A New Framework for Adaptive Sampling and Analysis During Long-Term Monitoring and Remedial Action Management (Project 87023)

Barbara S. Minsker, Principal Investigator

Department of Civil and Environmental Engineering
University of Illinois
Urbana, Illinois

Acknowledgments

- This project is a collaboration with:
  - Argonne National Laboratory-East, Argonne, IL (Robert Johnson and John Quinn)
  - UIUC Department of Civil and Environmental Engineering (Albert Valocchi)
  - UIUC Department of Statistics (Barbara Bailey)
  - UIUC National Center for Supercomputing Applications, Automated Learning Group (David Tcheng)

- William Michael (UIUC MS 2002) prepared the initial results shown, with funding from:
  - U.S. Department of Energy - Argonne National Laboratory (ANL) under Grant No. IF-01628
  - National Center for Supercomputing Applications (NCSA)
  - Walter E. Deuchler Fellowship

Outline

- Introduction and motivation
- Overview of project objectives and tasks
- Preliminary results

Motivation

- Long-term monitoring (LTM) of remedial actions is becoming increasingly important
  - More sites moving toward closure or completion of active cleanup
  - Changing regulatory environment allowing more contaminants to remain in place
  - LTM costs will be significant for foreseeable future
    - In 2001 report to Congress, DOE estimated it will spend:
      - $5.5 billion on long-term site management ("stewardship") between 2000 and 2006
      - More than $100 million/year over next 70 years

Motivation, Cont'd.

- To reduce the burden of LTM, DOE and other Federal agencies are making significant investments in
  - New field analytical techniques
  - Nonintrusive technologies
  - Sensor technologies
- These technologies will shift the way LTM is performed
  - Most likely, suites of *in situ* sensors will trigger more limited physical sampling
- To design and use these types of LTM networks effectively, a new adaptive paradigm for sampling and analysis is needed

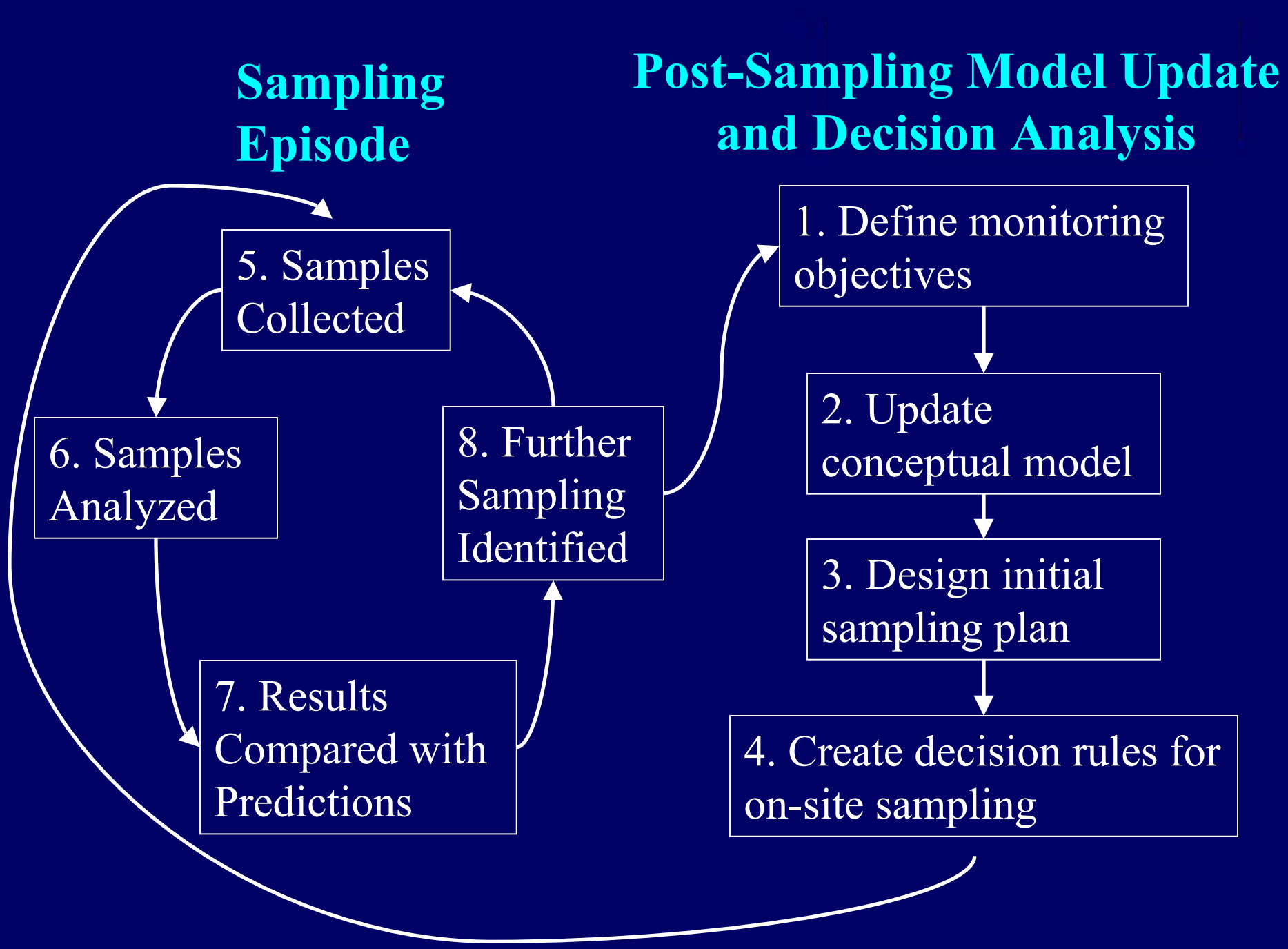
Overview of Project Objectives and Tasks

