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
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."

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

Bredehoef (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

Which model is correct?
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

(\textit{Draper, 1995; Hoeting et al., 1999; Neuman, 2002; Ye et al., 2004})

\[ p(\Delta|D) = \sum_{k=1}^{K} p(\Delta|M_k, D) p(M_k|D) \]

- 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

Model Probability:
\[
p(M_k | D) = \frac{p(D | M_k) p(M_k)}{\sum_{l=1}^{K} p(D | M_l) p(M_l)}
\]

Model Evidence
\[
p(D | M_k) = \int p(D | \theta_k, M_k) p(\theta_k | M_k) d\theta_k
\]

- \( \theta_k \): multiple-dimensional parameters of model \( M_k \)
- Need to calculate the multivariate integration.
  Computationally expensive!!!
Computational Challenge

\[ p(D|M) = \int p(D|\theta, M) p(\theta|M) \, d\theta \]

Arithmetic mean \( \hat{p}_{AM}(D|M) \approx \frac{1}{n} \sum_{i=1}^{n} p(D|\theta_i, M) \)

Harmonic mean \( \hat{p}_{HM}(D|M) \approx n \left/ \sum_{i=1}^{n} \frac{1}{p(D|\theta_i, M)} \right. \)

Straightforward MC implementation, but do not work well.

Nested sampling \( p(D|M) = \int_{0}^{1} L(X|D,M) \, dX \)

Thermodynamic Integration \( p(D|M) = \exp \int_{0}^{1} E_{\beta} \left[ \ln p(D|\theta, M) \right] d\beta \)

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

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

**Cartoon Description:**
- A man is fishing on a small island.
- Two sharks are in the water, one holding a sign saying "Sea-Level Rise" and the other saying "Good news! At the current rate of global warming we should be able to just swim over there and eat him in under five years...!"
Sea-Level Rise Scenarios

Projecting Future Sea-Level Rise

\[ S - S_0 = a (Y - Y_0) + b (Y - Y_0)^2 \]

- 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

\[ E(\Delta | D) = \sum_{i=1}^{I} E(\Delta | D, S_i) p(S_i) \]

- 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.

Conceptual Model Challenges in Groundwater Reactive Transport Modeling

Uncertainty in model parameters, structures, and Scenarios

Darcy's Law:
- $q = -K_{eff} \frac{dh}{dX}$
- $K_{eff} = f(K_1, K_2, K_3)$

K_1: silt & clay
K_2: medium sand
K_3: gravel & cobbles

Adsorption:
- $>SOH + H_2O + UO_2^{+2} \leftrightarrow >SOUO_2OH + 2H^+$

Coprecipitation:
- $(1-x)Ca^{+2} + xUO_2^{+2} + CO_3^{-2} \leftrightarrow Ca_{(1-x)}(UO_2)_xCO_3$

Ion Exchange:
- $Ca^{+2} + 2NaX \leftrightarrow CaX_2 + 2Na^+$

Intraparticle Diffusion:
- Flux = - D dC/dr

Speciation:
- $Ca^{+2} + UO_2^{+2} + 3CO_3^{-2} \leftrightarrow CaUO_2(CO_3)_3^{-2}$

Uranium(IV) Oxidative Dissolution (post remediation):
- $UO_2 + H^+ + 1/2 O_2 = UO_2^{+2} + H_2O$

Microbial Growth / Uranium Reduction:
- $CH_3COO^- + 6.2UO_2^{+2} + 2.8H_2O + .1NH_4 \rightarrow 6.2UO_2(s) + .1 bacteria + 6.8 H^+ + 1.6HCO_3^-$
Surface Complexation Models

- 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!

