Developing Hydraulic Conductivity Distributions for Use in Hydrologic Modeling

Amy Jordan¹, Doug S. Anderson², Leslie Gains-Germain³, Dylan B Boyle³, Lauren M Foster³, and Paul Black²

¹Carbon Solutions LLC
²Neptune and Company, Inc.
³Neptune and Co

April 12, 2024

Abstract

We present a methodology that uses pilot and anchor points with probability distributions for saturated hydraulic conductivity in a groundwater contaminant transport model. This approach directly links locations with calibration target data (e.g., water levels and drawdown at monitoring wells) to the most relevant physical parameter(s) that drive behavior, in a way that promotes model parsimony. Distributions for hydraulic conductivity are developed for monitoring well locations with pumping tests in order to reflect the state of uncertainty in the local estimates; these locations are called anchor points. Pilot points are placed between monitoring wells, and because they have more uncertainty these are generally assigned wider distributions that reflect plausible hydraulic conductivity values for the geologic material in which they are located. Scaling issues are considered in the development of these distributions. Pilot points are not randomly or uniformly distributed in the domain; rather they are considered connectors between locations with data (anchor points) and placed strategically between them. For a given model realization, hydraulic conductivity values at both pilot and anchor points are sampled from their respective distributions and all remaining locations are derived using an interpolation scheme (e.g., kriging). This approach to hydraulic conductivity assignment honors location-specific data, geologic heterogeneity, and spatial patterns. Given that inverse analysis of high-dimensional models tends to be ill-posed and thus sensitive to initialization of parameters, the distribution development process plays a critical role in driving the outcome of model calibration.

Hosted file

Developing Hydraulic Conductivity Distributions for Use in Hydrologic Modeling

A.B. Jordan¹, D.S. Anderson², L. Gains-Germain², D.B. Boyle², L.M. Foster², P. Black²

¹Carbon Solutions LLC.
²Neptune and Company, Inc.

Corresponding author: Amy Jordan (amy.jordan@carbonsolutionsllc.com)

Key Points:

- Hydraulic conductivity distributions are needed to represent parameter uncertainty and heterogeneity in stochastic groundwater models
- A method using pilot and anchor points is demonstrated for a groundwater model at a site contaminated with high explosives
- Parameter distribution development includes setting a distributional goal, data collection, weighting, and statistical analysis
Abstract

We present a methodology that uses pilot and anchor points with probability distributions for saturated hydraulic conductivity in a groundwater contaminant transport model. This approach directly links locations with calibration target data (e.g., water levels and drawdown at monitoring wells) to the most relevant physical parameter(s) that drive behavior, in a way that promotes model parsimony. Distributions for hydraulic conductivity are developed for monitoring well locations with pumping tests in order to reflect the state of uncertainty in the local estimates; these locations are called anchor points. Pilot points are placed between monitoring wells, and because they have more uncertainty these are generally assigned wider distributions that reflect plausible hydraulic conductivity values for the geologic material in which they are located. Scaling issues are considered in the development of these distributions. Pilot points are not randomly or uniformly distributed in the domain; rather they are considered connectors between locations with data (anchor points) and placed strategically between them. For a given model realization, hydraulic conductivity values at both pilot and anchor points are sampled from their respective distributions and all remaining locations are derived using an interpolation scheme (e.g., kriging). This approach to hydraulic conductivity assignment honors location-specific data, geologic heterogeneity, and spatial patterns. Given that inverse analysis of high-dimensional models tends to be ill-posed and thus sensitive to initialization of parameters, the distribution development process plays a critical role in driving the outcome of model calibration.

Plain Language Summary

Hydraulic conductivity of materials is a key input to most groundwater models. A method is presented that generates data-based probability distributions for hydraulic conductivity in a groundwater model. Heterogeneity and uncertainty are represented using a rigorous and defensible distribution development process.

