In Situ Monitoring of Groundwater Contamination Using the Kalman Filter

Franziska Schmidt[†], Haruko M. Wainwright*[‡], Boris Faybishenko[§], Miles Denham[∥], and Carol Eddy-Dilek[⊥]

[†] Department of Nuclear Engineering, University of California Berkeley, Etcheverry Hall, 2521 Hearst Avenue, Berkeley, California 94709, United States

[‡] Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, MS 74R-316C, Berkeley, California 94720-8126, United States

[§] Energy Geosciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720-8126, United States

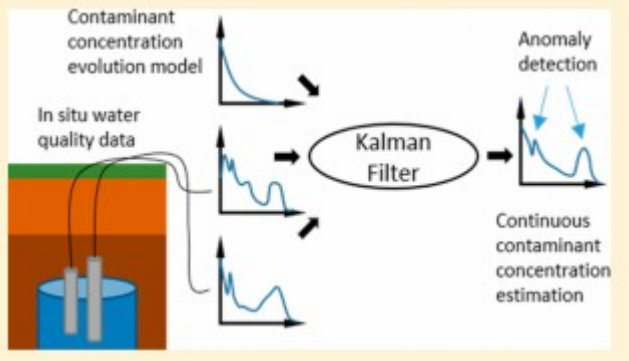
[∥] Panoramic Environmental Consulting, LLC, P.O. Box 906, Aiken, South Carolina 29802, United States

[⊥] Savannah River National Laboratory, Savannah River Site, Aiken, South Carolina 29808, United States

*E-mail: hmwainwright@lbl.gov.

Abstract

This study presents a Kalman filter-based framework to establish a real-time *in situ* monitoring system for groundwater contamination based on *in situ* measurable water quality variables, such as specific conductance (SC) and pH. First, this framework uses principal component analysis (PCA) to identify correlations between the contaminant concentrations of interest and *in situ*measurable variables. It then applies the Kalman filter to estimate contaminant concentrations continuously and in real-time by coupling data-driven concentration-decay models with the previously identified data correlations. We demonstrate our approach with historical groundwater data from the Savannah River Site F-Area: We use SC and pH data to estimate tritium and uranium concentrations over time. Results show that the developed method can estimate these contaminant concentrations based on *in situ* measurable variables. The estimates remain reliable with less frequent or no direct measurements of the contaminant concentrations, while capturing the dynamics of short- and long-term contaminant concentration changes. In addition, we show that data mining, such as PCA, is useful to understand correlations in groundwater data and to design long-term monitoring systems. The developed *in situ* monitoring methodology is expected to improve long-term groundwater monitoring by continuously confirming the contaminant plume’s stability and by providing an early warning system for unexpected changes in the plume’s migration.



Introduction

Sustainable remediation of soil and groundwater involves balanced decision-making with respect to environmental, social, and economic sustainability criteria and their acceptable ranges.(1,2)Overall environmental benefit, disturbances, and footprints, including waste, construction noise/traffic, ecological disturbance, and water/energy use, need to be considered. Sustainable remediation is often coupled with attractive end-use scenarios with restrictive use of the subsurface, such as the installation of solar power stations under long-term institutional control.(3)It usually involves attenuation-based remedies – such as enhanced natural attenuation (ENA) and monitored natural attenuation (MNA) – that remove or immobilize contaminants in the subsurface through natural or engineered processes.(4) Attenuation-based approaches are particularly effective at managing large volumes of contaminated soil and groundwater with relatively low concentrations, where soil removal is not practical due to its large volume, and active treatment systems are no longer effective due to relatively low contaminant concentrations.

Environmental monitoring has become increasingly important for implementations of ENA and MNA.(5,6) Attenuation-based remedies carry the increased burden of proof that the remaining contaminants will stay on-site, that the contaminant plume is stable, and that the residual contaminants do not pose a significant risk to public health. Regulatory approval of ENA or MNA requires long-term monitoring programs to ensure plume stability and demonstrate compliance with approved levels of risk reduction. This is particularly important for metals and radionuclides, because these contaminants will remain in the subsurface for an extended time. In addition, a hydrological shift or an extreme event, for example due to climate change, could alter the plume’s mobility in the future due to a change in the hydraulic gradient or geochemical conditions, particularly for redox-sensitive species like uranium, chromium, and arsenic.(7) Proper monitoring provides assurance for local communities in terms of health concerns and improves the negative perception often associated with contaminated sites.(5,8)

The challenge is to develop cost-effective strategies for long-term groundwater monitoring because it could become a significant portion of the overall life-cycle cost at contaminated sites. Although there have been significant advances in the development of sensor networks and data analytics technologies in the past decade,(9−11) they have rarely been used in groundwater monitoring, and most contaminants still cannot be detected directly. The current standard practice at contaminated sites is to sample and analyze groundwater samples from multiple wells quarterly or yearly. Such discrete measurements could miss significant changes in plume mobility and contaminant transport. In addition, current practices can lead to a “compliance only”-focus, meaning that contaminant concentrations are assessed for regulatory compliance alone and little effort is put into understanding the system and plume behavior. Lack of understanding could potentially result in increased liability, when data anomalies are misinterpreted.

Recently, Eddy-Dilek et al. (2014) have explored a new approach for the long-term monitoring of metal and radionuclide plumes.(12) Their method proposes to monitor the changes of *in situ*measurable groundwater variables (such as pH, temperature, specific conductance (SC), and water level), which are indicative of plume behavior. This approach is similar to previous hydrogeophysical studies, which have shown that, for example, electrical properties are correlated to contaminant concentrations.(13,14) This *in situ* monitoring approach has great potential to monitor contaminant concentrations continuously over time. In addition, when an event occurs that could potentially change plume mobility, such a monitoring system enables a swift response and a rapid adjustment of remedial actions.

In this study, we present a framework for the continuous real-time estimation of groundwater contaminant concentrations based on *in situ* measurable variables. First, we apply a data mining approach – here, principal component analysis (PCA) – to historical monitoring data sets to improve our system understanding by identifying possible correlations between different groundwater variables and contaminant concentrations. Once identified, the correlations need to be quantified with an appropriate correlation coefficient. After this exploratory data analysis, we apply the Kalman filter algorithm, which can integrate mixed time-series data (e.g., various data types and accuracies) with a temporal evolution model, i.e. physical or empirical models, continuously in real-time.(15) The Kalman filter has been applied in numerous fields, particularly in navigation,(16) traffic prediction,(17,18) remote sensing,(19) and sensor network applications.(20)In the context of groundwater monitoring, the Kalman filter enables us to estimate contaminant concentrations in real-time as well as to decrease the frequency of direct contaminant sampling.