<table>
<thead>
<tr>
<th>Model</th>
<th>Reactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>$S_1\text{OH} + UO_2^{+2} + H_2O = S_1\text{OUO}_2\text{OH} + 2H^+$</td>
</tr>
<tr>
<td>C2</td>
<td>$S_1\text{OH} + UO_2^{+2} + H_2O = S_1\text{OUO}_2\text{OH} + 2H^+$ $S_2\text{OH} + UO_2^{+2} + H_2O = S_2\text{OUO}_2\text{OH} + 2H^+$</td>
</tr>
<tr>
<td>C3</td>
<td>$S_1\text{OH} + UO_2^{+2} + H_2O = S_1\text{OUO}_2\text{OH} + 2H^+$ $S_2\text{OH} + UO_2^{+2} = S_2\text{OUO}_2^{+2} + H^+$</td>
</tr>
<tr>
<td>C4</td>
<td>$S_1\text{OH} + UO_2^{+2} + H_2O = S_1\text{OUO}_2\text{OH} + 2H^+$ $S_2\text{OH} + UO_2^{+2} + H_2O = S_2\text{OUO}_2\text{OH} + 2H^+$ $S_2\text{OH} + UO_2^{+2} = S_2\text{OUO}_2^{+2} + H^+$</td>
</tr>
<tr>
<td>C5</td>
<td>$S_1\text{OH} + UO_2^{+2} + H_2O = S_1\text{OUO}_2\text{OH} + 2H^+$ $S_2\text{OH} + UO_2^{+2} + H_2O = S_2\text{OUO}_2\text{OH} + 2H^+$ $S_2\text{OH} + UO_2^{+2} = S_2\text{OUO}_2^{+2} + H^+$ $S_3\text{OH} + UO_2^{+2} + H_2O = S_3\text{OUO}_2\text{OH} + 2H^+$</td>
</tr>
<tr>
<td>C6</td>
<td>$S_1\text{OH} + UO_2^{+2} + H_2O = S_1\text{OUO}_2\text{OH} + 2H^+$ $S_2\text{OH} + UO_2^{+2} + H_2O = S_2\text{OUO}_2\text{OH} + 2H^+$ $S_2\text{OH} + UO_2^{+2} = S_2\text{OUO}_2^{+2} + H^+$ $S_3\text{OH} + UO_2^{+2} + H_2O = S_3\text{OUO}_2\text{OH} + 2H^+$</td>
</tr>
<tr>
<td>C7</td>
<td>$S_1\text{OH} + UO_2^{+2} + H_2O = S_1\text{OUO}_2\text{OH} + 2H^+$ $S_2\text{OH} + UO_2^{+2} + H_2O = S_2\text{OUO}_2\text{OH} + 2H^+$ $S_3\text{OH} + UO_2^{+2} + H_2O = S_3\text{OUO}_2\text{OH} + 2H^+$</td>
</tr>
</tbody>
</table>
A model is composed of seven different components:
- System boundary ($B$),
- Forcing ($u$),
- Initial states ($x_0$),
- Parameters ($\theta$),
- 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 ($\theta$), and
(4) **Data uncertainty** in observations.
A Comprehensive and Hierarchical Framework

Scenario $S_1$
- Model $M_1$
  - Parameter $P_1$
    - $UQ(P_1|M_1,S_1)$

Scenario $S_2$
- Model $M_2$
  - Parameter $P_2$
    - $UQ(P_2|M_2,S_1)$

Same or Different Models and Model Parameters

Quantify Parametric Uncertainty
- $UQ(P,M|S_1)$
- $UQ(P,M|S_2)$
- $UQ(P,M,S)$

Address Uncertainty
- $UQ(P_1|M_1,S_1)$
- $UQ(P_2|M_2,S_1)$

Quantify Parametric and Model Uncertainty
- $UQ(P,M,S)$

Quantify Parametric, Model, and Scenario Uncertainty
- $UQ(P,M|S_1)$
- $UQ(P,M|S_2)$
Bayesian UQ Framework: Scenario, Model, and Parametric Uncertainty

- **Scenario uncertainty**
  \[
  \text{Var}(\Delta) = E_s \text{Var}(\Delta | S) + \text{Var}_s E(\Delta | S)
  \]
  Within-scenario  Between-scenario

- **Model uncertainty** of a given scenario
  \[
  \text{Var}(\Delta | S) = E_{M | S} \text{Var}(\Delta | M, S) + \text{Var}_{M | S} E(\Delta | M, S)
  \]
  Within-model  Between-model

- **Parametric uncertainty** of a given scenario and a given model
  \[
  \text{Var}(\Delta | M, S) = E_{\theta | M, S} \text{Var}(\Delta | \theta, M, S) + \text{Var}_{\theta | M, S} E(\Delta | \theta, M, S)
  \]
  Data  Parametric
Uncertainty Decomposition

\[ \text{Var}(\Delta) = E_S E_{M|S} E_{\theta|M,S} \text{Var}(\Delta | \theta, M, S) + E_S E_{M|S} \text{Var}_{\theta|M,S} E(\Delta | \theta, M, S) + E_S \text{Var}_{M|S} E_{\theta|M,S} E(\Delta | \theta, M, S) + \text{Var}_{S} E_{M|S} E_{\theta|M,S} E(\Delta | \theta, M, S) \]

Data uncertainty in \( \Delta \)
Scenario uncertainty in \( S \)
Model uncertainty in \( M \)
Parametric uncertainty in \( \theta \)
Groundwater Sensitivity Analysis under Model and Scenario Uncertainty: Not Look for Keys under the Lamppost
Sensitivity Analysis

