Reduced Order Models for Decision Analysis and Upscaling of Aquifer Heterogeneity

Velimir V. Vesselinov, Daniel O'Malley
Boian S. Alexandrov, Bryan Moore

Los Alamos National Laboratory, NM 87545, USA

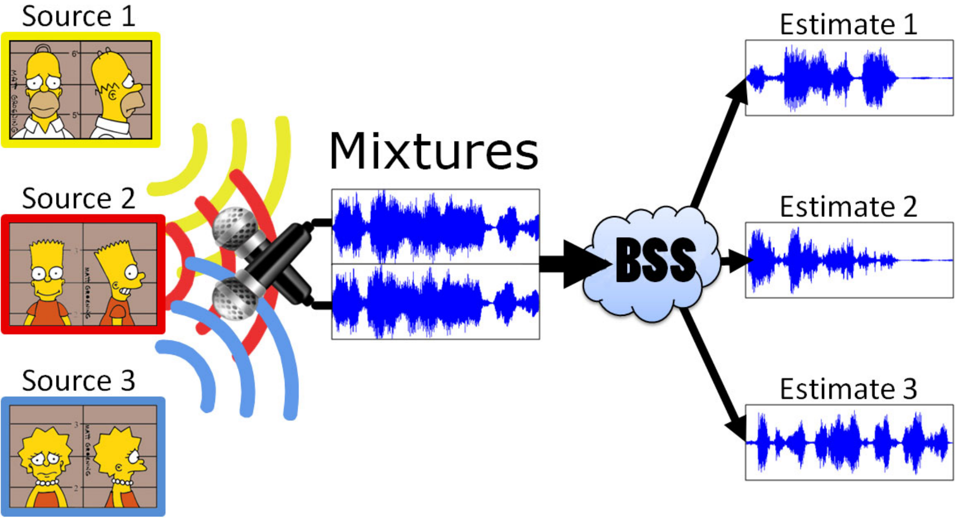
LA-UR-16-29305

Overview

- Blind source separation applied to hydrogeochemistry (Contaminant source identification)
- Reduced order modeling for contaminant transport (Upscaling of contaminant transport properties)

Blind Source Separation (BSS)

- BSS: an objective machine-learning method for source identification without a model (**model-free** analysis/inversion)



Blind Source Separation (BSS)

- Provides characterization of the physical sources causing spatial and temporal variation of observed state variables (e.g. pressures, concentrations, etc.)
- Avoids model errors
- Accounts for measurement errors
- Identification of the sources (forcings) can be crucial for conceptualization and model development
- If the sources are successfully "**unmixed**" from the observations, decoupled physics models may then be applied to analyze the propagation of each source independently
- Widely applicable

Blind Source Separation $\rightarrow$ Matrix Factorization

- Invert for the **unknown** sources $\mathbf{S}$ $[p \times r]$ that have produced **known** observation records, $\mathbf{H}$ $[p \times m]$, with **unknown** noise (measurement errors), $\mathbf{E}$ $[p \times m]$:

$$\mathbf{H} = \mathbf{S}\mathbf{A} + \mathbf{E}$$

  - $\mathbf{A}$ $[r \times m]$ is **unknown** "**mixing**" matrix
  - p is the number of observation points (wells)
  - m is the number of observed components
  - r is the number of **unknown** sources ($r < m$)
- The problem is ill-posed and the solutions are non-unique
- There are various methods to resolve this applying different "regularization" terms:
  - maximum variability
  - statistical independence
  - non-negativity
  - smoothness
  - simplicity, *etc.*