Project Objectives

- Enable effective interpretation of non-intrusive monitoring data
- Improve predictions and assessment of remediation performance
- Develop decision rules for on-site adaptive sampling and analysis
- Enable more informed decision making and risk analysis of LTM systems

Overview of Research Approach

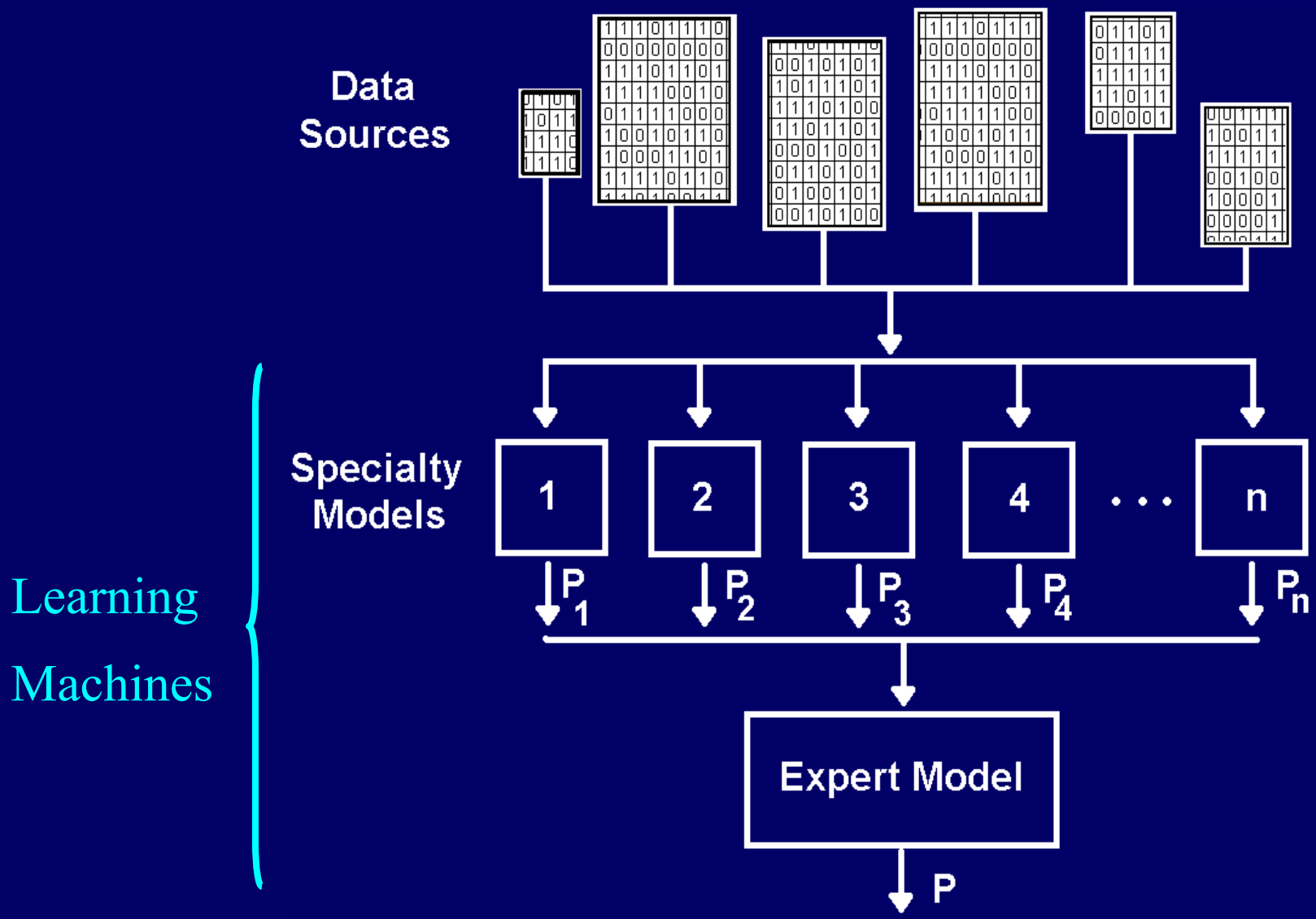
- Develop a new framework for adaptive sampling and analysis, decision making, and risk assessment
- Framework will use "living" models that
  - Readily analyze and incorporate new data
  - Adaptively identify further data needs