1 Introduction

Saturated hydraulic conductivity (K) is a measure of a porous material’s ability to conduct water under fully saturated conditions. It is specific to each combination of material and fluid properties (e.g., density, viscosity of water). For example, values of hydraulic conductivity
for clays and well-sorted gravel typically range from $10^{-6}$ to $10^{-3}$ ft/d and 10 to $10^3$ ft/d, respectively (Fetter, 1994). Hydraulic conductivity of materials is a key input to most flow and transport models, and often one of the most sensitive parameters.

In homogenous, isotropic materials, $K$ is the same in all directions. In contrast, $K$ within geologic media is often directionally dependent due to the depositional history of the site. In this paper, the directional dependence, or anisotropy, of $K$ is not discussed. However, the process for representing heterogeneity here is easily extendable to determining unique distributions for $K_h$ and $K_v$ (K in the horizontal and vertical directions, respectively); $K$ with axes that relate to the direction of flow (e.g., longitudinal or transverse); or $K_x$, $K_y$, and $K_z$ in a Cartesian coordinate system.

Values of $K$ in geologic media may show considerable spatial variability. Temporal variability in $K$ is usually not an issue for the timescales used in most groundwater models, unless biotic or chemical reactions are present that contribute to significant modifications to the solid matrix. In addition to spatial variability, estimates of hydraulic conductivity may be subject to considerable uncertainty in estimation, depending on the methods used to determine them. Finally, hydraulic conductivity is scale-dependent (Clauser, 1992; Neuman, 1990; Schulze-Makuch et al., 1999). The scale at which heterogeneity is represented in a model can have a significant impact on predicted contaminant transport (Carrera, 1993). These aspects of hydraulic conductivity make it difficult to determine appropriate values to use in modeling (Foster & Maxwell, 2019). Depending on the decision context for which the model will be used, the model’s structure must be designed to simulate appropriate hydraulic conductivities, and a defensible statistical approach is required to determine realistic draws for the associated input parameter(s) from a probabilistic distribution.

There are three typical approaches used to assign hydraulic conductivity in groundwater models, with methods that may overlap between the three. All methods may be used in a deterministic or stochastic modeling framework, and each have strengths and weaknesses (de Marsily et al., 2005). An example of heterogeneous geologic materials is shown in Figure 1a, followed by depictions of three common approaches to hydraulic conductivity representation applied to this example in Figure 1b to 1d.
In the first approach, distributions of K values (isotropic or anisotropic) may be assigned to geologic zones represented in the model (Figure 1b, a “layered cake” conceptual model), where single draws for K are applied within each zone. At its extreme, a high-resolution, detailed facies model may be developed under this paradigm (de Marsily et al., 2005). This method may also be expanded to multiple interacting continua (Berkowitz, 2002) within a zone in order to capture smaller-scale transport processes, if necessary for modeling the geologic media.

In the second approach (Figure 1c), geostatistical methods may be used to randomly generate heterogeneous K fields, with or without spatial covariance, where parameters may be used to differentiate between geologic media in a zoned model. The end result is a model domain with K heterogeneity at scales smaller than the strata represented in the model, e.g., (Rubin et al., 2010; Tompson & Gelhar, 1990).

A third approach, and the one that is described in this paper, uses values of K drawn from distributions for pilot and anchor points to generate a K field, e.g., (Certes & de Marsily, 1991; Doherty, 2003; LaVenue et al., 1995; LaVenue & Pickens, 1992; deMarsily, 1978; RamaRao et al., 1995). Distributions may be specific to the geologic materials depicted in the model, allowing for the representation of expected differences in properties between materials. Interpolation is used to fill in the model nodes between pilot and anchor points (Figure 1d). Optionally, boundaries between strata may also be enforced by regularization approaches that link pilot and anchor points within a material but not across materials (Doherty, 2003).
Figure 1. Typical approaches to representing hydraulic conductivity in hydrologic modeling. (a) An example geologic cross-section. (b) Homogeneous $K$ within layers or zones that define geologic strata. (c) Randomly generated fields with heterogeneity at scales smaller than the geologic strata. (d) Heterogeneous $K$ field determined by interpolation between pilot and/or anchor points, drawn from distributions that are appropriate for the geologic materials.