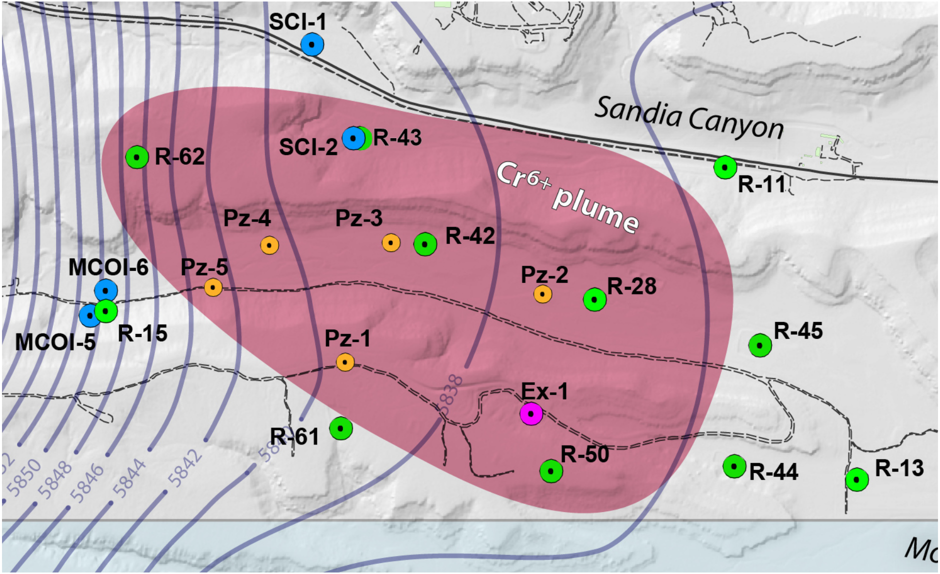
Blind source separation methods

- **ICA**: Independent Component Analysis
  - **Maximizing the statistical independence** of the retrieved forcings signals in $\mathbf{S}$ (i.e. the matrix columns are expected to be independent) by maximizing some high-order statistics for each source signal (e.g. kurtosis) or minimizing information entropy
  - The main idea behind **ICA** is that, while the probability distribution of a linear mixture of sources in $\mathbf{H}$ is expected to be close to a Gaussian (the Central Limit Theorem), the probability distribution of the original independent sources is expected to be non-Gaussian.
- **NMF**: Non-negative Matrix Factorization
  - **Non-negativity constraint** on the components of both the signal $\mathbf{S}$ and mixing $\mathbf{A}$ matrices
  - As a result, the observed data are representing only **additive signals that cannot cancel mutually** (suitable for many applications)
  - Additivity and non-negativity requirements may lead to sparseness in both the signal $\mathbf{S}$ and mixing $\mathbf{A}$ matrices

NMFk: Non-negative Matrix Factorization + k-means

- **NMFk**: we have developed a novel machine learning method for BSS coupling two machine-learning techniques:
  - Non-negative Matrix Factorization (NMF)
  - k-means clustering
- **NMFk** applies two constraints:
  - non-negativity
  - parsimony (simplicity)
- Implemented in **MADS** (Model Analysis & Decision Support)
- Coded in julia

LANL Chromium site (2015)



Hydrogeochemical data [29×6]

In the microphone analogy, this is what is recorded by the microphones.

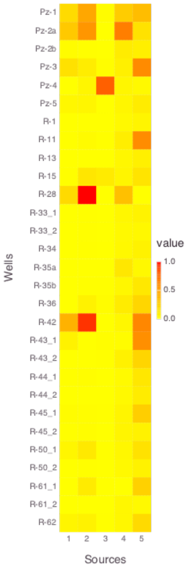
Well	Cr^{6+}	ClO_4^-	SO_4^{2-}	NO_3^-	Cl^-	^3H
Pz-1	406.22	1.84	47.846	17.07	35.401	101.397
Pz-2a	83.89	0.88	71.155	14.42	66.436	121.013
Pz-2b	35.01	0.419	6.2918	4.24	7.582	2.061
Pz-3	338.88	1.21	33.967	23.60	21.853	24.184
Pz-4	5.69	63.7	5.8175	17.90	3.0975	11.346
Pz-5	89.26	0.44	8.7896	4.98	7.8321	11.807
R-1	5.68	0.351	2.19	2.26	2	0.5
R-11	20.8	0.83	13.1	20.60	5.15	4.9
R-13	3.81	0.4	3.12	3.22	2.49	0.2
R-15	12.5	8.93	6.22	7.97	3.99	29
R-28	407	1.0	55.1	4.91	38.5	211
R-33#1	4.89	0.398	3.32	2.41	2.29	2
R-33#2	5.52	0.35	2.3	1.64	2.0	1.2
R-34	4.26	0.333	2.66	2.76	2.42	1.2
R-35a	4.3	0.422	5.62	2.10	6.74	0.6
R-35b	6.98	0.579	3.48	4.84	2.88	1.3
R-36	5.29	1.55	7.35	8.69	6.1	16
R-42	835	1.24	80.9	27.04	45.2	201
R-43#1	146	1.02	16.9	21.27	8.59	1.3
R-43#2	8.13	0.751	5.87	8.52	4.66	1.1
R-44#1	15.6	0.435	3.56	4.85	2.42	3.2
R-44#2	7.72	0.358	2.95	4.00	2.37	0.8
R-45#1	35.7	0.597	7.37	9.76	4.77	3.6
R-45#2	18.4	0.4	4.32	3.04	3.72	3.3
R-50#1	103	0.586	11.5	6.85	8.13	26
R-50#2	3.73	0.307	2.25	2.79	2.0	1.2
R-61#1	10.0	0.195	1.77	9.84	1.84	24
R-61#2	1	0.198	2.2334	1.51	2.4858	1

Identified groundwater types / contaminant sources [5×6]

In the microphone analogy, this is what was said by each person. Each person's speech corresponds to one row of this table.