We demonstrate this approach using the extensive historical data sets from the Department of Energy (DOE) Savannah River Site (SRS) F-Area (South Carolina, United States). The site is contaminated with radionuclides, such as

tritium (H-3), iodine-129 (I-129), and uranium-238 (U-238), and a significant amount of these contaminants remains in the vadose zone. (21)Approximately 200 groundwater monitoring wells have been installed since 1955. The vast amount of available historical data makes this site an excellent location to test and demonstrate our approach which can potentially be transferred to other contaminated sites. The proposed method is a general framework that could possibly contribute to improved environmental protection as well as reduced long-term monitoring costs.

Site Description

The SRS F-Area has three unlined seepage basins, which received approximately 7.1 million $m^3$of acidic, low-level radioactive waste solutions between 1955 and 1988.(22) Currently, an acidic contaminant plume extends from the basins to about 600 m downgradient, moving toward the Fourmile Branch creek. The plume contains various radionuclides (e.g., H-3, I-129, U-238) as well as nitric acid. The plume penetrates the upper and lower aquifer,(23) which are hydrologically connected, despite being separated by a thin clay-rich layer.(24,25)

In order to reduce the effect of infiltration, the basins were capped in 1991. (26) In addition, a pump-and-treat system was installed in 1997 and then replaced by ENA in 2004, using a hybrid funnel-and-gate system.(12,23) This system consists of low-permeability engineered flow barriers (Figure 1) and injections of alkaline solutions, which are considered effective at neutralizing the acidic groundwater and enhancing uranium retardation.(23)

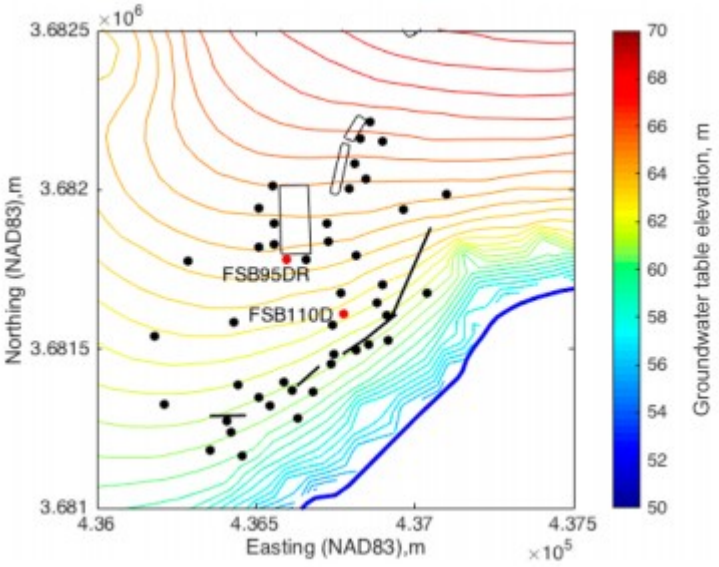


Figure 1. Site map of the SRS F-Area. The contour lines indicate the elevation of the groundwater table in meters. The contaminant plume originates from the three basins (rectangular boxes) and moves toward the Fourmile Branch creek in the southeast (thick blue line). Thick black lines indicate the location of low-permeability flow barriers. Black circles indicate well locations used for PCA analysis, and the red circles are FSB-95DR and FSB-110D, the wells used for the further demonstration of our framework.

The current ENA strategy is supposed to be replaced by MNA, based on the hypothesis that rainwater infiltration and groundwater mixing will eventually dilute the plume and neutralize the pH. Radioactive constituents have been monitored since the basins began operation in 1955, while quarterly groundwater sampling for nonradioactive contaminants began in 1982.(22)

Methodology

Exploratory Data Analysis

First, we apply PCA to the historical data sets to improve our system understanding by visualizing the correlations between the different groundwater quality variables and contaminant concentrations. Although such understanding can also be obtained through laboratory experiments and modeling,(23,27) it is important to use field data for validation. PCA converts multivariate data into a set of uncorrelated principal components (PCs), which are linear combinations of all original variables.(28,29) It is effective at visualizing and identifying data correlations and patterns, especially when the number of parameters is too large to assess the correlations with individual cross plots. In the context of hydrology and remediation, there have been several applications of PCA: Brown (1993) applied PCA to assess relationships between physical and hydraulic properties in a carbonate-rock aquifer.(30) Barker et al. (1988) used PCA to delineate the plume extent at a landfill site based on multivariate groundwater data.(31) In this study, we use PCA to support and improve long-term monitoring. The results of the PCA analysis inform our selection of *in situ* measurable variables suitable for the estimation of contaminant concentrations.

After PCA, we quantify the correlations between the different contaminant concentrations and relevant *in situ* measurable variables, such as the H^+ activity (calculated from pH values) and the SC, using the Pearson correlation coefficient.(32) We evaluate the time series of multiple variables at individual wells independently and investigate the spatial consistency across the site. This kind of correlation analysis is commonly used to estimate various hydrological and biogeochemical properties based on *in situ*-measured electrical properties.(33,34) The exploratory data analysis is done in RStudio 1.0.153.

Before applying PCA, we remove outliers from all time series, including the contaminant concentration data c_t, linearly interpolate them, and then log-normalize them using historical minima and maxima:

$$c_{t,norm} = \log\left(\frac{c_t - c_{\min}}{c_{\max} - c_{\min}} + 1\right) \qquad (1)$$

This log-normalization is useful to facilitate the visualization of multiple time series whose values vary in scale. The log-normalized values are shown in

the Supporting Information (SI) in Figure S1. We use Excel 2016 for the data preparation.

Estimation Strategy

Our goal is to estimate contaminant concentrations at groundwater wells continuously over time by using *in situ* measurable groundwater quality variables (e.g., SC and H^+ activity). We aim to use the contaminant concentration estimates to significantly reduce the frequency of direct measurements, i.e. groundwater sampling. Our method is based on two data-driven models: (1) the *temporal evolution model* of the contaminant concentrations and (2) the *data correlation model*between contaminant concentrations and *in situ* measurable variables.

