Realistic Geochemical Parameter Uncertainty for Performance Assessment Modeling – 17066

Amy Jordan*, Stephanie Fitchett*, Kate Catlett*, Daniel Levitt*, Gregg Occhiogrosso*, John Tauxe*, Patti Meeks*, Megan Higgs*, Paul Black*

* Neptune and Company, Inc., Los Alamos, New Mexico 87544

ABSTRACT

Performance assessments (PAs) include uncertain geochemical parameters that affect the fate and transport of contaminants in environmental media, such as partitioning coefficients for water/rock (K_d) and water/air (K_H), diffusion coefficients, and aqueous solubilities. A statistical approach to realistic, as opposed to "conservative," stochastic distribution development for geochemical parameters is presented here. For contaminants not identified as focus elements during screening, broad generic distributions are developed. If any parameter is flagged in the subsequent model sensitivity analysis (SA), it becomes a focus element for further research during the next round of distribution development. This iterative approach is efficient and technically defensible at each stage.

Data are collected in databases and may be appropriately weighted or filtered. In the relatively few cases where data exist from local or analogous rocks, those data may be prioritized. Once established, the databases can be used for future PAs or other environmental models and can be expanded as needed based on model SA results, additional screening, and as additional relevant data become available.

The iterative method is demonstrated for K_d and K_H in a PA model for a radioactive waste disposal facility. An SA was performed based on preliminary geochemical parameter distributions that showed certain model endpoints are sensitive to K_d and K_H values for some contaminants. This underscores the importance of including realistic uncertainty in these geochemical parameters.

INTRODUCTION

Modeling the long-term performance of radioactive waste disposal in geologic media has been an active area of research with continual improvement in methods and techniques. Typically, a site-specific performance assessment (PA) is a systems-level model composed of suitably coupled sub-models that simulate the processes leading to the transport and potential exposure of populations to hazardous constituents in waste. Geochemical parameters affect many of these processes, including contaminant transport in porous media, plant and animal uptake and redistribution of radionuclides, waste leaching, waste form degradation, and receptor exposure in scenarios such as gardening and gas inhalation.

Development of a sound geochemical approach for site-specific PA models can be costly and time-consuming. Locally sourced data from the same or analogous geologic material and aqueous environment are typically scarce, except for the largest-scale PA efforts in the country. When available, local data usually assess only a few elements and may not be truly representative of the full range of site conditions possible over the time evolution of the repository.

In most PA model environments, contaminants may be transported as gases, as aqueous solutes or as colloids in water, or may be retained when bound to rock and cement. "Partitioning" refers to the calculation of the fractions of a constituent found in each of these media: air, water, and solid. Air/water partitioning is frequently modeled using Henry's Law, and water/solid partitioning using a linear sorption approach. Both of these are simplified models that rely on assumptions of equilibrium and linear, reversible behavior. More complex models are available that could be employed if data are available to support and parameterize them. The benefit of these simplified models is they require few parameters and

are easy to implement, while generally describing appropriate behavior in the relevant and sensitive range of conditions (relatively dilute systems, weak sorption [1-3]).

The parameter of interest for water/solid partitioning is K_d, which represents the ratio of contaminant sorbed on rock to the amount in aqueous solution. K_d values are different for each elemental, ionic, or molecular species present, as determined by the species' relative affinity for the mineral phases as compared to water. Aqueous environmental factors that influence K_d values include redox potential, pH, pCO_2, ionic strength, dissolved organic carbon, and the presence of other species; however, there may be great uncertainty in subsurface water chemistry, especially in proximity to waste forms. Aspects of the solid material that affect K_d values are mineralogical fraction, which may be heterogeneous within a material unit (i.e., percent and type of clays, sand, etc.); pore size and surface area (e.g., sand versus gravels, crushed or intact material); and the presence of organic matter. Finally, laboratory K_d experiments are notoriously complex to interpret in the context of heterogeneous field sites [3]. The type of experiment (e.g., batch versus column) and methods of sample preparation (e.g., crushing, sieving), solution treatment, and experiment run time are all important factors that affect the results and their applicability to a particular site-specific scenario. All of the factors listed above contribute to the great difficulty in determining and applying K_d distributions.

Air/water partitioning following Henry's Law has temperature-dependent coefficients, $K_H(T)$. As with K_ds, the local aqueous geochemical environment influences K_H values, including ionic strength and solution composition, and dissolved organic matter [1].

Historically, many PAs used deterministic and so-called "conservative" values for parameters. However, it is now clear that when all of the processes in a PA model are considered, what is conservative for transport is not necessarily conservative for dose [4]. For example, the K_d of Tc-99 is frequently set to 0, with the justification that the value is conservative (i.e., the radionuclide is in a highly mobile form). While this may be realistic under some conditions, there are cases where the Tc K_d deviates from 0. Assuming a lower K_d will keep the radionuclide in solution and transport it throughout the model, instead of binding it to the solid material. This could result in a *higher* predicted groundwater concentration but *lower* predicted dose for the farmer or gardener who comes in contact with contaminated soil. A better approach is to use available information to develop *realistic* distributions that capture relevant sources of variability and uncertainty at a scale consistent with how the parameter is used within the PA model [5].

Several classic references are often used to assign K_d values, particularly for radionuclides where few data exist in the literature. Sheppard and Thibault [2] performed a literature review for K_d values for 49 elements in various media (sand, loam, clay, and organic-rich soil), with data from the U.S. and Canada providing lognormal K_d distributions for 48% of the results. For the remaining values, the tables are populated using soil/plant partitioning ratios, converted to K_ds following an empirical relationship [6]. Despite the stated limitations of this survey, these values often form the basis for K_d values for PAs through a chain of commonly-cited references, e.g., [7]. A number of other literature reviews, compilations, and critical analyses have also been published that cover high-priority elements (e.g., [8-12]) as well as environments of importance to waste disposal, such as cementitious materials (e.g., [13, 14]). A free online K_d database that allows user submissions is also available [15]. For major PA efforts, K_d databases have been established with site-specific data (e.g., [16-19]). Thermodynamic sorption modeling can be used to inform site-specific K_ds, with limitations [13, 20]. For K_H, a thorough literature review by Sander [21] has been updated [1] with additional data.

As PAs have been developed that are large in scope and have spurred the advancement of modeling techniques, probabilistic PAs have become more common, including for smaller-scale PA sites. Yet it is still common to see deterministic values for many contaminants that are assumed to not influence dose, and application of default distributions to key elements, such as uniform or triangular, often applied

without sound justification. A great improvement in technically defensible stochastic geochemical parameter development is possible, even in the absence of local data. While simple distributions may be adequate for assessing the parameters to which the model is sensitive and offer an improvement over deterministic values, more realistic distributions should be made for performance prediction purposes.

The iterative approach presented here focuses effort where it will have the greatest benefit for reducing uncertainty in risk and dose. Preliminary results have identified K_H and K_d for some elements as sensitive parameters, providing ample motivation for advancing our methods for their distribution development, as well as for other parameters. This geochemical parameter development approach uses appropriate statistical techniques, allows PA modelers to take advantage of the growing body of data available, including site-specific data, and the approach is easily extendable to other PA sites once the databases and framework are in place. An example of use in a PA model is provided in the Results section.

