Machine Learning, Uncertainty Quantification, and Decision Analyses: LANL Groundwater Contaminant Remediation Projects

Velimir V Vesselinov ("monty") vvV@lanl.gov
Daniel O’Malley, Sachin Pandey, Scott Hansen, Satish Karra, Boain Alexandrov, Zhiming Lu, Terry Miller, Danny Katzman

P&RA CoP Annual Technical Exchange Meeting, October 17-19, 2017
Chromium site

Model predicted plume shape (~2012)
Cr⁶⁺ MCL 50 ppb

1000 ppb
50 ppb

Mortandad Canyon
Sandia Canyon

Single-screen aquifer monitoring wells
Two-screen aquifer monitoring wells

Vadose zone (~300 m)
Chromium site conceptual model (3D)

- Chromium (Cr⁶⁺) of approximately 54,000 kg
- Los Alamos Canyon
- Sandia Canyon
- Mortandad Canyon
- Perched Groundwater
- Regional Aquifer
- Vadose zone (300 m)
- LANL boundary
- Water supply well PM-3

Chromium Models

Machine Learning

Decisions

Summary
Chromium site conceptual model (3D)

- Chromium
- Models
- Machine Learning
- Decisions
- Summary
Chromium site conceptual model (3D)
Chromium site

- High visibility (DOE/LANL/NMED/San Ildefonso/New Mexico)
- ~54,000 kg of Cr$^{6+}$ released in Sandia Canyon between 1956 and 1972 (uncertainties)
- Cr$^{6+}$ detected above MCL (50 ppb; NM standard) in several monitoring wells in the regional aquifer
- Cr$^{6+}$ plume size is about 2 km$^2$ (region above MCL)
- Cr$^{6+}$ plume is located near LANL site boundary
- Series of water-supply wells are located nearby (less than km)
- Limited remedial options due to aquifer depth (~300 m below the ground surface) and complexities in the subsurface processes
- Current conceptual model for chromium migration in the subsurface is supported by multiple lines of evidence
- So far all the “blind” model predictions have been generally consistent with all the new observations
Chromium project goals

- **GOAL #1**: apply modeling to support conceptualization of the site geologic, hydrologic and biogeochemical conditions
- **GOAL #2**: data- and model-based decision analyses for chromium remediation taking into account existing processes and uncertainties/unknowns
- **Processes**: flow, advective transport, dispersion, diffusion, biogeochemical reactions in the vadose zone and the regional aquifer
- **Uncertainties and unknowns**: ...
- **Remedial options**:
  - Natural attenuation (NA)
  - Enhanced attenuation (EA; biogeochemical processes)
  - Active remediation including mass removal in the vadose zone and the aquifer (pump-and-treat, etc.)
  - Combination of all above at different stages or at different locations
LANL site models

Chromium models

Machine Learning decisions

Summary
Chromium model: Predicted plume transients

Year 2005.10

Chromium
Models
Machine Learning
Decisions
Summary
Chromium model: Pumping drawdowns
Chromium model: pumping/extraction test predictions (Test 1)
Chromium model: pumping/extraction test predictions (Test 2)
CHROTRAN represents full 3D dynamics of processes involved in situ biochemical remediation of heavy metals in heterogeneous aquifers. CHROTRAN accounts for spatial and temporal transients of heavy metal to be remediated, introduced amendments (e.g., through well injection), bio-mass (bio-film) growth and decay (indirect Monod kinetics), direct abiotic reduction by donor-metal interaction, donor-driven biomass growth and bio-reduction, bio-fouling / bio-clogging / bio-mass crowding, and multiple donor consumption pathways.

