ZEM: Integrated Framework for Real-Time Data and Model Analyses for Robust Environmental Management Decision Making – 16438

Velimir V. Vesselinov*, Dan O'Malley*, Danny Katzman*
* Los Alamos National Laboratory (LANL), Los Alamos, NM, USA

ABSTRACT

Typically environmental management (EM) problems require analysis of large and complex data sets originating from concurrent data streams with different data collection frequencies and pedigree. These big data sets require on-the-fly analysis, assimilation and integration into a series of models with different complexity for various types of model-based analyses. This is needed to provide fast iterative model analyses based on the latest available data to guide decision-making. Furthermore, the data and model are associated with uncertainties. The uncertainties are probabilistic (e.g. measurement errors) and non-probabilistic (unknowns, e.g. alternative conceptual models characterizing site conditions). To address all of these issues, we have developed an integrated framework for real-time data and model analyses for environmental decision-making called ZEM. The framework allows for seamless and on-the-fly integration of data and modeling results for robust and scientifically-defensible decision-making applying advanced decision analyses tools such as Bayesian-Information-Gap Decision Theory (BIG-DT). The framework also includes advanced methods for optimization that are capable of dealing with a large number of unknown model parameters, and surrogate (reduced order) modeling capabilities based on support vector regression techniques. The framework is predominantly coded in Julia, a state-of-the-art high-performance programing language (http://julialang.org) and incorporates the MADS computational tools (Model Analysis & Decision Support; http://mads.lanl.gov). We demonstrate the application of ZEM for data and model analyses at the LANL Chromium site. The ZEM framework will be released as open-source code and can be applied to any environmental management site.

INTRODUCTION

Los Alamos National Laboratory (LANL) is a research facility operated by the U.S. Department of Energy (DOE) in north-central New Mexico. LANL is currently investigating a chromium (Cr^{6+}) plume in the regional aquifer beneath Sandia and Mortandad Canyons (Fig.1) to ensure contaminants do not threaten human health or the environment. A comprehensive investigation of this plume has been ongoing since 2005 [8,9,10,11]. The Cr^{6+} contamination resulted from infiltration of liquid effluents released from an electric power plant. The contaminated effluents were discharged in Sandia Canyon between 1956 and 1972, when treated sanitary wastewater was used for power-plant cooling; the water was treated with potassium dichromate, phosphate, zinc, and sulfuric acid. The total Cr^{6+} mass release into the Sandia Canyon is estimated at about 54,000 kg with uncertainty bounds between 31,000 to 72,000 kg. During that period, the water flux released in the canyon was about 500 to 1000 m^3/d; similar effluent volumes were probably discharged through 1992. Since 1972, mu-

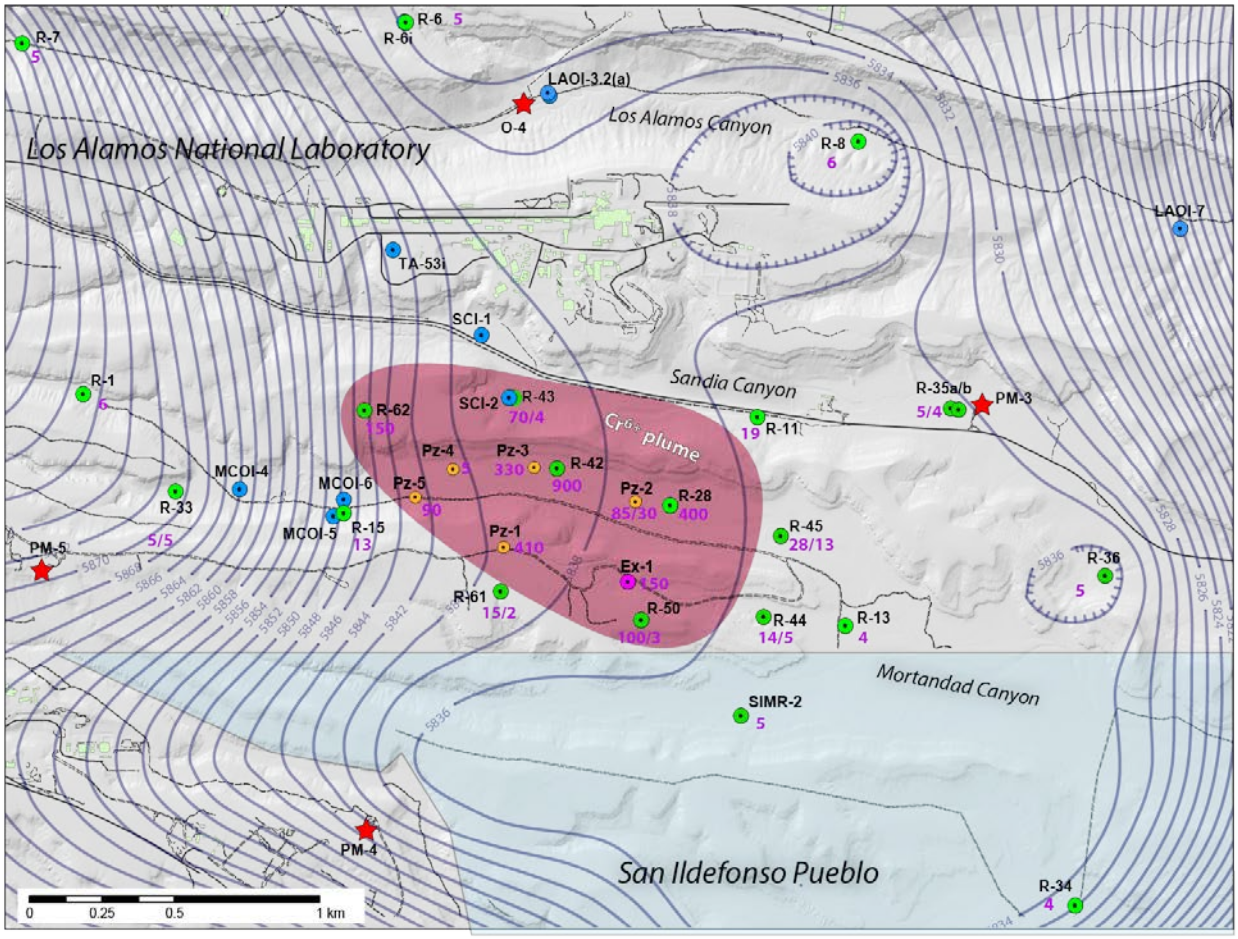

Fig1: Map of the Chromium site representing site conditions circa 2015. Green and blue dots show the locations of aquifer and vadose zone monitoring wells, respectively. Red stars are municipal water supply wells. Recently observed Cr^{6+} concentrations at each regional well are shown in purple [µg/ℓ]; the concentrations at two-screen wells (e.g. R-61) are shown as upper / lower screen values. The Cr^{6+} plume represents an area where concentrations are higher than 50 µg/ℓ (ppb).

