Data-driven discovery and approaches to model evaluation

Louis J. Gross

National Institute for Mathematical and Biological Synthesis

Departments of Ecology and Evolutionary Biology and Mathematics and The Institute for Environmental Modeling

University of Tennessee, Knoxville

Supported by NSF Award DBI-1300426, DOE ORNL#4000110008
• Foster new collaborative efforts to investigate fundamental and applied questions arising in biology using appropriate mathematical and computational methods

• Enhance the essential human capacity to analyze complex biological questions and develop necessary new mathematics

• Encourage broader public appreciation of the unity of science and mathematics.

**Deadlines for support requests: March 1, Sept. 1**

NIMBioS.org
NIMBioS fosters cross-disciplinary research

Node size
number of Working Group participants in a given research area, where the node radius is the log number of participants

Line size
number of collaborations between research areas within Working Groups

Working Groups focus on major scientific questions at the interface between biology and mathematics that require insights from diverse researchers who meet several times over a two-year period.
This workshop will bring together a multi-disciplinary group of molecular and cell biologists, physiologists, ecologists, mathematicians, computational biologists, and statisticians to explore the challenges and opportunities for developing and implementing models specifically designed to mechanistically link between levels of biological organization so as to inform ecological risk assessment and ultimately environmental policy and management. The focus will be on predictive systems models in which properties at higher levels of organization emerge from the dynamics of processes occurring at lower levels of organization.
Synthesis Centers around the world
Overview

• Models and science
• Objectives of models
• Data driven discovery
• Constraints on models
• Model evaluation
• Risk assessment examples
  • RAIS – Risk Assessment Information System
  • SADA - Spatial Analysis and Decision Assistance
• Take-home lessons
Science is thought to be a process of pure reductionism, taking the meaning out of mystery, explaining everything away, concentrating all our attention on measuring things and counting them up. It is not like this at all. The scientific method is guesswork, the making up of stories. The difference between this and other imaginative works of the human mind is that science is then obliged to find out whether the guesses are correct, the stories true. Curiosity drives the enterprise, and the open acknowledgement of ignorance.

Lewis Thomas - Sierra Club Bulletin, March/April 1982, P. 52
What is science?

Science is thought to be a process of pure reductionism, taking the meaning out of mystery, explaining everything away, concentrating all our attention on measuring things and counting them up. It is not like this at all. The scientific method is guesswork, the making up of stories. The difference between this and other imaginative works of the human mind is that science is then obliged to find out whether the guesses are correct, the stories true. Curiosity drives the enterprise, and the open acknowledgement of ignorance.

Lewis Thomas - Sierra Club Bulletin, March/April 1982, P. 52
Science is thought to be a process of pure reductionism, taking the meaning out of mystery, explaining everything away, concentrating all our attention on measuring things and counting them up. It is not like this at all. The scientific method is guesswork, the making up of stories. The difference between this and other imaginative works of the human mind is that science is then obliged to find out whether the guesses are correct, the stories true. Curiosity drives the enterprise, and the open acknowledgement of ignorance.

Lewis Thomas - Sierra Club Bulletin, March/April 1982, P. 52
The “stories” in science are models

A model is a simplification of reality. Think of it as a map - it includes some features that represent what we observe but not others. Modeling is the process of selective ignorance - we select what to include and what to ignore.
The “stories” in science are models

A model is a simplification of reality. Think of it as a map - it includes some features that represent what we observe but not others. Modeling is the process of selective ignorance - we select what to include and what to ignore.
You make models all the time:

What decision do you make when faced with:

The “best” model for you may not be the “best” model for someone else.
Models in Biology