The introduced amendments can include an electron donor (e.g., molasses), a nontoxic conservative bio-inhibitor (e.g., ethanol), and a biocide (e.g., dithionite).

http://chrotran.lanl.gov
Chrotran: model of well bio clogging
Chrotran: model of chromium bio-remediation using dithionite
Particle-based modeling of contaminant reduction: $A + B = C$

- Chromium
- Models
- Machine Learning
- Decisions
- Summary
Particle-based modeling of contaminant reduction: A + B = C
Particle-based modeling of contaminant reduction: $A + B = C$
• Blind Source Separation: a machine-learning method for source identification without a model

Source 1

Source 2

Source 3

Mixtures

BSS

Estimate 1

Estimate 2

Estimate 3

Chromium

Models

Machine Learning

Decisions

Summary
Let us assume there are 4 buckets representing 4 different groundwater types / contaminant sources ...
The water from the 4 buckets is mixed in unknown fashion in the aquifer ...
Machine Learning:

... we only know the groundwater mixtures observed in the monitoring wells
Machine Learning:

... based on the observed groundwater mixtures, we want to define the unknown water composition of the buckets (sources)
<table>
<thead>
<tr>
<th>Well</th>
<th>$Cr^{6+}$</th>
<th>$ClO_4^-$</th>
<th>$SO_4^{2-}$</th>
<th>$NO_3^-$</th>
<th>$Cl^-$</th>
<th>$^3H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pz-1</td>
<td>406.22</td>
<td>1.84</td>
<td>47.846</td>
<td>17.07</td>
<td>35.401</td>
<td>101.397</td>
</tr>
<tr>
<td>Pz-2a</td>
<td>83.89</td>
<td>0.88</td>
<td>71.155</td>
<td>14.42</td>
<td>66.436</td>
<td>121.013</td>
</tr>
<tr>
<td>Pz-2b</td>
<td>35.01</td>
<td>0.419</td>
<td>6.2918</td>
<td>4.24</td>
<td>7.582</td>
<td>2.061</td>
</tr>
<tr>
<td>Pz-3</td>
<td>338.88</td>
<td>1.21</td>
<td>33.967</td>
<td>23.60</td>
<td>21.853</td>
<td>24.184</td>
</tr>
<tr>
<td>Pz-4</td>
<td>5.69</td>
<td>63.7</td>
<td>5.8175</td>
<td>17.90</td>
<td>3.0975</td>
<td>11.346</td>
</tr>
<tr>
<td>Pz-5</td>
<td>89.26</td>
<td>0.44</td>
<td>8.7896</td>
<td>4.98</td>
<td>7.8321</td>
<td>11.807</td>
</tr>
<tr>
<td>R-1</td>
<td>5.68</td>
<td>0.351</td>
<td>2.19</td>
<td>2.26</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>R-11</td>
<td>20.8</td>
<td>0.83</td>
<td>13.1</td>
<td>20.60</td>
<td>5.15</td>
<td>4.9</td>
</tr>
<tr>
<td>R-13</td>
<td>3.81</td>
<td>0.4</td>
<td>3.12</td>
<td>3.22</td>
<td>2.49</td>
<td>0.2</td>
</tr>
<tr>
<td>R-15</td>
<td>12.5</td>
<td>8.93</td>
<td>6.22</td>
<td>7.97</td>
<td>3.99</td>
<td>29</td>
</tr>
<tr>
<td>R-28</td>
<td>407</td>
<td>1.0</td>
<td>55.1</td>
<td>4.91</td>
<td>38.5</td>
<td>211</td>
</tr>
<tr>
<td>R-33#1</td>
<td>4.89</td>
<td>0.398</td>
<td>3.32</td>
<td>2.41</td>
<td>2.29</td>
<td>2</td>
</tr>
<tr>
<td>R-33#2</td>
<td>5.52</td>