Temporal Evolution Model

A data-driven exponential decay model is often used to describe the temporal evolution of contaminant concentrations in a plume's trailing edge.(21,35−38) It approximates the combined effects of dilution, advective transport, and, if applicable, biodegradation and radioactive decay with an effective exponential decay constant. The model is known to have good predictive power, despite its simplicity. Since exponential decay becomes linear in log-normal concentrations $c_{t,norm}$, a linear model can describe its temporal evolution

$$\frac{\mathrm{d}c_{t,norm}}{\mathrm{d}t} = \dot{c}_{t,norm} \qquad (2)$$

where $\dot{c}_{t,norm}$ is the rate of concentration change or the concentration decay rate.

Data Correlation Model

Multiple studies have documented significant correlations between contaminant concentrations and *in situ* measurable variables.(14,21,33) Such correlations are physically explained by the contaminant transport conditions (e.g., cocontaminants, geochemistry). At the SRS F-Area, the SC is primarily controlled by nitrate, which was released together with other contaminants.(21)Uranium mobility is primarily influenced by the pH.(23,27) In this study, we assume that the correlations between contaminant concentrations $c_{t,norm}$ and *in situ* measurable variables are linear (SI, Figure S2)

$$SC_{t,norm} = a_{SC}c_{t,norm} + b_{SC} \qquad (3)$$

$$H^{+}{}_{t,norm} = a_{H^+}c_{t,norm} + b_{H^+} \qquad (4)$$

where analogous to the contaminant concentrations, SC_{norm} and $H^{+}{}_{t,norm}$ are the log-normalized SC and H^+ activity values, respectively. The linear regression parameters include the slopes ($a_{SC}$and a_{H^+}) and the intercepts (b_{SC} and b_{H^+}).

Kalman Filter

In this study, we formulate our system as a discrete time-dependent system based on a state-space model.(39) We define a system state vector at time step t as $x_t = [c_{t,norm}, \dot{c}_{t,norm}]^T$, where $c_{t,norm}$ represents the contaminant concentrations, $\dot{c}_{t,norm}$ is the decay rate, and T indicates a vector transpose. The temporal evolution model (eq 2) is translated into the state-transition equation to describe the change of x_t within the discrete time interval Δt:

$$\boldsymbol{x}_t = \boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{w} \quad (5)$$

The state transition matrix F is defined as

$$\boldsymbol{F} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \quad (6)$$

We assume that w is a system noise vector, that follows a zero-mean Gaussian distribution with covariance matrix Q and is associated with the uncertainty in the temporal evolution model due to various hydrological and geochemical fluctuations, as well as sampling and analytical errors.

In addition, we define the observation vector z_t that represents measured data (direct measurements of the contaminant concentrations $c_{t,norm,direct}$ and the *in situ* measurable variables $SC_{t,norm}$ and $H^+_{t,norm}$) at time t as $z_t = [c_{t,norm,direct}, SC_{t,norm}, H^+_{t,norm}]^T$. We expect that, to confirm the algorithm's estimation performance, direct measurements of the contaminant concentrations will be taken occasionally. However, these measurements would be less frequent than the current quarterly or yearly sampling standard. To include these infrequent measurements in a consistent manner, missing data are dealt with by ignoring the respective terms during the computation. The data correlation model (eqs 3 and 4) is translated into the state-observation equation, which describes the relationship between the state and observation vectors

$$\boldsymbol{z}_t = \boldsymbol{H}\boldsymbol{x}_t + \boldsymbol{u} + \boldsymbol{v} \quad (7)$$

where the observation matrix H and the intercept term u are defined as

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 \\ a_{SC} & 0 \\ a_{H+} & 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{u} = \begin{bmatrix} 0 \\ b_{SC} \\ b_{H+} \end{bmatrix} \quad (8)$$

The observation noise vector v is assumed to follow a zero-mean Gaussian distribution with covariance matrix R. It includes not only the instrument noise but also the uncertainty in the linear correlations (eqs 3 and 4).

The parameters in eqs 5 through 8 are estimated based on historical data sets before applying the Kalman filter. The covariance matrix Q is

determined based on past contaminant concentrations and their variability. The parameters in H and u are determined by the linear regression analysis between contaminant concentrations and *in situ* parameters (eqs 3 and 4). The covariance matrix R is defined based on the residuals of the linear regression. Although the parameters can also be updated during the estimation, we assume that they are constant for the sake of simplicity in this demonstration. We further assume that there is sufficient historical data available to determine the parameters.

Once all parameters are determined, we can apply the standard linear Kalman filtering algorithm to estimate the contaminant concentrations continuously in real-time (SI, Text S1 and Figure S3). The algorithm is recursive; at each time step, the state vector x_t is first predicted based on the previous time step x_{t-1} and the state-transition model (eq 5) and then updated based on the state-observation model (eq 7). We use MATLAB R2016a for the implementation of the Kalman filter.

Results

Exploratory Data Analysis

Figure 2 shows the biplot from the PCA analysis based on time-series data from 42 wells. The first and second PCs account for most of the variability (>80%). The biplot reveals that, except for I-129, all variables are positively correlated to each other, particularly the nonreactive tracers (H-3, SC, and nitrate). I-129 is isolated, possibly because it has more complex speciation and geochemical behavior than other species.(40) The SC is strongly correlated with the nitrate concentrations, since nitrate dominates the total dissolved solids at this site. The U-238 concentration has a positive correlation with the H^+ activity, which is consistent with laboratory experiments and modeling studies.(23,27) This plot suggests that the SC and the H^+ activity (both measurable *in situ*) can be used to estimate U-238 and H-3 concentrations.

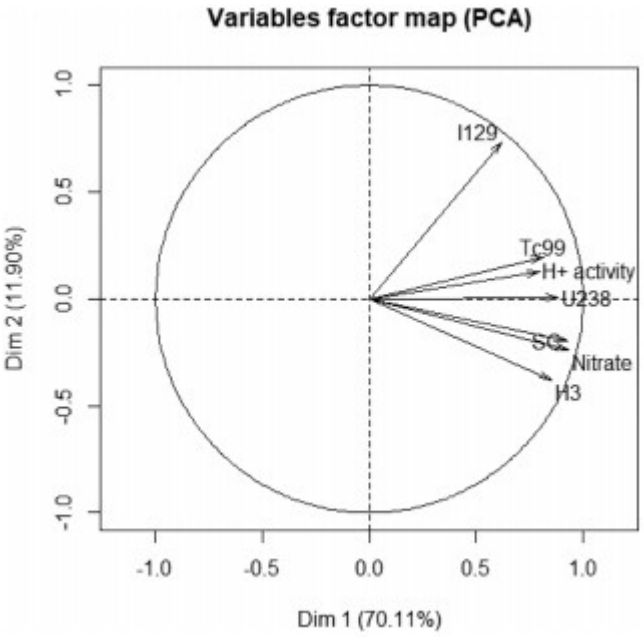


Figure 2. PCA biplot showing the second PC vs the first PC. Each arrow represents the loading of each variable on the PCs. Two variables with arrows pointing in the same/opposite direction are strongly positively/negatively correlated. Variables with orthogonal arrows are likely to be uncorrelated.