nicipal water was used for power-plant cooling and additives containing no chromium were used. The Sandia Canyon effluent also had elevated tritium (^3H) ranging between 4200 pCi/ℓ and 38,500 pCi/ℓ in 1976-77[11]. Since 1992, the effluent released in Sandia Canyon has been 1000 to 1500 m^3/d. Currently, the Sandia Canyon effluent volumes are substantially less than they were and are on the order of 200-400 m^3/d. Cr^{6+} migration in the subsurface is also influenced by effluent discharges in two neighboring canyons: Mortandad and Los Alamos (Fig.1). Since 1963, wastewater was released to Mortandad Canyon containing perchlorate (ClO_4^-), nitrate (NO_3^-), 1,4 dioxane ($C_4H_8O_2$), and tritium (^3H). These tracers were detected in perched vadose zones beneath the canyon. Some of these tracers were collocated with contaminant released in Sandia Canyon and detected in the regional aquifer. The Mortandad releases have historical volumes about 10 times less than the Sandia releases. Since August 2010, Mortandad Canyon discharges have effectively stopped. Past effluent releases in Los Alamos Canyon are characterized by elevated ^3H concentrations. ^3H with a potential Los Alamos Canyon origin has been co-detected with Cr^{6+} in the

perched and regional groundwater beneath Sandia and Mortandad Canyons [11]. Currently, no LANL outfalls release water into Los Alamos Canyon watershed, but significant natural infiltration occurs. The natural infiltration recharges along Sandia and Mortandad Canyons are substantially less than the natural infiltration recharge along Los Alamos Canyon.

Cr^{6+} concentrations in the regional aquifer exceed 50 µg/ℓ (the New Mexico groundwater standard) at several monitoring wells (Fig.1); the highest concentration is ~1000 µg/ℓ (ppb). ClO_4^- is also present in the aquifer at concentrations exceeding the established screening level of 4 µg/ℓ (Fig.1). 1,4 dioxane is present in perched-intermediate groundwater beneath Mortandad Canyon at concentrations exceeding the groundwater standard, but the contaminant is currently not detected in the regional aquifer [11].

A map of the LANL contamination related to a chromium (Cr^{6+}) plume in the regional aquifer beneath Sandia and Mortandad Canyons is presented in Fig. 1 [11].

SITE CONCEPTUAL MODEL

The site conceptual model describes the physical and biogeochemical processes controlling the movement of groundwater and contaminants in the environment. The current conceptual model is explained in detail in [11], and supported by multiples lines of evidence; in general, the site conceptual model that was proposed in the 2012 report is still consistent with the recently collected data. However, there are also uncertainties in the conceptual model which are further discussed and analyzed below. The establishment and the ongoing testing of the current conceptual model involved a series of field, laboratory and modeling analyses [11].

At least 25% of the released chromium mass is currently in Sandia Canyon sediments (~18,000 kg) with about 15,000 kg of this mass residing in the wetland at the head of the canyon. The chromium in the canyon sediments (including the wetland) is predominantly in reduced Cr^{3+} form [11]. Alluvial groundwater in Sandia Canyon is recharged daily by effluent discharges and periodically by stormwater; most of the infiltration occurs downgradient from SCI-1 (Fig.1) with rates potentially as high as 1-6 m/a. Currently, the groundwater infiltrating into the subsurface along the canyon bottom has Cr^{6+} concentrations substantially below the groundwater standard (50 µg/ℓ). The infiltration along Mortandad Canyon occurs predominantly in the area near MCOI-4 (Fig.1).

Water infiltrating along the three canyons migrates through the vadose zone mostly vertically driven by gravity. However, perching horizons divert some of the percolating water laterally along saturated zones extending between the three canyons. The perching horizons are observed at the top, within, and at the bottom of lavas embedded between vadose-zone sediments (Puye Formation). The flow direction in the perched zones is predominantly to the south/southwest; however, some of the flow in the vadose zone may also have been occurring to the north from Mortandad Canyon towards Sandia Canyon. Flows with opposite orientations might be occurring along different pathways at different elevations within the vadose zone. The flows are also expected to be influenced by the long-term variability in the infiltration rates along the canyons. The travel-times from the ground-surface to the perched horizons might be as short as 1 to 5 years. Between the canyons, some of the laterally-diverted

water continues to percolate vertically below the perched zone toward the regional aquifer; the groundwater potentially flows along preferential flowpaths that occur along some kind of hydraulic windows through the perched zone. The locations of these penetration zones are unknown. As a result of lateral diversion along perching horizons, the greatest mass Cr^{6+} contamination reaches the aquifer to the south of Sandia Canyon, where it initially infiltrated in the subsurface. The reduction of water released in Sandia Canyon in the next few years will not have an immediate effect on aquifer recharge; the current rate of recharge is expected to continue for 10 years or more due to the thickness (~300 m) and flow complexity of the vadose zone. There is limited information about Cr^{6+} concentrations in the vadose zone; currently, the highest observed Cr^{6+} concentration in the vadose zone is less than 600 µg/ℓ, lower than the highest concentration observed in the regional aquifer (~1000 µg/ℓ). Considering the contaminant mixing in the aquifer, the Cr^{6+} concentrations in the lower vadose zone that cause the contamination in the aquifer must be substantially higher than 1000 µg/ℓ.