There is some overlap in nomenclature between approaches; note that the term “anchor point” is used in the random spatial fields method described in Rubin et al. (2010). In that case, the scale of the data (local or nonlocal) determines how the anchor points are treated. The pilot point method as outlined by Doherty (2003) does not use the term “anchor point,” but its use is promoted here as a way to differentiate between locations with hydraulic conductivity estimates...
(where points are “anchored” to the data) and locations without them (pilot points). Pilot points are typically assigned wider distributions based broadly on their geologic unit or other spatial information. A discussion of the differences between the pilot point method described here and other methods using similar nomenclature is found in Rubin et al. (2010). Overlap between methods, such as using both pilot/anchor points as well as spatial random fields, is also done, e.g., (Lavenue & de Marsily, 2001; Murakami et al., 2010).

The benefit of the pilot point/anchor point approach (Figure 1d) is that it provides a particular flexibility for model calibration that may be absent from the other two common approaches. If calibration target data include observations such as water levels, drawdown response to pumping or injection (as in a pump-and-treat system), or contaminant concentrations, including pilot points between anchor points allows the calibration to link observations by varying spatially-explicit properties that drive hydraulic response, groundwater flow, and contaminant migration. This allows the model to approach a desired level of parsimony which is neither overly simplistic nor, at the other extreme, over-fitted based on the sparse field data available (Hill, 2006). Although values estimated in a calibration using pilot points are unlikely to be extremely accurate to “true” property values at those locations (Moore & Doherty, 2006), due to upscaling, homogenization, etc., another strength of the pilot point approach is that it provides the modeler with a broad picture of the hydraulic conductivity field at the site, which may help inform the conceptual site model.

Hydraulic conductivity “data” all represent estimates, not direct measurements, of the parameter. Estimates of hydraulic conductivity can be obtained in many ways, including: grain-size analysis of sampled materials using methods such as Kozeny-Carman (Bear, 1972); laboratory-scale fluid flow experiments on aquifer materials (Klute, 1965); in-situ pumping or slug tests at well locations (Schwartz & Zhang, 2002); borehole geophysical tools (Maliva et al., 2009); or estimates based on large-scale geophysical methods (Singha et al., 2007). Each type of estimate has an associated spatial scale and uncertainty that affects its applicability at the model scale.

The pilot point/anchor point method is presented here with an example from modeling a plume of 1,3,5-trinitro-1,3,5-triazinane (known as Royal Demolition Explosive or RDX) groundwater contamination at a Los Alamos National Laboratory (LANL) site. Distributions are
developed for the two primary geologic materials that affect the plume in the aquifer, the Puye (Tpf) and Tschicoma (Tvt) formations. The older Tschicoma formation is comprised of dacite lava flows from the eastern Jemez Mountains (N3B, 2019) (Figure 2). The Puye formation is composed of alluvial-fan deposits sourced from rhyolitic dome complexes of the Tschicoma formation. While there are other geologic materials in the model domain shown in Figure 2, the plume is expected to remain within these two materials (N3B, 2020). Figure 2 also shows the mesh used for the RDX groundwater model, with the plume area in a zoom inset.

After locating pilot and anchor points in the RDX groundwater model domain, the distribution development process follows a protocol designed to rigorously assess realistic parameter uncertainty (Brittingham et al., 2020; Gains-Germain et al., 2018; Higgs et al., 2017; Jordan et al., 2017). These steps are outlined in Section 2.0: setting a distributional goal, data collection and filtering, and statistical analysis. The results are summarized with discussion in Section 3.0.
Figure 2. RDX fate and transport model domain, mesh, geology, and approximate plume extent at the water table based on a modeled RDX 9.66 ppb contour from N3B (2020). Regional aquifer monitoring well locations are shown. The primary geologic units of interest for the RDX plume model are the Puye (Tpf) and Tschicoma (Tvt) formations.