- Physiology
- Neurobiology
- Development
- Microbiology
- Genetics
- Ecosystems
- Disease
Spatial-epidemiology model with vaccination

\[
\frac{dS_i}{dt} = -\beta_i S_i I_i - \gamma \sigma_i S_i + \sum_{j, j \neq i}^n a_{ji} S_j - \sum_{j, j \neq i}^n a_{ij} S_i - \mu_S S_i
\]

\[
\frac{dI_i}{dt} = \beta_i S_i I_i + \sum_{j, j \neq i}^n c_{ji} I_j - \sum_{j, j \neq i}^n c_{ij} I_i - \mu_I I_i
\]

\[
\frac{dR_i}{dt} = \gamma \sigma_i S_i + \sum_{j, j \neq i}^n a_{ji} R_j - \sum_{j, j \neq i}^n a_{ij} R_i - \mu_R R_i
\]

ICs:

\[(4) \quad S_i(0) = S_{i0}, \quad I_i(0) = I_{i0}, \quad R_i(0) = R_{i0}\]

Control (Vaccination):

\[(5) \quad 0 \leq \sigma_i \leq \sigma_{\text{max}} \quad \text{for } i = 1, 2, \ldots, n.\]
Models across multiple scales

$\frac{dQ}{dt} = f(Q,M)$

$\frac{dM}{dt} = g(Q,M)$

$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial t^2}$

$V = \frac{V_{\text{max}}[s]}{k_m + [s]}$

$\Delta p = sp(1-p)$

$\frac{dQ}{dt} = \max[s]$

$V = \frac{k_m + [s]}{s}$
Environmental Modeling

**Data sources**
- GIS map layers (Vegetation, hydrology, elevation), Weather, Roads, Species densities

**Models**
- Statistical
- Differential equations
- Matrix
- Agent-based

**Evaluation/Analysis**
- Visualization, corroboration, sensitivity, uncertainty

**Simulation**
- Matlab, C++, Distributed, Parallel

**Management input**
- Harvest regulation
- Water control
- Reserve design

**Monitoring**
- Species densities
- Animal telemetry
- Physical conditions
Objectives of Models

There are many reasons to use a model aside from prediction:

1. Suggest observations and experiments

2. Provide a framework to assemble bodies of facts/observations - standardize data collection

3. "Allows us to imagine and explore a wider range of worlds than ours, giving new perceptions and questions about how our world came to be as it is" F. Jacob - The Possible and the Actual, 1982

4. Clarifies hypotheses and chains of argument

5. Identifies key components in systems

6. Allow investigation while accounting for societal or ethical constraints
Objectives of Models

7. Allows simultaneous consideration of spatial and temporal change

8. Provides a means to extrapolate or interpolate to situations for which data can not easily be obtained

9. Prompts tentative and testable hypotheses

10. Serves as a guide to decision making in circumstances where action cannot wait for detailed studies or those studies are not feasible

11. Provides an antidote to the helpless feeling that the world is too complex to understand in any generality - provides a means to get at general patterns and trends
But this view is being challenged.
The essence is: Let the Data Tell it’s Own Story – who needs generality!
The stages through which well-managed data passes from project inception to conclusion.

The Data Lifecycle

- Discover
- Describe
- Collect
- Plan
- Integrate
- Analyze
- Assure
- Preserve

From DataONE.org
Cyclical Aspects of Data Models

The stages through which well-managed data passes from project inception to conclusion.

From DataONE.org
If you think Big Data is challenging, what about Big Models!

All of the concerns regarding big data (heterogeneity, data quantity, data quality, curation, metadata characterization) also apply to the complex models applied to provide regulatory guidance.
Science Paradigms

- Thousand years ago: science was empirical, describing natural phenomena
- Last few hundred years: theoretical branch using models, generalizations
- Last few decades: a computational branch simulating complex phenomena
- Today: data exploration (eScience) unifying theory, experiment, and simulation
  - Data captured by instruments or generated by simulator
  - Processed by software
  - Information/knowledge stored in computer
  - Scientist analyzes database/files using data management and statistics

The FOURTH PARADIGM
Data-Intensive Scientific Discovery

Edited by Tony Hey, Stewart Tansley, and Kristin Tolle
Constraints on models

Data constraints: Available data may not be sufficient to specify appropriate functional forms, interrelationships, or parameters. May force aggregation of components. May not be sufficient to elaborate criteria for evaluation of model performance.

