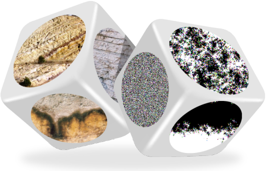
ZEM: Integrated Framework for Real-Time Data and Model Analyses for Robust Environmental Management Decision Making

Velimir V. Vesselinov, Dan O'Malley, Danny Katzman

Computational Earth Science, Los Alamos National Laboratory

Waste Management Symposium, March 8, 2016
LA-UR-16-21469



ZEM framework

- **ZEM** provides automated and reproducible workflow interconnecting Data $\Leftrightarrow$ Models $\Leftrightarrow$ Decisions
- **ZEM** is designed for **high-performance computing** and **big-data** analysis
- **ZEM** employs community software (**git**/**gitlab**) for **version control**, **team collaboration** and **project management** using cloud-based repositories (**gitlab.com** / **git.lanl.gov**) $\Rightarrow$ all past model inputs and obtained outputs are stored and can be reproduced
- **ZEM** provides quality assurance of the performance assessment process
- **ZEM** is written predominantly in julia
- julia: novel high-performance/dynamic language for technical computing (developed at MIT)

ZEM components

- **MADS** (Model Analysis & Decision Support): actively developed open-source high-performance computational framework for data- & model-based analyses in julia (madsjulia.lanl.gov)
- **MySQL** (www.mysql.com): open-source relational database management system stores all the site data (more than 10^7 entries)
- Web interfaces (for data queries and exploratory model analyses)
- Various simulators
- Visualization tools (matplotlib, gnuplot, Gadfly, Paraview, VisIt)
- julia/Python scripts to couple all the **ZEM** components
- For example, a single julia script can:
  - perform automated data query from the **ZEM** database
  - place the data in the model input files
  - initiate the simulations on HPC clusters
  - generate plots and movies with the final results

ZEM: Analytical simulators

- Analytical solutions for **groundwater flow** (implemented in **MADS** and **Wells**)
- Analytical solutions for Fickian (classical) and non-Fickian (anomalous) **contaminant transport** (implemented in **MADS**)
- Analytical simulator of groundwater flow and contaminant transport associated with infiltration recharge and perched horizons in the **vadose zone** (a fast screening tool) (implemented in **MADS**)
- Semi-analytical simulator for **capture zone** estimation and **tracer test** interpretation (push-and-pull and cross-well tracer tests; **MADS**)
- Analytical method for removal of **barometric pressure** and **tidal effects** in the water-level data (**CHipBeta**):

ZEM: Numerical simulators

- **FEHM**: groundwater flow and contaminant transport; geochemical reactions (LANL developed code)
- **PFloTran**: groundwater flow and contaminant transport; biogeochemical reactions (LANL developed open-source code)
- **LaGriT**: grid generation (LANL developed open-source code)
- **Ashley**: particle-based geochemical reactions (LANL developed code in julia)
- **FEniCS**: automated and efficient differential-equation solver (open-source community code)
- **libMesh**: advanced parallel partial-differential-equation solver (open-source community code)
- **Amanzi**: groundwater flow and contaminant transport; geochemical reactions (LANL developed code; future work)

- **Drawdown estimator**: tool for data- and model-based analysis for identification and deconstruction of pumping drawdowns (typically, drawdowns are smaller than the barometric pressure fluctuations and caused by overlapping pumping events)
- **RMF** (**R**obust **M**atrix **F**actorization): novel methodology for model-free inversion and data analysis
- Unsupervised objective **machine-learning methods** for data, model and decision analyses
- **Surrogate modeling** using state-of-the-art and newly developed methods (SVR, Bayesian)
- **Various data-analysis tools** such as principle and independent component analysis, trend analysis, spatial interpolation, etc.

(utilizing third-party julia community modules).

ZEM utilizes state-of-the-art and novel advanced methods for characterization of aquifer heterogeneity

- **Pilot-point**-based methods
- **Fourier**-based stochastic methods
- **Regularization**-based methods
- **Level-set** tomography (geologic facies reconstruction)
- **"Honest"** tomography (accounting for uncertainties and unknowns)
- **P**rincipal **C**omponent **G**eostatistical **A**analysis (**PCGA**; Kitanidis et al., 2014)
- **R**andom **G**eostatistical **A**analysis (**RGA**) for **big-data** tomography (Le et al., 2016)

ZEM: Analyses

ZEM have been successfully applied to support development of the site conceptual model representing hydrogeological and biogeochemical processes in the subsurface

- Contaminant source identification
- Contaminant source characterization (based on geochemical data and model-free inversion using unsupervised objective machine learning)
- Monitoring network design
- Evaluation of remediation scenarios
- Sensitivity and uncertainty quantification analyses
- Decision analyses
- In the last **3** years, **ZEM** analyses have accumulated more than **350** CPU-years of wall-clock computational time utilizing simultaneously up to **4096** processors on the LANL HPC clusters
- ... so far, all the **ZEM** blind predictions have been consistent with the new observations