METHODS

After defining the target or goal for the distribution [5], the iterative approach has four main steps that result in model implementation (Fig. 1). In the first step, Screening, key questions include: which contaminants are present at the site? What quantities are present in the inventory, how mobile are they, and how much impact do they have on dose? Expert judgment or prior sensitivity analysis (SA) results [22] can also be used to screen focus elements. Unlike the common use of deterministic values for contaminants not thought to be important, "generic" distributions are generally *broader* for elements not identified as focus elements and apply to multiple elements. Any of these contaminants may be flagged during SA if sensitive and returned to the beginning as a new focus element. However, the modeler must be aware that wider distributions may not be an appropriate substitute for missing model structure or physical processes. The risk in that case is the SA could fail to catch sensitive parameters, and incorrect model results could show a diluted risk to certain populations.

During Data Collection, the most relevant available sources are gathered. To determine relevance, some site assessment is performed first to identify the expected media types and aqueous environmental conditions at the site. In Filtering, data are culled based on the site assessment with consideration to the amount of data available, its quality and relevance to the site. Exploratory data analysis (EDA) is extremely useful for determining whether categorizations of the data have been appropriate, and new choices may be made if necessary. Finally, the data subsets are ready for Statistical Analysis, where distributions are developed for focus elements and other contaminants. The distributions are implemented in the model, and SA identifies new focus elements. These steps are discussed in more detail in the following sections for K_d and K_H distribution development, but can also be applied to other parameter development, such as solubility, diffusivity, and material properties.
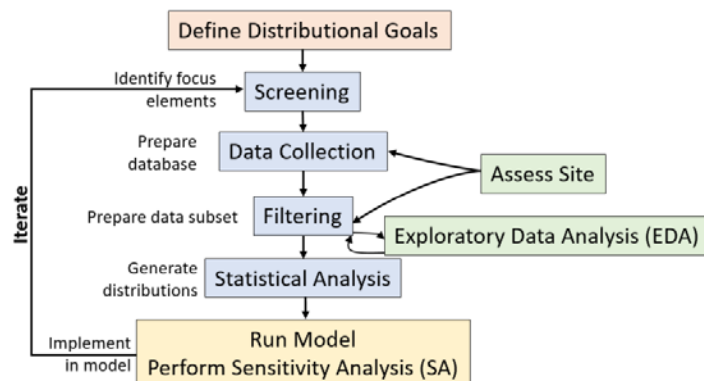

Fig. 1. Overview of the iterative method for geochemical parameter distribution development.

Radionuclide Screening

The process of reviewing and collecting sufficient data for geochemical parameter development may be time-consuming, especially with adequate quality assurance (QA) and deciphering the complicated dependence of geochemical data on the aqueous environment and the methods used to collect the data. At a typical waste site, after inventory analysis, a number of contaminants of potential concern are retained for the PA model and thus require geochemical parameter distributions. Not all of these will contribute strongly to dose, which is revealed in subsequent SA. However, a preliminary screening may be performed at the outset to reduce the list of contaminants that require increased attention.

The preliminary screening process involves sorting the list of all contaminants retained for the model into high, medium, and low bins for the three categories of amount (activity or mass of contaminant present), mobility (based on K_d), and dose (based on dose conversion factors [DCFs]). The bin endpoints are subjective and may be changed between models based on expert judgement, particularly for inventory. The high and medium bins are compared and contaminants that appear in all three lists are compiled. For example, I-129 falls into the "high" bin for K_d (i.e., highly mobile) and for dose, and is in the "medium" bin for inventory at one site. Progeny of screened radioelements and of high-mobility radiological parents are added (if they are not already included). Screening can also include consideration of other criteria, site-specific knowledge, prior sensitivity results, and expert judgment to add or remove focus elements.

For radiological contaminants, an important distinction must be made between radionuclides, elements, and species. While radiological PA models simulate transport of each radionuclide identified as a contaminant of potential concern, geochemical properties do not depend on isotope (except where molecular mass becomes important, as for tritiated versus untritiated water vapor) and so distributions are developed by element. For example, U-233, U-234, U-235, U-236, and U-238 are grouped into the *element* or *radioelement* uranium for geochemical parameter assignment. Each element may be found in multiple molecular or ionic *species*. For example, U in oxidizing environments is most commonly found as UO_2^{2+}, which hydrolyzes to form predominantly $UO_2(OH)_2$ from pH 5–9; depending on species present, complexes with carbonates, chloride, sulfate, and phosphate are also possible [10]. Each of these species will have different geochemical parameters due to size, charge, and affinity for other molecules. The problem of speciation is especially complex for carbon/C-14. Carbon migrates in the environment in many forms: as CO_2 and CH_4 gases, as part of organic compounds (both aqueous and gaseous), and in solution as carbonate, bicarbonate, and carbonic acid, and bound with other species such as U.

Data Collection

To produce the database for K_d distribution development, an extensive review of both peer-reviewed and white paper literature was performed, incorporating K_d data from over 60 sources. The intention was to include a wide range of available data for each key element, so distributions could be built that reflect the uncertainty in the subsurface environment at the site and to allow flexibility in including or excluding data, based on expert judgement of the applicability of those data to the site. The database sources included the following types of documents, in generally decreasing order of completeness of information:

1) Papers containing original data, typically found in peer-reviewed journals and including extensive information about experimental conditions, methodology, and uncertainty.
2) Literature reviews, which provide references to the original data. This category includes commonly cited sources, e.g., [2, 9, 12-14].
3) Model description white papers, reports, or data packages, which give K_d values determined for a particular modeling application. These reports sometimes provide an explanation of the methods used to determine these values, often based on a combination of literature review, such as the papers in category 2 above, site-specific data if available, and expert judgement.

All data that could be obtained from these sources are entered into the database, including (when available) the information listed in Table I. Care is taken to avoid entering different experimental references that are actually repeats of the same data, but data duplication will occur when literature reviews are entered. The database identifies whether each source is primary experimental data, literature review, mechanistic sorption model results, expert judgement, or values used in another model. Having this information helps avoid repeated measurements and bias, while still allowing the user to easily compare resulting distributions with previously used values for PAs or with commonly accepted distributions. The type of source can also be used as a user-defined proxy for the quality of the data.

Geochemical speciation modeling is sometimes used to determine the breakdown of species for each element of interest under the aqueous environmental conditions at a site. In the iterative approach presented here, data for multiple species may be entered, as well as data from natural geochemical environments, which can produce broader distributions than those focused on an individual species. As contaminants are identified as sensitive in the SA, further study or geochemical speciation modeling can be performed to determine which species are likely to be present in what percentages. This can then be used to weight or filter the data in the database.

Despite the importance of experimental conditions in interpreting K_d values, many literature sources do not include complete information about experimental factors such as pH and Eh, particularly in large literature reviews. Some sources give only a coarsely defined differentiation of the relevant porous media, such as "sand," "clay," "basalt," while others provide more detailed mineralogical and/or textural conditions. How incomplete data were dealt with depended on how many data were available, and the extent of impact of the particular condition on the partitioning behavior of that element, as described in the discussion of filtering below.