Sampling Episode
Post-Sampling Model Update and Decision Analysis
5. Samples Collected
6. Samples Analyzed
7. Results Compared with Predictions
8. Further Sampling Identified
1. Define monitoring objectives
2. Update conceptual model
3. Design initial sampling plan
4. Create decision rules for on-site sampling

Tasks

1) Develop synthetic datasets for developing and testing the framework
2) Create hierarchical models that integrate available process knowledge at all scales, combining
   - Data-driven models (neural networks, decision trees, etc.)
   - A suite of physically-based models (analytical or simple numerical models)

General Hierarchical Framework



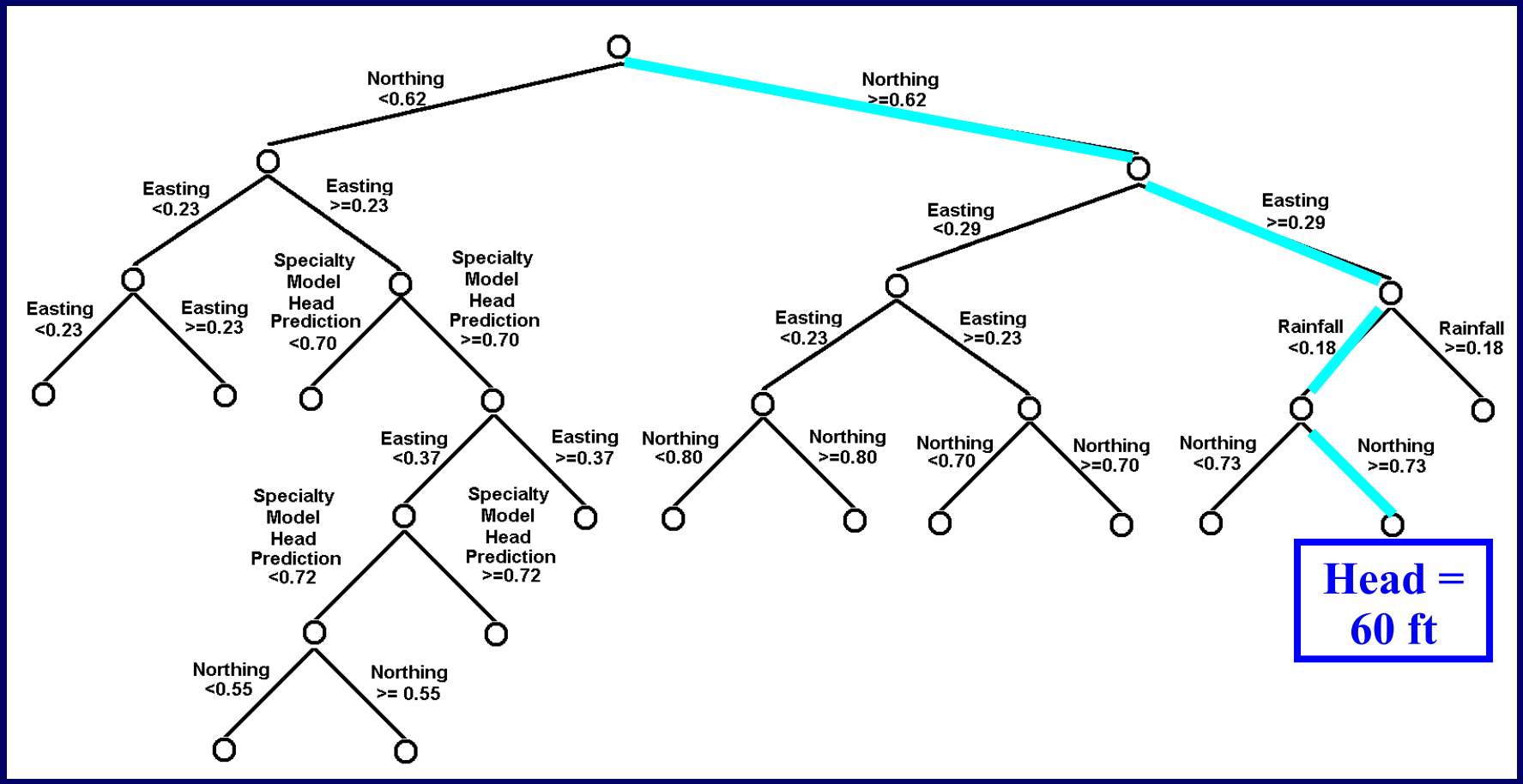
Tasks, Cont'd.

3) Develop methods for updating the conceptual models
   - Create Bayesian updating approaches for hierarchical models using Markov chain Monte Carlo methods
   - Create an interactive, automated model and parameter identification system
     - To initially identify appropriate hierarchical models or make major model updates
     - Using interactive genetic algorithms, in which the user evaluates candidate solutions and guides the model search

Tasks, Cont'd.

4) Identify optimal initial sampling plans
   - Use interactive, multiobjective genetic algorithms to decide where, when, and what to sample initially
5) Develop decision rules for on-site adaptive sampling and analysis
   - After the initial samples are collected, analyze the data on-site to decide what additional data are needed
   - Compare 2 approaches:
     - Simple scenario-based approaches developed using professional judgment
     - Automated decision tree approaches

Example Decision Tree (Simplified for Viewing)



Uses historical data to predict hydraulic heads at different locations under different conditions. If new data show substantial deviation, collect more data.

Tasks, Cont'd.

5) Evaluate long-term monitoring and management of pump-and-treat containment and phytoremediation at the Argonne 317/319 Site

