

Coupling Big Data Analytics and Reactive Transport Modeling for Cost-effective Groundwater Monitoring – 17163

Haruko Wainwright *, Boris Faybishenko *, Sergi Molins *, George Pau *, Miles Denham **, Mark Amidon **, David Moulton ***, Konstantin Lipnikov ***, Vicky Freedman ****, Mark Freshley ****, Carol Eddy-Dilek **

* Lawrence Berkeley National Laboratory

** Savannah River National Laboratory

*** Los Alamos National Laboratory

**** Pacific Northwest National Laboratory

ABSTRACT

This study presents an innovative approach for sustainable and cost-effective groundwater monitoring. This approach takes advantage of recent advances in various technologies: (1) *in situ* autonomous sensors, (2) big data analytics, and (3) parallel high-performance computing for flow and reactive transport modeling. *In situ* sensors are used to periodically measure the key variables (such as pH, redox potential, electrical conductivity, and groundwater level), which control contaminant mobility and the plume spatial and temporal distribution. Based on a limited number of groundwater sampling, the data analytics methods—data mining and machine learning—allow us to identify and quantify the correlations between the *in situ*-measured variables and contaminant concentrations, and also to detect significant changes associated with the plume mobility. In addition, a state-of-the-art parallel numerical flow and reactive transport simulator Amanzi, and uncertainty quantification software Agni are used to provide an improved physical and mechanistic understanding of the contaminated groundwater system behavior, and to predict the long-term plume distribution for optimizing and adapting the monitoring strategy. Amanzi and Agni were developed as part of the Advanced Simulation Capability for Environmental Management (ASCEM) program of the DOE Office of Environmental Management. Such modeling is critical, particularly, for assessing the impact of climate change and associated hydrological shifts. The developed approach is expected to significantly reduce the groundwater sampling frequency and associated cost. In addition, the real-time information on plume mobility serves as an early warning system, improving the resiliency of contaminated or potentially contaminated sites. We demonstrate this approach using as an example the Savannah River Site (SRS) F-Area, where groundwater is contaminated by various radionuclides, including uranium, tritium and technetium.

INTRODUCTION

Nuclear weapon production during the Cold War has resulted in groundwater contamination at many locations in the United States. Low-level radioactive waste solutions were often disposed into unlined seepage basins with minimal or no engineered barriers. Some of high-level radioactive waste tanks have been reported to have leaks, which led to widespread soil and groundwater contamination. The

U.S. Department of Energy, Office of Environmental Management (DOE-EM) manages the remediation of many contaminated sites, which is considered one of the most technically complex cleanup challenges in the world [1]. The overall cost is predicted to exceed \$200 billion over the next few decades.

There have been a variety of remediation techniques deployed at the DOE-EM sites, such as soil excavation, pump-and-treat, and bioremediation. Although many remediation efforts have been successful or are expected to finish in the next several decades, there are still dozens of sites that have contaminant concentrations above regulatory standards. The biggest challenge is to address a large volume of contaminated soil and groundwater with relatively low concentrations, cases where soil removal is not practical due to its large volume, and pump-and-treat systems are not effective due to relatively low contaminant concentration. For such sites, monitored natural attenuation (MNA) has been considered as a preferred alternative, with the premise that the contamination would not compromise the public health, and that natural flow and geochemical process would be sufficient to reduce concentrations in the future.

In addition, Enhanced attenuation (EA) has become a recent focus to achieve cost-effective and sustainable remediation. EA introduces an amendment to enhance the attenuation to bring concentrations down to regulatory levels until natural processes take over. For example, at the SRS F-Area, base injection was performed to increase pH and immobilize uranium, which resulted in concentration lower than the regulatory standard and cost savings of approximately \$9 million per year compared to the existing pump-and-treat system [2].