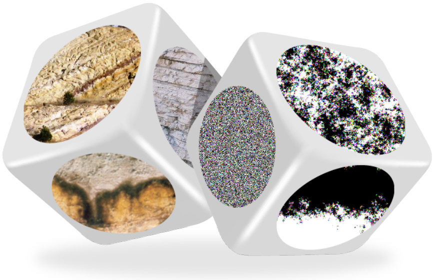
ZEM ⇔ MADS (Model Analysis & Decision Support)

- open-source, version-controlled, high-performance computing framework implementing state-of-the-art and novel adaptive computational techniques for:
  - sensitivity analysis (local / global)
  - uncertainty quantification (local / global)
  - optimization / calibration / parameter estimation (local / global) parallel Krylov-space methods for **big-data** analyses
  - model ranking & selection
  - decision analysis (GLUE, information gap, Bayesian, **B**ayesian - **I**nformation **G**ap **D**ecision **T**heory (**BIG-DT**), **Measure-Theoretic**-based approaches)
  - decision-based experimental design

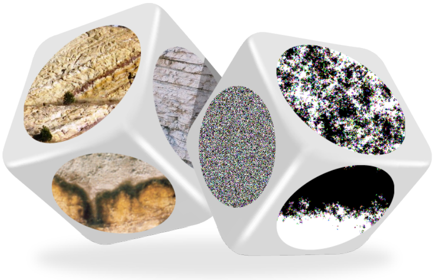
ZEM ⇔ MADS (Model Analysis & Decision Support)

- provides **internal** coupling with analytical groundwater flow and contaminant transport solvers
- allow **external** coupling with any existing physics simulator
- coded in julia
- source code, examples, test problems, performance comparisons, and tutorials are available at:
  - `http://madsjulia.lanl.gov`
  - `http://madsjl.readthedocs.org/`

- Probabilistic methods work very well for dice-rolling experiments

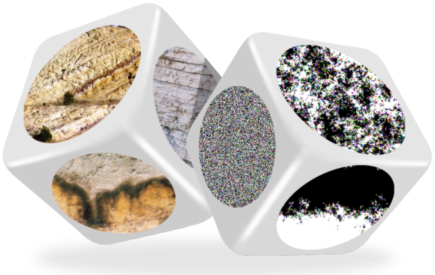


- Probabilistic methods work very well for dice-rolling experiments
- However, many earth-science uncertainties cannot be represented probabilistically (for example, using GoldSim)

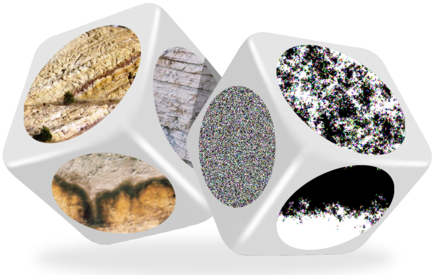


- Probabilistic methods work very well for dice-rolling experiments
- However, many earth-science uncertainties cannot be represented probabilistically (for example, using GoldSim)
- Actual geologic heterogeneity is typically unknown (**left die**)

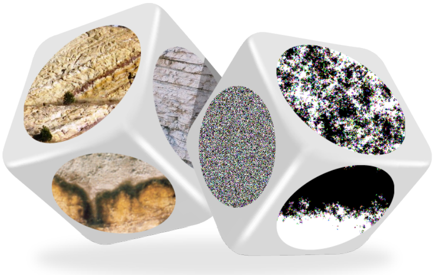
MADS: Bayesian-Information-Gap Decision Theory (BIG-DT)



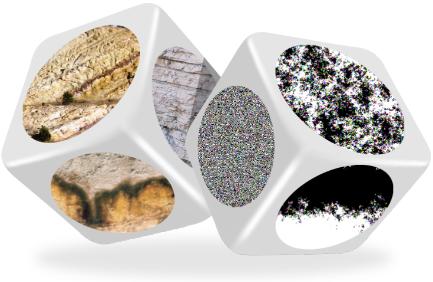
- Probabilistic methods work very well for dice-rolling experiments
- However, many earth-science uncertainties cannot be represented probabilistically (for example, using GoldSim)
- Actual geologic heterogeneity is typically unknown (**left die**)
- We also do not know which of the possible models of geologic heterogeneity is representative (**right die**), but probabilistic methods require to choose a single representative model conditioned on the available data



- We also do not know what all the sides of the dice look like, and how many sides there are

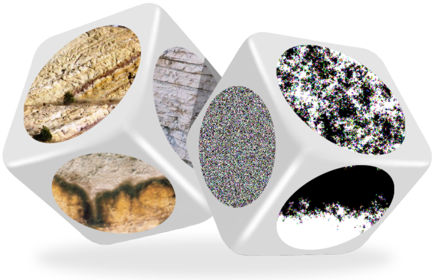


- We also do not know what all the sides of the dice look like, and how many sides there are
- Therefore, we cannot **enumerate all possible outcomes**