The regional aquifer is complex and heterogeneous including confined and unconfined zones [11]. The shallow portion of the aquifer (near the water table) is predominantly under phreatic (unconfined) conditions; the thickness of the phreatic zone is variable and approximately 30 to 50 m below the regional water table. The deep portion of the aquifer is predominantly under confined conditions, and it is stressed by municipal water-supply pumping (the nearby pumping wells are shown in Fig.1). The pumping appears to have little impact on the groundwater flow and transport directions along the regional water table in the study area. Model analyses indicate full to partial confinement of the deep portion of the aquifer at the chromium site is required to calibrate to hydraulic heads observed in the monitoring wells. However, the poor hydraulic communication between the deep and shallow aquifer section does not preclude the possibility of Cr^{6+} migration between the shallow and deep aquifer zones. Between the two zones, the hydraulic gradient has a strong downward vertical component because of the water-supply pumping, creating the possibility for downward Cr^{6+} flow via “hydraulic windows,” although these have not been observed.

The groundwater flow in the aquifer is generally from west to east. The hydraulic gradients are high to the west (due to elevated mountain front recharge) and low to the east (close to the Rio Grande); however, the gradients are relatively low in the Cr^{6+} plume area (~0.001 m/m). The juxtaposition of zones with steep and flat hydraulic gradients complicates the analyses of the flow directions in the study area. The contour lines of the elevation of the water table (circa 2012) in the study area are shown in Fig.1. The shape of the regional water table is complex; it is affected by aquifer heterogeneity and infiltration recharge (considering the large volumes of surface water flowing along Sandia and Los Alamos Canyons). The water table in the study area is also slightly influenced by water-supply pumping; however, these effects are relatively small and seasonal (the biggest water-level declines are observed in the summer months when the water demand is the highest). However, current water-level records suggest limited pumping impact on the current shape of the regional water table. The low hydraulic gradient near R-28(Fig.1) is potentially caused by the high effective permeability of the sediments in this portion of the aquifer. The water-level data suggest infiltration mounding of the water table near wells R-8, R-36,

and R-42 (Fig.1). The R-8 and R-36 mounds are potentially caused by Los Alamos Canyon infiltration. The mounding near R-42 is expected to be related to infiltration recharge of the aquifer carrying the historic Cr^{6+} contamination from Sandia Canyon. The mounding may cause a portion of the Cr^{6+} plume to migrate to the north, west and south of contaminant arrival locations, enhancing the contaminant mixing in the aquifer. The size and magnitude of the water-table mounds depend on the distribution of infiltration flux and local hydrogeological properties.

Groundwater flow and contaminant transport directions in the phreatic zone are generally determined by the gradient of the regional water table (Fig.1). However, the aquifer appears to be highly heterogeneous and anisotropic due to hydrostratigraphic bedding and channelization of sedimentary deposits. Aquifer heterogeneity is expected to cause local 3D deviations in flow directions not captured by the water-level data. Aquifer heterogeneity may focus a predominant portion of the contaminant flow into aquifer areas having high permeability. In addition, advective flow paths of contaminant migration may not be strictly perpendicular to the equipotential water-table lines (i.e., parallel to the hydraulic gradients). Deviations may occur because of anisotropy of aquifer materials caused by sedimentation processes and tilting of the hydrostratigraphic units to the south-southwest with dip of ~10^{o}. These factors may cause the flow and hydraulic-gradient vectors to deviate (if the hydraulic gradient is not coincident with the principal directions of the anisotropic permeability tensor) [11]. Analyses of long- and short-term pumping tests conducted in the study area demonstrate the aquifer is heterogeneous. Recent pumping test data suggest the existence of high- and low-permeable zones; however, the interpretation is uncertain because of the small magnitude of the pumping-test drawdowns in observation wells [11]. As shown in Fig.1, the hydraulic gradient suggests the flow direction near R-42 and R-28 is east-southeast (with possible diversions to the southeast or northeast). However, there is uncertainty in this flow-direction estimate because of: (1) the relatively small magnitude of the hydraulic gradient in this area and (2) the 3D and heterogeneity effects that impact the pressure measurements at individual well screens (screens have different depths and lengths; none of the screens in the study area straddles the regional water table).

The footprint of contaminant arrival locations from the vadose zone to the top of the aquifer is unknown. All the existing site information including the shape of the Cr^{6+} plume (Fig.1) suggests that in the area near R-42 and R-28 either there is (1) one relatively large and continuous breakthrough area (potentially with nonuniformly distributed contaminant fluxes) or there are (2) several breakthrough areas (with different contaminant fluxes). From the breakthrough areas, the Cr^{6+} plume migration is expected to be predominantly east-southeast toward wells R-44 and R-45 (Fig.1). This general direction of Cr^{6+} transport appears to parallel the separate ClO_4^- plume originating from Mortandad Canyon infiltration. The ClO_4^- plume is defined by observations in wells R-15, R-61, and R-50 (Fig.1). However, the shapes of the contaminant plumes and the groundwater flow directions based on hydraulic gradients in Fig.1 are not entirely consistent; this suggests that aquifer anisotropy and heterogeneity and/or infiltration mounding effects not captured in the existing water-level data may have important impacts on contaminant migration and plume shapes.

