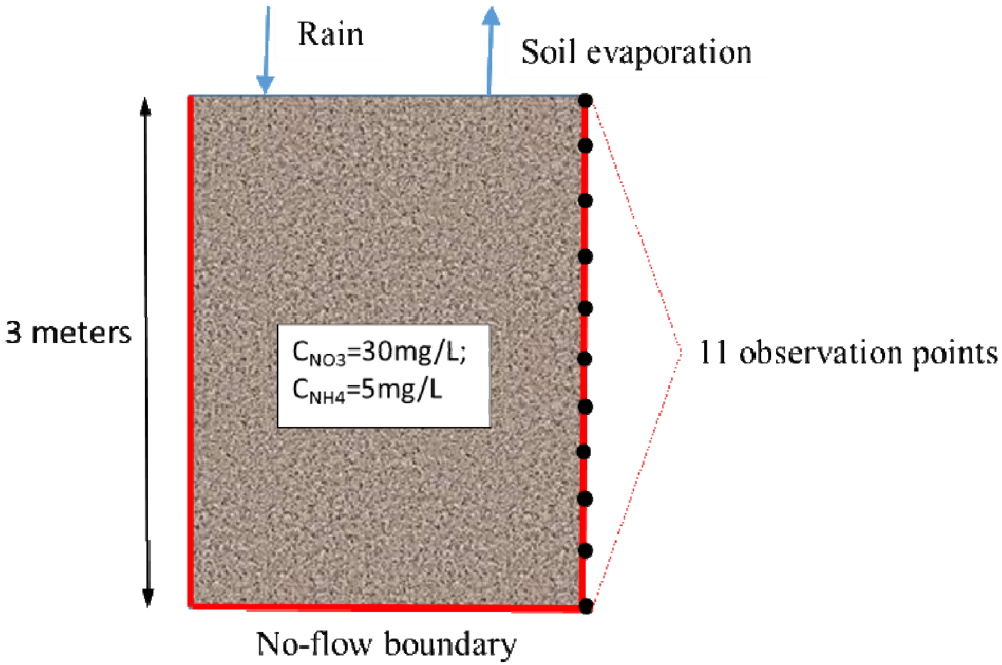
Scientific Questions:

- If a parameter (e.g., K_{nit}) is important in one model and/or scenario, will it be also important in another model and/or scenario?
- If not, how to identify important parameters under model and scenario uncertainty?

$$R_{nit} = \frac{\partial \theta N}{\partial t} = -K_{nit}(\theta + \rho K_d) f_m f_T N$$

	Nitrification
T1	$f_T = Q_{10}^{(T-T_r)/10}$
T2	$f_T = \begin{cases} 0, (T \le 2℃) \\ 0.15(T-2), (2℃ < T \le 6℃) \\ 0.1T, (6℃ < T \le 20℃) \\ e^{(0.47-0.027T+0.00193T^2)}, (20℃ < T \le 40) \end{cases}$
M1	$f_m = \begin{cases} \left(\frac{\theta-\theta_w}{\theta_{lo}-\theta_w}\right)^m, \theta_w \le \theta < \theta_{lo} \\ 1, \theta_{lo} \le \theta < \theta_{hi} \\ e_s + (1-e_s)\left(\frac{\theta_s-\theta}{\theta_s-\theta_{hi}}\right)^m, \theta_{hi} \le \theta < \theta_s \end{cases}$
M2	$f_m = \begin{cases} \frac{\mathrm{pF}}{1.5}, \mathrm{pF} < 1.5 \\ 1, 1.5 \le \mathrm{pF} < 2.5 \\ 2 - \frac{2\mathrm{pF}}{5}, 2.5 \le \mathrm{pF} < 5 \\ 0, \mathrm{pF} > 5 \end{cases}$