The application of EA and MNA, however, create a new challenge in site closure, since contaminants are sequestered in subsurface rather than removed. Continued groundwater monitoring will be required for several decades to ensure long-term stability and safety. The current practice of monitoring is based on the regular sampling and analyzing the contaminant concentrations in groundwater samples at numerous wells, which becomes costly over a long time frame. Although the monitoring cost is a fraction of active remediation cost, the long-term monitoring is still expected to account for a large fraction of the life-cycle cleanup costs at the DOE sites.

To reduce the long-term monitoring cost, Savannah River National Laboratory (SRNL) and Lawrence Berkeley National Laboratory (LBNL) have been developing an innovative approach, using *in situ* automated sensors to replace and reduce groundwater sampling [2,3]. Recent advances in *in situ* sensors allow us to continuously monitor groundwater parameters, and to stream data through wireless or phone networks. Although *in situ* measurable properties, such as pH, dissolved oxygen, nitrate concentration, redox potential, groundwater level, and electrical conductivity, are not the contaminant concentrations of interest, many of them are the key properties that control plume mobility and its spatial and temporal distributions. At the SRS, for example, *in situ* measurable properties are found to be almost perfectly correlated to the contaminant concentrations [3]. Since these *in situ* variables are also leading indicators of the plume mobility, the *in situ* sensors

can serve as an early warning system so that practical actions can be taken to manage the plume migration. For this purpose, the application of data analytics methods is one of the key approaches in monitoring strategy to detect meaningful changes in the plume migration and to identify the noise that is inherent in environmental data.

The current study focuses on two aspects of advances in advanced long-term monitoring: (1) big data analytics, and (2) reactive transport modeling. Data analytics capabilities have been rapidly expanding in other areas (such as computer science, artificial intelligence, climate science), which can be easily transferred to the DOE-EM applications. Data analytics enables us to extract critical information from historical datasets accumulated over many years at the DOE-EM sites, and is important for system understanding and decision-making, as well as for improving sampling and monitoring strategies. Modeling of flow and contaminant transport enables the prediction of contaminant plume evolution under various conditions for many years.

Over the last five years, the ASCEM project, funded by DOE-EM, has made a significant advancement in the development of a flow reactive transport code, Amanzi, taking an advantage of DOE's state-of-the-art high performance computing capabilities. The developed 3D reactive transport simulator can now include realistic geological features, boundary conditions and artificial structure (such as engineered barrier systems), as well as complex geochemical reactions (such as the interactions between contaminant concentrations and controlling variables) [3].

We demonstrate the effective integration of these components at the SRS F-Area, where contaminant plumes and subsurface structure are well characterized and historical monitoring and geological datasets have been effectively curated. Having real data and established models, the SRS F-Area can be considered as a testbed to achieve sustainable and cost-effective remediation and monitoring at the DOE-EM sites. Such an integrated technology can also be transferable to other types of groundwater contamination such as the DOE Office of Legacy Management sites, nuclear power plants, and other contaminated sites.

SITE DESCRIPTION

The Savannah River Site (SRS) is located in south-central South Carolina, near Aiken, approximately 160 kilometers from the Atlantic Coast. It covers about 800 km² (300 mi²) and contains facilities constructed in the early 1950s to produce special radioactive isotopes (e.g., plutonium and tritium) for the U.S. nuclear weapons stockpile. The SRS F-Area seepage basins were constructed as unlined, earthen surface impoundments that received ~7.1 billion liters of acidic, low-level waste solutions from the processing of irradiated uranium in the F-Area Separations facility from 1955 through 1988 [4]. Currently, an acidic contaminant plume extends from the basins ~600 m downgradient to the Four Mile Branch creek, including various radionuclides, such as uranium isotopes, Sr-90, I-129, Tc-99, tritium, and other contaminants, such as nitrate.