2 Methods

2.1 Locating Pilot and Anchor Points

In the method presented here, anchor points are first determined by selecting monitoring well locations with reliable hydraulic conductivity estimates. Not all monitoring well locations are necessarily anchor points if hydraulic conductivity estimates are not available at these wells. Next, pilot points are located between anchor points. If it is later found that the model calibration struggles to find a good match to field data, additional pilot points may be added to increase heterogeneity, but in this approach it is desired to represent the minimum amount of complexity needed to achieve a successful calibration and make useful predictions given the decision context (Hill, 2006). This method lends itself to encouraging parsimony, especially when combined with regularization (Doherty, 2003), by setting up pilot points as direct linkages between locations with observational data (e.g., water levels and drawdown due to pumping) that are calibrated in the model. While in theory every node that is not an anchor point could be a pilot point (to increase heterogeneity to the extreme), such an approach would vastly increase the number of parameters required and could lead to spurious heterogeneity that is not supported by the available information, related to model goals, or contributing to decision endpoints. More pilot points should be placed in areas of (a) suspected heterogeneity and (b) where observational data points are denser (Doherty, 2003). Moore & Doherty (2006) suggest the density of pilot points should be commensurate with the density of observations. Additional strategic methods for placing pilot points have been developed to improve the value of each pilot point location added (LaVenue & Pickens, 1992; Moore & Doherty, 2006; Yang et al., 2012).

The pilot and anchor point locations for the RDX groundwater model are shown in Figure 3. In each iteration of the calibration, the likelihood of candidate K values for pilot and anchor points are calculated using their respective distributions and the objective function.
Figure 3. Pilot point (gray) and anchor point (black) locations in the RDX groundwater model. Pilot and anchor points may be placed at different depths at the same location, e.g., pp16 and pp17. Additionally, a pilot point location (e.g. pp12) and anchor (CdV-R-15-3) may be at the same location at different depths.

Candidate K values for pilot and anchor points are interpolated to assign K values to every remaining node in the domain. Kriging is an effective choice (de Marsily et al., 2005). The RDX model uses standard three-dimensional kriging with a spherical variogram (Cressie, 1988), as implemented in the MADS kriging package in Julia (https://github.com/madsjulia/Kriging.jl). The spherical variogram parameters, sigma and scale, are also allowed to vary in the RDX model, to allow calibration to help determine the best interpolation scheme within the framework of ordinary kriging. Outside the region with pilot and anchor points the K field becomes homogeneous and assigned the domain global mean value, which is a function of both the pilot point and anchor point values, and the kriging parameters. Three separate kriged fields across the model domain are ultimately developed for Kx, Ky, and Kz.
2.2 Distributional Goal

Setting a goal for the distributions is one of the most fundamental and often overlooked steps of this process, because the distribution and methods must fit the model’s specific needs and spatial/temporal scales in order to serve the decision(s) that will be made based on results (Gains-Germain et al., 2018; Higgs et al., 2017). In this step of the distribution development, the modeling team must determine: (1) how the parameter uncertainty will be characterized in the model; (2) the spatial and temporal scales over which each distribution is applied; and (3) the sources of physical variability and uncertainty in the parameter K as they relate to the scales of the data and the model application. This process leads to (4) an explicit statement of the goal of statistical analysis, i.e., the outcome of the distribution development effort.

In this example, for the LANL RDX groundwater model, the distributional goal is written as follows, given the steps above:

(1) A value will be drawn from $K$ (horizontal) and $K_z$ (vertical) distributions to represent an average value of the parameter over a spatial extent (volume). Although a single $K$ distribution is used for both $K_x$ and $K_y$, the model calibrates $x$ and $y$ directions independently at all locations so final values may differ. For model calibration, an initial, minimum, and maximum value are required. The draws within that range are determined by the optimization algorithm. For uncertainty analysis and predictive modeling, developed distributions are used as priors in a Bayesian modeling framework. Distributions are needed for both pilot points and anchor points that represent an appropriate state of uncertainty, at the appropriate scale.