Cr^{6+} migration in the vadose zone and the regional aquifer is impacted predominantly by advection, dispersion, diffusion, retardation and redox reactions (specifically reduction). The contaminant mass moves predominantly by advection with the groundwater flow. Mechanical dispersion within the vadose zone and aquifer is enhanced by the pronounced heterogeneity of the subsurface materials. The diffusion of Cr^{6+} between zones with higher and lower permeability is expected to retard the effective transport velocity; after the Cr^{6+} concentrations in the more permeable zones decrease due to diminishing contaminant source (the release was between 1956 and 1972), back diffusion of Cr^{6+} from low into high permeability zones is expected to retard the long-term decline of Cr^{6+} concentrations in the vadose zone and the regional aquifer [11]. The geochemical interactions of chromium with subsurface materials are expected to have important effects on contaminant transport by causing Cr^{6+} reduction and retardation. The chromium retardation from adsorption or chemical reactions between Cr^{6+} in the solid and water phases does not reduce the contaminant mass but slows down Cr^{6+} contaminant transport. However, chromium reduction from chemical reactions converting Cr^{6+} into less toxic and less mobile Cr^{3+} effectively decreases the contaminant mass in groundwater. The reduced Cr^{3+} precipitates out and becomes functionally immobile; as a result, it is no longer considered a portion of the contaminant plume. Reverse oxidization reactions of Cr^{3+} to Cr^{6+} at the site are not well understood; currently, they are considered to be negligible because of potentially slow kinetics of the involved geochemical reactions and lack of oxidant minerals such as MnO_2 in the aquifer sediments. The Cr^{6+} reduction in the environment is demonstrated using Cr stable isotope analyses [11,5]. The isotope data indicate reduction has occurred between the surface and the regional aquifer, but they do not show significant Cr^{6+} reduction in the aquifer; however, analytical uncertainties may account for the anthropogenic Cr^{3+} mass due to reduction in the range between 1% and 10% of the current Cr^{6+} mass in the aquifer. Although the reducing capacity of the lavas in the vadose zone is significant, either the kinetics of reduction are relatively slow or the reduction capacity has been overwhelmed along the predominant flow paths by the large mass and high historic concentrations of Cr^{6+} that passed through the basalt. The analyses of stable isotope data for evaluation of the Cr^{6+} reduction are further complicated due to existence of natural and anthropogenic chromium in the subsurface; the natural (non-anthropogenic) Cr^{6+} concentrations in the regional aquifer are between 5-8 µg/ℓ. Further analyses are being conducted to better constrain uncertainties associated with Cr^{6+} reduction at the site. Estimates of chromium source and mass distribution in the environment including the predicted Cr speciation and uncertainty bounds are presented in [11]. The estimates for the contaminant mass in the vadose zone are based on analyses of pore water and rock samples from boreholes and numerical model simulations.

The site data suggest groundwater infiltrates from the ground surface to the top of the regional aquifer (vertical distance ~300 m) with infiltration rates on the order of 1 to 6 m/a, and travel times between 5-60 years; the effective groundwater flow velocity through the vadose zone is ~5-60 m/a. The uncertainty bound of groundwater travel time through the vadose zone accounts for potential complex flow paths with low and high groundwater flow velocities. The mean travel time for a dominant portion of the groundwater infiltrating trough the vadose zone is expected to be ~20-40 years. Cr^{6+} transport through the vadose zone is expected to be retarded because the Cr^{6+} is

slowed by diffusion between high and low permeability zones, and geochemical reactions with subsurface materials. Based on hydraulic gradients and properties of the aquifer, groundwater flow in the aquifer is expected to occur with pore-water (linear) velocity of about 40-50 m/a (discussed further below); the distance between R-28 and R-44/R-45 is ~400 m; therefore the travel time will be on the order of 10 years. These velocity estimates are generally consistent with the timing of the Cr^{6+} releases (1956-1972) and the Cr^{6+} detection in the aquifer (since 2004). The currently observed contamination in the aquifer (Fig.1) potentially represents a steady-shape plume caused by equilibrium between the contaminant flux arriving from the vadose zone and the groundwater flow in the aquifer. The concentrations at most of the wells have been relatively steady; e.g. R-28 concentration has been around 400 µg/ℓ since 2004. However, there are also concentration transients with differing magnitudes and trends that may be caused by local variations in the geochemical conditions (including borehole drilling effects) and plume non-uniformity. The concentration transients continue to be closely monitored, analyzed, and compared to model predictions to assess plume stability and detect any unexpected fluctuations.

METHODS

The ZEM framework is designed to provide on-the-fly analysis, assimilation and integration of the site data into a series of tools and models to perform various types of data- and model-based analyses. This is needed because environmental management (EM) at our site requires analysis of large and complex data sets originating from concurrent data streams with different data collection frequencies and pedigree. ZEM also represents a general decision analysis framework consistent with ideas discussed in [6,39,19].

The data streams fed into ZEM include (1) hydraulic and barometric pressure data collected by transducers and streamed in real-time to web servers, (2) pumping rate records provided on a monthly basis (with daily resolution), and (3) geochemical concentration data from groundwater samples provided with frequencies varying from weekly to quarterly (in the future, some of the geochemical data will be coming in real-time as well as using in-situ probes). Every well at the chromium site has a pressure transducer, and the groundwater samples are taken manually from all the wells with varying frequency.

The hydraulic data streams include information for the ambient and stressed flow conditions. The stresses are caused by on-going municipal water-supply pumping in the area and specially designed pumping tests conducted at the site. Similarly, the geochemical data provides information about the site conditions during ambient flow as well as during field pumping and tracer tests. In the future, the hydraulic and geochemical data will also provide information about the performance of interim pump and treat activities at the site.

The analyses of these data will provide information not only for further characterization of the aquifer properties but also about the effectiveness of remedial activities. The real-time monitoring and analyses of the hydraulic and geochemical data during the pump and treat activities can provide important information to optimize the remedial performance by adjusting pumping/injection rates as wells as pumping/injection location. The interim pump and treat system at the site is currently

designed in a way that allows different wells to be used for pumping or injection depending on the performance. All these data streams have to be processed and analyzed by the models as soon as possible to provide nearly real-time input to the site managers.

The tools and models applied for data- and model-based analyses within ZEM have different data needs and complexity depending on the performed task. Most of the model-based analyses are interfaced using the code MADS (Model Analyses for Decision Support; http://mads.lanl.gov; [18]). MADS is an open-source high-performance computational framework for model-based decision support employing system and physics simulation models. MADS performs various types of model analyses, including sensitivity analysis, parameter estimation, uncertainty quantification, model calibration, model selection, model averaging, and decision analyses. MADS also includes advanced methods for optimization that are capable of dealing with a large number of unknown model parameters (for example, on the order of 10^6) and a large number of model calibration targets (for example, on the order of 10^7). Since the analyses include large number of unknowns and targets, the applied optimization methods in MADS are applied in a way that accounts for the information content in the calibration data (including measurement errors) to avoid overfitting. Such problems may arise in the case of tomographic analysis for characterization of aquifer heterogeneity. To solve this kind of problems, we developed novel optimization methods based on Krylov space techniques [15,16] and tomographic techniques based on randomized order-reduction approaches [13]. MADS also incorporates surrogate (reduced order) modeling capabilities based on support vector regression techniques. MADS also includes tools for simulating groundwater flow [22,23,24] and contaminant transport [27,28]