Various remediation activities have been conducted at the site, including capping of the basins (1991) and pump-and-treat (1997–2004). A hybrid funnel-and-gate system has been in operation since 2004, which includes low-permeability engineered flow barriers, and injection of alkaline solutions. The base injections are considered to be effective in neutralizing the acidic groundwater and in greatly increasing uranium retardation, since uranium mobility is significantly influenced by pH (because higher pH values increase uranium sorption). At the same time, the barriers slow down plume migration and increase decay and mixing before the plume reaches the Four Mile Branch creek, a down-gradient stream that ultimately captures the plume. Monitored Natural Attenuation (MNA) is a desired closure strategy for the site, assuming that infiltration of rainwater will eventually increase the pH of the plume, causing much stronger retardation and dilution of the uranium plume.

METHODOLOGY

Big Data Analytics

The SRS database includes the results of monitoring of 422 analytes, which can be grouped into 4 main categories: radioactive, VOCs, heavy metals, and general groundwater quality data. The database includes the results of measurements in 165 wells and in 3 aquifers. In this study, we used the results of measurements in 39 typical monitoring wells over the last 30 years from January 1990 to September 2015.

The analyzed database contains both numerical and categorical data, with duplicates of time stamps, gaps in time series, and some negative values of concentrations, i.e., outliers. The QA/QC of time series data included removing outliers or noise (based on finding so-called additive outliers), using a cubic spline function. We also interpolated all variables at the same Quarterly time stamps, which is the necessary condition to provide further correlation analysis between the concentrations of radionuclides and controlling variables.

Principal component analysis (PCA) is a typical data mining method to quantify and visualize the variability and correlations among multi-dimensional variables. The goal of the PCA is to transform the initial variables into a new set of variables, which explain the variation in the data. These new variables, which are called principal components, represent a linear combination of the original variables. The PCA is used to reduce the dimensionality of multivariate data to 2 or 3 principal components that can be used to visualize graphically the dataset with minimal loss of information.

Using the SRS data, the PCA was conducted to analyze multivariate data to identify the correlations among different variables, with a particular focus on the correlations between in situ measurable parameters (pH, electrical conductivity, nitrate concentration) and contaminant concentrations (U-238, I-129, H-3). We would note that electrical conductivity at this site is dominated by the nitrate concentration.

Reactive Transport Modeling

A 3D reactive transport model has been continuously developed and improved over the last five years [3,5,6]. We describe the model briefly here for completeness. The 3D hydrogeological model was developed including the heterogeneous hydrostratigraphic interfaces and engineered barrier systems. There are three layers units within the Upper Three Runs Aquifer: an upper aquifer zone (UUTRA), a Tan Clay Confining Zone (TCCZ), and a lower aquifer zone (LUTRA). The domain also includes low-permeability engineered barriers, which are part of the funnel-and-gate system (Fig. 1). The unstructured 3D prismatic mesh was created using the Los Alamos Grid Toolbox (LaGriT; <http://lagrit.lanl.gov>; Fig. 1b-d). The mesh used in this study has 1,849,039 cells and 982,998 vertices