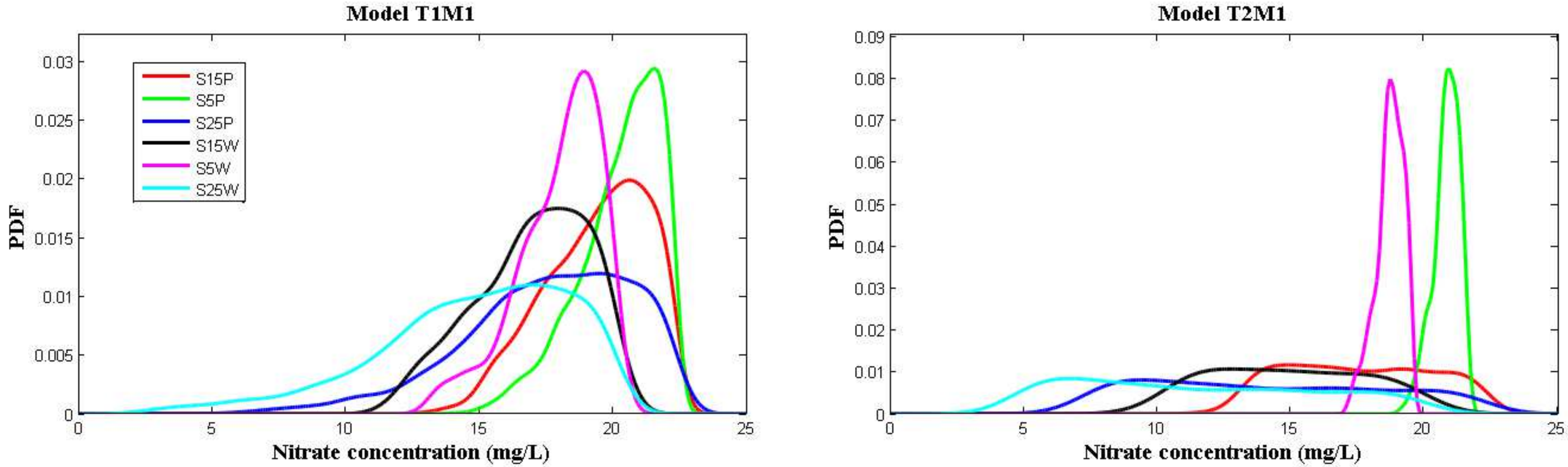
Nitrification and Denitrification Parameters



- Lysimeter-scale, 2-D flow and nitrogen reactive transport
- Parameters

  K_{nit}, K_{den}, T_r, Q_{10}, and m
- Parameter distributions are based on literature data.

Model and Scenario Uncertainty



- For a given model, simulated nitrate concentration varies dramatically under different scenarios.
- For a given scenario, simulated nitrate concentration varies significantly under different models.

Sensitivity Index for Individual Models and Scenarios

	Scenario 1 (S15P)				Scenario 2 (S5P)				Scenario 3 (S25P)			
	T1 M1	T2 M1	T1 M2	T2 M2	T1 M1	T2 M1	T1 M2	T2 M2	T1 M1	T2 M1	T1 M2	T2 M2
K_{den}	76.2	98.4	55.1	**91.5**	**63.3**	**48.4**	**27.5**	**5.91**	75.9	99.0	60.8	**97.7**
K_{nit}	4.25	1.35	25.5	9.09	9.97	53.6	47.2	94.5	1.08	0.32	3.73	1.64
T_r	23.5	0.00	21.5	0.00	10.5	0.00	14.3	0.00	27.7	0.00	41.5	0.00
Q_{10}	8.05	0.00	5.89	0.00	30.2	0.00	28.2	0.00	9.31	0.00	10.0	0.00
m	0.01	0.00	0.00	0.00	0.02	0.07	0.00	0.00	0.00	0.00	0.00	0.00
	Scenario 4 (S15W)				**Scenario 5 (S5W)**				**Scenario 6 (S25W)**			
	T1 M1	T2 M1	T1 M2	T2 M2	T1 M1	T2 M1	T1 M2	T2 M2	T1 M1	T2 M1	T1 M2	T2 M2
K_{den}	78.0	100	63.0	**96.6**	60.1	49.2	49.4	**5.71**	68.5	99.0	75.5	**99.2**
K_{nit}	2.93	1.39	6.12	1.57	7.64	52.4	18.2	100	0.79	0.29	2.29	0.32
T_r	23.2	0.00	26.1	0.00	11.4	0.00	11.0	0.00	36.1	0.00	31.1	0.00
Q_{10}	7.55	0.00	7.73	0.00	29.6	0.00	29.5	0.00	10.8	0.00	9.79	0.00
m	0.01	0.00	0.00	0.00	0.02	0.06	0.00	0.00	0.00	0.00	0.00	0.00

- Under scenario S5P, the total sensitivity index of parameter K_{den} changes from 63.3% for model T1M1 to 5.91% for model T2M2
- For model T2M2, the index of parameter K_{den} changes from 5.71% under Scenario S5W to 99.2% under Scenario S25W

Sensitivity Index for Multiple Models and Scenarios

	S15P	S5P	S25P	S15W	S5W	S25W	S
K_{den}	88.17	49.68	90.84	92.68	68.39	86.03	87.01
K_{nit}	5.14	28.79	1.07	3.34	12.85	1.26	3.20
T_r	7.68	8.84	9.70	8.08	6.39	15.24	10.86
Q_{10}	3.12	26.22	2.97	3.06	18.74	4.35	4.85
m	0.01	0.04	0.00	0.01	0.03	0.00	0.01

- The multi-model sensitivity index still varies considerably under different scenarios.
- Therefore, it is necessary to evaluate the sensitivity index for multiple models and multiple scenarios.
- The multi-model, multi-scenario sensitivity index gives the composite evaluation of parameter importance.
- The relative importance of the parameters is physically reasonable.