The existing tools in MADS are designed to be adaptive allowing for minimum user inputs to perform the selected analyses. Furthermore, most of the tools can perform both global (exploring the entire parameter space) and local (exploring only the parameter space in the vicinity of a particular optimal solution) analyses. This is important because frequently the EM problems are ill-posed; these are problems that can have multiple plausible solutions. In this way, the global analyses can be applied to explore the entire parameters space (for example, to search for a global optimum), while the local analyses can explore local optima in the parameter space. The ill-posedness of the EM problems present substantial challenges for uncertainty and decision analyses. Soon, the MADS code will also include capabilities for performing a decision-based experimental design. In addition, MADS includes tools for data-based analyses which do not use models. For example, MADS can be applied to perform model-free inversion of data based on the NMFk (Nonnegative Matrix Factorization coupled with the *k*-means algorithm) machine-learning algorithm [1]; we call this approach model-free inversion.

Various demos, execution examples, and a tutorial are available at the MADS website [18]. There are two code versions of MADS that are currently maintained: one is written in C/C++, and the other is written in Julia (both are open source). Most of the model analyses can be performed with both versions of the code. However, the Julia version code is being developed more actively at the moment. Julia is a novel, high-level, high-performance dynamic programming language for technical compu-

ting (http://julialang.org). It provides substantial benefits over the other existing programming languages and allows for relatively easy and fast code development.

In the ZEM framework, the data can be applied as both soft (e.g. acceptable bounds) and hard (e.g. calibration targets) constraints in the models. The data and models in the ZEM framework are associated with uncertainties and these uncertainties can be probabilistic (e.g. measurement errors) and non-probabilistic (unknowns, e.g. alternative conceptual models characterizing site conditions). To address all of these issues, we have developed robust and scientifically-defensible decision analyses tools based Bayesian-Information-Gap Decision Theory (BIG-DT) [29,26,25]. The BIG-DT is implemented in MADS as well.

Similarly to MADS, the ZEM framework is coded predominantly in Julia. Currently, ZEM also includes Perl, Python and BASH scripts. The long-term goal is that all the coding will be transferred to Julia.

RESULTS AND DISCUSSIONS

Three-dimensional tomography of aquifer heterogeneity

One of the important aspects of any hydrogeologic study is the characterization of the aquifer heterogeneity. Any hydrogeologic problem related to environmental management requires information about the uncertainty in the groundwater flow and transport that is caused by the aquifer heterogeneity. The aquifer heterogeneity is very difficult to estimate. It is even more difficult to bound the uncertainties in these estimates. At the next level, it is even more difficult to bound the uncertainties in the model predictions about contaminant transport taking into account the uncertainties in the aquifer heterogeneity. In general, we are addressing all these issues applying various different approaches [2,3,4,25,26,29]. Here, we demonstrate a computationally challenging problem of the three-dimensional hydraulic tomography of aquifer heterogeneity based on observed pressures at monitoring wells caused by pumping at a series of water-supply and test wells. There are several wells located near the contaminated site that provide municipal water supply (Fig.1). These wells have irregular pumping regimes that depend on the well conditions and supply needs. In addition, a series of pumping tests are conducted at some of the site wells (e.g. CrEX-1, R-28, R-42, R-44, R-45, etc.; Fig.1). The temporal fluctuations of the pumping rates at all these wells are known. There are sufficient differences in the respective pumping regimes to allow for fingerprinting the pressure impacts observed in monitoring wells. This fingerprinting is a part of an inverse modeling process that targets the matching of the observed pressure transients at the monitoring wells by a 3D numerical groundwater-flow model. The model simulates the pumping-rate fluctuations at each pumping well for a period of 2 years. Daily pressure fluctuations observed at all the site monitoring wells are applied as calibration targets. There are more than 7 pumping wells and more than 20 observation points. In total, there are more than 18,000 calibration targets. Barometric and tidal pressure effects manifested in the collected pressure records are also automatically removed. The adjustable model parameters during the inversion are the spatial 3D fields of hydraulic conductivity and storativity. The unknown parameter fields are initially assumed to be homogenous and isotropic. Various approaches are applied to parameterize the 3D