- Test best approaches from previous tasks on historical data

Preliminary Results

Exploratory Research

Exploratory Research on Knowledge Integration

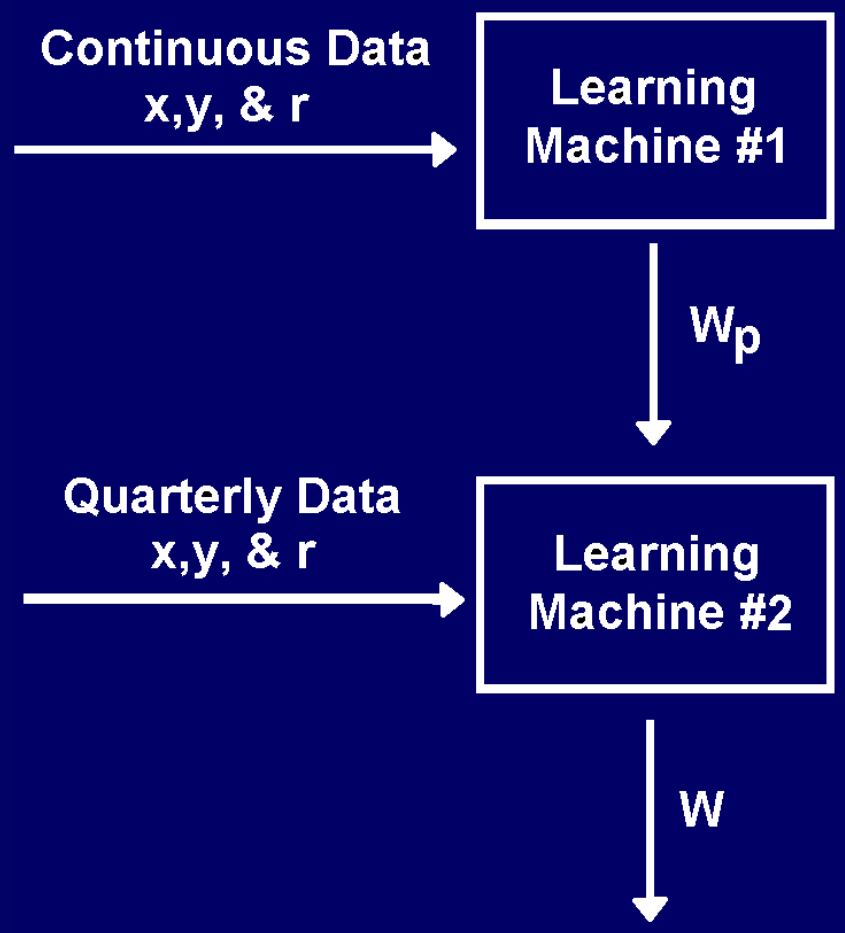
- Argonne National Laboratory East 317/319 Area
  - VOC and tritium contamination
  - Phytoremediation and pump & treat systems installed
  - Site wants to reduce monitoring costs but ensure remediation goals are met
- Initial study only focusing on hydraulic head data to ensure containment