TABLE I. K_d database components and description

Item	Description
Date added	Used for QA purposes. After initial QA, QA is performed on new entries only.
Element	Element symbol/name.
K_d value	Reported K_d values, converted to units of mL/g if necessary.
Value type	Values reported in the literature may be means across a wide range of studies (as in a literature review), means of replicates of samples in the same study, single measurement values, or values proposed based on expert judgement.
Data type	Type of study: literature review, experiment, geochemical model, or values used in other models (e.g., in a previous PA).
Min/max	Where ranges are given, the min and max columns are populated.
Spread	If spread information is given (e.g., standard deviation), it is listed here.
Spread type	Identifies the type of spread or distribution, if any. Common examples include one or more standard deviations, standard deviation of logarithms, and variance.
N samples	The number of samples represented by the value given on a line in the database.
Soil type	The soil, rock, or engineered material type exactly as given in the source.
Location	If the geographic location of the sample is provided, it is recorded.
pH	pH or pH range.
Eh	Numerical (in mV) or descriptive redox information, e.g., "oxic" or "anoxic."
Species	If known, the species is given exactly as in the source, e.g., TcO_4^-, Np(V).
Source	An ID code linking the value to a particular reference.
Additional information	Used for pointing where in the reference the values or information may be found (e.g., "Table 3," "text p.33"). Also used for any other information that may affect K_d values, such as "high ionic strength solution."

To construct the K_H database, Henry's Law coefficients are gathered from the literature, primarily [1], but also other primary sources not found in the former. Values for partitioning of tritiated water vapor from the liquid at various temperatures are also collected [23-25]; these data are analyzed separately.

Literature values for K_H spanned a range of temperature T (0 to 30°C) based on the experimental conditions. The values were adjusted to $K_{H,0}$ at a reference temperature of $T_0 = 25$°C using [21]:

$$K_{H,0} = K_H(T)\exp\left[-\left(\frac{-d(\ln K_H)}{d(1/T)}\right)\left(\frac{1}{T_0+273.15} - \frac{1}{T+273.15}\right)\right] \qquad \text{(Eq. 1)}$$

The temperature dependence parameter $-d(\ln K_H)/d(1/T)$ is also found in the literature, and is entered in the database when available.

Henry's Law applies for dilute solutions [1]. However, groundwater, pore waters, and waste-altered waters at disposal sites may have high dissolved solids content due to interactions with waste forms and geologic media. These effects on K_Hs were not considered in the first round of distribution development, but may be an area for future research for sensitive elements.

K_d Approach – Filtering and Statistical Analysis

The following methods are used to prepare K_d distributions (Filtering and Statistical Analysis steps in Fig. 1). First, site-specific environmental conditions are assessed and the types of materials and water present are identified. Typical solids include cover layers composed of different rocks and soils, the waste, disturbed native material (e.g., crushed rock), intact geologic media, cement, and bentonite/clay systems. In some cases, materials may be utilized specifically for their effect on geochemical behavior, e.g., reducing agents to diminish mobility.

Water "types" at a site may include pore water, near-surface meteoric water, groundwater, and waters altered by leaching from the wastes or waste forms (e.g., cementitious water). The aqueous environmental conditions of these water types are assessed: what is the expected pH of pore water in the native soil? In intact rock? Around the waste and waste forms? What are the redox conditions and ionic strength? What ions are likely to be present in solution (e.g., carbonates)? Some of these answers will be used to filter the data, *if* enough information is available in the database to do so.

A list is constructed during this assessment that includes each combination of soil/rock type and associated water chemistry that will be considered in the model. Due to data limitations, it is usually necessary to make the groupings broad and have as few combinations as is practical. In the example PA model below, the solid material groupings are Sand, Clay, and Cement; the water types are pore water (used for Sand and Clay) and oxidizing or reducing cementitious water.

For each solid material, determinations must be made about which rock types in the database are applicable. For example, the category "Sand" may be applied in the model to sandstone, gravel cover layers, and topsoil (i.e., the category is much more broad than the name implies). Therefore, to ensure the distribution is appropriately broad, a range of rock types in the database may be selected for inclusion in this category: e.g., sand, loam, soil, sediment, while for Clay, rock type descriptors in the database such as clay, silty clay, clay/sand, and silty clay loam are selected. If a K_d in Sand or Clay is flagged as sensitive when used with a broad distribution, additional research can refine the category or characterize different layers separately. Later in the iterative process, for elements that are sensitive, mineralogical considerations can also be included in filtering. For example, high Tc K_d values can be due to iron oxide surface coatings, which might suggest excluding data by the Location field for sites known to have such conditions in some sediments, e.g., Savannah River Site [26], if they are not present at the study site.

Similarly, for each water type, determinations are made about which data are applicable. If water conditions in a layer are expected to be oxidizing with pH > 7, only data with known oxic conditions and neutral or greater pH could be selected. In practice, because many data lack specific information about redox conditions and pH, if there are very few data for an important element, the database user may wish to make assumptions (e.g., data without pH information may be used). This is especially reasonable for elements for which these conditions are not thought to significantly impact K_ds. For example, for elements with K_ds sensitive to redox conditions (e.g., Am, Cr, Np, Pu, Tc, U [9]), the user may choose to exclude data where redox conditions in the dataset are unknown, while for other elements, those data could still be used in the first round of distribution development.

Some elements demonstrate particular K_d characteristics. Negative K_d values are found in the literature for some species (such as TcO_4^- and Cl^-) in some materials [9]. The explanation given for this seemingly non-physical result is anion exclusion [27], caused by repulsion between charges in the material surface and the contaminant. This causes higher sampled concentrations than expected from initial solution concentrations. Noble gases dissolved in aqueous solution, such as Ar, Kr, and Rn, have $K_d = 0$ based on their inert geochemical behavior. Hydrogen (tritium) is generally thought to be non-partitioning, but as HTO it may replace bound H_2O and show a small apparent partitioning effect. Therefore, some nonzero hydrogen values are found in the data [10].

Some elements, such as Ra, undergo coprecipitation, e.g., to $(Ba,Ra)SO_4$ in sulfate-bearing waters. This shows up as very high K_d values despite actual K_ds based on sorption being rather low. This affects the database values as some experiments control for the effect while others do not [11]. Because this process is expected to take place in natural environments, these unrealistically high apparent K_d values may be retained for distribution development.

Carbon is another element affected by precipitation, which makes describing the complex partitioning and transport of C-14 in a simple K_d model problematic. At some sites, a significant fraction of mobile C-14 in the vicinity of buried waste is inorganic carbon (e.g., HCO_3^- and CO_3^{2-}) [28, 29]. While precipitation is a different physical process from sorption, it may affect experimental results, giving very high apparent inorganic carbon K_d values in cement [28, 30]. For some elements, e.g., U, soluble carbonate complexes are *less* sorbing than other U species, so the presence of carbonate generally *decreases* U K_d values and thereby increases mobility.

After relevant environmental conditions and rock types are selected, EDA can help review and assess the quantity and quality of data for each element and rock type (Fig. 1). Outlying data points can be easily identified and researched further. For example, in one case several unusual data points were identified during EDA and a quick review of the original source revealed high ionic strength conditions in the water used in the experiment. The database was then filtered to exclude high ionic strength solutions beyond the expected site values.