<td>0.35</td>
<td>2.3</td>
<td>1.64</td>
<td>2.0</td>
<td>1.2</td>
</tr>
<tr>
<td>R-34</td>
<td>4.26</td>
<td>0.333</td>
<td>2.66</td>
<td>2.76</td>
<td>2.42</td>
<td>1.2</td>
</tr>
<tr>
<td>R-35a</td>
<td>4.3</td>
<td>0.422</td>
<td>5.62</td>
<td>2.10</td>
<td>6.74</td>
<td>0.6</td>
</tr>
<tr>
<td>R-35b</td>
<td>6.98</td>
<td>0.579</td>
<td>3.48</td>
<td>4.84</td>
<td>2.88</td>
<td>1.3</td>
</tr>
<tr>
<td>R-36</td>
<td>5.29</td>
<td>1.55</td>
<td>7.35</td>
<td>8.69</td>
<td>6.1</td>
<td>16</td>
</tr>
<tr>
<td>R-42</td>
<td>835</td>
<td>1.24</td>
<td>80.9</td>
<td>27.04</td>
<td>45.2</td>
<td>201</td>
</tr>
<tr>
<td>R-61#1</td>
<td>10.0</td>
<td>0.195</td>
<td>1.77</td>
<td>9.84</td>
<td>1.84</td>
<td>24</td>
</tr>
<tr>
<td>R-61#2</td>
<td>1</td>
<td>0.198</td>
<td>2.23</td>
<td>1.51</td>
<td>2.48</td>
<td>1</td>
</tr>
<tr>
<td>R-44#1</td>
<td>15.6</td>
<td>0.435</td>
<td>3.56</td>
<td>4.85</td>
<td>2.42</td>
<td>3.2</td>
</tr>
<tr>
<td>R-44#2</td>
<td>7.72</td>
<td>0.358</td>
<td>2.95</td>
<td>4.00</td>
<td>2.37</td>
<td>0.8</td>
</tr>
<tr>
<td>R-45#1</td>
<td>35.7</td>
<td>0.597</td>
<td>7.37</td>
<td>9.76</td>
<td>4.77</td>
<td>3.6</td>
</tr>
<tr>
<td>R-45#2</td>
<td>18.4</td>
<td>0.4</td>
<td>4.32</td>
<td>3.04</td>
<td>3.72</td>
<td>3.3</td>
</tr>
<tr>
<td>R-50#1</td>
<td>103</td>
<td>0.586</td>
<td>11.5</td>
<td>6.85</td>
<td>8.13</td>
<td>26</td>
</tr>
<tr>
<td>R-50#2</td>
<td>3.73</td>
<td>0.307</td>
<td>2.25</td>
<td>2.79</td>
<td>2.0</td>
<td>1.2</td>
</tr>
<tr>
<td>R-43#1</td>
<td>146</td>
<td>1.02</td>
<td>16.9</td>
<td>21.27</td>
<td>8.59</td>
<td>1.3</td>
</tr>
<tr>
<td>R-43#2</td>
<td>8.13</td>
<td>0.751</td>
<td>5.87</td>
<td>8.52</td>
<td>4.66</td>
<td>1.1</td>
</tr>
<tr>
<td>R-42</td>
<td>835</td>
<td>1.24</td>
<td>80.9</td>
<td>27.04</td>
<td>45.2</td>
<td>201</td>
</tr>
<tr>
<td>Source</td>
<td>$Cr^{6+}$ µg/L</td>
<td>$ClO_4^-$ µg/L</td>
<td>$SO_4^{2-}$ mg/L</td>
<td>$NO_3^-$ mg/L</td>
<td>$Cl^-$ mg/L</td>
<td>$^3H$ pCi/L</td>
</tr>
<tr>
<td>--------</td>
<td>----------------</td>
<td>-----------------</td>
<td>------------------</td>
<td>---------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>1</td>
<td>1300</td>
<td>0</td>
<td>87</td>
<td>8.8</td>
<td>66</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>0.21</td>
<td>0.56</td>
<td>11</td>
<td>0</td>
<td>0.021</td>
<td>130</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>51</td>
<td>2</td>
<td>13</td>
<td>0.094</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0.24</td>
<td>0</td>
<td>19</td>
<td>4</td>
<td>33</td>
<td>0.069</td>
</tr>
<tr>
<td>5</td>
<td>0.009</td>
<td>0</td>
<td>7</td>
<td>21</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Machine Learning: groundwater type / source maps