Effort constraints: Resource constraints may limit the amount of detail it is feasible to include. Limits time modelers and collaborators may invest as well as pressure to produce results.

Computational constraints: Despite great enhancements in computational resources, there are many problems still not feasible to carry out computationally.

Other constraints: ethical or other societal considerations.
No one model can do everything!
Verification - model behaves as intended, i.e. equations correctly represent assumptions; equations are self-consistent and dimensionally correct. Analysis is correct. Coding is correct - there are no bugs.

Calibration - use of data to determine parameters so the model "agrees" with data. This is specific to a given criteria for accuracy. Some call this Tuning or Curve-fitting.

Corroboration - model is in agreement with a set of data independent from that used to construct and calibrate it.

Validation - model is in agreement with real system it represents with respect to the specific purposes for which it was constructed. Thus there is an implied notion of accuracy and domain of applicability.

Evaluation (testing) - appropriateness to objectives; utility; plausibility; elegance; simplicity; flexibility.
Evaluating different types of models

Models for theory development –

General, some realism, little precision.

Make qualitative comparisons to patterns, not quantitative ones, over some parameter space. No calibration or corroboration performed, except theoretical corroboration (meaning that model agrees with the general body of theory in the field).
Evaluating different types of models

Descriptive models -
Precise, little realism, not general

Statistical hypothesis testing; time series analysis methods applied.

Models for specific systems -
Realism, some precision, not general

Quantitative comparisons, constrained by available data. Compare component-by-component if data are available.
Evaluation rather than “Validation”

The NRC report on *Models in Environmental Regulatory Decision Making* avoided the use of the term validation for several reasons including:

- The confusion over the term in different communities
- The prevalent perception that a “valid” model exists outside of the objective for which it was developed – and these objectives may not be assessed through “domains of applicability”
- The implication that validation is “static” whereas the report recommends a life-cycle for ongoing model evaluation and a plan for carrying this out
Models and evaluation

- Given the many objectives for models, we should expect many diverse criteria for evaluating whether a model is useful.
- Before developing a model in any detail, criteria should be established for evaluating its use.
- Evaluation should account for constraints of Data Availability, Effort and Resources, Computation.
- Include evaluation of alternative approaches based on these constraints to assess most appropriate methods, decide level of detail, scale, and what to ignore (e.g. modeling is a process of “selective ignorance” and the art is in deciding what to include and what to exclude).
Animal models

The most frequent use of the term “model” in connection with biology concerns animal models used as proxies for humans to investigate medical questions. General guidance on evaluating animal models is prevalent throughout the literature (a specific disease model faithfully mimics the human disease, a model system is appropriate for the human system being modeled) but there is little direct methodology to evaluate a particular animal model.
A recent detailed analysis of the use of animal models (Wall and Shani, 2008) argues that “on average, the extrapolated results from studies using tens of millions of animals fail to accurately predict human responses” despite the fact that these studies have been invaluable for investigating general processes and biological pathways. They note that the recommendations on animal models make theoretical sense but often lack practicality.

R. Wall and M. Shani (2008) Are animal models as good as we think? Theriogenology 69: 2–9
Why so little emphasis on evaluation?

1. It’s difficult and requires potentially different skill sets from those constructing and using models.

2. Science is very much a human enterprise and it is natural that once one has devoted considerable effort to developing a particular model, it is difficult to critique yourself.