Once several rounds of EDA are completed and the data filtering selections finalized, K_d value datasets are prepared for distribution development. The filtering, EDA, and distribution calculations are performed using the R statistical computing software [31]. Two categories of distributions are developed: those for the focus elements identified in the screening, each of which receive a distribution for each rock and water type; and "generic" distributions for all other elements (described below).

Most records report a mean K_d or a single value for K_d, though the data also include geometric means, medians, and most likely values. In some cases, only a range of values is given. When only a range is given, the midpoint of the range is used. If distributional information is also provided, a mean or geometric mean is computed. Eventually, the distribution development process will be updated to better account for distributional and spread information in the sources. For now, modeling is based on the provided individual K_d values, except in a few cases where data are scarce and a distribution is provided.

There is often considerable variation in reported values between sources, compared to relatively small variation in reported values within a source. To accommodate the differences while not allowing sources with the largest numbers of recorded K_d values to dominate the distributions, a double bootstrapping approach is taken for estimating distribution parameters.

Distributions of K_d values for individual element/rock type/water type combinations developed for the focus elements are often approximately lognormal. Ideally, all distributions used in the PA model would represent *average K_d* values, where the average is appropriate over the duration of the model and over the spatial cells used in the model [5]. If the data are assumed to represent K_d values at points in time and space that are applicable to the site, and there are sufficient data, a normal distribution may be appropriate for average K_d values for each element, using information from the data and decisions about the scale of the average. When data are sparse, distributions of average K_d values are right-skewed and might be more reasonably modeled by lognormal distributions. For each distribution associated with a focus element, when sufficient data are available, both a normal and lognormal distribution are considered. Typically, the distribution that allows greater variability on the low end of the distribution is chosen.

The following basic approach is used in this example for specifying the normal and lognormal distributions of focus elements based on bootstrapped data:

- If the data include non-positive values, a normal distribution is used, and the distribution is truncated at the minimum data value. Negative random draws from such distributions are replaced by 0 in the PA model, resulting in a point mass at 0.
- When the data include only positive values, both a normal and lognormal distribution are initially considered. The one deemed more appropriate by professional judgement is chosen, where the goal was typically to retain variability on the low end of the distribution.
- In cases where data include only positive values and a normal distribution is chosen, the distribution is truncated at one tenth the minimum K_d value in the dataset that informs the distribution development. The truncations were adopted because element/soil combinations with high K_d values also typically have considerable variation in K_d values, so normal distributions based on the parameter values obtained from the bootstrap distribution would sometimes allow negative values to be randomly drawn. Truncating in these cases disallows very small K_d values when larger ones are expected, and yet the chosen truncation allows for a reasonably wide distribution.

Generic K_d distributions are developed to represent the non-focus elements, and are intended to encompass the variability in the data for the different elements. The generic distributions do not describe distributions of averages, but rather distributions of K_d values, since the variability between different elements should not be averaged away. They are informed by all data for all elements, including the focus elements. The same criteria for pH and redox conditions are applied for filtering data for the generic distributions as the focus elements.

The generic distributions include a low, medium, and high distribution for each of the rock and water type combinations. Low K_d elements have arithmetic average of available K_d values less than 5 mL/g; medium K_d elements between 5 mL/g and 50 mL/g, inclusive; and high K_d elements, greater than 50 mL/g.

The generic K_d distributions are assumed to be lognormal. The lognormal parameters μ and σ are estimated as the mean and standard deviation, respectively, of the natural logarithm of the available K_d values, when those values are entirely or predominantly nonnegative. Exponentiating μ and σ gives the geometric mean and geometric standard deviation required for the model. For the low K_d distributions, due to nonpositive values, the arithmetic mean and standard deviation are found using all the data, and the lognormal distribution with the corresponding mean and standard deviation are used. (This allowed all of the data to inform the distribution, without having to make a substitution choice for nonpositive values.)

K_H **Approach – Filtering and Statistical Analysis**

All of the data in the K_H database that apply to dilute solution concentrations are typically used. If high ionic strength solutions are to be considered in a PA model, data can be collected for those conditions as well and filtered appropriately. Because Henry's coefficients are strongly temperature-dependent, the approach taken to sample them involves uncertainty in subsurface temperature. The soil temperature distribution developed for the site and the reference temperature $K_{H,0}$ distributions for each element, except H, are used to produce appropriate $K_H(T)$ values for each realization using the following steps:

1. Sample for $K_{H,0}$ from the given distributions for each element other than H.
2. Sample for soil temperature from the site-specific distribution.
3. Apply Eq. (1), but with roles of K_H and $K_{H,0}$ reversed, to determine $K_H(T)$ at soil temperature T.

The statistical approach for $K_{H,0}$ distributions for all elements except tritium is to develop a distribution for the individual data values, rather than the average; either normal or gamma distributions were selected. Values for the constant -d ln K_H/d($1/T$) presented in the literature show variability, but a deterministic value is used in step 3 above because holding -d ln K_H/d($1/T$) fixed has minimal effect on uncertainty in $K_H(T)$ compared to T and $K_{H,0}$.

For tritium, there are no temperature dependence parameters available. A plot of K_H versus T from the available data suggests the relationship is approximately linear over a temperature range overlapping the range of expected subsurface soil temperatures. Thus a regression is fit, and the predictive distribution for an individual value of K_H for tritium, given a specific soil temperature, is used as the distribution for K_H. The approach is the following:

1. Fit a regression for predicting K_H for tritium as a function of soil temperature T. This results in a regression equation $\hat{y}(T) = a + bT$, and an estimate s of the standard deviation of the regression residuals.
2. Sample for soil temperature from the site distribution (same value as step 2 above).
3. Use the predictive distribution for an individual value of K_H, conditional on the sampled soil temperature $T = T_p$, as the distribution describing $K_H(T_p)$.

The mean of the predictive distribution is given by the regression estimate $\hat{y}(T_p)$, and a random sample for $K_H(T_p)$ can be drawn by first obtaining a random draw t^*_{n-2} from a t-distribution with $n-2$ degrees of freedom, and then transforming that draw to get a predicted value of $K_H(T_p)$ by computing

$$\hat{y}(T_p) + t^*_{n-2}s\sqrt{1 + \frac{1}{n} + \frac{(T_p - \bar{T})^2}{\sum_{i=1}^{n}(T_i - \bar{T})^2}} \qquad \text{(Eq. 2)},$$

where the data used to fit the regression line are the pairs (T_i, K_{Hi}), $i = 1, \ldots, n$.

RESULTS

The process outlined above is illustrated with an example from a PA model that simulates long-term dose and exposure to receptors using the GoldSim contaminant transport platform [32]. The disposal facilities are excavated in a dense clay. Waste is disposed within layers of steel-reinforced concrete. Above the waste is a cover constructed of concrete, native clays, sands and soils. The mineralogy of the site is dominated by three distinct porous material types, grouped into three modeled materials: sands, clays, and cementitious materials. The materials in the model were grouped based on mineralogical and textural

similarity to the actual engineered or environmental media being modeled. For example, the Sand grouping is used for topsoil, all non-clay cover layers, and sandstone strata. The Cement grouping is used for actual cement layers as well as the waste materials in the model.