  - Objective: Demonstrate that data-driven models can integrate diverse data and models to improve predictions of hydraulic heads

Data Sources

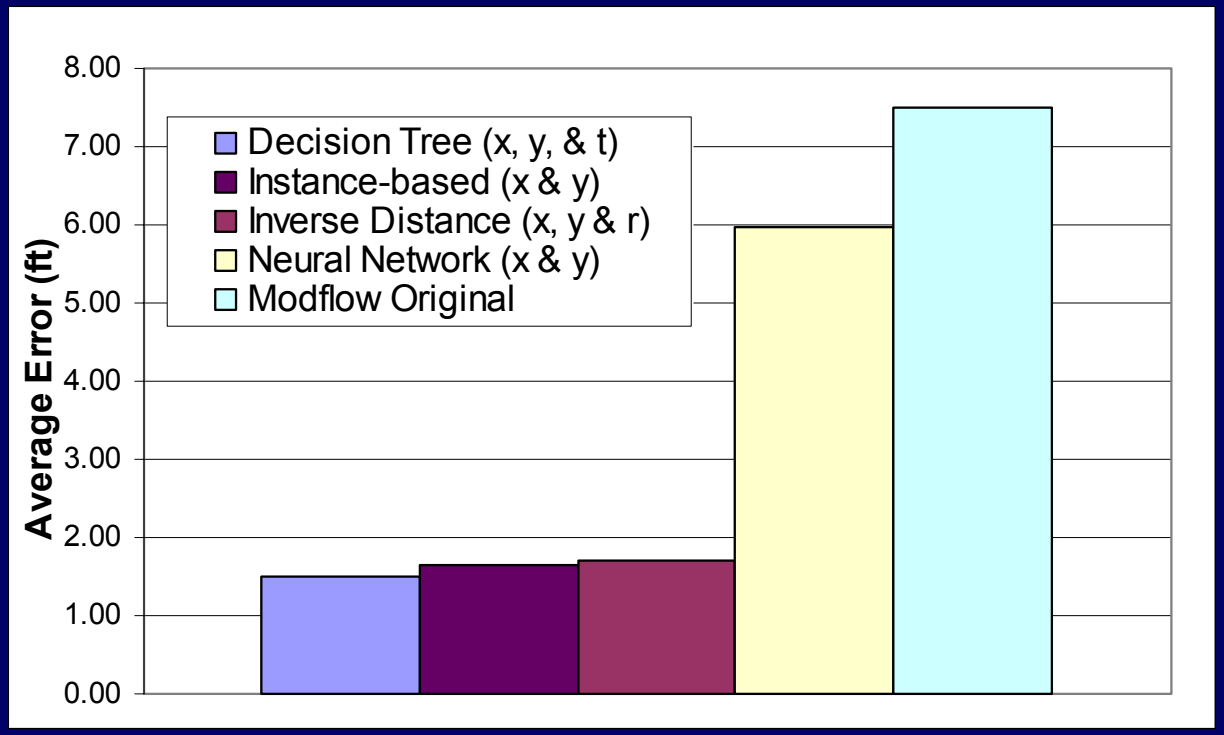
- Predicting head levels at 22 routinely-measured wells
  - May 2001
- Multiple data sources used to make predictions
  - 147 historical quarterly hydraulic head measurements
  - 5,047 hourly hydraulic head readings ("continuous" data)
  - Total precipitation just prior to head measurements
  - Modflow model