3. Modern settings with a great amount of team effort to develop models or experimental protocols can constrain individuals who do not wish to be an outcast in a lab.
Take home lessons

- Model evaluation for all types of biological models is relatively rare.
- Set criteria for model evaluation prior to expending a lot of effort on a model.
- Tie evaluation criteria to model objectives.
- Encourage consideration of evaluation in all your educational initiatives.
- Multiple models are good – encourage this.
- Consider whether an evaluation has been done or discussed whenever you review a paper or grant proposal.
Risk assessment is a process for characterizing the nature and magnitude of health risks to humans (e.g., residents, workers, recreational visitors) and ecological receptors (e.g., bird, fish, wildlife) from chemical contaminants and other stressors that may be present in the environment. Risk assessment involves four major steps:

1. **Hazard Identification**—an examination of whether a stressor has the potential to cause harm to humans or ecological systems
2. **Dose-Response Assessment**—an examination of the numerical relationship between exposure and effects
3. **Exposure Assessment**—an examination of what is known about the frequency, timing, and levels of contact with a stressor
4. **Risk Characterization**—an examination of what is known about the frequency, timing, and levels of contact with a stressor
Risk Assessment Information System (RAIS) - Public access website for all things concerning environmental risk assessment: toxicity values, chemical parameters, PRG calculation, risk calculation, ARARs, ecological benchmarks.

Fred Dolislager
Leslie Galloway
Debra Stewart
The University of Tennessee
The Institute for Environmental Modeling
• ORNL developed the first instance with sponsorship from DOE in 1996.

• RAIS consolidated and unified risk assessment procedures and practices by hosting documents, equations and databases on the Oak Ridge Reservation.

• Developed our tools through IAG with EPA to provide national guidance.

• Gave the tools utility for international use by allowing users to modify all of our parameters.
Welcome to
The Risk Assessment Information System

About the RAIS

This work has been sponsored by the U.S. Department of Energy (DOE), Office of Environmental Management, Oak Ridge Operations (ORO) Office through a contract between URS | CH2M Oak Ridge LLC (UCOR) and the University of Tennessee.

These web pages are under configuration management and are subject to quality assurance review before being published.
# RAIS Usage

<table>
<thead>
<tr>
<th>Month</th>
<th>Unique visitors</th>
<th>Number of visits</th>
<th>Pages</th>
<th>Hits</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 2014</td>
<td>5,365</td>
<td>8,652</td>
<td>59,963</td>
<td>225,362</td>
<td>29.08 GB</td>
</tr>
<tr>
<td>Feb 2014</td>
<td>5,823</td>
<td>9,151</td>
<td>56,396</td>
<td>230,608</td>
<td>28.56 GB</td>
</tr>
<tr>
<td>Mar 2014</td>
<td>6,575</td>
<td>10,304</td>
<td>59,400</td>
<td>254,756</td>
<td>31.69 GB</td>
</tr>
<tr>
<td>Apr 2014</td>
<td>5,925</td>
<td>9,476</td>
<td>63,577</td>
<td>236,714</td>
<td>31.81 GB</td>
</tr>
<tr>
<td>May 2014</td>
<td>5,266</td>
<td>8,886</td>
<td>66,772</td>
<td>229,446</td>
<td>33.83 GB</td>
</tr>
<tr>
<td>Jun 2014</td>
<td>4,811</td>
<td>8,433</td>
<td>48,588</td>
<td>197,129</td>
<td>28.53 GB</td>
</tr>
<tr>
<td>Jul 2014</td>
<td>4,319</td>
<td>8,077</td>
<td>48,542</td>
<td>186,123</td>
<td>26.75 GB</td>
</tr>
<tr>
<td>Aug 2014</td>
<td>4,276</td>
<td>7,698</td>
<td>45,936</td>
<td>173,105</td>
<td>26.13 GB</td>
</tr>
<tr>
<td>Sep 2014</td>
<td>5,190</td>
<td>9,579</td>
<td>42,845</td>
<td>183,718</td>
<td>30.46 GB</td>
</tr>
<tr>
<td>Oct 2014</td>
<td>5,753</td>
<td>10,462</td>
<td>54,556</td>
<td>203,850</td>
<td>33.99 GB</td>
</tr>
<tr>
<td>Nov 2014</td>
<td>5,671</td>
<td>9,280</td>
<td>49,101</td>
<td>209,911</td>
<td>32.63 GB</td>
</tr>
<tr>
<td>Dec 2014</td>
<td>5,039</td>
<td>8,537</td>
<td>47,103</td>
<td>175,697</td>
<td>30.71 GB</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>64,013</strong></td>
<td><strong>108,535</strong></td>
<td><strong>642,779</strong></td>
<td><strong>2,506,419</strong></td>
<td><strong>364.17 GB</strong></td>
</tr>
</tbody>
</table>
RAIS Tool Bar Highlights