- We also do not know what all the sides of the dice look like, and how many sides there are
- Therefore, we cannot **enumerate all possible outcomes**
- All these issues make purely probabilistic analyses **flawed** for many earth-science problems

MADS: Bayesian-Information-Gap Decision Theory (BIG-DT)

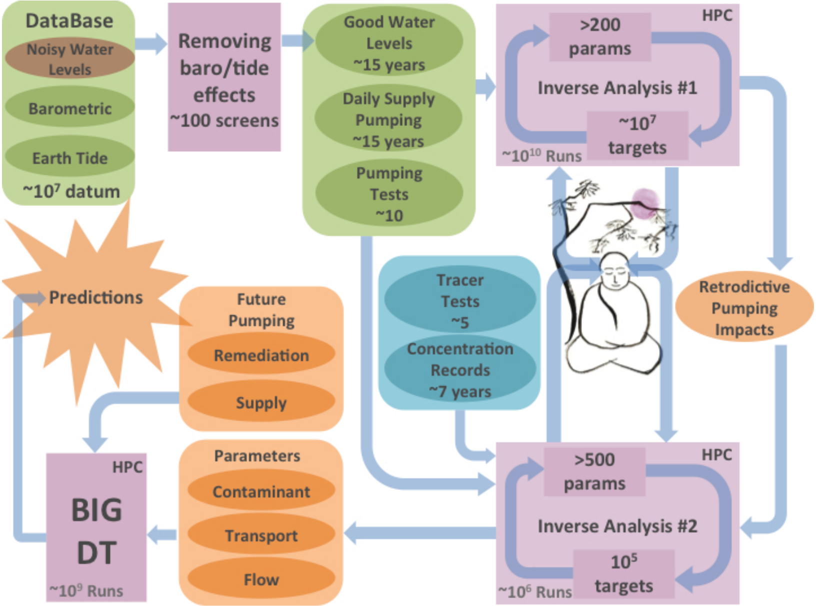


- We also do not know what all the sides of the dice look like, and how many sides there are
- Therefore, we cannot **enumerate all possible outcomes**
- All these issues make purely probabilistic analyses **flawed** for many earth-science problems
- **B**ayesian - **I**nformation **G**ap **D**ecision **T**heory (BIG-DT) for Uncertainty Quantification & Decision Analysis has been developed to address these issues (O'Malley & Vesselinov 2014 SIAM UQ)

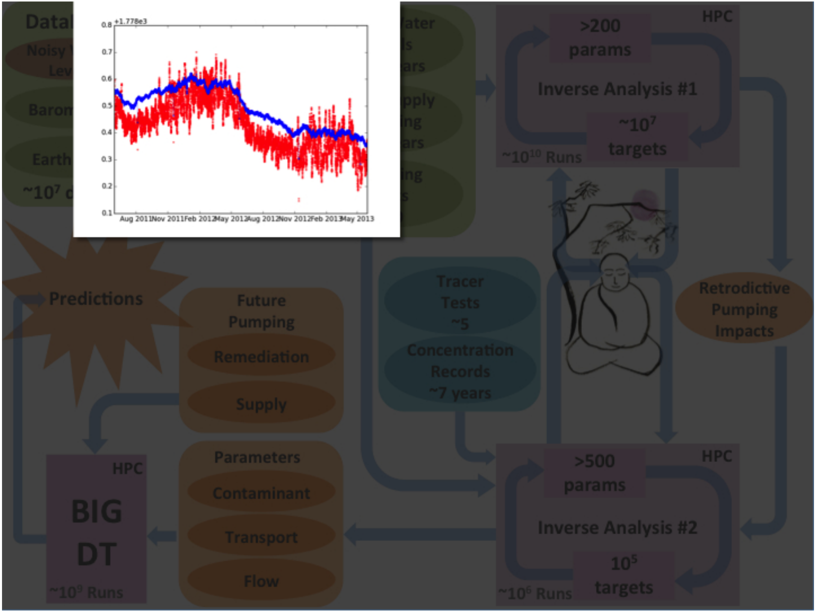
ZEM development support

- **LANL Environmental Projects**
- **DiaMonD Project**:
  - DiaMonD: Integrated Multifaceted Approach to Mathematics at the Interfaces of Data, Models, and Decisions
    - University of Texas at Austin
    - Massachusetts Institute of Technology (MIT)
    - Stanford University
    - Colorado State University
    - Florida State University
    - Los Alamos National Laboratory
    - Oak Ridge National Laboratory
  - Funded by DOE Office of Science
  - `http://dmd.mit.edu`