• 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

Razavi and Gupta (2015, WRR)
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?

\[
\Delta = f(\theta_A, \theta_B, \theta_C, \theta_D)
\]

Feedback for model development and data collection

Sobol' first-order effect sensitivity index

\[
S_i = \frac{V_{\theta_i} \left( E_{\theta_{-i}}(\Delta | \theta_i) \right)}{V(\Delta)}
\]
Challenges of Global Sensitivity Analysis under Model Uncertainty

\[ \Delta = f(\theta_A, \theta_B, \theta_C, \theta_D) \]

\[ S_i = \frac{V_{\theta_i}(E_{\theta_{-i}}(\Delta | \theta_i))}{V(\Delta)} \]

<table>
<thead>
<tr>
<th>( S_A )</th>
<th>( S_B )</th>
<th>( S_C )</th>
<th>( S_D )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>20%</td>
<td>20%</td>
<td>50%</td>
</tr>
</tbody>
</table>

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

\[
\Delta = M_1(\theta_A, \theta_B, \theta_C, \theta_D) \quad \Delta = M_2(\theta_A, \theta_B, \theta_C)
\]

\[
S_i = \frac{V_{\theta_i}(E_{\theta_{-i}}(\Delta|\theta_i))}{V_1(\Delta)} \quad S_i = \frac{V_{\theta_i}(E_{\theta_{-i}}(\Delta|\theta_i))}{V_2(\Delta)}
\]

<table>
<thead>
<tr>
<th></th>
<th>$S_A$</th>
<th>$S_B$</th>
<th>$S_C$</th>
<th>$S_D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model M_1</td>
<td>10%</td>
<td>20%</td>
<td>20%</td>
<td>50%</td>
</tr>
<tr>
<td>Model M_2</td>
<td>50%</td>
<td>20%</td>
<td>10%</td>
<td>N/A</td>
</tr>
</tbody>
</table>

- Different models may have different parameters
- Sensitivity index is not comparable across 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, $V_1(\Delta)$.

Parametric Uncertainty Under Model and Scenario Uncertainty

\[ \text{Var}(\Delta) = E_S E_{M|S} E_{\theta|M,S} \text{Var}(\Delta | \theta, M, S) \]

\[ + E_S E_{M|S} \text{Var}_{\theta|M,S} E(\Delta | \theta, M, S) \]

\[ + E_S \text{Var}_{M|S} E_{\theta|M,S} E(\Delta | \theta, M, S) \]

\[ + \text{Var}_S E_{M|S} E_{\theta|M,S} E(\Delta | \theta, M, S) \]

**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

\[ Var(\Delta) = E_S E_{M|S} E_{\theta|M,S} Var(\Delta | \theta, M, S) \]

\[ + E_S E_{M|S} Var_{\theta|M,S} E(\Delta | \theta, M, S) \]

\[ + E_S Var_M E_{\theta|M,S} E(\Delta | \theta, M, S) \]

\[ + Var_S E_{M|S} E_{\theta|M,S} E(\Delta | \theta, M, S) \]

For a single model and a single scenario
(no model and scenario uncertainty)

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

Sobol’ method:

\[ S_{T_i} = \frac{E_{\theta_{\sim i}|M,S} V_{\theta_i|M,S}(E(\Delta | \theta, M, S) | \theta_{\sim i})}{V_{\theta|M,S} E(\Delta | \theta, M, S)} \]
New Sensitivity Index

For multiple models but a single scenario (model uncertainty and no scenario uncertainty)

\[ E_S E_{M|S} \text{Var}_{\theta|M,S} E(\Delta \mid \theta, M, S) = E_{M|S} \text{Var}_{\theta|M,S} E(\Delta \mid \theta, M, S) \]

\[ S_{Ti} = \frac{E_{M|S} E_{\theta_{-i}|M,S} \text{Var}_{\theta|M,S}(E(\Delta \mid \theta, M, S) \mid \theta_{-i})}{E_{M|S} \text{Var}_{\theta|M,S} E(\Delta \mid \theta, M, S)} \]

\[ = \sum_{M} P(M \mid S) \text{Var}_{\theta|M,S}(E(\Delta \mid \theta, M, S) \mid \theta_{-i}) \]

\[ = \sum_{M} P(M \mid S) \text{Var}_{\theta|M,S} E(\Delta \mid \theta, M, S) \]
New Sensitivity Index

For multiple models and multiple scenarios (model uncertainty and scenario uncertainty)

\[ S_{Ti} = \frac{\sum \sum P(S) P(M \mid S) E_{\theta \sim_i \mid M, S} \text{Var}_{\theta \mid M, S} (E(\Delta \mid \theta, M, S) \mid \theta \sim_i)}{\sum \sum P(S) P(M \mid S) \text{Var}_{\theta \mid M, S} E(\Delta \mid \theta, M, S)} \]

\[ E_S E_{M \mid S} \text{Var}_{\theta \mid M, S} E(\Delta \mid \theta, M, S) \]
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

C.6 S.1 Drawing and labelling a diagram of the nitrogen cycle.