- Toxicity Profiles (70 Analytes)
- Toxicity Values (800 Chemicals and 1368 Radionuclides)
- Toxicity Metadata (37 Pieces of Study Information)
- Chemical Factors (22 variables)
- PRGs (5 land uses) (5 media)
- Risk models (5 land uses)
- Ecological Benchmarks (4 media) (82 benchmarks)
- ARAR (Federal + 13 US States)
- Soil to Groundwater
- Background Values for metals
- Radionuclide decay chain generator
- Air and soil transport models
Risk Assessment Steps: Data Evaluation

RAIS Preliminary Remediation Goals (PRGs) Calculator

Fields that are highlighted are required.

Select Scenario
- Resident
- Indoor Worker
- Outdoor Worker
- Recreator
- Excavation Worker
- Farmer

Select PRG type
- Defaults
- Site Specific

RAIS Soil Screening Levels Calculator

\[ \text{SSL} (\text{mg/kg}) = C_w \left( \frac{\text{mg}}{\text{L}} \right) \times \left[ K_d \left( \frac{\text{L}}{\text{kg}} \right) + \left( \theta_w \left( \frac{\text{L}_{\text{water}}}{\text{L}_{\text{soil}}} \right) + \theta_a \left( \frac{\text{L}_{\text{air}}}{\text{L}_{\text{soil}}} \right) \times H' \right) \right] \]

where:

\[ \theta_a \left( \frac{\text{L}_{\text{air}}}{\text{L}_{\text{soil}}} \right) = \left( \frac{\text{L}_{\text{water}}}{\text{L}_{\text{soil}}} \right) - \theta_w \left( \frac{0.3 \text{L}_{\text{water}}}{\text{L}_{\text{soil}}} \right) \]

\[ n \left( \frac{\text{L}_{\text{pore}}}{\text{L}_{\text{soil}}} \right) = 1 - \left( \frac{\rho_b (1.5 \text{L})}{\rho_s (2.65 \text{L})} \right) \]

\[ K_d \left( \frac{\text{L}}{\text{kg}} \right) = K_{oc} \left( \frac{\text{L}}{\text{kg}} \right) \times f_{oc} (0.002 \text{ unitless}) \]

Inorganic Soil Background Selection

Results

All units are ppm except where noted

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Soil-type</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aluminum</td>
<td>Soils over granites and gneisses</td>
<td>7.2-8.2%</td>
</tr>
<tr>
<td>Aluminum</td>
<td>Sandy and lithosols on sandstones</td>
<td>2.5-4.3%</td>
</tr>
<tr>
<td>Aluminum</td>
<td>Various soils</td>
<td>0.45-10%</td>
</tr>
<tr>
<td>Aluminum</td>
<td>Soils over limestones and calcareous rocks</td>
<td>0.43-1.3%</td>
</tr>
<tr>
<td>Aluminum</td>
<td>Soils over volcanic rocks (or ash*)</td>
<td>6.9-8.1%</td>
</tr>
</tbody>
</table>
Risk Assessment Steps: Exposure Assessment