In some of the cover layers, pore waters are expected to have the native pore water chemistry, as influenced by the water source, residence time, rock mineralogy, pCO_2, Eh, pH, and temperature. Very near the surface, pore water in the cover layers (such as topsoil) may be more strongly influenced by low-pH meteoric water. In the waste disposal zone, pore waters are expected to be in equilibrium with the cementitious materials. Below the waste and concrete layers, pore water is again assumed to have a "native" composition unaffected by cement and waste materials. The selected water types for distribution development were simply pore water, assumed to be the same in Sand and Clay for lack of finer distinctions in the data; and cementitious water, applied only to the Cement material, which could be oxidizing or reducing (for future consideration of adding reductants to the waste form materials).

Screening Results

Following inventory analysis, a total of 244 radionuclides were retained for inclusion in the PA model; these represent 62 unique elements, of which 6 are volatile. To develop K_d distributions for 62 elements in the four selected combinations of rock and water type would be extremely time-consuming; therefore, the screening process was initiated as described in the Methods section.

Elements were considered for inclusion in the focus group when they appeared in two or three of the medium or high bin screened lists (quantity, mobility, and dose). Expert judgement was supplied based on prior modeling results to identify additional elements. The focus elements for K_d development selected during the first screening were Am, C, I, Np, Pu, Ra, Tc, and U. All volatile elements (Ar, C, H, I, Kr, and Rn) were considered focus elements for K_H distribution development.

K_d Distributions

Once the solid material categories were determined, collection of K_ds for these rock types was emphasized, but the database includes many other related rock/soil types as well and is continually expanding. Site conditions and availability of data dictated the filtering rules used to prepare the datasets for each media type. For elements with redox-sensitive K_ds, when information about these conditions was unavailable, those data were excluded. Otherwise, the data were retained for oxidizing distributions. Focus elements assumed sensitive to redox conditions for K_d were Am, Np, Pu, Tc, and U [9].

Soil materials at the site provide a generally neutral to alkaline aqueous environment. When pH was not specified in the original source, the data were assumed to fall under the neutral to alkaline category (pH 7 to 9) for all materials except cement, in which case the pH was assumed >9.

Although many of the soil type descriptors in the database are imprecise and are *not* expected to accurately represent the site-specific conditions, they are expected to cover a wide range of mineralogical examples and support broad distributions when filtered appropriately for the given categories in the model (note that each category in the model itself covers a range of modeled real materials). When elements are flagged as sensitive, additional research can be performed to selectively narrow the filtering rules to more media-specific values.

Fig. 2 shows an example of the filtered data and corresponding distributions developed for Tc in Sand and U in Clay. Data values are shown as points below the distribution, with values/spreads colored by source name in the database. Ranges are shown when such information was provided. When mean/geometric mean and standard deviation/geometric standard deviation (or equivalents) are provided, the 5th and 95th

percentiles of the corresponding normal/lognormal distributions are used to represent range.

Negative K_d values for Tc seen in Fig. 2(a) are common in the filtered data for sand and sand-like materials. If a negative value is sampled for a probabilistic realization in the site model, the value is changed to 0 mL/g. The GoldSim contaminant transport module does not allow for K_d values to be negative, though anion exclusion could be modeled as a separate process. Replacing negative draws in GoldSim by 0s is *not* represented in Fig. 2(a) (the scale for a point mass at 0 and a scale for the continuous density functions are different, and not accurately representable simultaneously).

Typically, there were significantly fewer data available for clay-like materials than for Sand. Fig. 2(b) illustrates some of the issues associated with distribution development for K_ds, with much of the data scattered on the low end of the range and one value at 1600 mL/g. While there are a number of sources for U in clay-like materials in the pH range 7 to 9, only two of those references provide information about the redox conditions. Because U K_ds are redox-sensitive, the only data retained are from [2] and [10]. While the point at 1600 mL/g appears to be an outlier in the selected data set, the full data (not controlling for unknown redox conditions) shows additional data in the high K_d range. Furthermore, 1600 mL/g is a value used in previous modeling for this type of clay, e.g., [33].
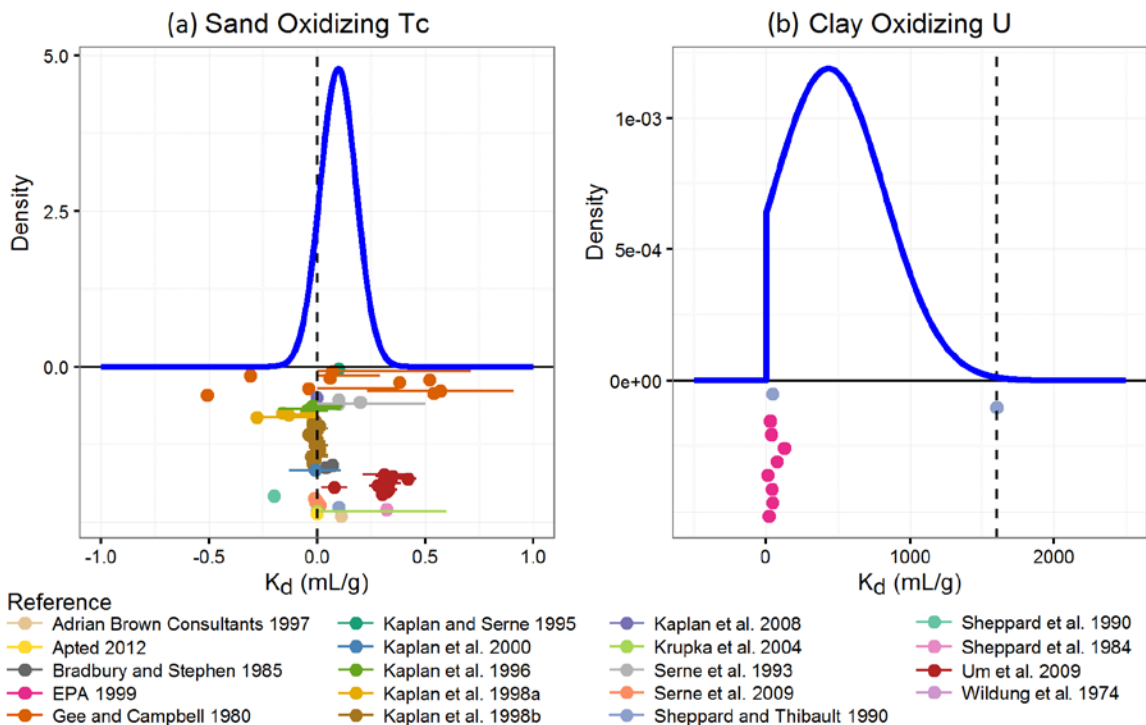

Fig. 2. (a) K_d distribution for Tc in sand-like materials. (b) K_d distribution for U in clay. Raw data that meet the screening criteria are shown on the lower portion of the plot, along with associated spread or ranges. Limited horizontal scale (not all data/spreads are shown). Dashed lines show previously-used deterministic values in prior modeling for the site.