The Pearson correlation coefficients between the SC and U-238 are positive up to 0.96 with an average of 0.55 (Figure 3a), except for a few low correlations at some wells at the edge of the plume. The correlation between the H^+ activity and U-238 is also positive up to 0.91 with an average of 0.58 (Figure 3b). The results for H-3 are similar to positive correlations for both the SC (maximum: 0.98, average: 0.73) and the H^+ activity (maximum: 0.93, average: 0.51). Overall, Figure 3 shows that the correlation coefficients are spatially consistent across the site. The statistically insignificant correlations and outliers are due to low contaminant concentrations and/or occur at wells at the edge or outside the plume.

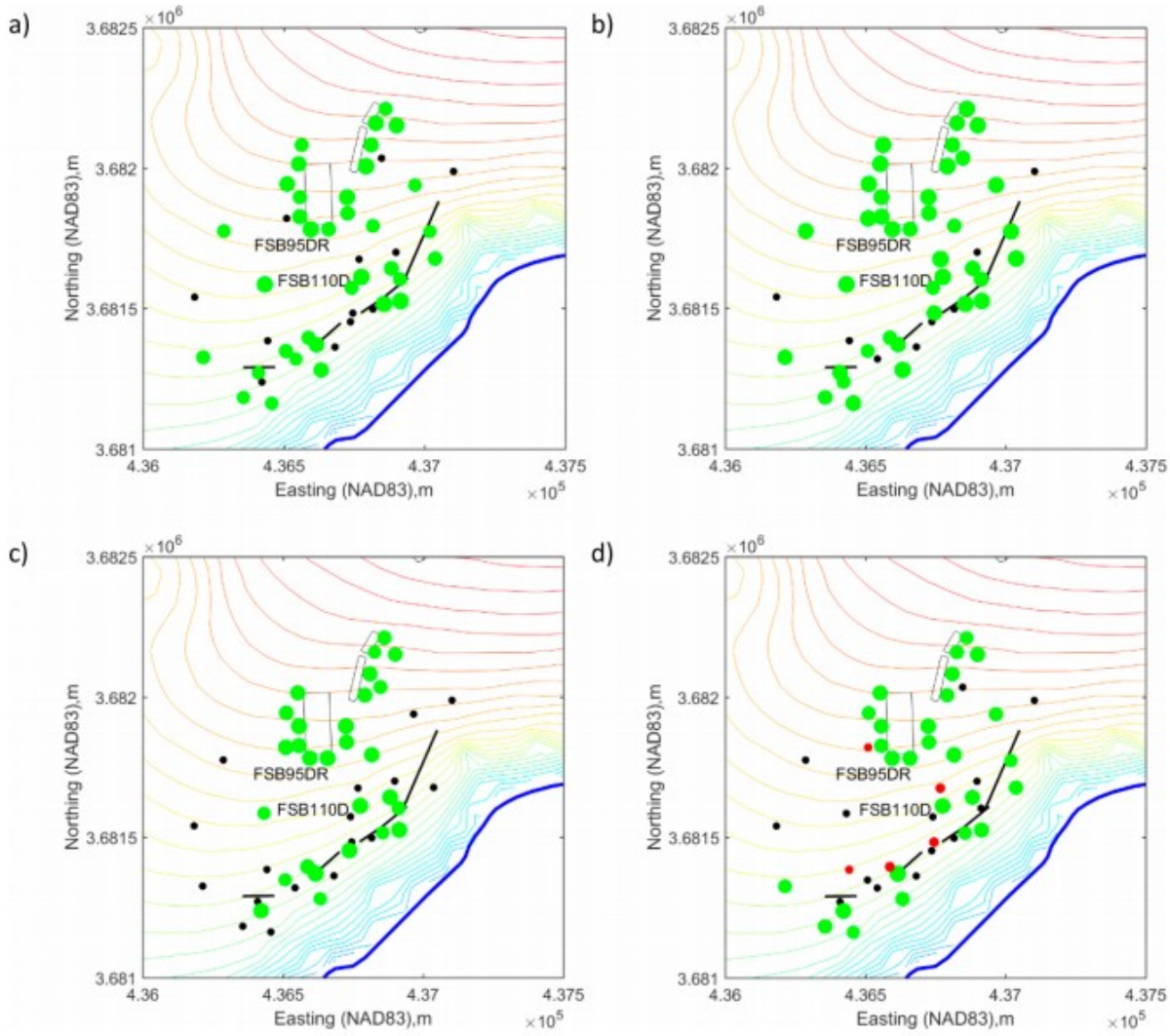


Figure 3. Bubble plots of the Pearson correlation coefficients for a) SC and U-238, b) H^+ and U-238, c) SC and H-3, and d) H^+ and H-3. All concentrations are log-normalized (eq 1). Positive correlations are shown in green circles, i.e. bubbles, and negative ones are shown in red. A black dot marks a well with an insignificant p-value (>0.05). The size of a bubble represents the absolute value of the coefficient. All wells shown are fully screened in the upper aquifer.

We use two monitoring wells for a further demonstration: FSB-95DR, located immediately downgradient (21 m) of Basin 3, and FSB-110D, located 220 m downgradient. Both wells are screened in the upper aquifer and have been used in previous studies.(21,23) Significant positive correlations between *in situ* variables (H^+ activity and SC) and contaminant concentrations (U-238 and H-3) are observed at both wells (Table 1). The correlations are linear or close to linear with few outliers (SI, Figure S2). The correlation coefficients are higher than 0.72 (p-value <0.05) between the SC and both contaminants, while those for the H^+ activity are generally lower. The correlation coefficients with the H^+ activity are higher for uranium than for tritium at both wells. In general, the correlations are stronger further downgradient at well FSB-110D compared to well FSB-95DR.

Table 1. Pearson Correlation Coefficients for the Entire Time Series for the SC and H^+ vs U-238 and H-3 at Wells FSB-95DR and FSB-110D[a]

	U-238, 95DR	H-3, 95DR	U-238, 110D	H-3, 110D
SC	0.89	0.90	0.94	0.95
H^+ activity	0.76	0.72	0.84	0.75

[a]All correlations have a p-value smaller than 0.05, i.e. they are statistically significant.