Global Sensitivity Analysis for Process Identification under Model Uncertainty

Heng Dai, Pacific Northwest National Laboratory
Ming Ye (mye@fsu.edu), Florida State University
Anthony Walker, Oak Ridge National Laboratory
Xingyuan Chen, Pacific Northwest National Laboratory

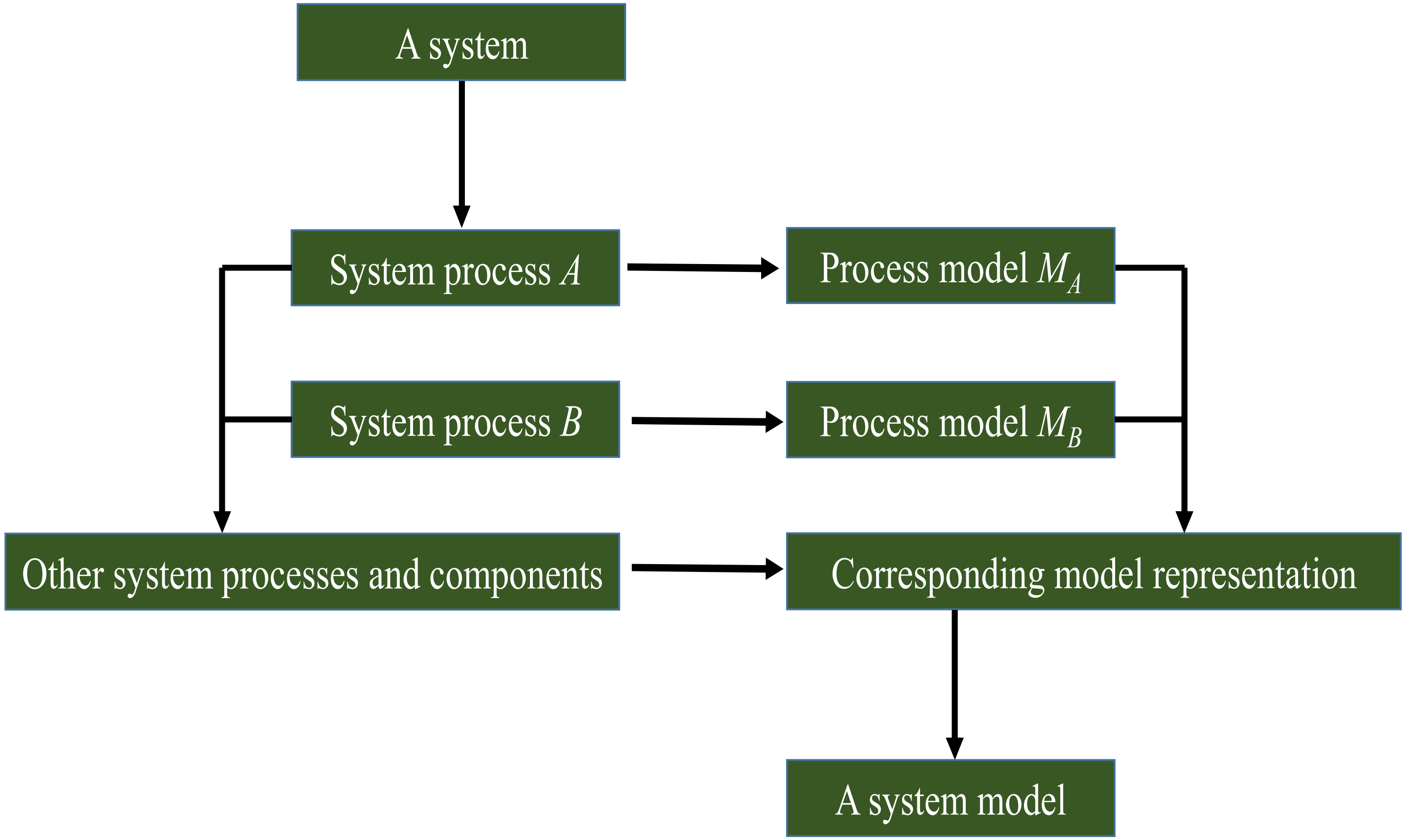
Dai et al. (2017, WRR)

Identify Controlling Processes

- Identify the controlling processes that determine system dynamics at various scales, because
  - It is difficult, if not impossible, to model everything.
  - Only a few dominant processes control system dynamics.
  - Numerical models including the dominant processes could capture system dynamics.
- Many methods for identifying dominant processes:
  - Model-driven/Data-driven/Hybrid methods
  - Statistical analysis (PCA, FFT, information criteria, Sensitivity analysis)