Source	Cr^{6+} $\mu g/L$	ClO_4^- $\mu g/L$	SO_4^{2-} mg/L	NO_3^- mg/L	Cl^- mg/L	^3H pCi/L
1	**1300**	0	**87**	8.8	**66**	11
2	0.21	0.56	11	0	0.021	**130**
3	0.25	**51**	2	**13**	0.094	0
4	0.24	0	**19**	4	**33**	0.069
5	0.009	0	7	**21**	0	0

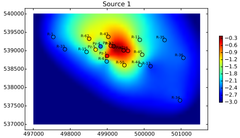
Estimated mixtures at the wells [29×5]



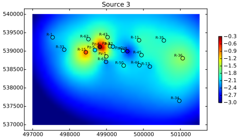
In the microphone analogy, this is how loud each person's voice (column) is when recorded by each microphone (row).

Maps of groundwater types / sources

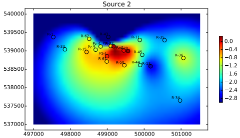
Cr^{6+}, SO_4^{2-}, Cl^-



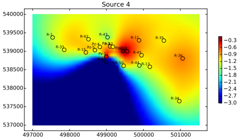
ClO_4^-, NO_3^-



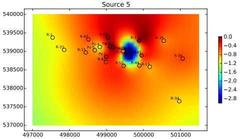
^3H



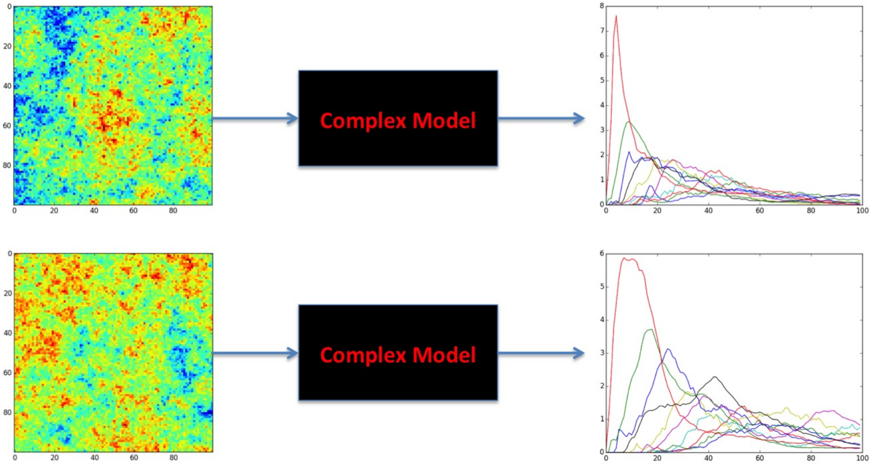
Cl^-, SO_4^{2-}



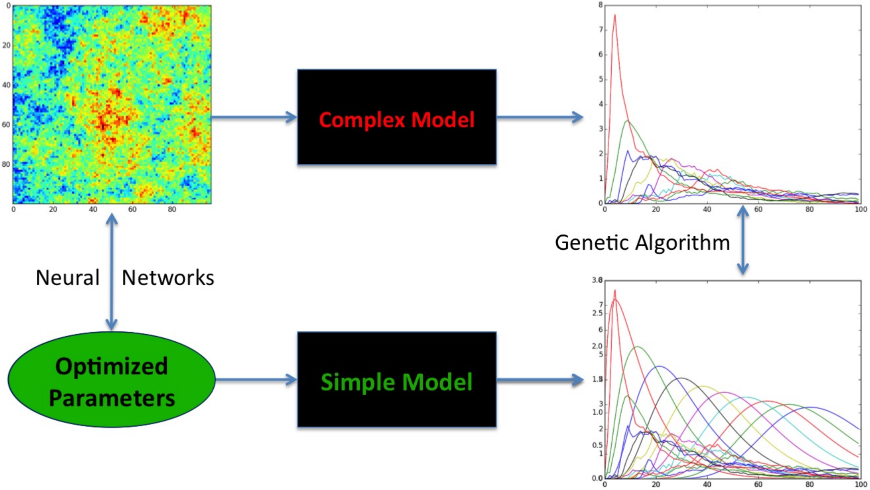
NO_3^-



Complex transport modeling

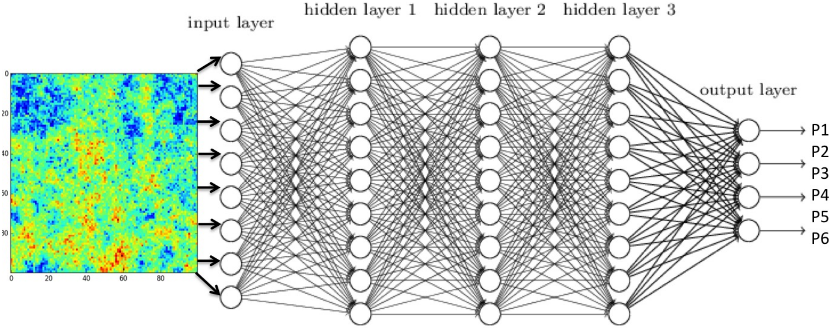


Reduced-order transport modeling



Neural network + analytical solutions

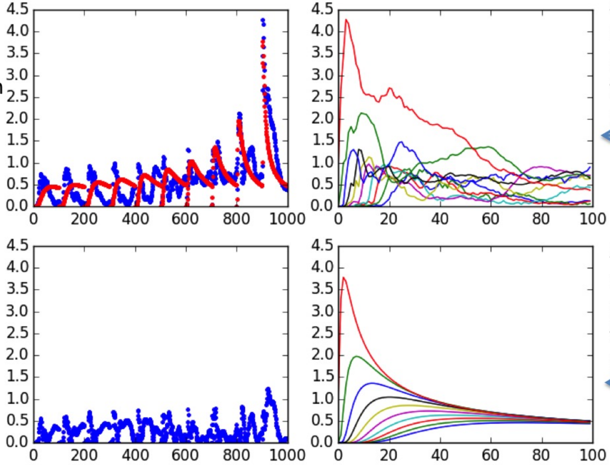
- We use analytical solutions from O'Malley & Vesselinov (AWR, 2014)
- These solutions are implemented in Anasol.jl, part of **MADS**
- A permeability field is fed into a neural network, and the neural network produces a small set of inputs to the analytical model



Results

The two concentration time series superimposed on each other

All the sensors concentrations plotted as a function of time: Complex model



Error between two concentration time series

All the sensors concentrations plotted as a function of time: Simple model

- NMFk applied to groundwater mixing
- Neural networks applied to groundwater transport