Since an untruncated normal distribution would allow inappropriately low random draws (including negative values), and the lowest value in the filtered data set was 12 mL/g, the U distribution was truncated at 10% of the minimum value, or 1.2 mL/g. Both a normal and lognormal distribution were considered. A normal distribution was chosen because draws less than the data minimum would be rare in a lognormal distribution, and the decision was made to allow smaller average K_d values to be drawn than the individual data values. This allows a more robust sensitivity analysis. If the model is sensitive to

this parameter, more information will be sought to refine the distribution.

The non-focus elements were sorted into ranges as described in the Methods section (Low, Medium, and High) and generic distributions were developed for each media type in each range (Fig. 3). A lognormal distribution was fit to each of the groups. The distributions are shown with the horizontal axis on a logarithmic scale (so the distributions themselves look normal). On a standard scale, the High distribution spans so many orders of magnitude that a visual comparison is difficult.
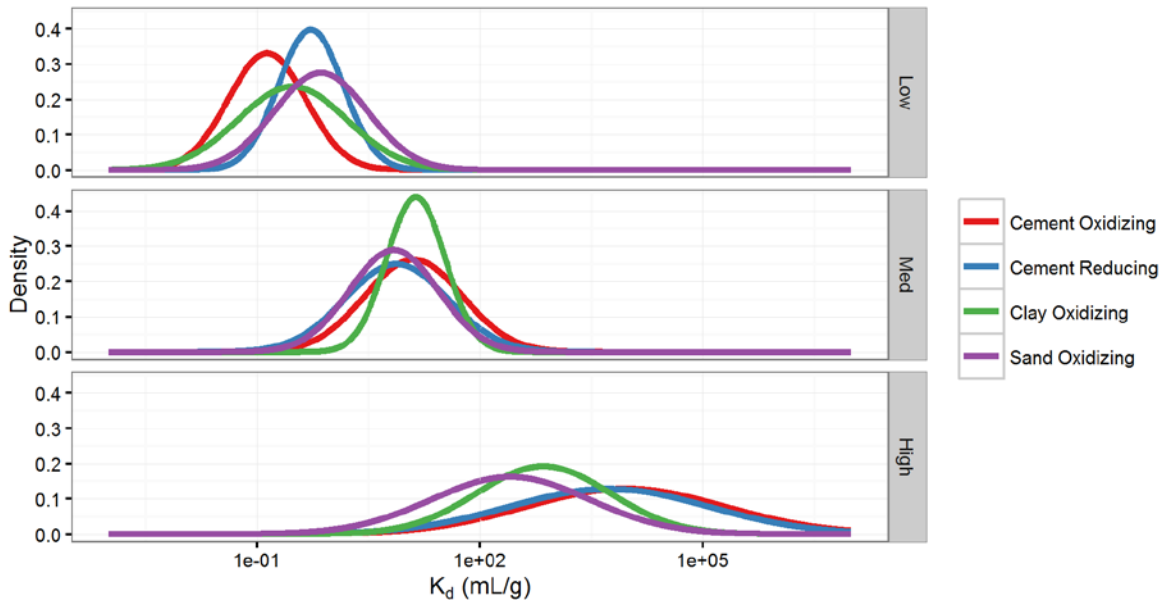

Fig. 3. Generic lognormal K_d distributions for Sand, Clay, and Cement, shown on a logarithmic horizontal axis scale.

All final distributions are given in Table II. The K_d distributions chosen for the 8 focus elements considered individually in Sand, Clay, and Cement (oxidizing and reducing conditions) are all either truncated normal or lognormal distributions. In the PA model, truncated distributions must have specified values for both minimum and maximum; an arbitrarily large value (denoted "Large") is used as the upper bound and the lower bound is either 10% of the minimum value in the dataset (if positive) or the minimum value in the dataset (if negative). The values used for the distributions are derived from the double bootstrap approach described in the Methods section. For Np and Pu in clay soils and oxidizing conditions, lognormal distributions based on [2] are used instead, and for C in cement and reducing conditions, the generic distribution for high K_d elements in cement was used. These encompass a wide range of potential values, and do not describe a distribution of averages. The wide distributions are used in these cases because of the scarcity of information and the corresponding uncertainty in K_d values.

Table II. K_d distributions for media represented in the PA model

Element	Soil/water partition coefficients (K_d) (mL/g)[a]			
	Sand	**Clay**	**Cement (oxidizing)**	**Cement (reducing)**
Ar, Kr, Rn	0	0	0	0
Am	N(2.1E3, 8.7E2, [1.2E1, Large])	LN(4.8E3, 1.4E0)	LN(7.3E3, 1.4E0)	LN(1.0E4, 1.7E0)
C	N(2.4E0, 9.9E-1, [-3.9E-1, Large])	N(1.1E0, 9.8E-1, [-6.0E-1, Large])	N(1.3E4, 8.9E3, [2.0E0, Large])	LN(5.4E3, 2.2E1)
I	N(3.3E0, 1.1E0, [-3.0E-2, Large])	N(2.9E0, 2.3E0, [-1.3E-1, Large])	N(1.8E1, 6.8E0, [0.0E0, Large])	LN(1.2E1, 3.1E0)
Np	N(8.1E1, 6.3E1, [6.3E-2, Large])	LN(5.5E1, 3.8E0)	LN(7.4E4, 7.7E0)	LN(3.3E5, 6.6E0)
Pu	N(6.4E2, 1.7E2, [1.0E1, Large])	LN(5.1E3, 2.1E0)	LN(7.1E4, 3.7E0)	LN(7.5E4, 3.8E0)
Ra	N(6.5E3, 6.6E3, [1.3E0, Large])	N(6.1E4, 5.6E4, [1.2E1, Large])	N(7.5E1, 1.0E1, [5.0E0, Large])	LN(2.2E2, 2.3E0)
Tc	N(9.8E-2, 8.3E-2, [-2.8E0, Large])	N(2.5E-1, 3.4E-1, [-1.5E-1, Large])	N(4.7E-1, 2.1E-1, [-4.6E-1, Large])	N(1.5E3, 5.6E2, [8.1E-1, Large])
U	N(9.2E0, 4.2E0, [1.4E-2, Large])	N(4.3E2, 3.9E2, [1.2E0, Large])	LN(2.7E4, 2.6E0)	N(1.1E4, 7.8E3, [1.0E1, Large])
Low Generic	LN(7.2E-1, 4.2E0)	LN(3.0E-1, 5.4E0)	LN(1.4E-1, 3.3E0)	LN(5.2E-1, 2.7E0)
Med. Generic	LN(6.9E0, 4.0E0)	LN(1.4E1, 2.5E0)	LN(1.3E1, 4.6E0)	LN(7.4E0, 5.0E0)
High Generic	LN(2.5E2, 1.2E1)	LN(7.1E2, 7.9E0)	LN(9.1E3, 2.2E1)	LN(5.4E3, 2.2E1)

[a] N = Truncated Normal(Mean, Standard Deviation, [Min, Max]); LN = Lognormal(Geometric Mean, Geometric Standard Deviation); "Large" = 1e30 mL/g

K_H Distributions

The Henry's coefficient distributions were developed as discussed in the Methods section. Carbon is currently only considered as CO_2 in vapor at the site, but in places where site-specific knowledge suggests methane and other volatile organic compounds are also important, these species may be included in the K_H database and weighted according to their expected percentages. H_2 is the primary species in which tritium is expected to be released from the waste, but it partitions into water. For the site model, tritium is simulated as tritiated H_2O vapor. Limited knowledge of I suggests developing a wide distribution that captures the literature values for I as both I_2 and CH_3I.