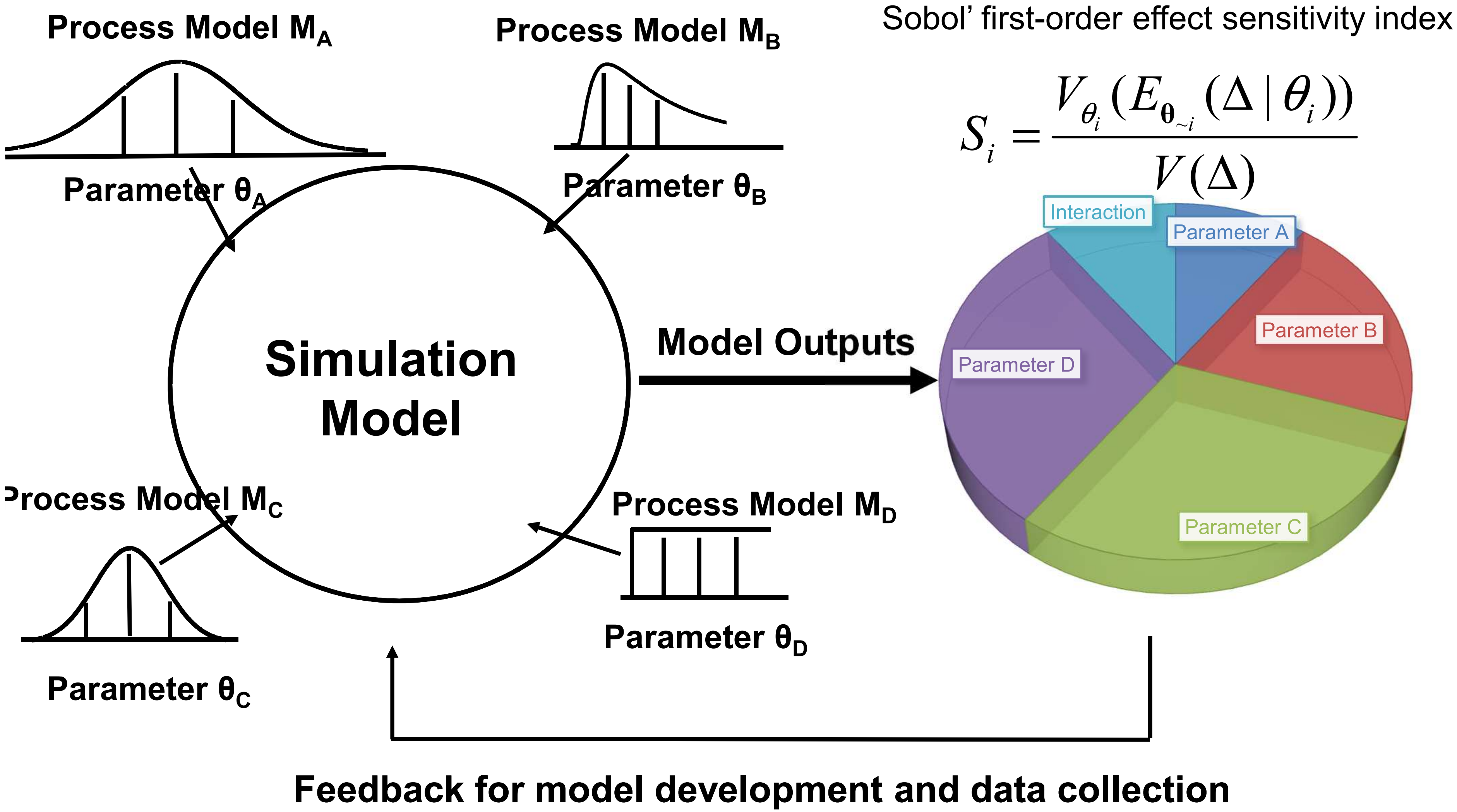
Potential Problem with a Single Model

Building a process-based system model with a single representation of each process



Global Sensitivity Analysis for Process Identification

- Develop a single model for each process
- Conduct sensitivity analysis for the process model parameters
- Identify dominant processes, if their parameters are influential to model outputs