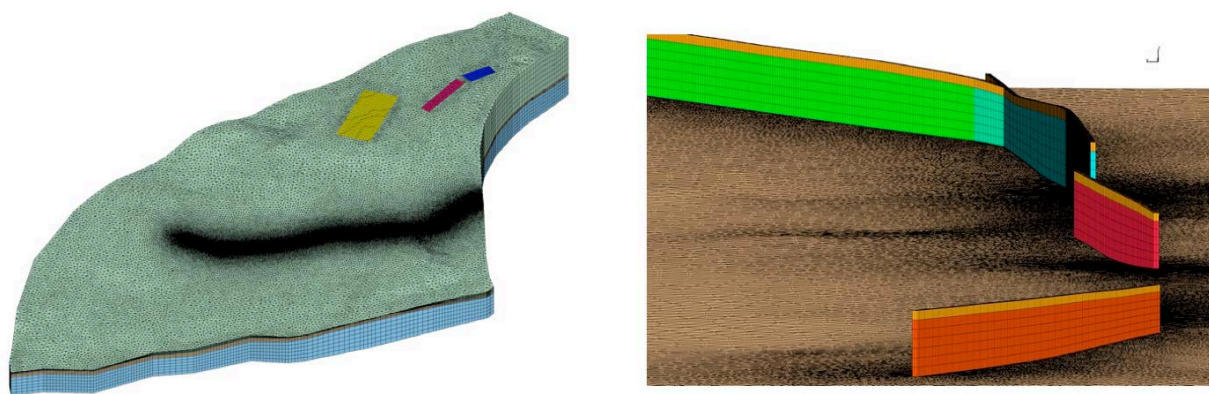


Figure 1. (a) 3D prismatic mesh generated by LaGriT is shown using an exaggerated vertical scaling to highlight the three stratigraphic layers, and (b) the barrier wall representation. In (a), the green region is the upper aquifer, the middle brown layer is the Tan Clay confining zone, and the blue region is the lower aquifer. The yellow, red and blue surface areas are the three basin locations.

The 3D flow and transport simulations were performed using the Richards equation and advective transport Process Kernels in Amanzi on the National Energy Research Scientific Computing Center (NERSC) high performance computing (HPC) platforms. The aquifer properties that were developed for the two-dimensional flow model by Bea et al. [5] were used in the current study. A very low permeability of $1 \times 10^{-17} \text{ m}^2$ was assumed for the engineered barriers. The upstream model boundary was treated as fixed pressure based on the measured water-table height. The upper groundwater boundary condition was assigned as a constant recharge boundary with a moving seepage face to allow groundwater flow to upwell and discharge to the Four Mile Branch creek. The other boundary conditions were treated as no-flow, including the vertical boundary below the stream, the bottom boundary of the model, which is a relatively continuous low-permeability clay layer, and sides of the model which are parallel to flow.

The geochemical conditions have been extensively characterized through many field and laboratory experiments, particularly for uranium geochemistry. In this paper,

we focus on uranium chemistry, based on the mechanistic models that have been developed to describe its sorption and pH-dependent behaviors. As described in [5], the natural attenuation of the acidic-U(VI) plume in the F-Area is likely to be affected mainly by a combination of the following processes: (1) adsorption/desorption of U(VI) onto/from the surface of different minerals (mainly kaolinite and goethite at this site) under different mechanisms (i.e., electrostatic surface complexation and/or ion exchange) [7]; (2) pH effects related to H⁺ sorption and/or Al mineral dissolution and precipitation; (3) mixing of the plume groundwater with clean (and higher pH) background groundwater. A three-dimensional reactive transport model was assembled by combining the flow and transport model and the geochemical model. The combination of the flow and transport portions of the model and the geochemical model was enabled by the Alquimia interface in Amanzi. This interface makes it possible to use existing geochemical codes (e.g., PFLOTRAN, CrunchFlow) within the ASCEM HPC infrastructure through a generic coupling.

To account for the uncertainty in model parameters, we used ASCEM's uncertainty quantification (UQ) toolset, Agni. First, we performed Monte-Carlo simulations by randomly generating parameter sets and running reactive transport simulations with given parameter sets. In addition, the Sobol' global sensitivity indices were computed based on the MC-based approach developed by Wainwright et al. [8]. The sensitivity analysis enables us to identify the controlling parameters that cause variability and uncertainty as well as key parameters of long-term monitoring.

RESULTS AND DISCUSSION

Data Analytics

Fig. 2 illustrates the graphs of original (open circles) and QA/QC corrected (i.e., with the outliers removed and aligned to the same time stamps (90-day difference)) of the water table and concentrations. Our QA/QC procedure successfully removes the outliers, particularly in the water table and tritium concentration data. These corrected data were then used for the PCA analysis and determination of the correlation between the variables.