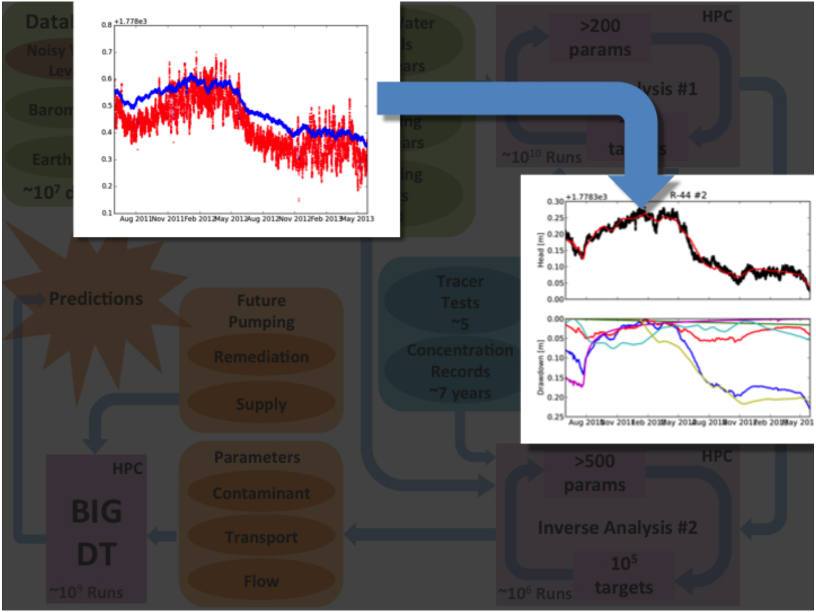
ZEM workflow: Data ⇔ Models ⇔ Decisions



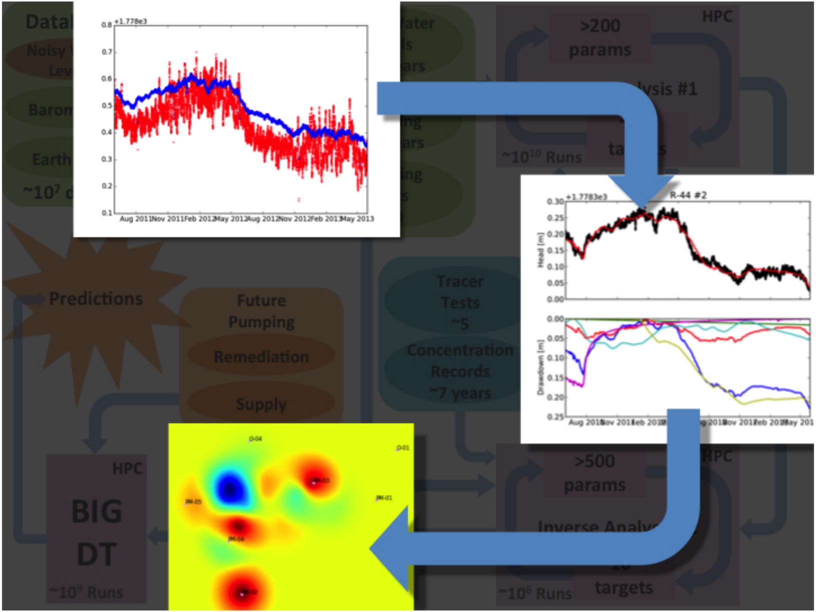
ZEM workflow: Data ⇔ Models ⇔ Decisions



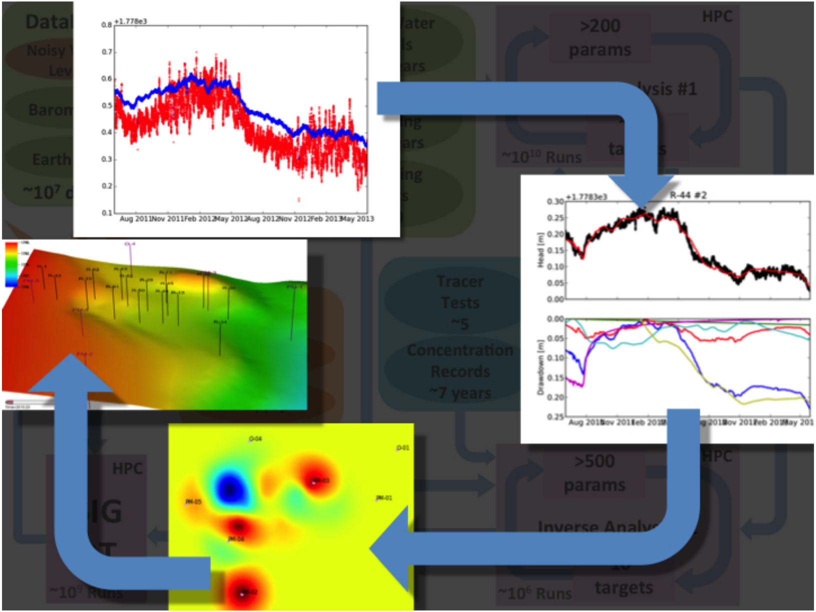
ZEM workflow: Data ⇔ Models ⇔ Decisions



ZEM workflow: Data ⇔ Models ⇔ Decisions



ZEM workflow: Data ⇔ Models ⇔ Decisions



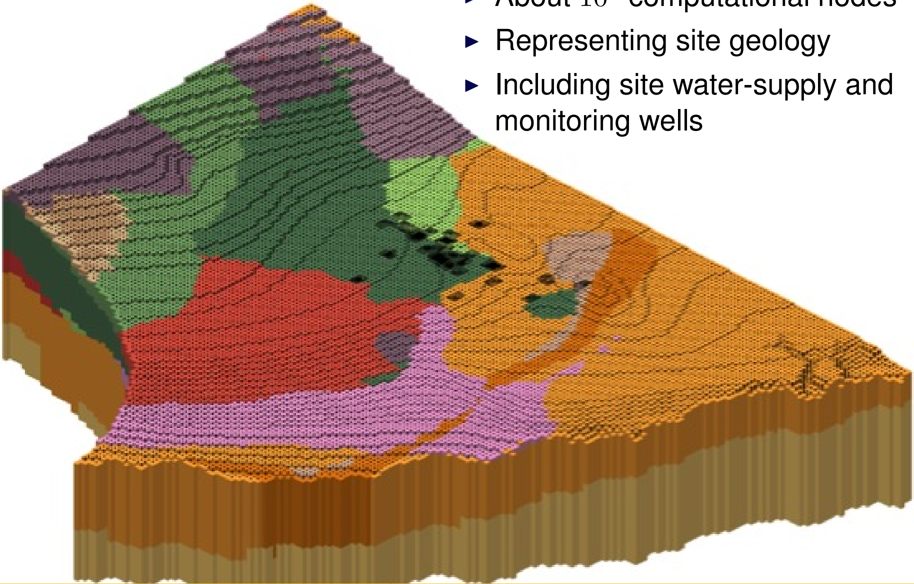
Chromium site high-level summary

- High visibility project
- ~54,000 kg of Cr^{6+} released in Sandia Canyon between 1956 and 1972 (with substantial **uncertainties** and **unknowns**)
- Cr^{6+} detected above MCL (50 ppb; NM standard) at 6 monitoring wells in the regional aquifer beneath LANL
- 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)
- Contaminant mass distribution in the subsurface in **unknown**
- Contaminant source location and mass flux at the top of the regional aquifer are **unknown** due to **complex** 3D pathways through the vadose zone