The data and fitted distributions for $K_{H,0}$ for all elements except tritium are shown in Fig. 4(a–e). Because the values for I are right-skewed, a gamma distribution is used. For the remaining elements, normal distributions are used. Means and standard deviations are calculated from the literature data. The calculated standard deviations are doubled to produce sufficiently wide distributions to capture the more

13

extreme values. All of the normal distributions are truncated because $K_{H,0}$ cannot be negative. To avoid divide by zero errors, the minimum is set to a very small value instead of zero. The maximum is set to a very large value, as in Table II above.

The procedure for determining the tritium K_H distribution is different than the other radioelements. Data collected for K_H versus temperature are shown in Fig. 4(f). The fitted regression line is given by

$$\hat{y} = -2.04 * 10^{-4} + 7.53 * 10^{-7}[1/K] * T_p \quad \text{(Eq. 3)},$$

which is used to calculate the mean of the distribution of K_H at sampled temperature T_p. Using the development described in the Methods section (Eq. 2), and simplifying to avoid an additional model parameter, the distribution for $K_H(T_p)$ is approximated by a normal distribution with mean $\hat{y}(T_p)$ and standard deviation 2.6×10^{-6}. These are shown in Figure 4(f) with two examples of the computed normal distribution centered on the regression line. As before, distributions are truncated at 0 to avoid negatives, though such draws are very rare.
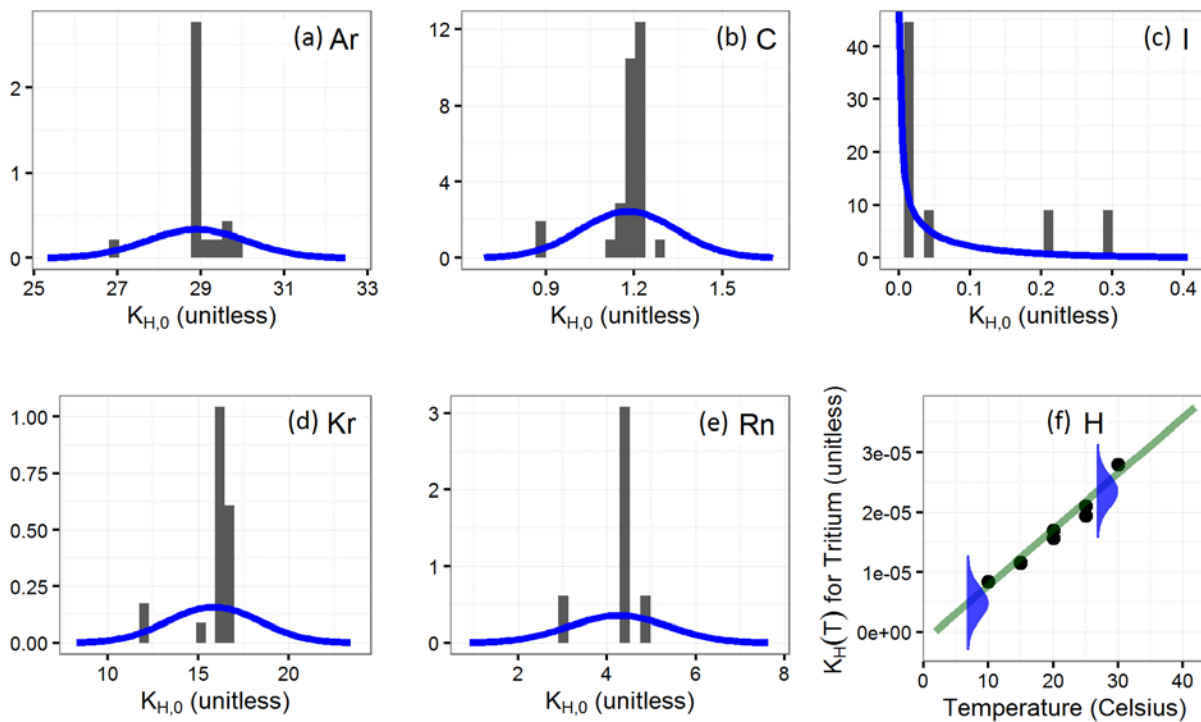

Fig. 4. (a–e) Fitted distributions for Henry's Law coefficients at reference temperature, $K_{H,0}$. (f) Tritium data as a function of temperature, with distributions shown for $T_p = 7°C$ and $27°C$.

Discussion

The distributions developed above were implemented in a GoldSim site model, and an SA was performed to demonstrate which parameters were responsible for most of the uncertainty for dose endpoints [22]. Frequently-observed sensitive parameters included inventory amounts, biotic model parameters (e.g., root depth), site geometry uncertainties, receptor behaviors and layout of their facilities, and geochemical parameters, including K_H, K_d, and solubility for particular element/media combinations. Some receptor doses were sensitive to K_ds for I and Th in Cement, Tc in Clay, and K_H for I and Rn.

These results will inform the next iteration of geochemical parameter development for the site, including

additional work on K_d and K_H for sensitive elements. The results also provide an indication of potential areas of focus for parameter distributions for PAs at other sites, i.e., the relative importance of uncertainty in geochemical parameters compared to other uncertain receptor, site, and inventory parameters.

CONCLUSIONS

The geochemical parameter development approach presented here is a rigorous and technically defensible framework that may be followed for many modeled values, including partitioning coefficients, other transport variables, material properties, and biotic parameters. The steps of the method are screening, data collection informed by site assessment, filtering of data (also informed by the site assessment, as well as EDA to evaluate data quantity and quality), statistical analysis, and model implementation/SA. Sensitivity results are used to refine the model as needed by returning to the screening or filtering steps. This iterative approach is taken so efficient use of time and budget is focused on parameters likely to have the greatest impact on risk calculations and, thereby, on decisions regarding waste disposal.

The K_d database currently includes over 2,400 entries from 69 references and is continually growing as it is used for PA projects in disparate rock types. The K_H database covers several volatile radioelements expected at disposal facilities, and can be expanded to include other gaseous species. Distributions developed from these databases were implemented in a site PA model, and certain elements were flagged as sensitive during SA, indicating new focus elements for the next round of distribution development.

The geochemistry approach presented here, once established for a site, may be used with minor modifications elsewhere (new rock type groupings, additional data collection for local or site-specific values, aqueous condition filtering, and review of final distributions). Each new site leverages data collection and statistical analysis performed for all of the prior sites. In theory, without considering aqueous conditions or speciation, $K_{H,0}$ distributions are not site-specific, and only a new temperature distribution is required for the next site (usually developed during the site assessment phase).

While the development of geochemical parameter distributions by collecting data and producing appropriate statistical representations is not new, the benefit to this updated approach is that it considers all available knowledge (site-specific and generic), statistical best practices, and has great flexibility for the user in culling data based on the amount available and quality of the literature source.