Nitrification

\[ \text{Ammonium} + \text{Oxygen} \rightarrow \text{Nitrate} \]

\[ \text{Nitrite} + \text{Oxygen} \rightarrow \text{Nitrate} \]

Denitrification

\[ \text{Nitrate} \rightarrow \text{Nitric oxide} \]
\[ \text{Nitric oxide} \rightarrow \text{Nitrogen dioxide} \]
\[ \text{Nitrogen dioxide} \rightarrow \text{Nitrous oxide} \]
\[ \text{Nitrous oxide} \rightarrow \text{Nitrogen} \]
Nitrogen Reactive Transport Modeling

CONCEPT OF NITROGEN REMOVAL

Tallahassee Wastewater Treatment Plant ($270M)

ADE (for transport)
+ODE (for reactions)

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

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

\(K_{nit}\) Optimal nitrification rate
\(f_m\) Reduction factors of soil moisture (m)
\(f_T\) Reduction factors of soil temperature (T)
Model Uncertainty in Literature


• *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]

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

<table>
<thead>
<tr>
<th></th>
<th>Nitrification</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>$f_T = Q_{10}^{(T-T_r)/10}$</td>
</tr>
<tr>
<td>T2</td>
<td>$f_T = \begin{cases} 0, &amp; (T \leq 2°C) \ 0.15(T-2), &amp; (2°C &lt; T \leq 6°C) \ 0.1T, &amp; (6°C &lt; T \leq 20°C) \ e^{(0.47-0.027T+0.00193T^2)}, &amp; (20°C &lt; T \leq 40) \end{cases}$</td>
</tr>
<tr>
<td>M1</td>
<td>$f_m = \begin{cases} \left( \frac{\theta - \theta_w}{\theta_{lo} - \theta_w} \right)^m, &amp; \theta_w \leq \theta &lt; \theta_{lo} \ 1, &amp; \theta_{lo} \leq \theta &lt; \theta_{hi} \ e_s + (1-e_s)\left( \frac{\theta - \theta_s}{\theta_s - \theta_{hi}} \right)^m, &amp; \theta_{hi} \leq \theta &lt; \theta_s \end{cases}$</td>
</tr>
<tr>
<td>M2</td>
<td>$f_m = \begin{cases} \frac{pF}{1.5}, &amp; pF &lt; 1.5 \ 1, &amp; 1.5 \leq pF &lt; 2.5 \ 2 - \frac{2pF}{5}, &amp; 2.5 \leq pF &lt; 5 \ 0, &amp; pF &gt; 5 \end{cases}$</td>
</tr>
</tbody>
</table>
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?

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

<table>
<thead>
<tr>
<th>( f_T = Q_{10}^{(T-T_r)/10} )</th>
</tr>
</thead>
</table>
| \( f_T = \begin{cases} 
0, & (T \leq 2^\circ C) \\
0.15(T - 2), & (2^\circ C < T \leq 6^\circ C) \\
0.1T, & (6^\circ C < T \leq 20^\circ C) \\
e^{(0.47 - 0.027T + 0.00193T^2)}, & (20^\circ C < T \leq 40) 
\end{cases} \) |

| \( f_m = \begin{cases} 
\left( \frac{\theta - \theta_w}{\theta_{lo} - \theta_w} \right)^m, & \theta_w \leq \theta < \theta_{lo} \\
1, & \theta_{lo} \leq \theta < \theta_{hi} \\
e_s + (1 - e_s) \left( \frac{\theta_s - \theta}{\theta_s - \theta_{hi}} \right)^m, & \theta_{hi} \leq \theta < \theta_s 
\end{cases} \) |

| \( f_m = \begin{cases} 
pF, & pF < 1.5 \\
1, & 1.5 \leq pF < 2.5 \\
2 - \frac{2pF}{5}, & 2.5 \leq pF < 5 \\
0, & pF > 5 
\end{cases} \) |
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