- **Limited remedial options** due to aquifer depth (~300 m below the ground surface) and **complexities** in the subsurface processes
- Current conceptual model for chromium transport in the subsurface is supported by **multiple lines of evidence**

Chromium project goals

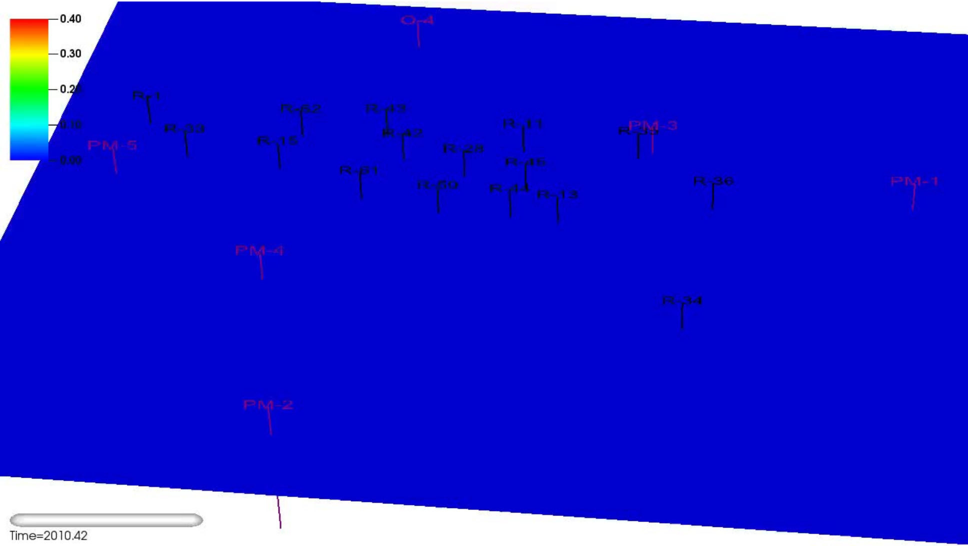
- **GOAL #1**: apply modeling to support **conceptualization** of the site geologic, hydrologic and biogeochemical conditions
- **GOAL #2**: perform data- and model-based **decision analyses** for chromium remediation taking into account existing processes and **uncertainties**/**unknowns**
- **Remedial scenarios**:
  - Natural attenuation (**NA**)
  - Enhanced attenuation (**EA**; biogeochemical processes)
  - Active remediation including mass removal in the vadose zone and the aquifer (**pump-and-treat**, etc.)
  - **Combinations** of all above at different times/locations