REFERENCES

1. Sander, R., *Compilation of Henry's Law Constants (version 4.0) for Water as Solvent.* Atmospheric Chemistry and Physics, 2015. 15(2015): p. 4399–4981.
2. Sheppard, M.I. and D.H. Thibault, *Default Soil Solid/Liquid Partition Coefficients, K_dS, for Four Major Soil Types: A Compendium.* Health Physics,, 1990. 59(4): p. 471–482.
3. Brady, P.V. and C.M. Bethke, *Beyond the K_d Approach.* Ground Water, 2000. 38(3): p. 321–322.
4. Tauxe, J., et al., *Improving Site-Specific Radiological Performance Assessments—13431*, in *WM2013 Conference, February 24–28*. 2013: Phoenix AZ.
5. Higgs, M., et al., *Statistical Methods for Effective Spatial and Temporal Scaling in Support of Probabilistic Performance Assessments—17340*, in *WM2017 Conference, March 5–9*. 2017: Phoenix AZ.
6. Baes, C.F., III, et al., *A Review and Analysis of Parameters for Assessing Transport of Environmentally Released Radionuclides through Agriculture*. 1984, Oak Ridge National Laboratory: Oak Ridge TN.
7. Kennedy, W.E. and D.L. Strenge, *Residual Radioactive Contamination from Decommissioning— Technical Basis for Translating Contamination Levels to Annual Total Effective Dose Equivalent*.

1992, United States Nuclear Regulatory Commission: Washington DC.

8. Yu, C., et al., *User's Manual for RESRAD Version 6*. 2001, Argonne National Laboratory: Argonne IL.

9. EPA, *Understanding Variation in Partition Coefficient, K_d, Values, Volume I: The K_d Model, Methods of Measurement, and Application of Chemical Reaction Codes*. 1999, United States Environmental Protection Agency: Washington DC.

10. EPA, *Understanding Variation in Partition Coefficient, K_d, Values, Volume II: Review of Geochemistry and Available K_d Values for Cadmium, Cesium, Chromium, Lead, Plutonium, Radon, Strontium, Thorium, Tritium (^3H), and Uranium*. 1999, United States Environmental Protection Agency: Washington DC.

11. EPA, *Understanding Variation in Partition Coefficient, K_d, Values, Volume III: Review of Geochemistry and Available K_d Values for Americium, Arsenic, Curium, Iodine, Neptunium, Radium, and Technetium*. 2004, United States Environmental Protection Agency: Washington DC.

12. IAEA, *Handbook of Parameter Values for the Prediction of Radionuclide Transfer in Terrestrial and Freshwater Environments, Technical Reports Series No. 472*. 2010, Vienna Austria: International Atomic Energy Agency.

13. Krupka, K.M., R.J. Serne, and D.I. Kaplan, *Geochemical Data Package for the 2005 Hanford Integrated Disposal Facility Performance Assessment*. 2004, Pacific Northwest National Laboratory: Richland WA.

14. Wang, L., et al., *A New Radionuclide Sorption Database for Benchmark Cement Accounting for Geochemical Evolution of Cement*, in *Cement-Based Materials for Nuclear Waste Storage*, F. Bart, Editor. 2013, Springer Science+Business Media: New York NY. p. 103–112.

15. EarthRef. *GERM Partition Coefficient (Kd) Database*. 2016; Available from: https://earthref.org/KDD/.

16. Bechtel SAIC, *Radionuclide Transport Models Under Ambient Conditions*. 2007, Bechtel SAIC Company LLC: Las Vegas NV.

17. Stoller-Navarro, *Phase II Contaminant Transport Parameters for the Groundwater Flow and Contaminant Transport Model of Corrective Action Unit 98: Frenchman Flat, Nye County, Nevada*. 2005, Stoller-Navarro Joint Venture: Las Vegas NV.

18. Sohlenius, G., et al., *Inferences About Radionuclide Mobility in Soils Based on the Solid/Liquid Partition Coefficients and Soil Properties.* Ambio, 2013. 2013(42): p. 414–424.

19. Cantrell, K.J., R.J. Serne, and G.V. Last, *Hanford Contaminant Distribution Coefficient Database and Users Guide*. 2003, Pacific Northwest National Laboratory: Richland WA.

20. Altmann, S., J. Bruno, and C. Tweed, *Using Thermodynamic Sorption Models for Guiding Radioelement Distribution Coefficient (Kd) Investigations: A Status Report*. 2001, Nuclear Energy Agency: Paris France.

21. Sander, R., *Compilation of Henry's Law Constants for Inorganic and Organic Species of Potential Importance in Environmental Chemistry*. 1999, Max-Planck Institute of Chemistry: Mainz Germany.

22. Stockton, T., P. Black, and P. Duffy, *Modern Sensitivity Analysis—15651*, in *WM2015 Conference, March 15–19*. 2015: Phoenix AZ.

23. Ogram, G.L., *Precipitation Scavenging of Tritiated Water Vapour (HTO)*. 1985, Ontario Hydro Research Division.

24. Smiles, D.E., W.R. Gardner, and R.K. Schulz, *Diffusion of Tritium in Arid Disposal Sites.* Water Resources Research, 1995. 31(6): p. 1483–1488.

25. Denham, M., *Vapor—Aqueous Solution Partition Coefficients for Radionuclides Pertinent to High Level Waste Tank Closure*. 2010, Savannah River National Laboratory: Aiken SC.

26. Kaplan, D.I., et al., *Range and Distribution of Technetium K_d Values in the SRS Subsurface*

Environment. 2008, Savannah River National Laboratory: Aiken SC.

27.   Kaplan, D.I., K.E. Parker, and R.D. Orr, *Effects of High-pH and High-Ionic-Strength Groundwater on Iodide, Pertechnetate, and Selenate Sorption to Hanford Sediments: Final Report for Subtask 3a*. 1998, Pacific Northwest National Laboratory: Richland WA.

28.   Serne, R.J., et al., *Site-Specific K_d Values for the Waste Control Specialists Disposal Facilities in Andrews County, Texas*. 2009, Battelle, Pacific Northwest Division: Richland WA.

29.   Wang, L., et al., *Review of Sorption Values for the Cementitious Near Field of a Near Surface Radioactive Waste Disposal Facility, Project Near Surface Disposal of Category A Waste at Dessel*. 2009, ONDRAF/NIRAS: Brussels Belgium.

30.   Krupka, K.M. and R.J. Serne, *Effects on Radionuclide Concentrations by Cement/Ground-Water Interactions in Support of Performance Assessment of Low-Level Radioactive Waste Disposal Facilities*. 1998, Pacific Northwest National Laboratory: Richland WA.

31.   R Core Team. *R: A Language and Environment for Statistical Computing*. 2015; Available from: https://www.r-project.org/.

32.   GTG, *GoldSim Contaminant Transport Module User's Guide*. 2014, GoldSim Technology Group: Issaquah WA.

33.   WCS, *Radioactive Material License No. R04100 CN600616890/RN101702439, Updated Performance Assessment for the Low-Level Waste Facility*. 2011, Waste Control Specialists LLC: Andrews TX.