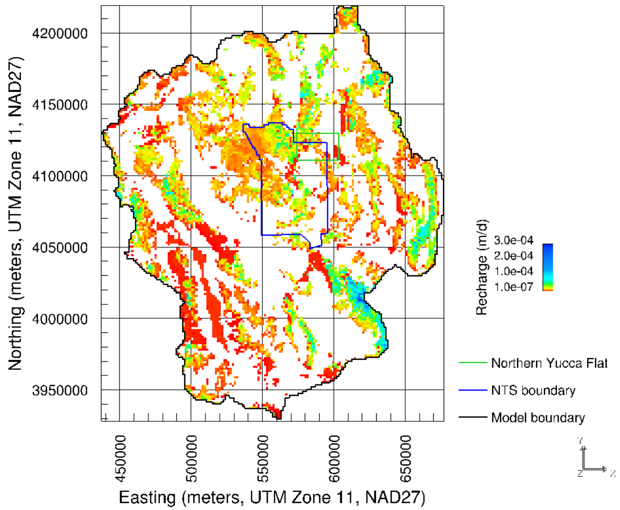


Uncertainty in Process Models

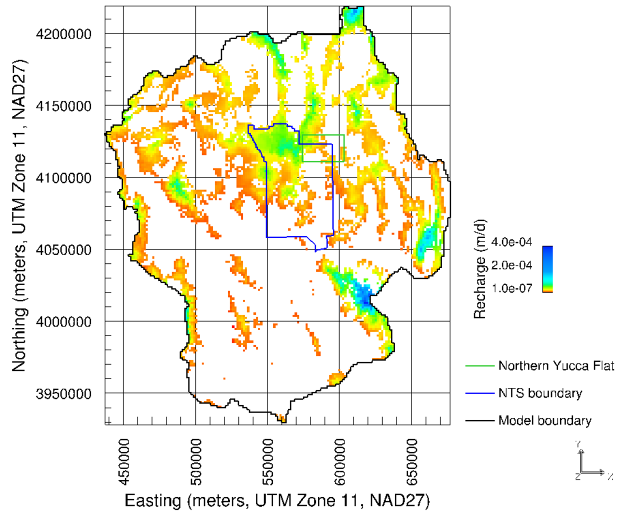
Death Valley Regional Flow System (DVRFS) Model

- Recharge process: six alternative recharge models
- Geological process: six hydrostratigraphic frameworks

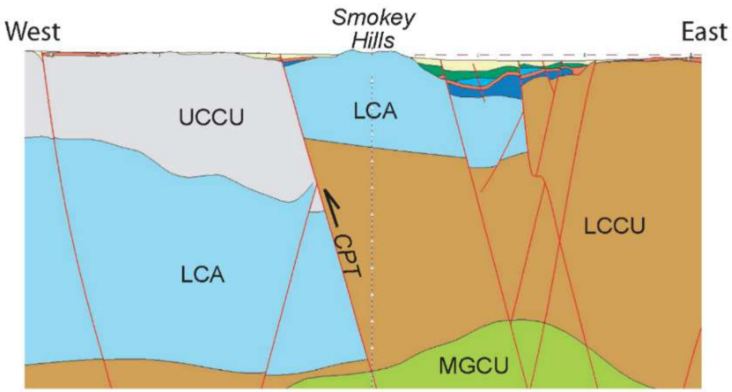
Net infiltration model (NIM1)



Chloride mass balance model (CMB2)