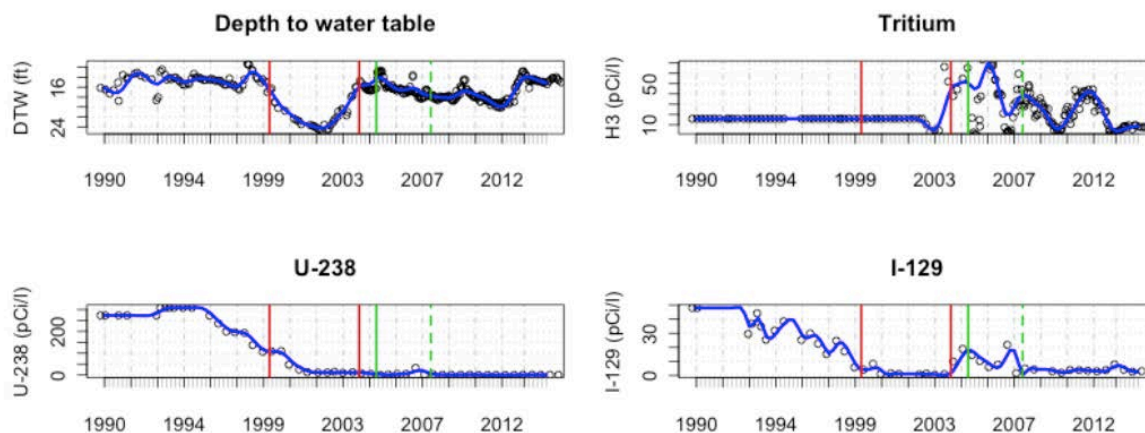


Figure 2. QA/QC results. Each plot includes the original data (open circles) and QA/QC corrected data. The 1st vertical red line in 1999 indicates the time of the basin closure; the 2nd vertical red line in 2003 indicates the end of the pump, treat, and reinject operations, which were conducted in 1999-2003; the 1st green line indicates the time of the Installation of Engineered Subsurface Barrier (ESB) in 2004. The base injections began in the ESB gates in 2005, and base injections in the wetlands in 2008 (dashed vertical green line).

Fig. 3 shows the results of the PCA analysis of the dataset of 39 wells. We divided those wells into two groups located in upgradient and downgradient regions from the barrier walls. Since the first and second principle components account for most of the variability (>60%), we only show those two components. In the bi-plots of Fig. 3, each arrow represents the loading of each variable on the principle components. When two arrows are pointing in the same or opposite direction, those two variables are strongly correlated. When the two arrows are orthogonal, those two variables are likely to be uncorrelated.

PCA reveals that (1) pH is strongly correlated with I-129 and U-238 concentrations, and (2) nitrate concentrations are strongly correlated with Tc-99 and H-3, and weakly correlated with I-129 and U-238. This would be because U-238 and I-129 are affected by pH and other geochemical conditions, while Tc-99 and H-3 are non-reactive tracers, behaving similar to nitrate transport. In addition, we can see that specific conductance (SC) is correlated with nitrate concentrations, since nitrate dominates total dissolved solid (TDS) at this site. These findings are consistent for both regions, suggesting that the correlations can be used over the site.