RAIS Preliminary Remediation Goals (PRGs) Calculator

Soil Non-Carcinogenic Ingestion Equation

\[
\text{PRG}_{\text{res-soil-nc-ing}} (\text{mg/kg}) = \left( \frac{\text{THQ} \times \text{AT}_{\text{ress}} (\frac{365 \text{ days}}{\text{year}}) \times \text{ED}_{\text{ress}} (26 \text{ years})}{1 \times \text{RfD} \left( \frac{\text{mg}}{(\text{kg-day})} \right) \times \text{IFS}_{\text{res-adj}} \left( \frac{36,750 \text{ mg}}{\text{kg}} \right) \times 10^{-6} \text{ kg}} \right)
\]

where:

\[
\text{IFS}_{\text{res-adj}} \left( \frac{36,750 \text{ mg}}{\text{kg}} \right) = \left( \frac{\text{ED}_{\text{ressc}} (6 \text{ years}) \times \text{EF}_{\text{ressc}} (\frac{350 \text{ days}}{\text{year}}) \times \text{IRS}_{\text{ressc}} (200 \text{ mg})}{\text{BW}_{\text{ressc}} (15 \text{ kg})} \right) + \left( \frac{\text{ED}_{\text{ressc}} (26 \text{ years}) - \text{ED}_{\text{ressc}} (6 \text{ years}) \times \text{EF}_{\text{ressc}} (\frac{350 \text{ days}}{\text{year}}) \times \text{IRS}_{\text{ressc}} (100 \text{ mg})}{\text{BW}_{\text{ressc}} (80 \text{ kg})} \right)
\]

Risk Assessment Steps: Exposure Assessment

<table>
<thead>
<tr>
<th>Age Segment (yr)</th>
<th>AF (mg/cm²)</th>
<th>BW (kg)</th>
<th>ED (yr)</th>
<th>EF (day/yr)</th>
<th>ET (hr/event)</th>
<th>IRS (mg/day)</th>
<th>SA (cm²/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-2</td>
<td>0.2</td>
<td>15</td>
<td>2</td>
<td>350</td>
<td>24</td>
<td>200</td>
<td>2690</td>
</tr>
<tr>
<td>2-6</td>
<td>0.2</td>
<td>15</td>
<td>4</td>
<td>350</td>
<td>24</td>
<td>200</td>
<td>2690</td>
</tr>
<tr>
<td>6-16</td>
<td>0.07</td>
<td>80</td>
<td>10</td>
<td>350</td>
<td>24</td>
<td>100</td>
<td>6032</td>
</tr>
<tr>
<td>16-30</td>
<td>0.07</td>
<td>80</td>
<td>10</td>
<td>350</td>
<td>24</td>
<td>100</td>
<td>6032</td>
</tr>
<tr>
<td>Child (0-6)</td>
<td>0.2</td>
<td>15</td>
<td>6</td>
<td>350</td>
<td>24</td>
<td>200</td>
<td>2690</td>
</tr>
<tr>
<td>Adult (6-70)</td>
<td>0.07</td>
<td>80</td>
<td>20</td>
<td>350</td>
<td>24</td>
<td>100</td>
<td>6032</td>
</tr>
</tbody>
</table>
Risk Assessment Steps: Toxicity Assessment

![RAIS](https://example.com) - The Risk Assessment Information System

<table>
<thead>
<tr>
<th>Chemical</th>
<th>CAS</th>
<th>Chronic Oral Reference Dose (mg/kg-day)</th>
<th>RFDOCREF</th>
<th>Chronic Inhalation Reference Concentration (mg/m³)</th>
<th>RFCICREF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benzene</td>
<td>000071-43-2</td>
<td>4.00E-03</td>
<td>IRIS</td>
<td>3.00E-02</td>
<td>IRIS</td>
</tr>
<tr>
<td>Benzidine</td>
<td>000092-87-5</td>
<td>3.00E-03</td>
<td>IRIS</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Risk Assessment Steps: Risk Characterization

The image shows a webpage with a section titled "Resident Equation Inputs for Ambient Air." It includes a table with variables and their values for exposure time, frequency, duration, and risk. The table is used in the calculation of Ambient Air Risk. The webpage also includes links to various tools and tutorials such as Toxicity Profiles, Ecological Benchmarks, Soil to Groundwater Background Values, ARAR Search, and Gamma Radiation Instrument Response Tool.