<table>
<thead>
<tr>
<th>Scenario 1 (S15P)</th>
<th>Scenario 2 (S5P)</th>
<th>Scenario 3 (S25P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1M1</td>
<td>T2M1</td>
<td>T1M2</td>
</tr>
<tr>
<td>$K_{den}$</td>
<td>76.2</td>
<td>98.4</td>
</tr>
<tr>
<td>$K_{nit}$</td>
<td>4.25</td>
<td>1.35</td>
</tr>
<tr>
<td>$T_r$</td>
<td>23.5</td>
<td>0.00</td>
</tr>
<tr>
<td>$Q_{10}$</td>
<td>8.05</td>
<td>0.00</td>
</tr>
<tr>
<td>$m$</td>
<td>0.01</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario 4 (S15W)</th>
<th>Scenario 5 (S5W)</th>
<th>Scenario 6 (S25W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1M1</td>
<td>T2M1</td>
<td>T1M2</td>
</tr>
<tr>
<td>$K_{den}$</td>
<td>78.0</td>
<td>100</td>
</tr>
<tr>
<td>$K_{nit}$</td>
<td>2.93</td>
<td>1.39</td>
</tr>
<tr>
<td>$T_r$</td>
<td>23.2</td>
<td>0.00</td>
</tr>
<tr>
<td>$Q_{10}$</td>
<td>7.55</td>
<td>0.00</td>
</tr>
<tr>
<td>$m$</td>
<td>0.01</td>
<td>0.00</td>
</tr>
</tbody>
</table>

- 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

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

• 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

Sobol’ first-order effect sensitivity index

\[
S_i = \frac{V_{\theta_i}(E_{\theta_{-i}}(\Delta | \theta_i))}{V(\Delta)}
\]

Feedback for model development and data collection
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?
Process Sensitivity Index: Basic Idea

Uncertainty of model simulation $\Delta$ involves two uncertain processes: $A$ and $B$

Uncertainty of $\Delta$: $V(\Delta)$

Reduction: $V(\Delta) - V(\Delta|A)$

Uncertainty of $\Delta$: $V(\Delta|A)$

Uncertainty of $\Delta$: $V(\Delta|B)$

Reduction: $V(\Delta) - V(\Delta|B)$

An important process corresponds to larger uncertainty reduction.

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

- Variance decomposition (law of total variance)
  \[ V(\Delta) = E_K(V_{\sim K}[\Delta \mid K]) + V_K(E_{\sim K}[\Delta \mid K]) \]
  \[ = E_M(K)(V_{\sim M_K}[\Delta \mid M_K]) + V_M(K)(E_{\sim M_K}[\Delta \mid M_K]) \]
  Variance after fixing models of K  Variance reduction after fixing models of K

- Process sensitivity indices
  \[ PS_K = \frac{V_M(K)(E_{\sim M_K}[\Delta \mid M_K])}{V(\Delta)} \]

- Variance reduction: \( \text{Var}(X) = E(X^2) - (E(X))^2 \)
  \[ V_M(K)(E_{\sim M_K}[\Delta \mid M_K]) = E_M(K)(E_{\sim M_K}[\Delta \mid M_K])^2 - (E_M(K)E_{\sim M_K}[\Delta \mid M_K])^2 \]

- Model averaging
  \[ E_M(K)(X_{M_K}) = \sum_{M_K} X_{M_K} P(M_K) \]  Model probability
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)

\[ G_1 : K \] 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 \( K_2 \)

• Groundwater flow in an unconfined aquifer
• Multispecies reactive transport
Scientific Question:
Which process (recharge or geology) is more important to the state variables (head and concentration)?

<table>
<thead>
<tr>
<th>Si</th>
<th>R1G1</th>
<th>R1G2</th>
<th>R2G1</th>
<th>R2G2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>G</td>
<td>R</td>
<td>G</td>
</tr>
<tr>
<td>Head (x=6,000m)</td>
<td>94.9</td>
<td>4.78</td>
<td>88.7</td>
<td>10.6</td>
</tr>
<tr>
<td>Conc. (x=5,500m)</td>
<td>21.3</td>
<td>78.5</td>
<td>1.48</td>
<td>98.5</td>
</tr>
</tbody>
</table>

Considering model uncertainty:

<table>
<thead>
<tr>
<th>PS_K</th>
<th>Recharge</th>
<th>Geology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head</td>
<td>28.43</td>
<td>67.94</td>
</tr>
<tr>
<td>Conc. (x=5,500m)</td>
<td>0.11</td>
<td>90.99</td>
</tr>
</tbody>
</table>

- 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

\[ y = M(\theta) + r = M(\theta) + b(x, \phi) + \epsilon \]

**Observations**

\[ y \]

**Physical model**

\[ M(\theta) \]

**Error model**

\[ b(x, \phi) \]

**Observation error**

\[ \epsilon \]

---

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