\[ Cr^{6+}, SO_4^{2-}, Cl^- \]

Source 1

Chromium

Models

Machine Learning

Decisions

Summary
Machine Learning: groundwater type / source maps

$ClO_4^-, NO_3^-$

Source 3

Chromium
Models
Machine Learning
Decisions
Summary
Machine Learning: groundwater type / source maps

$NO_3^-$

Source 5

Chromium

Models

Machine Learning

Decisions

Summary
Probabilistic methods work very well for dice-rolling experiments
- Probabilistic methods work very well for dice-rolling experiments
- However, many earth-science uncertainties cannot be represented probabilistically
- To address these issues, recently we have developed, published and applied Bayesian-Information-Gap Decision Theory (BIG-DT)
Bayesian Information Gap Decision Analysis: Setup

**Unknowns:**
- contaminant mass release, source location \((x, y)\) and size
- hydraulic conductivity
- porosity
- dispersivity (longitudinal and transverse)
- contaminant transport parameters (mean mobile/immobile times of pore-scale mixing)

**Knowns:**
- well locations
- well pumping rates
- ambient hydraulic gradient
- location of compliance boundary
- hydraulic heads at the monitoring wells
- contaminant concentrations at the monitoring wells
- 30 monitoring wells
- 10 annual observations (heads/concentrations) per well (600 in total)
Bayesian Information Gap Decision Analysis: No action
Bayesian Information Gap Decision Analysis: Pumping Chromium Models

Machine Learning Decisions

Summary
Bayesian Information Gap Decision Analysis: Results

- **Chromium**
- **Models**
- **Machine Learning**
- **Decisions**
- **Summary**

Graph showing the robustness versus the maximum probability of failure for different scenarios:
- Red: Use all 3 extraction wells
- Blue: Use the outer 2 extraction wells
- Black: Use only the middle extraction well

Inset map showing the locations of supply wells, observation wells, contaminant source, extraction wells, and the compliance boundary.
Key Advancements/Developments: Modeling

- New model of biogeochemical reactions for in-situ remediation with unprecedented capabilities (Chrotran)
- Novel approach for geochemical reaction modeling using particle tracking (free of numerical diffusion/dispersion)
- Novel methods for fast big-data model analyses of groundwater flow & transport using adjoints
- Advanced analytical solutions of contaminant transport (including non-Fickian dispersion processes)
- Novel analytical solutions of groundwater flow towards a pumping well under complex hydrogeological conditions
- Advanced surrogate (reduced-order) and deep-learning modeling
Novel methods for model analyses (model calibration, uncertainty quantification (UQ), sensitivity analysis, optimal experimental design (OED), risk assessment, decision analysis)

Novel methods for tracer-test characterization of plume dispersion & aquifer heterogeneity

Novel methods for contaminant source identification

Novel unsupervised (objective) machine learning for blind source separation (patented)

Novel Bayesian-Information-Gap Decision Theory (BIG-DT)
Modeling have been successfully applied to support development of the site conceptual model representing hydrogeological and biogeochemical processes.

Monitoring network at the site was successfully augmented over the years using model-assisted decision analyses.

So far all the “blind” model predictions have been generally consistent with all the new observations.

Model-assisted decision analyses are currently performed to design site remediation activities.

Various novel methods and techniques have been developed and applied to address the project needs (dispersion characterization, heterogeneity characterization, modeling biogeochemical processes, decision analyses, etc.).

In the last 6 years, we accumulated more than 4,000 CPU-years of wall-clock computational time utilizing simultaneously up to 4096 processors on the LANL HPC clusters.
MADS (Model Analysis & Decision Support)

- high-performance computational framework for big-data and complex model analyses
- provides state-of-the-art and novel methods for:
  - sensitivity analyses
  - uncertainty quantification
  - model calibration / parameter estimation
  - surrogate modeling
  - machine learning
  - risk assessment
  - decision analysis
  - decision-based optimal experimental design
- open-source & version-controlled code
- fully documented, QA-ed, verified and tested in real time
- http://mads.lanl.gov
Chrotran

- a massively parallel numerical simulator for in situ biogeochemical remediation of heavy metals in heterogeneous aquifers
- unique capabilities to simulate sophisticated, multi-scale biogeochemical remediation processes
- based upon the existing PFLOTRAN code framework
- coupled to MADS for various types of model analyses
- open-source & version-controlled code
- documented, verified and tested
- http://chrotran.lanl.gov
ZEM framework

- **ZEM** provides automated and reproducible workflow interconnecting Data ⇔ Models ⇔ Decisions
- Designed for high-performance computing and big-data analysis
- **ZEM** employs git for version control, team collaboration and project management using cloud-based services
- All past and intermediate model states/inputs and obtained key model outputs are tracked, stored and can be reproduced
- Provides quality assurance in the performance assessment process
- Data/model processing is highly automated; a single **ZEM** command produced all the model outputs presented here
- **ZEM** is written predominantly in **Julia**