Chromium site model



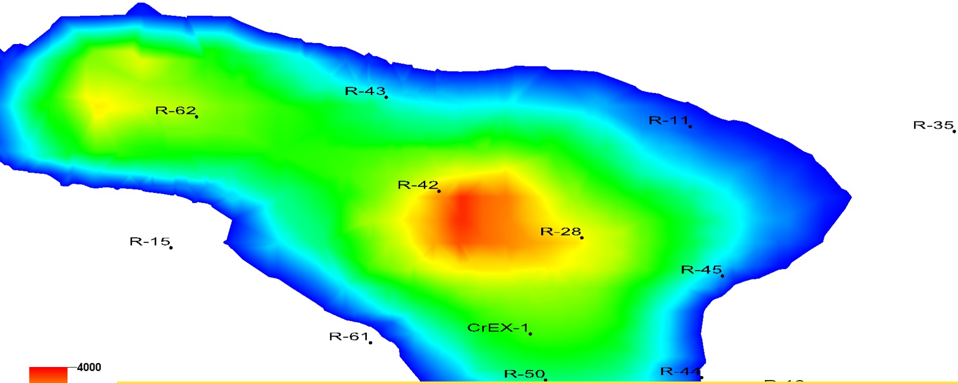
- About 10^6 computational nodes
- Representing site geology
- Including site water-supply and monitoring wells

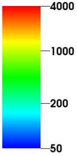
Drawdowns from the existing supply wells



Play/Pause

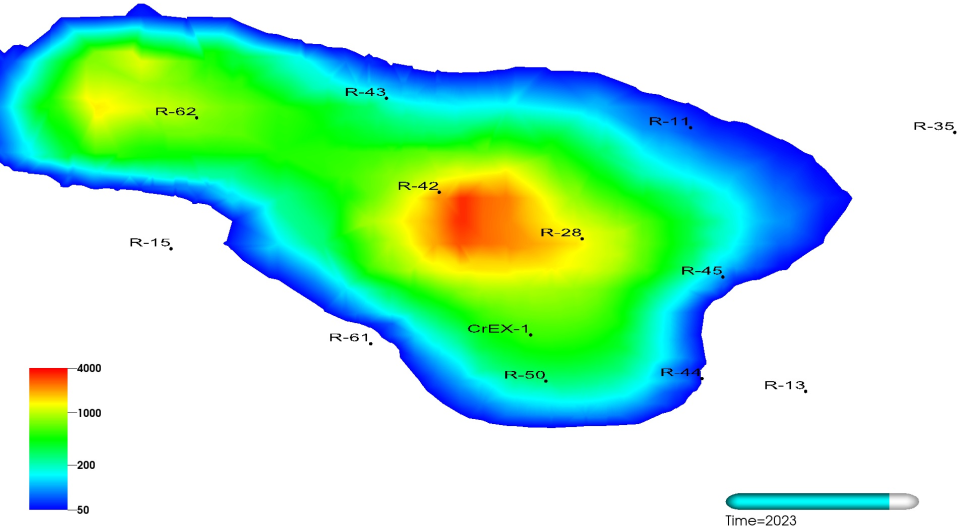
Chromium plume transients



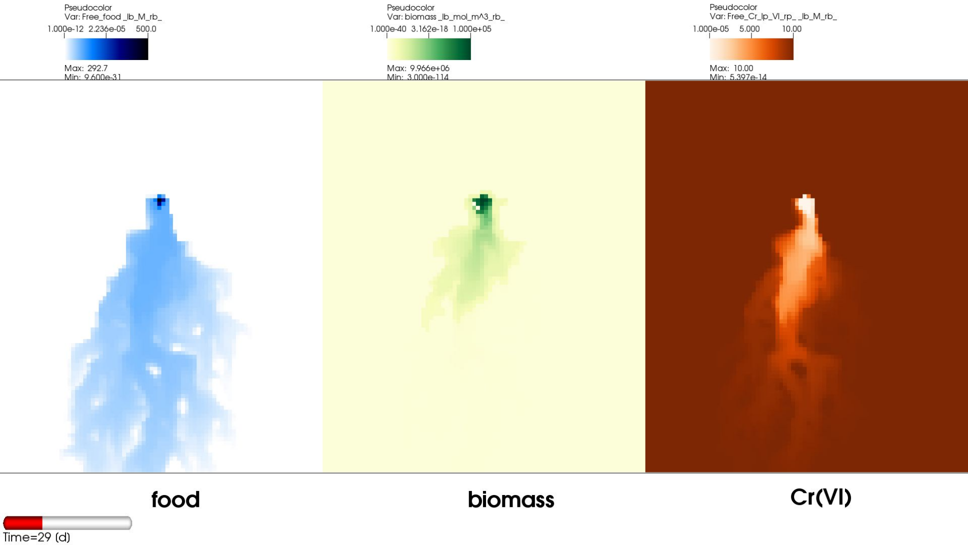


- Model is calibrated against all the pressure and concentration transients
- ... so far, ~20 CPU-years of wall-clock computational time are accumulated
- ... additional model improvements are still needed