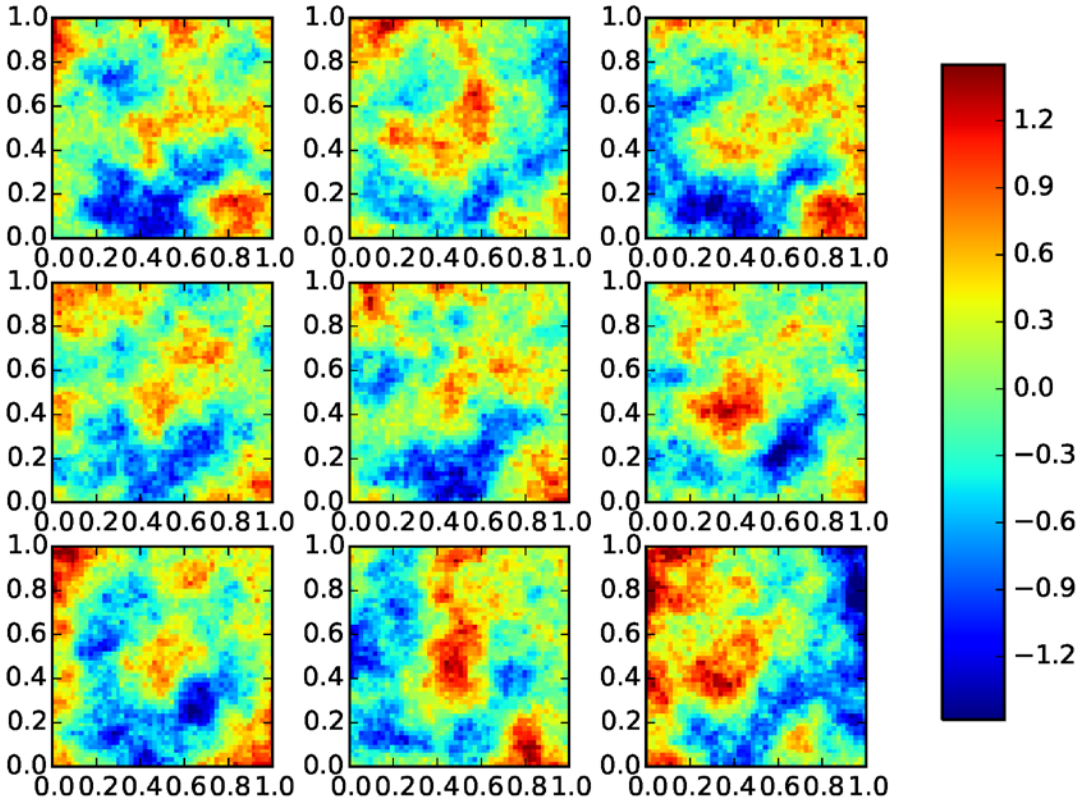

Fig.3: A synthetic example of a series of tomographic images of different hydraulic conductivity ($\log_{10}$[m/s]) fields that are equally consistent with the observed pressure transients.

aquifer heterogeneity including pilot points [12], principle component geostatistical analysis [14], and randomized geostatistical analysis [13]. More than 200 model parameters are calibrated in some of the inverse analyses. The goal of this analysis is not only to characterize aquifer heterogeneity but also to capture the temporal variability in the hydraulic gradients due to pumping. This provides information about the potential impact of the aquifer pumping on the site contaminant transport. The research is ongoing. Figure 2 shows example model-analysis results where observed pressures are reproduced by a calibrated numerical model for a subset of observation wells. The applied modeling approach was capable of providing excellent matches between the observed and simulated pressures. However, this does not suggest that the estimated model parameters are well constrained. Additional site work demonstrated that numerous plausible parameter sets are capable of reproducing the observed pressure data equally well (where the plausible parameter sets represent the different plausible heterogeneous fields of hydraulic conductivity and plausible storativity). For example Fig.3, shows the results from a synthetic tomographic study consistent with the site conditions. To produce the results in Fig.3, a number or pressure observations are synthetically conducted and are applied to characterize aquifer heterogeneity. However, more than 100 permeability fields (some of which are very different as seen in Fig.3) are estimated and all of them are similarly consistent with the observed pressure transients. This demonstrates the unidentifiability

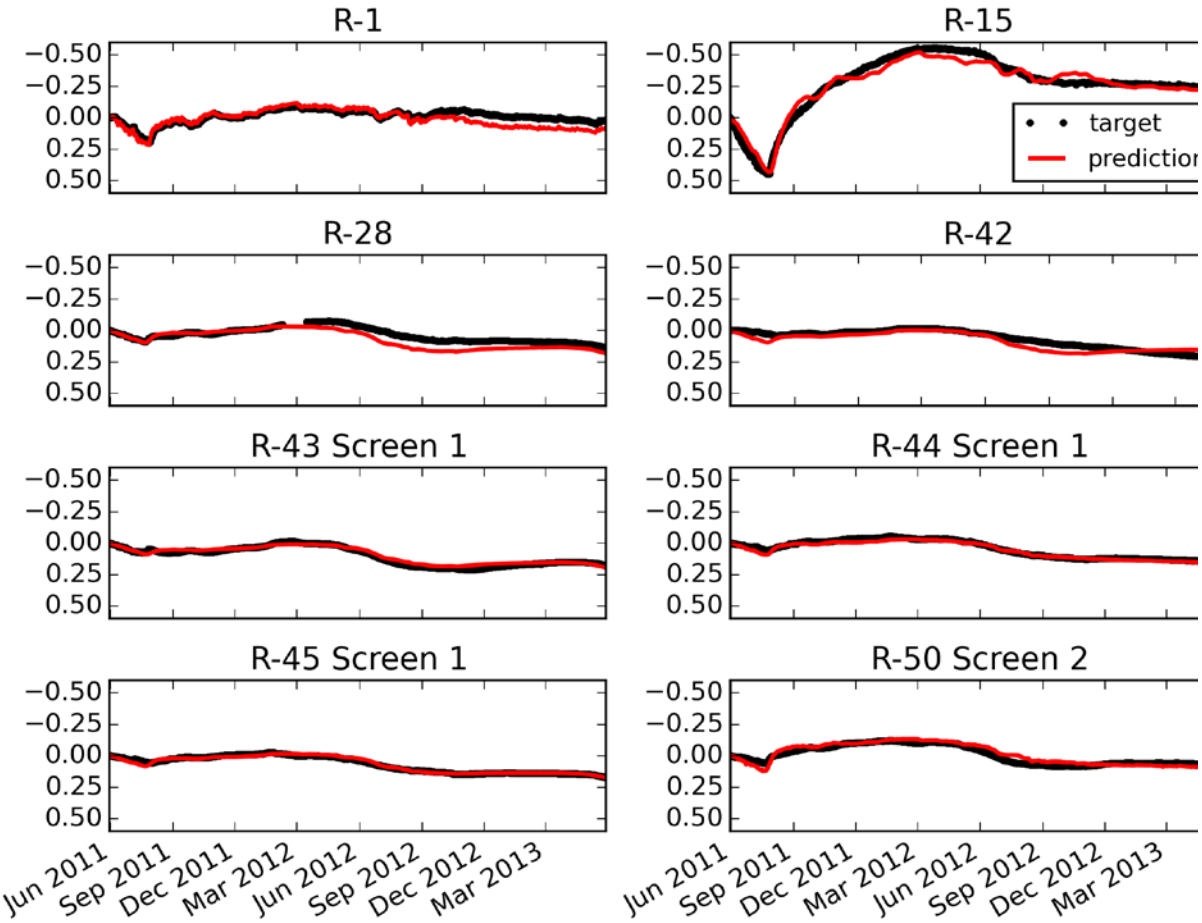

Fig.2: Predicted (red lines) vs observed (black dots) pressures at various monitoring locations.

of the inverse problem, and suggests further challenges related to ongoing model analyses for uncertainty quantification and remedy selection.