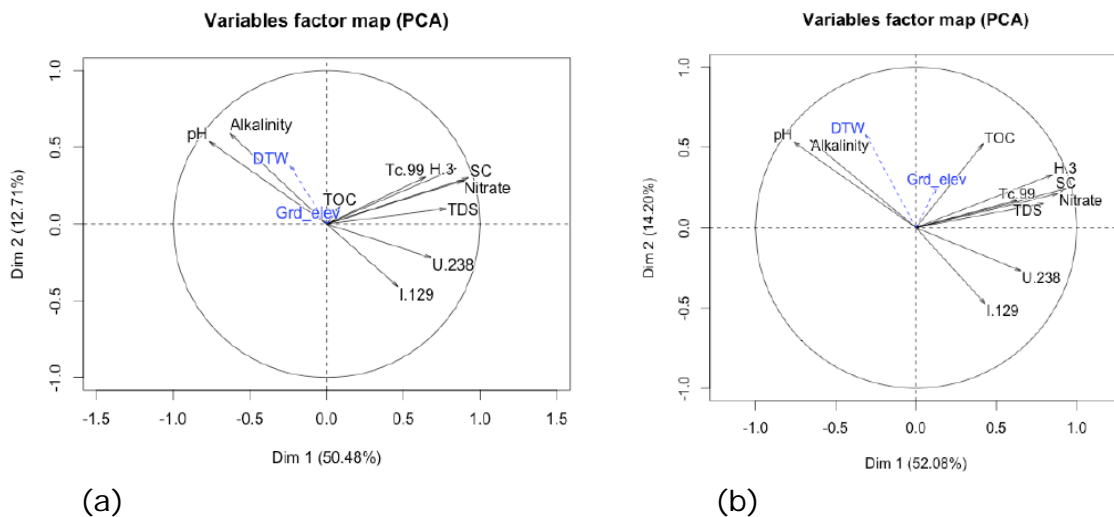


Figure 3. Results of the PCA analysis of the water table depth and concentrations in the wells: (a) downstream of the wall, and (b) upstream of the wall. These figures visualize the loading of each variables on the first and second components.

Reactive Transport Modeling

Fig. 4 shows the simulated evolution of the low-pH and uranium plumes. This simulation includes capping of the seepage basin, which limited the infiltration after the basin operation stopped. It does not include simulations of other remediation treatments at the site. The plumes initially move straight down vertically until they hit the water table, and then migrate laterally mainly within the upper aquifer (Figs. 4a and 4d). The low-pH plume moves more quickly downgradient (Figs. 4a and 4b), increasing the mobility of uranium and creating a way for the uranium plume to follow (Figs. 4d and 4e). As the plume migrates downgradient toward the creek, the plume goes through the troughs in the bottom of the upper aquifer (Fig. 4b). The model predicts that a significant amount of uranium is expected to trap in the vadose zone (Fig. 4f) in 2050 even though pH would be neutralized (Fig. 4c), which suggests the long-term effect of capping the basin.