Experiment 1: Data-Driven Hierarchical Model

- Low-level "specialty" model
  - Predicts heads from continuous data alone
- High-level "expert" model
  - Learns which data are most accurate when and where



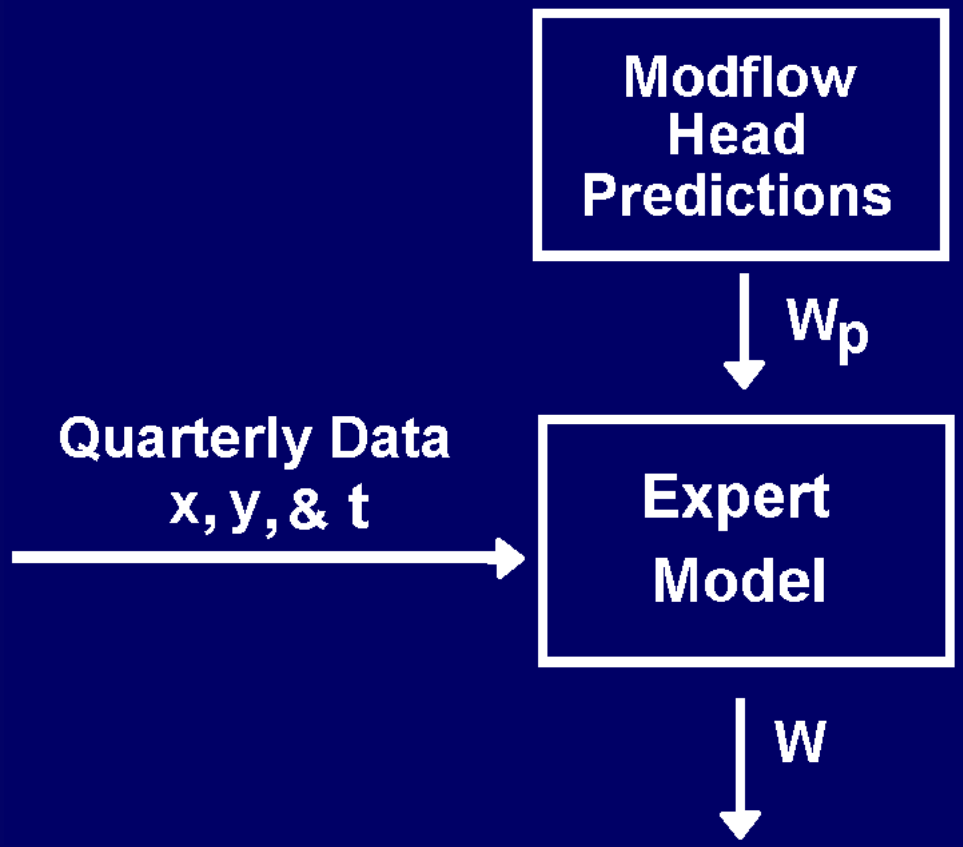
Comparison of Learning Machine Methods



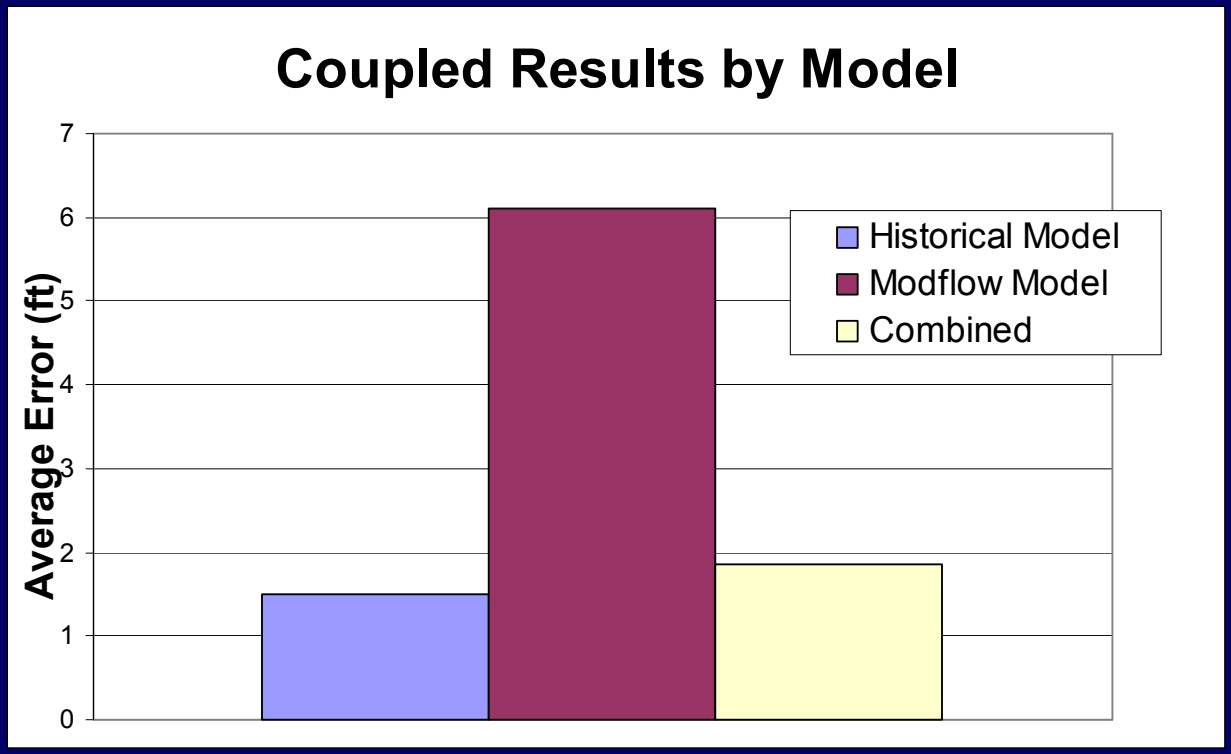
- Many combinations of methods and data sources were explored
- Decision trees yield the best predictions
- Historical quarterly data was most helpful
- Continuous data were not helpful for prediction

Experiment 2: Coupled Hierarchical Model

- Low-level "specialty" model
  - Modflow
- High-level "expert" model
  - Learns which data are most accurate when and where



Coupled Comparison



- Error in the Modflow model is greatly reduced

Conclusions

- New paradigms are needed for effectively managing long-term restoration sites
- Hierarchical data-driven models hold substantial promise for effectively integrating data sources and models, even at multiple scales
- Combining data driven and numerical models (like Modflow) can
  - Improve accuracy of numerical models in areas with substantial data
  - Improve spatial coverage of data-driven models
  - Allow off-the-shelf, historical numerical models to be updated automatically (creating "living" models)