Kalman Filter for Contaminant Concentration Estimation

We demonstrate the Kalman filter monitoring approach using historical data sets. The U-238 and H-3 concentrations are estimated over time based on both SC and H^+ activity (Figure 4). The direct measurements of contaminant concentrations are used only once every two years. The 95% confidence interval shrinks at the time of the direct measurements because their uncertainty is very low. The estimated concentrations generally follow the validation data (not used for the estimation but set aside for validation purposes) for both U-238 and H-3 at both wells. Most of the validation data points are within the 95% confidence interval. At well FSB-95DR, both small contaminant concentration peaks (around 2008 and 2011) are successfully detected (Figure 4a, b) because the *in situ* measurable SC and H^+ activity increase at the same time (Figure S1). Although the contaminant concentrations at this site are expected to decrease over time due to natural attenuation, natural fluctuations or temporary concentration increases, e.g., due to water table changes, are expected.

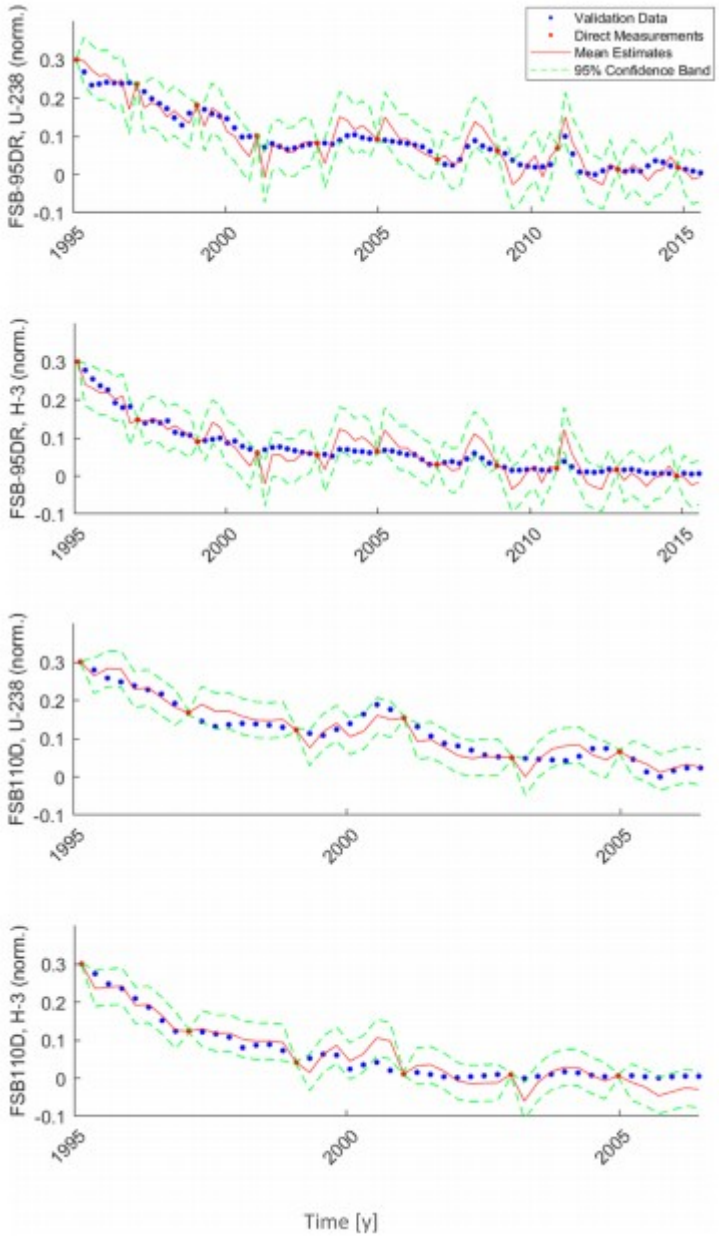


Figure 4. Kalman filter results: (a) U-238 concentrations at well FSB-95DR, (b) H-3 concentrations at well FSB-95DR, (c) U-238 concentrations at well FSB-110DR, and (d) H-3 concentrations at well FSB-110DR. In the plots, measured historical contaminant concentrations used for validation, but not for the estimation, are shown as blue dots, and direct measurements used as filter inputs for performance confirmation are shown as red dots. The estimated mean concentrations returned by the Kalman filter are shown as red lines, and the 95% confidence intervals are shown as green dashed lines. All estimates are only based on the quasi-continuous measurements (1 per 90 days) of SC and H^+ activity and occasional direct measurements (red dots, every 8 time steps or 720 days).

To analyze the sensitivity of the results to the two *in situ* variables, we compare the root-mean-square error (RMSE) of the estimation results based on only the SC, the H^+ activity, and both (Table 2). In general, the SC-based estimation is more accurate than the H^+ activity-based one, because the correlation with the SC is better (Table 1). Using both variables improves the

estimation results in most cases, except for H-3 at FSB-110D (Table 2). The estimation accuracy for both wells is comparable, although the results for FSB-110D are better due to the slightly higher correlation coefficients (Table 1).

Table 2. RMSE of the Kalman Filter Predictions Based on One or Both in Situ Variables[a]

	U-238, 95DR [$\times 10^{-03}$]	H-3, 95DR [$\times 10^{-03}$]	U-238, 110D [$\times 10^{-03}$]	H-3, 110D [$\times 10^{-03}$]
SC & H^+	**0.8**	**0.8**	**0.5**	0.7
SC	1.0	0.9	0.6	**0.6**
H^+	3.9	3.8	1.8	4.1

[a]All predictions are based on the continuous measurements (1 per 90 days) of SC and H^+ activity and occasional direct measurements of the contaminant concentrations (every 8 time steps or 720 days). Best performance for a given contaminant and well is marked in bold.

We also compare the estimation results with respect to different direct measurement, i.e. groundwater sampling, frequencies (Table 3). In general, the estimation accuracy decreases with decreasing direct sampling frequency. However, the change in the RMSE appears to level off with larger sampling intervals. The results based on groundwater sampling every 540 and 720 days show negligible differences, and the RMSE is only slightly higher when no direct measurements are used. This indicates that the filter still gives reasonable results without groundwater sampling if the data-driven correlation coefficients and temporal evolution models are reliable over time.