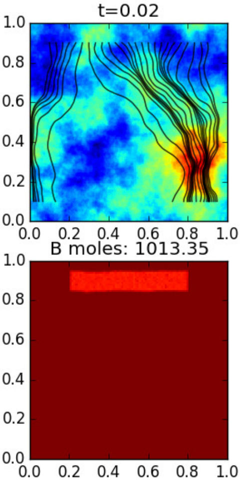
Chromium plume transients



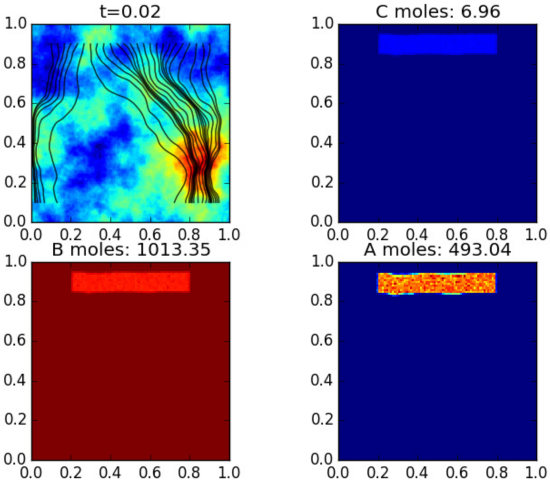
Chromium bio-remediation modeling (PFloTran)



Geochemical particle-based model (Ashley)



- A + B = C
- Reduction of contaminant B by injecting A
- Reduction of contaminant A by interacting with B
- A instantaneously released (500 moles)
- B uniformly distributed in the aquifer (1000 moles)



- **20%** of **A** did not react

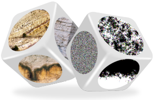
Play/Pause

Highlights

- **ZEM** provides automated and reproducible workflow interconnecting Data ⇔ Models ⇔ Decisions using **high-performance computing** and **big-data** analysis tools
- **ZEM** have been successfully applied to perform various data- and model-based analyses at the LANL Chromium site.
- In the last **3** years, **ZEM** analyses have accumulated more than **350** CPU-years of wall-clock computational time utilizing simultaneously up to **4096** processors on the LANL HPC clusters
- ... so far, all the **ZEM** blind predictions have been consistent with the new observations

Highlights

- Many uncertainties in the environmental management problems **cannot** be represented probabilistically
- Newly developed methodology **BIG-DT** (**B**ayesian-**I**nformation **G**ap **D**ecision **T**heory) is developed to address this issue (O'Malley & Vesselinov 2014 SIAM UQ)
- **BIG-DT** is applicable to any real-world engineering problems
- **BIG-DT** is available in **MADS** (open source code written in julia)
  - `http://madsjulia.lanl.gov`
  - `http://madsjl.readthedocs.org/`