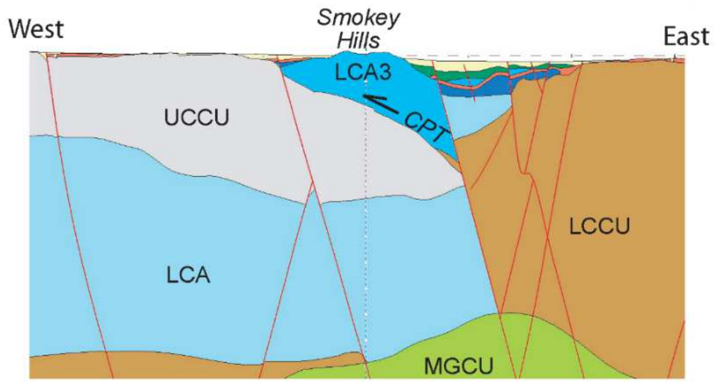


Base Model



Profile through Northern Yucca Flat-Base Model

CP Thrust Alternative

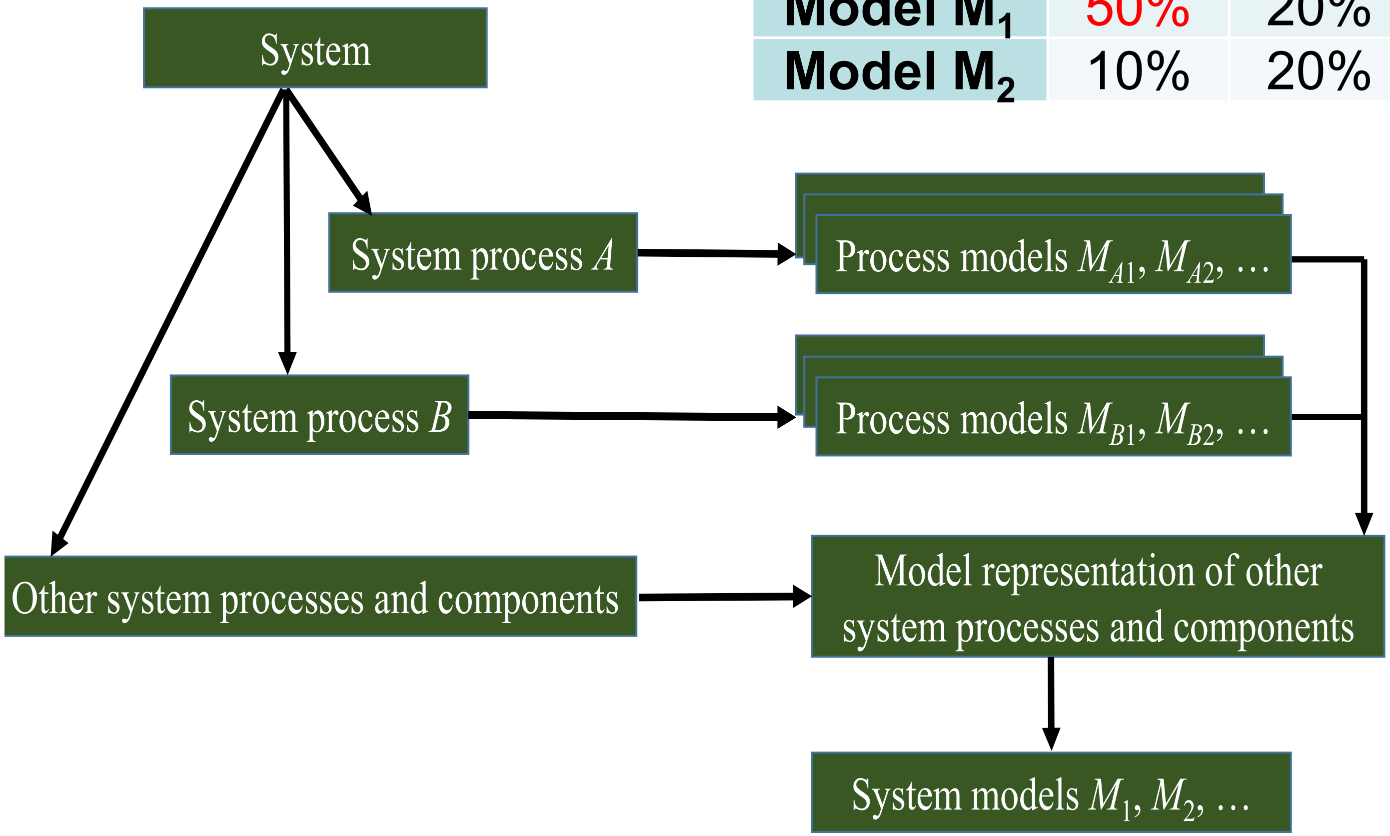


Profile through Northern Yucca Flat-CP Thrust Alternative

Scientific Question

If we are not certain about the choice of process models and model parameters, can we correctly identify the controlling processes of a complex system?