Related model and decision analyses presentations at AGU 2016

- Lu, Vesselinov, Lei: Identifying Aquifer Heterogeneities using the Level Set Method (**poster**, Wednesday, 8:00 - 12:00, **H31F-1462**)
- Zhang, Vesselinov: Bi-Level Decision Making for Supporting Energy and Water Nexus (**West 3016**: Wednesday, 09:15 - 09:30, **H31J-06**)
- Vesselinov, O'Malley: Model Analysis of Complex Systems Behavior using MADS (**West 3024**: Wednesday, 15:06 - 15:18, **H33Q-08**)
- Hansen, Vesselinov: Analysis of hydrologic time series reconstruction uncertainty due to inverse model inadequacy using Laguerre expansion method (**West 3024**: Wednesday, 16:30 - 16:45, **H34E-03**)
- Lin, O'Malley, Vesselinov: Hydraulic Inverse Modeling with Modified Total-Variation Regularization with Relaxed Variable-Splitting (**poster**, Thursday, 8:00 - 12:00, **H41B-1301**)
- Pandey, Vesselinov, O'Malley, Karra, Hansen: Data and Model Uncertainties associated with Biogeochemical Groundwater Remediation and their impact on Decision Analysis (**poster**, Thursday, 8:00 - 12:00, **H41B-1307**)
- Hansen, Haslauer, Cirpka, Vesselinov: Prediction of Breakthrough Curves for Conservative and Reactive Transport from the Structural Parameters of Highly Heterogeneous Media (**West 3014**, Thursday, 14:25 - 14:40, **H43N-04**)
- O'Malley, Vesselinov: Groundwater Remediation using Bayesian Information-Gap Decision Theory (**West 3024**, Thursday, 17:00 - 17:15, **H44E-05**)
- Dawson, Butler, Mattis, Westerink, Vesselinov, Estep: Parameter Estimation for Geoscience Applications Using a Measure-Theoretic Approach (**West 3024**, Thursday, 17:30 - 17:45, **H44E-07**)