Relevant Publications

1. Grasinger, M., O'Malley, D., Vesselinov, V.V., Karra, S., Decision Analysis for Robust CO_2 Injection: Application of Bayesian-Information-Gap Decision Theory, IJGGC, doi: 10.1016/j.ijggc.2016.02.017, 2016.
2. Mattis, S.A., Butler, T.D. Dawson, C.N., Estep, D., Vesselinov, V.V., Parameter estimation and prediction for groundwater contamination based on measure theory, WRR, doi: 10.1002/2015WR017295, 2015
3. Barajas-Solano, D. A., Wohlberg, B., Vesselinov, V.V., Tartakovsky, D. M., Linear Functional Minimization for Inverse Modeling, WRR, doi: 10.1002/2014WR016179, 2015
4. O?Malley, D., Vesselinov, V.V., Bayesian-Information-Gap decision theory with an application to CO2 sequestration, Water Resources Research, doi: 10.1002/2015WR017413, 2015
5. Lu, Z., Vesselinov, V.V., Analytical Sensitivity Analysis of Transient Groundwater Flow in a Bounded Model Domain using Adjoint Method, WRR, doi: 10.1002/2014WR016819, 2015
6. O'Malley, D., Vesselinov, V.V., Cushman, J.H., Diffusive mixing and Tsallis entropy, Phys.Rev E, 91, 042143, 2015
7. Vesselinov, V.V., O'Malley, D., Katzman, D., Model-Assisted Decision Analyses Related to a Chromium Plume at Los Alamos National Laboratory, Waste Management, 2015
8. O'Malley, D., Vesselinov, V.V., A combined probabilistic/non-probabilistic decision analysis for contaminant remediation, SIAM-UQ, doi: 10.1137/140965132, 2014
9. O'Malley, D., Vesselinov, V.V., Cushman, J.H., A Method for Identifying Diffusive Trajectories with Stochastic Model, Journal of Statistical Physics, Springer, doi: 10.1007/s10955-014-1035-6, 2014
10. Alexandrov, B., Vesselinov, V.V., Blind source separation for groundwater level analysis based on non-negative matrix factorization, WRR, doi: 10.1002/2013WR015037, 2014
11. O'Malley, D., Vesselinov, V.V., Analytical solutions for anomalous dispersion transport, AWR, doi: 10.1016/j.advwatres.2014.02.006, 2014.
12. Heikoop, J.M., Johnson, T.M., Birdsell, K.H., Longmire, P., Hickmott, D.D., Jacobs, E.P., Broxton, D.E., Katzman, D., Vesselinov, V.V., Ding, M., Vaniman, D.T., Reneau, S.L., Goering, T.J., Glessner, J., Basu, A., Isotopic evidence for reduction of anthropogenic hexavalent chromium in LANL groundwater, Chemical Geology, doi: 10.1016/j.chemgeo.2014.02.022, 2014.
13. O'Malley, D., Vesselinov, V.V., Groundwater remediation using the information gap decision theory, WRR, doi: 10.1002/2013WR014718, 2014.
14. Harp, D.R., Vesselinov, V.V., Accounting for the influence of aquifer heterogeneity on spatial propagation of pumping drawdown, Journal of Water Resource and Hydraulic Engineering, 2(3), pp. 65-83, 2013.
15. Vesselinov, V.V., Katzman, D., Broxton, D., Birdsell, K., Reneau, S., Vaniman, D., Longmire, P., Fabryka-Martin, J., Heikoop, J., Ding, M., Hickmott, D., Jacobs, E., Goering, T., Harp, D.R., Mishra, P., Data and Model-Driven Decision Support for Environmental Management of a Chromium Plume at LANL, Waste Management, 2013.
16. Vesselinov, V.V., Harp, D.R., Adaptive hybrid optimization strategy for calibration and parameter estimation of physical process models, Computers & Geosciences, doi: 10.1016/j.cageo.2012.05.027, 2012.
17. Mishra, P.K., Vesselinov, V.V., Neuman, S.P., Radial flow to a partially penetrating well with storage in an anisotropic confined aquifer, JH, doi: 10.1016/j.jhydrol.2012.05.010, 2012.
18. Mishra, P.K., Vesselinov, V.V., Kuhlman, K.L., Saturated/unsaturated flow in a compressible leaky-unconfined aquifer, AWR, doi: 10.1016/j.advwatres.2012.03.007, 2012.
19. Mishra, P.K., Gupta, H.V., Vesselinov, V.V., On simulation and analysis of variable-rate pumping tests, Ground Water, doi: 10.1111/j.1745-6584.2012.00961.x, 2012.
20. Vesselinov, V.V., Harp, D.R., Model Analysis and Decision Support (MADS) for complex physics models, CMWR, 2012.
21. Harp, D.R., Vesselinov. V.V., Contaminant remediation decision analysis using information gap theory, Stochastic Environmental Research and Risk Assessment (SERRA), doi:10.1007/s00477-012-0573-1, 2012.
22. Harp, D.R., Vesselinov, V.V., An agent-based approach to global uncertainty and sensitivity analysis, Computers & Geosciences, doi: 10.1016/j.cageo.2011.06.025, 2011.
23. Harp, D.R., Vesselinov, V.V., Analysis of hydrogeological structure uncertainty by estimation of hydrogeological acceptance probability of geostatistical models, AWR, doi: 10.1016/j.advwatres.2011.06.007, 2011.
24. Harp, D.R., Vesselinov, V.V., Identification of Pumping Influences in Long-Term Water Level Fluctuations, Groundwater, doi: 10.1111/j.1745-6584.2010.00725.x, 2010.
25. Harp, D.R., Vesselinov, V.V., Stochastic inverse method for estimation of geostatistical representation of hydrogeologic stratigraphy using borehole logs and pressure, invited, SERRA, doi: 10.1007/s00477-010-0403-2, 2010.
26. Vesselinov, V.V., Uncertainties In Transient Capture-Zone Estimates, CMWR, ISBN 90-5809-124-4, 2006.

Team

- **Dan O'Malley**
- **Zhiming Lu**
- **Satish Karra**
- **Terry Miller**
- **Lucia Short**
- **Youzou Lin**
- **Boian Alexandrov**
- **Bhat Sham**
- **Xiaodong Zhang**
- **Scott Hansen**
- **Steve Mattis** (UT-Austin)
- **Matt Grasinger** (Pitt)
- **Ellen Le** (UT-Austin)
- **Justin Laughlin** (UCSD)
- **Natalia Siuliukina** (UCSD)
- **Filip Iliev** (UCSC)
- **Xi Chen** (UT-Austin)
- **Harriet Li** (MIT)
- **Eric Benner** (UNM)
- **David Barajas-Solano** (UCSD)

Why ZEM?

- **ZEM** $\approx$ **ZEN**
- **ZEM**: **Z**eitgeist (spirit of the time) **E**nvironmental **M**odeling
- **ZEM**: the Slavic root word for Earth