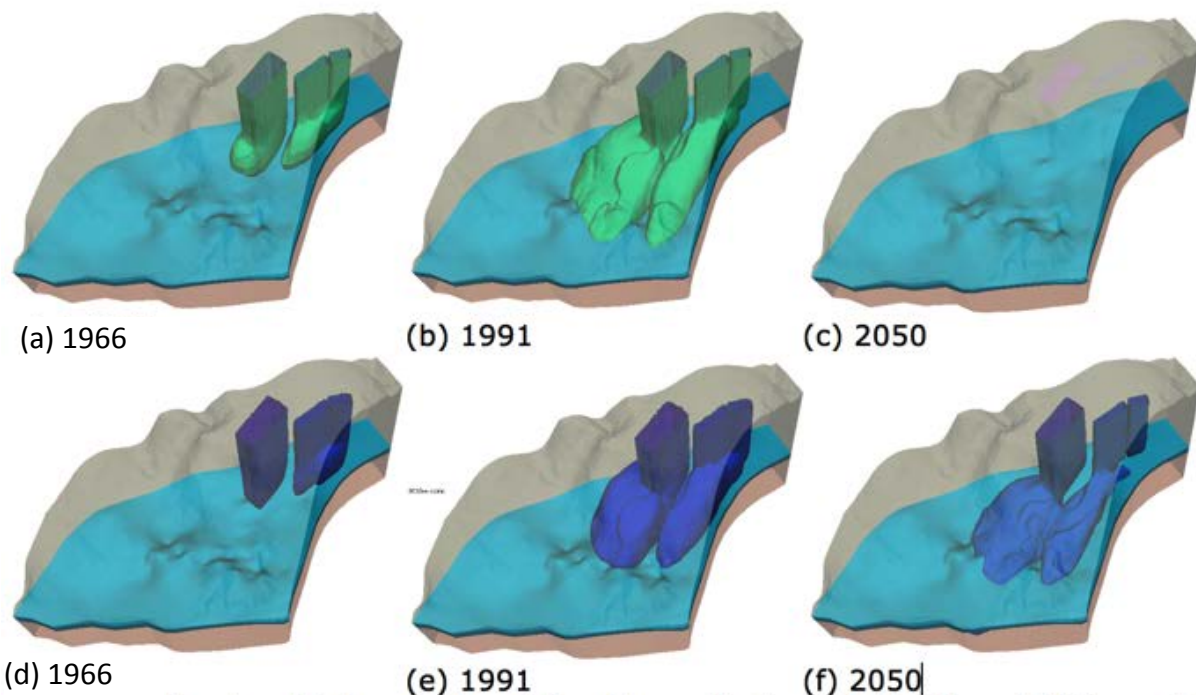


Figure 4. Upper figures (a-c): simulated evolution of low-pH plume ($\text{pH} > 4$); and lower figures (d-f): uranium plume ($\text{concentration} > 1 \times 10^{-6} \text{ mol/L}$). The sky blue region is the low permeable TCCZ, which separates the upper and lower aquifers. Vertical exaggeration=15X.

Modeling and UQ analysis allow us to simulate the correlations between the contaminant concentrations and *in situ* controlling variables under various hydrological and geochemical conditions. Fig. 5 demonstrates the correlations between U-nitrate and U-pH correlations determined for different conditions (100 realizations). We varied seven parameters (cation exchange capacity, sorption site density, source pH and U concentration, precipitation, discharge rate and permeability), which were identified to be important in the previous studies [5,6].

In addition, sensitivity analysis enables us to identify which parameters are influencing these correlations and creating variability more significantly than the others. Sobol' global sensitivity indices (Fig. 6) suggest that the key parameters are the Upper Aquifer permeability, cation exchange capacity (CEC), and source uranium concentrations. The precipitation has little effect on the correlations. These findings are useful for long-term monitoring. For example, the key parameters are mostly material properties, the correlations would not be sensitive to future climatic conditions.

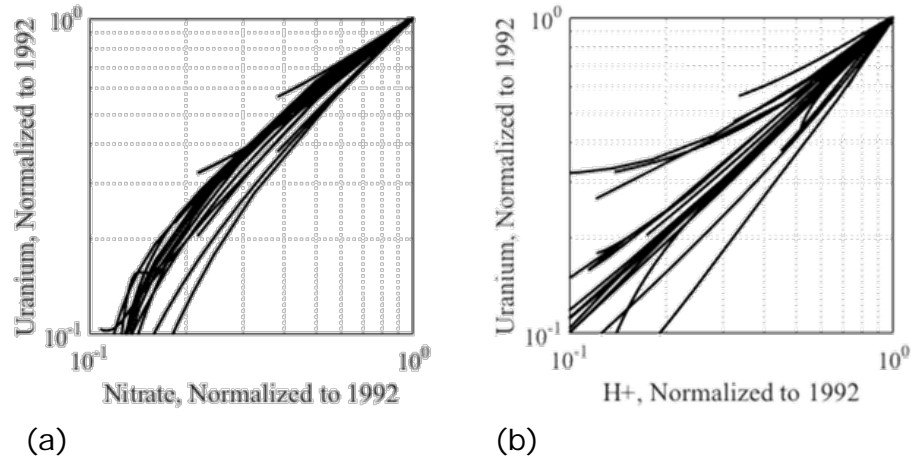


Figure 5. Simulated correlations between uranium (U) concentration (log-transformed mol/L) and *in situ* variables at FSB95D and Well FSB110D: (a) nitrate concentration (log-transformed mol/L) and (b) pH. In each plot, the correlations are normalized to the ones in 1992 in order to facilitate the visualization and comparison.