The page displays a sidebar with options for Select Chemicals Info Type and Select Individual Chemicals. The selected chemicals include ALAR, Acenaphthene, Acenaphthylene, Acephate, Acetaldehyde, Acetate, Acetone, Acetone Cyanohydrin, and Acetonitrile.
RAIS Working with SADA

- RAIS Toxicity database shared with SADA
- RAIS Chemical parameters database shared with SADA
- RAIS (and EPA) Risk and PRG models shared with SADA.
Spatial Analysis and Decision Assistance
SADA

Robert Stewart
Fred Dolislager
Tom Purucker

The Institute for Environmental Modeling, University of Tennessee
Geographic Information Science and Technology Group, Oak Ridge National Laboratory
SADA project engages research and development at the nexus of geospatial analytics, risk assessment, and decision analysis.

Goals are to embed risk assessment, uncertainty modeling, and downstream decision processes entirely within a spatial context.

Two lanes define project activities
- Advancing methods in a variety of areas particularly well connected to environmental regulatory community, characterization, remediation, RCRA, Superfund, MARSSIM, etc.
- Freeware desktop application (SADA) integrating environmental risk analytics, spatial modeling, and decision sciences.
Questions That SADA Addresses

- What exposure scenarios are likely dangerous?
- What contaminants are driving the risk?
- What pathways (ingestion, inhalation, etc)?
- What is the risk or concentration limit for an exposure time of 30 years?, 1 day? 1 hour?
- Where is exposure unsafe? Who might be in harms way? How sure are we?
- Where should we apply risk mitigation measures?
- Where and what type of additional information would support the model?
- What are our decision risks?
Answers that SADA V5 Provides

Initial Sample Designs
- Judgmental
- Random
- Simple Grid
- Standard Grid
- Unaligned Grids
- Search Grids
- MARSSIM
- 3D Search

Cost Benefit Analytics
Built on risk-space models
Permit what if’s
Quantify cost and decision risk reduction

Secondary Sample Designs
Sample where model needs most support....
Risk Based Decisions Over Time

Day 1

Day 2

Day 3
But that’s Not All

- Because SADA is an open spatiotemporal modeling environment, it can be used for numerous applications outside of toxicological and radiological risk.

- Examples include, engineering, geophysical, geological, ecosystems monitoring, epidemiological
Where

- In the 17th year of deployment (began ~1998)
- 18,000+ registered users
- 90+ scientific and regulatory communications (e.g. journal articles, reports, web pages, theses, etc.)
- User group, workshops, conferences, international presence etc.

www.sadaproject.net
Who

Mostly Government Subcontractors

- Private Company, 21%
- Faculty, 21%
- Students, 17%
- Outside US, 16%
- Inside US, 6%
- Federal Government, 6%
- State Government, 4%
- Other, 4%
- Local Government, 2%
- Citizen, 2%

Other 47%

- EPA 16%
- EU 12%
- DOE 8%
- DOD 6%
- DOH 1%
- DOEd 2%
- DOI 2%
- NRC 4%
What
### SADA Contributors/Sponsors/Advisors