(2) The spatial scale represented by a draw from the distribution is determined by pilot point density and kriging parameters, but is expected to be at a scale of similar order of magnitude to the spacing between monitoring well locations and pilot points (which itself may vary considerably based on the distances between monitoring wells). Temporally, draws from the distributions are used throughout the entire model run (tens to hundreds of years) under the assumption that material properties will stay near constant over time.

(3) Sources of spatial variability in $K$ are related to complex geologic depositional processes that generate preferential flow paths (high $K$ zones) and areas that resist flow (low $K$ zones) in subsurface sedimentary and volcanic materials at the site. It is expected that
heterogeneity in $K$ exists at scales much smaller than the values as represented in the model by draws from the distribution. The spatial scale used in the model generally corresponds to that represented by pilot point density in concert with kriging parameters (as discussed below). Sources of uncertainty in $K$ are due to the methods used to estimate their values, and the appropriateness of those estimates for the upscaling required for the model.

(4) Probability distributions will be developed for pilot points in the model domain based on the available data for the geologic materials they are located within. Distributions will be developed for anchor points based on the available data at those locations. The values must represent the plausible ranges for average values at the appropriate spatial scale given the state of uncertainty in $K$ values.

2.3 Data Collection

After the distributional goal is established, data gathering can begin. This may involve screening (the process of determining which data are relevant to the model, and prioritizing that information over all other possible sources of information). For example, if the model includes zones of gravel and sand, then data collection takes place for those two materials, and not for clays or other materials absent from the model. While this may seem obvious, sometimes the screening process warrants additional documentation to address why certain data have not been collected for completeness. For the LANL RDX groundwater model, screening meant seeking out data in the Puye and Tschicoma formations specifically. Later, additional filtering may be performed, as described in Section 2.4.

While peer-reviewed journal articles are the gold standard for references in the data collection process, in many cases site-specific experimental data are found in reports or other types of publications. “Values used” in previous model efforts may be included in the database as a lower-quality source of information, as it may already include expert opinion, bias, or other unknown modification based on the other model’s goals. Nonetheless, values used in similar models are occasionally helpful to include in the database for comparison.

The intention is to capture the most current state of knowledge about the parameter, with site- or material-specific information as much as possible, or using general information or
literature review papers for the material if specific data are unavailable. The data are collected into a database (e.g., a spreadsheet) that retains as much metadata as practical. The database identifies whether the sources represent primary experimental data, literature review, expert judgement, or values used in another model. The scale of measurement is included, where “small” represents core or laboratory scales, “intermediate” represents portions of a site but not the entire site, and “site” represents the entire site. For K estimates at LANL for the RDX site, most of the information falls into the first two scales (Table 1).

The two primary small-scale measurement techniques used at the site both occur during the drilling of boreholes. The first method is to retrieve core samples that have enough integrity to perform particle-size analysis, which are then used to estimate hydraulic conductivities using the Kozeny-Carman relationship (Bear, 1972).

This type of analysis was performed in the LANL Chromium project area, approximately 6 km downgradient of the RDX project area, with samples obtained from sonic coreholes CrCH-1, CrCH-2, CrCH-3, CrCH-4, and CrCH-5 (LANL, 2018). These data are included in the hydraulic database for completeness despite the small scale being potentially unrepresentative of the intermediate-scale hydraulic conductivities needed to match the scale of the model.

The other small-scale technique of determining hydraulic conductivity is using borehole geophysical methods. Results obtained from Combinable Magnetic Resonance (CMR) analysis for the Puye in well R-26 in the RDX area (Kleinfelder, 2005) is included in the database. Slug testing would also be considered small-scale (Gh de Marsily et al., 2005), but slug test estimates are not commonly available at LANL wells.