Contaminant source identification

Frequently, at the contamination sites, the groundwater in the aquifer is a mixture of waters with different origins (sources) that are commingled in the aquifer; several of these groundwater recharge sources might include contaminants. Typically, all these sources will have different geochemical signatures due to differences in their origins and flowpaths through the subsurface before infiltrating in the aquifer. The identification of the contamination/infiltration sources causing the observed geochemical concentrations in the aquifer can be very challenging at sites where complex physical and chemical processes occur.

Source identification can be complicated because (1) some of these sources may have similar geochemical signatures, (2) some of the sources may geochemically interfere with each other, and (3) groundwater transport through the subsurface (from the entry point at the ground surface to the observation point in the aquifer) may be impacted by various physical and chemical processes (e.g., diffusion, dispersion,

Table 1: Geochemical concentrations representing mixtures of groundwaters with unknown origins.

Well	Cr^6 [µg/L]	ClO_4^- [µg/L]	SO_4^{2-} [mg/L]	NO_3^- [mg/L]	Cl^- [mg/L]	^3H [pCi/L]
CRPZ-1	406.22	1.84	47.846	17.07	35.401	101.397
CRPZ-2a	83.89	0.88	71.155	14.42	66.436	121.013
CRPZ-2b	35.01	0.419	6.2918	4.24	7.582	2.061
CRPZ-3	338.88	1.21	33.967	23.60	21.853	24.184
CRPZ-4	5.69	63.7	5.8175	17.90	3.0975	11.346
CRPZ-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
R-62	116	0.828	13.7	7.83	7.87	16

sorption, retardation, precipitation, etc.). To address all these issues, the source characterization may require calibration of a numerical model simulating these complexities against the observed geochemical data. Here, we apply an alternative approach based on a novel model-free machine-learning algorithm for Blind Source Separation (BSS).

BSS methods are based on unsupervised (objective and adaptive) machine-learning algorithms. A classical BSS conundrum is the “cocktail-party'” problem [1]: several microphones are recording the sounds in a room (e.g., music, conversations, noise, etc.), and each microphone is recording a mixture of all the available sounds. The BSS goal is to “unmix” and reconstruct the original sound sources from the recorded mixtures. For our groundwater-contamination problem, the BSS goal is to retrieve the original geochemical concentrations of the contamination/infiltration sources, **S** (a matrix with size $\boldsymbol{p} \times \boldsymbol{k}$), based only on the observed geochemical concentrations in groundwater samples, **D** ($\boldsymbol{p} \times \boldsymbol{m}$), collected at a set of monitoring wells, where $\boldsymbol{p}$ is the number of the wells, $\boldsymbol{k}$ is the number of unknown sources, and $\boldsymbol{m}$ is the number of geochemical components observed at the wells. We initially know only the matrix **D** of the concentrations observed at the monitoring wells (Table 1); **D** is formed by mixing $\boldsymbol{k}$ original groundwater sources of unknown geochemical composition **S**. These original sources are blended by an unknown mixing matrix (representative of some unknown groundwater mixing model), **W** ($\boldsymbol{k} \times \boldsymbol{m}$) and produce the observations: $\mathbf{D} = \mathbf{S} * \mathbf{W} - \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ denotes a noise or errors in the measurements (also un-

Table 2: Identified groundwater types with different geochemical composition; the key components associated with each type are labeled; the first three types represent contaminant sources.

Source #	Cr^{6} [µg/L]	ClO_4^- [µg/L]	SO_4^{2-} [mg/L]	NO_3^- [mg/L]	Cl^- [mg/L]	^3H [pCi/L]
1	**2000**	0	**140**	14	**100**	18
2	0.33	0.88	17	0	0.333	**210**
3	0.39	**80**	3.1	**21**	0.15	0
4	0.36	0	**30**	6.8	**53**	0.11
5	0.014	0	11	**33**	0	0

known). Since both **S** and **W** are unknown (we do not know even the exact size of these matrices, because we do not know ***k***), the main difficulty is that our groundwater-contamination BSS problem is ill-posed. To address this, we have developed a novel BSS method coupling Non-negative Matrix Factorization (NMF) with *k*-means clustering algorithm, and called this new framework NMFk [1]. Here, we apply NMFk using the open-source code MADS [35]. NMFk is applied for analysis and deconstruction of groundwater geochemistry observed in the regional aquifer beneath the Los Alamos National Laboratory (LANL) site for characterization of contaminant sources. Our preliminary results suggest that 5 original groundwater sources with different geochemical composition are mixed in the aquifer; potentially 3 of them are associated with contaminant (Cr^{6+}, ClO_4^- and ^3H) releases (Table 2). The source concentrations estimated by NMFk are consistent with more complicated inverse analyses using numerical models applied to solve this problem. The results are also consistent with Principal Component Analysis of the same data. The NMFk results will be applied as input to numerical models. In this way, instead of calibrating against all the geochemical data, the numerical models would be calibrated against the NMFk predicted geochemical mixtures.

CONCLUSIONS

The ZEM framework has been developed for efficient coupling of data, models and decisions for environmental management problems. The framework includes various codes including MADS and PFLOTRAN. ZEM can be applied to perform various types of analyses including 3D tomographic imaging of aquifer heterogeneity, characterization of contaminant sources, and decision analyses for selection of remedial scenarios. ZEM has been applied to perform various types of data- and model-based analyses for the LANL Chromium site.

REFERENCES

1. Alexandrov, B., Vesselinov, V.V., Blind source separation for groundwater level analysis based on non-negative matrix factorization, *Water Resources Research*, doi: 10.1002/2013WR015037, 2014.
2. Barajas-Solano, D. A., Wohlberg, B., Vesselinov, V.V., Tartakovsky, D. M., Linear Functional Minimization for Inverse Modeling, *Water Resources Research*, doi: 10.1002/2014WR016179, 2015.
3. Harp, D., Vesselinov. V.V., An agent-based approach to global uncertainty and sensitivity analysis, Computers & Geosciences, doi:10.1016/j.cageo.2011.06.025, 2011.
4. Harp, D., 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.
5. 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 Los Alamos National Laboratory groundwater, Chemical Geology, doi: 10.1016/j.chemgeo.2014.02.022, 2014.
6. Janssen, Ron, and Marjan van Herwijnen, (1992), *Multiobjective decision support for environmental management*. Dordrecht: Kluwer.
7. Koch, R., and S. Schmeer, 2008, Groundwater level status report for 2008, Tech. Rep. LA-14397-PR, Los Alamos National Laboratory, Los Alamos, NM.