<table>
<thead>
<tr>
<th>Contributors/Sponsors/Advisors</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States Nuclear Regulatory Commission</td>
</tr>
<tr>
<td>United States Environmental Protection Agency</td>
</tr>
<tr>
<td>United States Department of Energy</td>
</tr>
<tr>
<td>Environmental Measurements Laboratory</td>
</tr>
<tr>
<td>Pacific Northwest National Laboratory</td>
</tr>
<tr>
<td>Oak Ridge National Laboratory</td>
</tr>
<tr>
<td>Michigan State University</td>
</tr>
<tr>
<td>San Diego State University</td>
</tr>
<tr>
<td>Stanford (GSLIB codes)</td>
</tr>
<tr>
<td>Swedish Geotechnical Institute</td>
</tr>
<tr>
<td>Numerous subcontractors and Private Sector Advising</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contributors/Sponsors/Advisors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Michelle Burgess (USEPA, Superfund)</td>
</tr>
<tr>
<td>John Bing-Canar (USEPA, Region 5)</td>
</tr>
<tr>
<td>Brian Cooper (USEPA, Region 5)</td>
</tr>
<tr>
<td>Tim Drexler (USEPA, Region 5)</td>
</tr>
<tr>
<td>Dr. Brenda Jones (USEPA, ERAF)</td>
</tr>
<tr>
<td>Leslie Galloway (TIEM/University of Tennessee)</td>
</tr>
<tr>
<td>Dr. Bruce Ralston (Geography/UT)</td>
</tr>
<tr>
<td>John Wilson (PNNL)</td>
</tr>
<tr>
<td>Rich Pawlowitz (Bechtel-Jacobs)</td>
</tr>
<tr>
<td>Dr. George Christakos (San Diego State University)</td>
</tr>
<tr>
<td>Dr. Hwa-Lung Yu (San Diego State University)</td>
</tr>
<tr>
<td>Jim Davidson (PNNL/Davidson &amp; Davidson)</td>
</tr>
<tr>
<td>Melinda Vzanna (student/University of Tennessee)</td>
</tr>
<tr>
<td>Rob Fletcher (student/University of Tennessee)</td>
</tr>
<tr>
<td>Peter Starzec (Swedish Geotechnical Institute)</td>
</tr>
<tr>
<td>Jerry Montgomery (US Army Corp Of Engineers)</td>
</tr>
<tr>
<td>Jenny Norman (Swedish Geotechnical Institute)</td>
</tr>
<tr>
<td>Tom Nicholson (USNRC)</td>
</tr>
<tr>
<td>Bobby Abueid (USNRC)</td>
</tr>
<tr>
<td>Jim Wulf (Tetratech)</td>
</tr>
<tr>
<td>Carl Stineman (Consultant)</td>
</tr>
<tr>
<td>Martin Bittens (University of Tuebingen, DE)</td>
</tr>
</tbody>
</table>

**Project Manager:** Dr. Robert Stewart, stewartrn@ornl.gov
Lessons - Doing the Modeling:

Work closely with those with long experience in the system being modeled.

Moderate the above based first on the availability of data to construct reasonable models, and secondly on the difficulty of constructing and calibrating the models.

Don't try to do it all at once - start small - but have a long-term plan for what you wish to include overall, given time and funding.
Lessons - Doing the Modeling:

Leave room for multiple approaches: don't limit your options.

In the face of limited or inappropriate data, use this as an opportunity to encourage further empirical investigations of key components of the system.

Build flexibility in as much as possible.

Be flexible about what counts as success.
Lessons - Personnel Matters:

Build a quality team who respect each other's abilities and won't second guess each other, but who accept criticism in a collegial manner.

Keep some part of the team out of the day-to-day political fray.

Be persistent, and have at least one member of the team who is totally dedicated to the project and willing to stake their future on it.

Do whatever you can to maintain continuity in the source of long-term support for the project.
Lessons - Interacting with Stakeholders

Constantly communicate with stakeholders.

Regularly explain the objectives of your modeling effort, as well as the limitations, to stakeholders. Be prepared to do this over and over for the same people, and do not get frustrated when they forget what you are doing and why.

Be prepared to regularly defend the scientific validity of your approach.
Lessons - Interacting with Stakeholders:

Don't limit your approach because one stakeholder/funding agency wants you to.

Be prepared for criticism based upon non-scientific criteria, including personal attacks.

Ignore any of the stakeholders at your peril.