	S_A	S_B	S_C	S_D
Model M_1	50%	20%	20%	10%
Model M_2	10%	20%	20%	50%

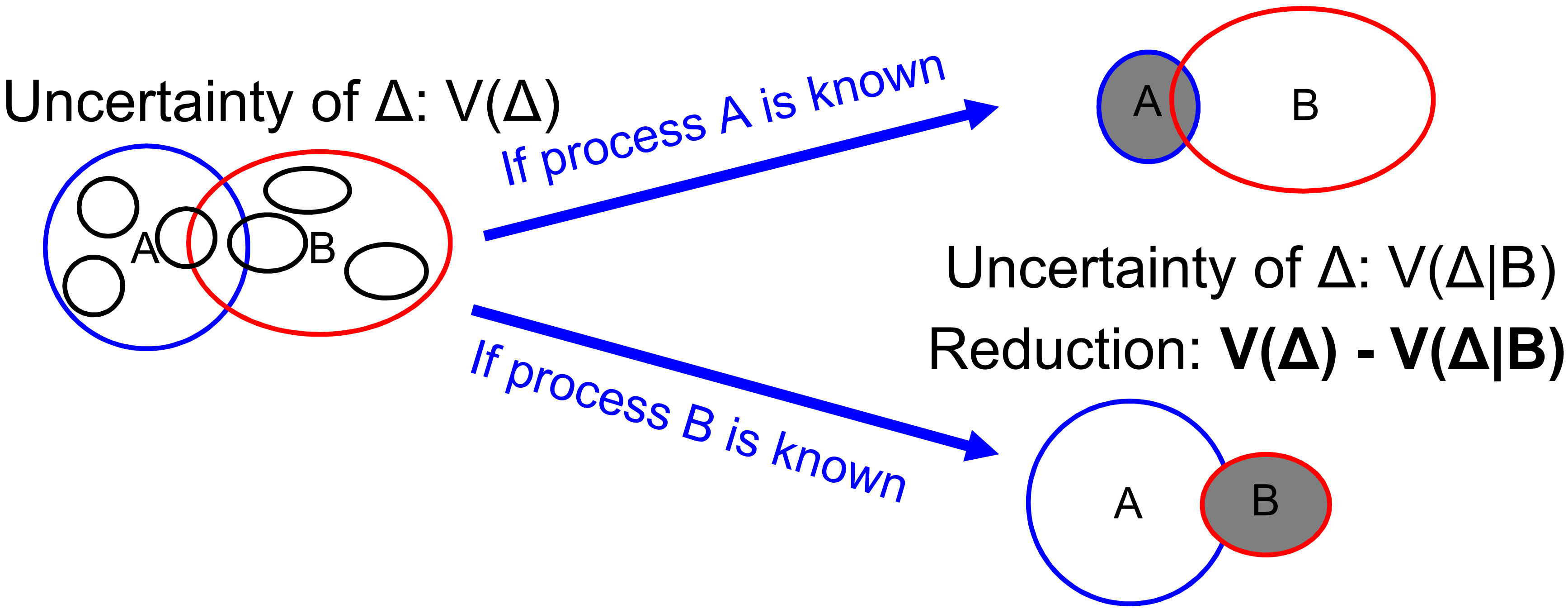


Process Sensitivity Index: Basic Idea

Uncertainty of model simulation Δ involves two uncertain processes: A and B

Uncertainty of Δ: V(Δ|A)

Reduction: **V(Δ) - V(Δ|A)**



An important process corresponds to larger uncertainty reduction.

Due to process model uncertainty, use <u>the average over all process models</u>, i.e., using $E_A V(\Delta|A)$ and $E_B V(\Delta|B)$.

Mathematical Formulation

- Variance decomposition (law of total variance)

$$V(\Delta) = E_{\mathbf{K}}(V_{\sim\mathbf{K}}[\Delta \mid K]) + V_{\mathbf{K}}(E_{\sim\mathbf{K}}[\Delta \mid K])$$

$$= E_{\mathbf{M}_K}(V_{\sim\mathbf{M}_K}[\Delta \mid M_K]) + \boxed{V_{\mathbf{M}_K}(E_{\sim\mathbf{M}_K}[\Delta \mid M_K])}$$

Variance after fixing models of K Variance reduction after fixing models of K

- **Process sensitivity indices**

$$PS_K = \frac{\boxed{V_{\mathbf{M}_K}(E_{\sim\mathbf{M}_K}[\Delta \mid M_K])}}{V(\Delta)}$$

- Variance reduction: $Var(X)=E(X^2)-(EX)^2$

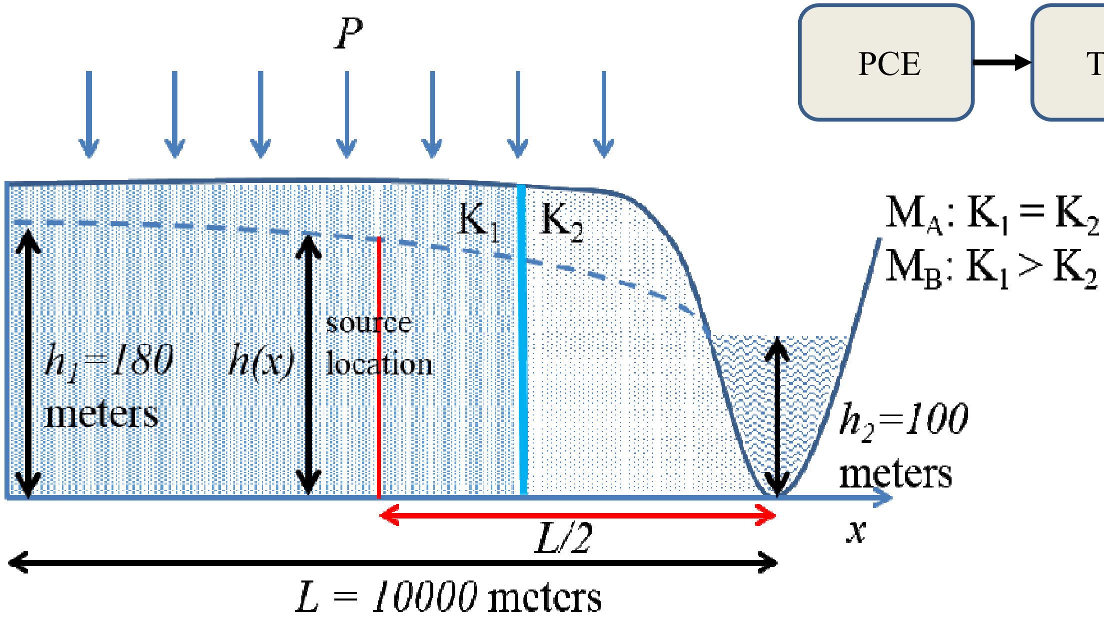
$$\boxed{V_{\mathbf{M}_K}(E_{\sim\mathbf{M}_K}[\Delta \mid M_K])} = \boxed{E_{\mathbf{M}_K}(E_{\sim\mathbf{M}_K}[\Delta \mid M_K])^2} - (E_{\mathbf{M}_K}E_{\sim\mathbf{M}_K}[\Delta \mid M_K])^2$$

- Model averaging

$$\boxed{E_{\mathbf{M}_K}(X_{M_K})} = \sum_{M_K} X_{M_K} \boxed{P(M_K)}$$ Model probability