Table 3. RMSE of the Kalman Filter Predictions for Different Contaminant Sampling Frequencies[a]

	U-238, 95DR [$\times 1 \times 10^{-03}$]	H-3, 95DR [$\times 1 \times 10^{-03}$]	U-238, 110D [$\times 1 \times 10^{-03}$]	H-3, 110D [$\times 1 \times 10^{-03}$]
180 days	0.3	0.5	0.4	0.5
360 days	0.7	0.7	0.4	0.6
540 days	0.8	0.8	0.5	0.6
720 days	0.8	0.8	0.5	0.7
900 days	0.9	0.8	0.5	0.7
never	1.0	0.9	0.6	0.7

[a]All predictions are based on the continuous measurements (1 per 90 days) of SC and H^+ activity and occasional contaminant groundwater samples with different frequencies.

Discussion

Our results demonstrate that (1) data mining (such as PCA and correlation analyses) is useful to understand a groundwater system and to identify correlations between different species for long-term monitoring, and (2) the Kalman filter enables the estimation of contaminant concentrations based on *in situ* measurable groundwater quality variables. Our results show that the proposed algorithm can be used to establish real-time monitoring systems and to reduce the groundwater sampling frequency, while continuously

monitoring the evolution of the plume. Such a monitoring system can confirm plume stability continuously as well as serve as an early warning system to detect plume remobilization. In addition, the correlations and the system understanding obtained through data mining can be used to explain data anomalies if detected.

The successful estimation is attributed to the fact that (1) SC and H^+ activity (converted from pH) follow similar temporal dynamics as the contaminant concentrations, and (2) the exponential decay model describes the temporal evolution of the concentrations reasonably well. Although these two assumptions have significant uncertainty, the Kalman filter can integrate both pieces of complementary information and provide reliable contaminant concentration estimates. Compared to typical time-series models, our approach has the flexibility to include continuous *in situ* data as well as sparse groundwater sampling data in a consistent manner. In addition, by integrating the temporal evolution model, the estimates are less susceptible to measurement errors or data noise, compared to regression models that exclusively use *in situ* data.

The estimation results are consistent for both the source-zone and the downgradient well (FSB-95DR and FSB-110D, respectively), since the data correlations between contaminants and *in situ*variables are consistent across the site. Although the data correlations and the estimation performance are slightly higher for the downgradient well, the performance is still comparable for the source-zone well. Such spatial consistency could enable us to deploy more extensive sensor network technologies at dozens or hundreds of wells to cover the entire plume.(41)

In addition, the framework shows good performance for both H-3 and U-238, even though their physical/geochemical properties and transport mechanisms are extremely different. H-3 is a nonreactive tracer with a short half-life, while U-238 has a complex geochemistry with a long half-life. At the source-zone well, the difference between the correlation coefficients for the two contaminants is small (Table 1), possibly because the source-zone well concentrations are influenced mainly by the source concentrations rather than by plume mobility. At the downgradient well, the correlation coefficients between the contaminants and the H^+ activity are very different. After applying the Kalman filter, however, the predictive performance is comparable for both contaminant concentrations, since the data correlation model and the temporal evolution model are integrated within the filter.

Our results show that, with few exceptions, having two *in situ* variables (H^+ activity and SC) generally improves the estimation performance compared to using only one of them (Table 2). This is because, although the pH is measurable *in situ* and is a key driver for uranium transport, pH measurements are known to be prone to calibration and other errors, leading to lower correlation coefficients. The SC, on the other hand, is easily measurable with higher reliability. Although there have been more *in situ*

sensors for groundwater quality developed in recent years,(42) their accuracy and reliability have to be carefully assessed before applying our method.

Our approach is likely to be transferable to other sites and other contaminants. Groundwater contamination by radionuclides (including U-238 and H-3) has been found at many locations in the world, including not only nuclear weapons production sites but also nuclear power plants,(43)uranium mining and processing sites,(44) and fertilizer production sites.(45) For other contaminant types, such as organic solvents, we expect different *in situ* measurable variables to be indicative of contaminant concentration changes, e.g., dissolved oxygen or carbon. New *in situ* sensors are currently being developed, e.g., optical sensors for dissolved oxygen, carbon, and nitrate measurements.(46) Because of its generality and flexibility, our approach will expand the applicability of these new sensors and possibly drive the development of others.

In this demonstration, we have estimated the Kalman filter parameters based on historical data sets. This means that the filter's estimation performance depends on the amount and quality of the available data. Most sites accumulate a significant amount of groundwater data before transitioning from active remediation to ENA or MNA. Proper archiving and management of this data(47) are critical for applying our method. For sites that do not have sufficiently large historical data sets, the Kalman filter can be expanded to accommodate time-variant parameters,(29) and the correlations can be updated while the filter is in use.

Furthermore, the Kalman filter approach can include other models in addition to the exponential decay model used in this study, e.g., numerical flow and transport simulation models.(48)Therefore, our approach provides a flexible framework for real-time continuous monitoring of contaminated sites that can confirm plume stability, detect contaminant concentration changes, and support the transition from active remediation to ENA and MNA.

Acknowledgments

This material is based upon work supported as part of the ASCEM project, which is funded by the U.S. Department of Energy Environmental Management, and as part of the Lawrence Berkeley National Laboratory Science Focus Area, which is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, both under Award Number DE-AC02- 05CH11231 to Lawrence Berkeley National Laboratory. F.S. was supported by the Jane Lewis Fellowship at UC Berkeley.

Abbreviations

DOE	United States Department Of Energy
ENA	enhanced natural attenuation
FSB	groundwater well series name

H-3	hydrogen-3, tritium
I-129	iodine-129
MNA	monitored natural attenuation
PC	principal component
PCA	principal component analysis
RMSE	root-mean-square error
SC	specific conductance
SI	Supporting Information
SRS	United States DOE Savannah River Site
Tc-99	technetium-99
U-238	uranium-238

References

This article references 48 other publications.

1.

Ellis, D. E.; Hadley, P. W. Sustainable Remediation White Paper - Integrating Sustainable Principles, Practices, and Metrics Into Remediation Projects. *Remediat. J.* 2009, *19* (3), 5– 114, DOI: 10.1002/rem.20210

2.

Bardos, P.; Bone, B.; Boyle, R.; Ellis, D.; Evans, F.; Harries, N. D.; Smith, J. W. N. Applying Sustainable Development Principles to Contaminated Land Management Using the SuRF-UK Framework. *Remediat. J.*2011, *21* (2), 77–100, DOI: 10.1002/rem.20283

3.

Wang, J.; McDaniel, P. *Smart Energy Resources Guide*; 2008.

4.