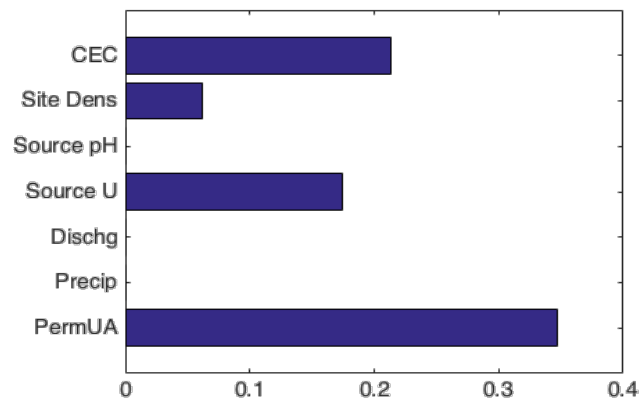


Figure 6. Sensitivity analysis results; Sobol' sensitivity index of each variable with respect to the correlations between pH and U-238 concentrations.

CONCLUSION

In this study, we demonstrated the key scientific advances for establishing cost-effective long-term monitoring, using the modern methods of data analytics and reactive transport modeling. The data analytics methods were used to perform the QA/QC statistical analysis, and to identify the correlations among *in situ* measured controlling parameters and contaminant concentrations. The PCA method is found to be a powerful tool to analyze multi-dimensional and large datasets. We also developed a 3D flow and reactive transport model to describe the contaminant plume evolution in a mechanistic manner, including the complex pH-dependent reactions of uranium. UQ results suggest that the correlations between *in situ* measurable concentrations (i.e., pH, EC, nitrate) and contaminant concentrations are variable influenced mainly by hydrological and geochemical properties.

Our future work will focus on evaluating the efficacy of the current or planned enhanced attenuation treatments, and their impact on the long-term monitoring strategy. We will also continue the spatio-temporal data analysis and modeling to identify an optimal number and the layout of groundwater monitoring wells and *in situ* sensors.

REFERENCES

1. NRC, U.S. Department of Energy's Environmental Management Science Program: Research Needs in Subsurface Science. National Academy of Sciences: Washington, DC, 2000.
2. Denham, M. and C. Eddy-Dilek (2016), Sustainable Remediation of Radionuclides By a Common Sense Approach to Enhanced Attenuation, Proceedings of WM2016 Conference, March 6 – 10, 2016, Phoenix, Arizona, USA.
3. Wainwright, H.M., B. Faybishenko, S. Molins, J.A. Davis, B. Arora, G. Pau, J. Johnson, G. Flach, M. Denham and C. Eddy-Dilek, J.D. Moulton, K. Lipnikov, C. Gable, T. Miller and M. Freshley (2016), Effective Long-term Monitoring Strategies by Integrating Reactive Transport Models with In situ Geochemical Measurements, Waste Management 2016, March 6 – 10, 2016, Phoenix, Arizona, USA.
4. Killian, T.H., N.L., Kolb, P. Corbo, I.W. Marine, Environmental information document, F-Area seepage basins. Report No. DPST 85-704.E.I. du Pont de Nemours & Co, Savannah River Laboratory, Aiken SC 29808, 1986.
5. Bea, S. A., Wainwright, H., Spycher, N., Faybishenko, B., Hubbard, S. S., & Denham, M. E., Identifying key controls on the behavior of an acidic-U (VI) plume in the Savannah River Site using reactive transport modeling. Journal of contaminant hydrology, 151, 34-54, 2013.
6. Freshley, M. et al. Advanced Simulation Capability for Environmental Management (ASCEM) Phase II Demonstration, ASCEM-SITE-2012-01, 2012.
7. Dong, W., T. Tokunaga, J. Davis, and J. Wan, Environmental Science & Technology, 46,1565–1571, 2012.
8. Wainwright, H.M., S. Finsterle, Y. Jung, Q. Zhou and J.T. Birkholzer, Making Sense of Global Sensitivity Analysis, Computers & Geosciences, ISSN 0098-3004, <http://dx.doi.org/10.1016/j.cageo.2013.06.006>, 2013.

ACKNOWLEDGEMENTS

This study was supported as part of the ASCEM project, which is supported by U.S. Department of Energy, Office of Environmental Management, under the award number DE-AC02-05CH11231 to LBNL. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.