Method Illustration

PCE → TCE → DCE → VC → ETH



- Groundwater flow in an unconfined aquifer
- Multispecies reactive transport

Recharge process (R)

$$R_1 : w = a(P - 355.6)^{0.50}$$

$$R_2 : w = b(P - 399.80)$$

Model R_1 has random parameter, *a*

Model R_2 has random parameter, *b*

Geology process (G)

$$G_1 : K \text{ for any } x$$

$$G_2 : K = \begin{cases} K_1 & \text{for } x < 7000 \\ K_2 & \text{for } x \geq 7000 \end{cases}$$

Model G_1 has random parameter *K*

Model G_2 has random parameters *K_1 and K_2*

Scientific Question:

Which process (recharge or geology) is more important to the state variables (head and concentration)?

	NOT considering model uncertainty							
S_i	R_1G_1		R_1G_2		R_2G_1		R_2G_2	
	R	G	R	G	R	G	R	G
Head (x=6,000m)	94.9	4.78	88.7	10.6	61.5	37.8	6.51	93.2
Conc. (x=5,500m)	21.3	78.5	1.48	98.5	67.7	25.5	16.5	81.0

	Considering model uncertainty	
PS_K	Recharge	Geology
Head	28.43	67.94
Conc. (x=5,500m)	0.11	90.99

- Biased process identification may be resulted, because the identified important process changes for different models.
- The new process sensitivity index is able to avoid the biased identification.

Conclusions

- Old Question: If we are not certain about what model parameter values to use, can we identify the important parameters/processes?
- Model uncertainty and scenario uncertainty force us to answer a new question: If we are not certain about
  - not only what model parameter values to use
  - but also what models and what model scenarios to use

  can we identify the important parameters and processes?
- The answer is yes, as along as we know what we are uncertain about and know how to quantify the corresponding uncertainty.
- The new sensitivity index is mathematically/statistically general, and can be used to a wide range of problems.
- The method implementation is computationally expensive, but the computational barrier can be broken.

How many lampposts do we need?

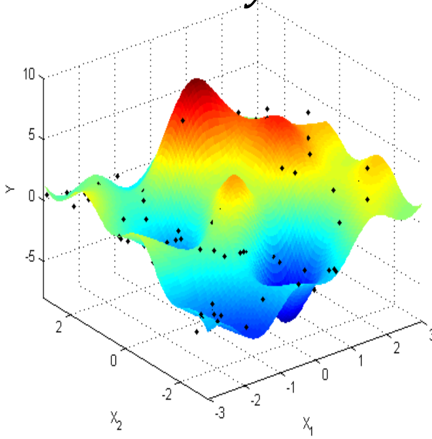
Developing conceptual models is more like an art!

What if the keys are not under any lampposts?

Start with a lamppost, but may go elsewhere

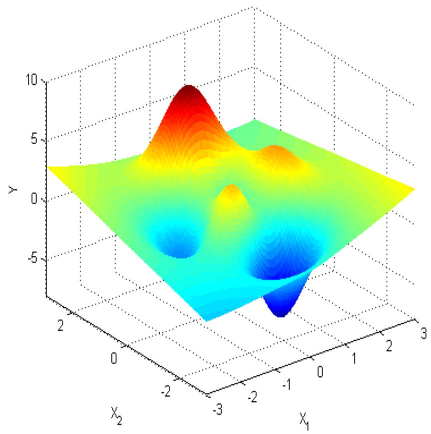
$$y = M(\theta) + r = M(\theta) + b(\boldsymbol{x}, \boldsymbol{\phi}) + \epsilon$$

Observations
y



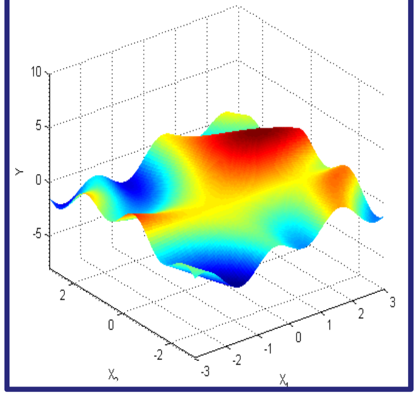
=

Physical model
$M(\boldsymbol{\theta})$



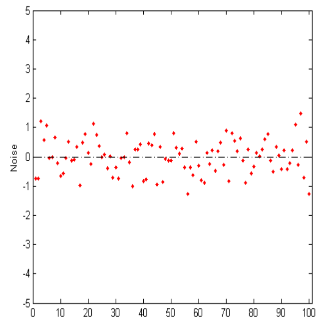
+

Error model
$b(\boldsymbol{x}, \boldsymbol{\phi})$



+

Observation
erro
$\boldsymbol{\epsilon}$



Data-driven

Xu, Valocchi, Ye, et al. (2017a, WRR)
Xu, Valocchi, Ye, et al. (2017b, WRR)

THANK YOU!