Denham, M. E.; Eddy-Dilek, C. Sustainable Remediation of Radionuclides By a Common Sense Approach to Enhanced Attenuation. In *Proceedings of WM2016 Conference, March 6–10, 2016, Phoenix, Arizona, USA*; 2016.

5.

Burger, J. Protective Sustainability of Ecosystems Using Department of Energy Buffer Lands as a Case Study. *J. Toxicol. Environ. Health, Part A* 2007, *70* (21), 1815– 1823, DOI: 10.1080/15287390701459205

6.

Pierce, E. M.; Freshley, M. D.; Hubbard, S. S.; Looney, B. B.; Zachara, J. M.; Liang, L.; Lesmes, D.;Chamberlain, G. H.; Skubal, K. L.; Adams, V.; *Scientific Opportunities to Reduce Risk in Groundwater and Soil Remediation*; Pacific Northwest National Lab: Springfield, VA, 2009; DOI: 10.2172/963597 .

7.

Li, L.; Steefel, C. I.; Williams, K. H.; Wilkins, M. J.; Hubbard, S. S. Mineral Transformation and Biomass Accumulation Associated With Uranium Bioremediation at Rifle, Colorado. *Environ. Sci. Technol.* 2009, *43*(14), 5429–5435, DOI: 10.1021/es900016v

8.

Adachi, N.; Adamovitch, V.; Adjovi, Y.; Aida, K.; Akamatsu, H.; Akiyama, S.; Akli, A.; Ando, A.; Andrault, T.;Antonietti, H.Measurement and Comparison of Individual External Doses of High-School Students Living in Japan, France, Poland and Belarus—the ‘D-Shuttle’ Project—. *J. Radiol. Prot.* 2016, *36* (1), 49–66, DOI: 10.1088/0952-4746/36/1/49

9.

Kerkez, B.; Glaser, S. D.; Bales, R. C.; Meadows, M. W. Design and Performance of a Wireless Sensor Network for Catchment-Scale Snow and Soil Moisture Measurements. *Water Resour. Res.* 2012, *48* (9),W09515, DOI: 10.1029/2011WR011214

10.

National Research Council. *Frontiers in Massive Data Analysis*; The National Academies Press:Washington, DC, 2013; DOI: 10.17226/18374 .

11.

Pellerin, B. A.; Bergamaschi, B. A.; Horsburgh, J. S. *In Situ Optical Water-Quality Sensor Networks - Workshop Summary Report*; U.S. Geological Survey: Reston, VA, 2012.

12.

Eddy-Dilek, C.; Millings, M. R.; Looney, B. B.; Denham, M. E. *Innovative Strategy for Long Term Monitoring of Metal and Radionuclide Plumes*; Savannah River Site (SRS): Aiken, SC (United States), 2014.

13.

Dafflon, B.; Wu, Y.; Hubbard, S. S.; Birkholzer, J. T.; Daley, T. M.; Pugh, J. D.; Peterson, J. E.; Trautz, R. C.Monitoring CO 2 Intrusion and Associated Geochemical Transformations in a Shallow Groundwater System Using Complex Electrical Methods. *Environ. Sci. Technol.* 2012, *47* (1), 314– 321, DOI: 10.1021/es301260e

14.

Gasperikova, E.; Hubbard, S. S.; Watson, D. B.; Baker, G. S.; Peterson, J. E.; Kowalsky, M. B.; Smith, M.;Brooks, S. Long-Term Electrical Resistivity Monitoring of Recharge-Induced Contaminant Plume Behavior.*J. Contam. Hydrol.* 2012, *142–143*, 33– 49, DOI: 10.1016/j.jconhyd.2012.09.007

15.

Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* 1960, *82* (1), 35–45, DOI: 10.1115/1.3662552

16.

Cooper, S.; Durrant-Whyte, H. A Kalman Filter Model for GPS Navigation of Land Vehicles. In *Proceedings of IEEE/RSJ. International Conference on Intelligent Robots and Systems (IROS'94)*; IEEE: 1994; Vol. 1, pp157– 163, DOI: 10.1109/IROS.1994.407396 .

17.

Work, D. B.; Blandin, S.; Tossavainen, O. P.; Piccoli, B.; Bayen, A. M. A Traffic Model for Velocity Data Assimilation. *Appl. Math. Res. eXpress* 2010, *39* (2), 141– 167, DOI: 10.1093/amrx/abq002

18.

Blandin, S.; Couque, A.; Bayen, A.; Work, D. On Sequential Data Assimilation for Scalar Macroscopic Traffic Flow Models. *Phys. D* 2012, *241* (17), 1421–1440, DOI: 10.1016/j.physd.2012.05.005

19.

Sahoo, A. K.; De Lannoy, G. J. M.; Reichle, R. H.; Houser, P. R. Assimilation and Downscaling of Satellite Observed Soil Moisture over the Little River Experimental Watershed in Georgia, USA. *Adv. Water Resour.*2013, *52*, 19–33, DOI: 10.1016/j.advwatres.2012.08.007

20.

Olfati-Saber, R. Distributed Kalman Filtering for Sensor Networks. In *2007 46th IEEE Conference on Decision and Control*; IEEE: 2007; pp 5492– 5498, DOI: 10.1109/CDC.2007.4434303 .

21.

Tokunaga, T. K.; Wan, J.; Denham, M. E. Estimates of Vadose Zone Drainage from a Capped Seepage Basin, F-Area, Savannah River Site. *Vadose Zone J.* 2012, *11* (3), 0, DOI: 10.2136/vzj2011.0131

22.

Killian, T. H.; Kolb, N. L.; Corbo, P.; Marine, I. W. *Environmental Information Document F-Area Seepage Basins*; DPST-85-706; Savannah River Laboratory: 1987.

23.

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. *J. Contam. Hydrol.* 2013, *151*, 34– 54, DOI: 10.1016/j.jconhyd.2013.04.005

24.

Sassen, D. S.; Hubbard, S. S.; Bea, S. A.; Chen, J.; Spycher, N.; Denham, M. E. Reactive Facies: An Approach for Parameterizing Field-Scale Reactive Transport Models Using Geophysical Methods. *Water Resour. Res.* 2012, *48* (10), W10526, DOI: 10.1029/2011WR011047

25.

Wainwright, H.; Chen, J.; Sassen, D.; Hubbard, S. Bayesian Hierarchical Approach and Geophysical Data Sets for Estimation of Reactive Facies over Plume Scales. *Water Resour. Res.* 2014, *50* (6), 4564– 4584,DOI: 10.1002/2013WR013842

26.