8. LANL, 2006. Interim Measures Investigation Report for Chromium Contamination in Groundwater, LA-UR-07-6018, Los Alamos, New Mexico.
9. LANL, 2008. Fate and Transport Investigations Update for Chromium Contamination from Sandia Canyon, LA-UR-08-4702, Los Alamos, New Mexico.
10. LANL, 2009. *Investigation Report for Sandia Canyon*, LA-UR-09-6450, Los Alamos, New Mexico.
11. LANL, 2012. *Phase II Investigation Report for Sandia Canyon*, LA-UR-12-24593, Los Alamos, New Mexico.
12. Lavenue, M., & de Marsily, G. Three-dimensional interference test interpretation in a fractured aquifer using the Pilot Point Inverse Method. *Water Resources Research*, doi:10.1029/2000WR000289, 2001.
13. Le, E., O'Malley, D., Vesselinov, V.V., Bui-Thanh, T., Characterization of medium heterogeneity with a fast randomized geostatistical approach (RGA) for inverse problems, *Water Resources Research*, (submitted), 2015.
14. Lee, J., Kitanidis, P., Large- scale hydraulic tomography and joint inversion of head and tracer data using the Principal Component Geostatistical Approach (PCGA). *Water Resources Research* 50, 2014.
15. Lin, Y., O'Malley, D., Vesselinov, V.V., Computationally Efficient Levenberg-Marquardt Algorithms Based on Krylov Subspace Approximation, *SIAM-CS*, (submitted), 2015.
16. Lin, Y., O'Malley, D., Vesselinov, V.V., A Computationally Efficient Parallel Levenberg-Marquardt Algorithm for Highly Parameterized Inverse Modeling, *Water Resources Research*, (submitted), 2015.
17. Lu, Z., Vesselinov, V.V., Analytical Sensitivity Analysis of Transient Groundwater Flow in a Bounded Model Domain using Adjoint Method, *Water Resources Research*, doi: 10.1002/2014WR016819, 2015.
18. MADS, v.1.1.15, 2015, Integrated open-source computational framework for Model Analysis & Decision Support (MADS), http://mads.lanl.gov.
19. Matthies, M., Giupponi, C., & Ostendorf, B. (2007). Environmental decision support systems: Current issues, methods and tools. *Environmental Modeling & Software*, *22*(2), 123-127.
20. 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, *Water Resources search*, doi: 10.1002/2015WR017295, 2015.
21. Miller, T.A., Vesselinov, V.V., et al., Integration of geologic frameworks in meshing and setup of computational hydrogeologic models, Pajarito Plateau, New Mexico Geologic Society Book, pp. 492-499, 2007.
22. 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.
23. Mishra, P.K., Vesselinov, V.V., Kuhlman, K.L., Saturated–unsaturated flow in a compressible leaky-unconfined aquifer, *Advances in Water Resources*, doi: 10.1016/j.advwatres.2012.03.007, 2012.
24. Mishra, P.K., Vesselinov, V.V., Neuman, S.P., Radial flow to a partially penetrating well with storage in an anisotropic confined aquifer, *Journal of Hydrology*, doi: 10.1016/j.jhydrol.2012.05.010, 2012.
25. O'Malley, D., & Vesselinov, V.V., A combined probabilistic/non-probabilistic decision analysis for contaminant remediation, Journal on Uncertainty Quantification, SIAM/ASA, doi: 10.1137/140965132, 2014.
26. O'Malley, D., & Vesselinov, V.V., Groundwater remediation using the information gap decision theory, Water Resources Research, doi: 10.1002/2013WR014718, 2014.
27. O'Malley, D., Vesselinov, V.V., Analytical solutions for anomalous dispersion transport, *Advances in Water Resources*, doi: 10.1016/j.advwatres.2014.02.006, 2014.
28. 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.
29. 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.
30. U.S. National Research Council, Alternatives for Managing the Nation's Complex Contaminated Groundwater Sites, National Academies Press, Washington, D.C. ISBN 978-0-309-27874-4, 2012.
31. Vesselinov, V.V., 2004, An alternative conceptual model of groundwater flow and transport in saturated zone beneath the Pajarito Plateau, LA-UR-05-6741, Los Alamos, New Mexico.
32. Vesselinov, V.V., D.H. Harp, Adaptive hybrid optimization strategy for calibration and parameter estimation of physical process models, Computers & Geosciences, doi: 10.1016/j.cageo.2012.05.027, 2012.
33. Vesselinov, V.V., Harp, D., *Decision support based on uncertainty quantification of model predictions of contaminant transport*, CMWR 2010: XVIII International Conference on Water Resources, J. Carrera (Ed), CIMNE, Barcelona 2010.
34. 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.
35. Vesselinov, V.V., Harp, D.R., Model Analysis and Decision Support (MADS) for complex physics models, Computational Methods in Water Resources XIX (CMWR 2012), online peer-reviewed publication at http://cmwr2012.cee.illinois.edu.
36. 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 Los Alamos National Laboratory (LANL), Waste Management Symposium, 2013.
37. 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.

38. Vesselinov, V.V., Pau. G., Finsterle, S., AGNI: Coupling model analysis tools and high-performance subsurface flow and transport simulators for risk and performance assessments, Computational Methods in Water Resources XIX (CMWR 2012), online peer-reviewed publication at http://cmwr2012.cee.illinois.edu.
39. Walker, W. E., Harremoës, P., Rotmans, J., Van der Sluijs, J. P., Van Asselt, M. B. A., Janssen, P., & Von Krauss, M. K. (2003). *Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support*. Integrated Assessment, 4(1), 5-17.

ACKNOWLEDGEMENTS

Presented work was supported by Los Alamos National Laboratory Environmental Programs Projects. Dan O'Malley was also supported by a Los Alamos National Laboratory (LANL) Director's Postdoctoral Fellowship. Velimir V. Vesselinov was also supported by the DiaMonD project: An Integrated Multifaceted Approach to Mathematics at the Interfaces of Data, Models, and Decisions, funded by the U.S. Department of Energy Office of Science, Grant #11145687.