USEPA. *Superfund Record of Decision: Savannah River (USDOE) (Operable Unit 6), SC1890008989*; Aiken, South Carolina, 1993.

27.

Dong, W.; Tokunaga, T. K.; Davis, J. A.; Wan, J. Uranium(VI) Adsorption and Surface Complexation Modeling onto Background Sediments from the F-Area Savannah River Site. *Environ. Sci. Technol.* 2012,*46* (3), 1565– 1571, DOI: 10.1021/es2036256

28.

Everitt, B.; Hothorn, T. *An Introduction to Applied Multivariate Analysis with R*; Springer Science & Business Media: 2011; DOI: 10.1007/978-1-4419-9650-3 .

29.

Hamilton, J. D. *Time Series Analysis*; Princeton University Press: Princeton, 1994; Vol. 2.

30.

Brown, C. E. Use of Principal-Component, Correlation, and Stepwise Multiple-Regression Analyses to Investigate Selected Physical and Hydraulic Properties of Carbonate-Rock Aquifers. *J. Hydrol.* 1993, *147*(1–4), 169– 195, DOI: 10.1016/0022-1694(93)90080-S

31.

Barker, J. F.; Barbash, J. E.; Labonte, M. Groundwater Contamination at a Landfill Sited on Fractured Carbonate and Shale. *J. Contam. Hydrol.* 1988, *3* (1), 1– 25, DOI: 10.1016/0169-7722(88)90014-9

32.

Davison, A. C. *Statistical Models*; Cambridge University Press: 2003; Vol. 11.

33.

Chen, J.; Hubbard, S. S.; Williams, K. H. Data-Driven Approach to Identify Field-Scale Biogeochemical Transitions Using Geochemical and Geophysical Data and Hidden Markov Models: Development and Application at a Uranium-Contaminated Aquifer. *Water Resour. Res.* 2013, *49* (10), 6412– 6424, DOI: 10.1002/wrcr.20524

34.

Wainwright, H. M.; Flores Orozco, A.; Bücker, M.; Dafflon, B.; Chen, J.; Hubbard, S. S.; Williams, K. H.Hierarchical Bayesian Method for Mapping Biogeochemical Hot Spots Using Induced Polarization Imaging.*Water Resour. Res.* 2016, *52* (1), 533– 551, DOI: 10.1002/2015WR017763

35.

Denham, M. E.; Eddy-Dilek, C. Influences on Effective Decay Rates of Radionuclides in Groundwater: F-Area Seepage Basins, Savannah River Site; WM2017 Conference, Phoenix, Arizona, USA, 2017.

36.

Kinase, S.; Takahashi, T.; Sato, S.; Sakamoto, R.; Saito, K. Development of Prediction Models for Radioactive Caesium Distribution within the 80-Km Radius of the Fukushima Daiichi Nuclear Power Plant.*Radiat. Prot. Dosim.* 2014, *160* (4), 318– 321, DOI: 10.1093/rpd/ncu014

37.

Newell, C. J.; Rifai, H. S.; Wilson, J. T.; Connor, J. A.; Aziz, J. A.; Suarez, M. P. *Calculation and Use of First-Order Rate Constants for Monitored Natural Attenuation Studies*; United States Environmental Protection Agency, National Risk Management Research Laboratory: 2002.

38.

Tardiff, M. F.; Katzman, D. Estimating Contaminant Attenuation Half-Lives in Alluvial Groundwater Systems.*J. Environ. Monit.* 2007, *9* (3), 266, DOI: 10.1039/b700494j

39.

Hinrichsen, D.; Pritchard, A. J. *Mathematical Systems Theory I*; Texts in Applied Mathematics; Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; Vol. 48, DOI: 10.1007/b137541 .

40.

Kaplan, D. I.; Roberts, K. A.; Schwehr, K. A.; Lilley, M. S.; Brinkmeyer, R.; Denham, M. E.; DiPrete, D.;Li, H.-P.; Powell, B. A.; Xu, C.Evaluation of a Radioiodine Plume Increasing in Concentration at the Savannah River Site. *Environ. Sci. Technol.* 2011, *45* (2), 489– 495, DOI: 10.1021/es103314n

41.

Zhang, Z.; Glaser, S. D.; Bales, R. C.; Conklin, M.; Rice, R.; Marks, D. G. Technical Report: The Design and Evaluation of a Basin-Scale Wireless Sensor Network for Mountain Hydrology. *Water Resour. Res.* 2017, *53*(5), 4487– 4498, DOI: 10.1002/2016WR019619

42.

Loden, P.; Han, Q.; Porta, L.; Illangasekare, T.; Jayasumana, A. P. A Wireless Sensor System for Validation of Real-Time Automatic Calibration of Groundwater Transport Models. *J. Syst. Softw.* 2009, *82*, 1859– 1868,DOI: 10.1016/j.jss.2009.05.049

43.

Richards, S.; Shannon, M.; Frye, T.; Keim, A.; Shepherd, J.; Klementowicz, S.; Nicholson, T.; Nimitz, R.Liquid Radioactive Release Lessons Learned Task Force Final Report Executive Summary; 2006.

44.

Abdelouas, A. Uranium Mill Tailings: Geochemistry, Mineralogy, and Environmental Impact. *Elements* 2006,*2* (6), 335– 341, DOI: 10.2113/gselements.2.6.335

45.

Bituh, T.; Marovic, G.; Franic, Z.; Sencar, J.; Bronzovic, M. Radioactive Contamination in Croatia by Phosphate Fertilizer Production. *J. Hazard. Mater.* 2009, *162* (2–3), 1199– 1203, DOI: 10.1016/j.jhazmat.2008.06.005

46.

Pellerin, B. A.; Bergamaschi, B. A.; Horsburgh, J. S. *In Situ Optical Water Quality Sensor Networks - Workshop Summary Report*; 2012.

47.

Agarwal, D. A.; Faybishenko, B.; Freedman, V. L.; Krishnan, H.; Kushner, G.; Lansing, C.; Porter, E.;Romosan, A.; Shoshani, A.; Wainwright, H.A Science Data Gateway for Environmental Management.*Concurr. Comput. Pract. Exp.* 2016, *28* (7), 1994– 2004, DOI: 10.1002/cpe.3697

48.

Schöniger, A.; Nowak, W.; Hendricks Franssen, H.-J. Parameter Estimation by Ensemble Kalman Filters with Transformed Data: Approach and Application to Hydraulic Tomography. *Water Resour. Res.* 2012, *48*(4), W04502, DOI: 10.1029/2011WR010462