The “intermediate” scale data are obtained from standard aquifer testing practices at the LANL site. Pumping test analyses can be performed in a single-well format (the pumped well is also monitored for drawdown) or in a multi-well format (a nearby monitoring well is used for drawdown rather than, or in addition to, the pumped well itself). Both single- and multi-well tests are considered intermediate scale for the purposes of this analysis, although the volume of aquifer interrogated by any pumping test depends on the time frame and rate of pumping. Most of the LANL pumping test analyses are of the single-well format. In either case, drawdown versus time is plotted over the course of a pumping test and the recovery period is fit by one of many empirical or semi-analytical solutions for aquifer testing that exist in the literature. The
appropriateness of the selected method depends on assumptions in that method relative to the
true behavior of the aquifer (e.g., homogeneity, isotropy, confinement, etc.) and the well
characteristics (e.g., fully or partially penetrating). Not all well completion reports include
hydraulic conductivity estimates; some only present transmissivity, which is related to hydraulic
conductivity by $K = T/b$, where $T$ is transmissivity and $b$ is effective aquifer thickness (Schwartz
& Zhang, 2002). Transmissivity estimates were not included in the database directly, but
hydraulic conductivity estimated from transmissivities in the literature are included.

Methods of fitting the time-series drawdown data that appear commonly in LANL well
completion reports include Theis and the related Cooper-Jacob method for confined aquifers, the
Hantush equation for partially-penetrating wells, and the Neuman method for unconfined
aquifers (Schwartz & Zhang, 2002). The exact methods used vary between analysts and pumping
test configurations in the well completion report collection at LANL. These analyses may
introduce subjectivity if the fitting is performed by eye, or even when using semi-automated
methods to minimize residuals. The time selected for curve-fitting along the drawdown or
recovery process is also a factor in the appropriateness of the hydraulic conductivity estimate and
its scale; in the case of single well tests, early-time drawdown data interrogates a smaller portion
of the aquifer located closer to the well screen, and late-time data represents a larger volume of
the aquifer system as pumping stresses reach further away from the well screen.

Another intermediate-scale single well pumping test method is referred to as the specific
capacity method. The specific capacity method described in McLin (2005), which estimates
lower bounds for hydraulic conductivities, is typically used in this dataset.

A reason that well K estimates may also be uncertain for appropriate use in the model
that is unrelated to upscaling has to do with potential bias in location of the well screens. In some
cases where K values are estimated from pumping tests, monitoring or infrastructure (pump and
treat) wells may have had higher K strata targeted for their well screen. Therefore, the K
estimates from pumping tests could overestimate an appropriate average value for the model
volume over which they are applied.
Table 1. Summary Data for K Estimates in the Puye and Tschicoma Formations

<table>
<thead>
<tr>
<th>Formation</th>
<th>Type</th>
<th>N</th>
<th>Scale</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Puye</td>
<td>Corehole grain-size analysis</td>
<td>4</td>
<td>Sm.</td>
<td>From Puye formation downgradient of RDX site</td>
</tr>
<tr>
<td>CMR</td>
<td></td>
<td>1</td>
<td>Sm.</td>
<td>R-26 geophysical logging</td>
</tr>
<tr>
<td></td>
<td>Pumping test analysis</td>
<td>34</td>
<td>Int.</td>
<td>Curve-fitting using, e.g., Theis, Neuman, Hantush, etc.</td>
</tr>
<tr>
<td></td>
<td>Specific capacity estimates</td>
<td>8</td>
<td>Int.</td>
<td>Method of McLin (2005)</td>
</tr>
<tr>
<td>Tschicoma</td>
<td>Pumping test analysis</td>
<td>4</td>
<td>Int.</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Summary of K estimates for Puye and Tschicoma formation well screens in the database collected for this analysis across the LANL site, including the RDX area. The number of independent estimates (N) represents the number of unique monitoring well/screen locations for each type of analysis.

2.4 Filtering

The data collection should aim to be comprehensive, although it may stay within the bounds of the screening identified earlier; the filtering step is used after evaluating all of the collected data for relevance, scale, and quality. Filtering may be thought of as applying a weight of zero to certain data in the database, while fractional “value” weights (between 0 and 1) may be used to address various aspects of data quality and relevance (Edwards, 1977).

Exploratory data analysis (EDA) is a tool used to investigate any patterns present among the data. It is an essential process that helps guide the filtering and/or data weighting procedures to be performed. EDA can include making visualizations of the data for each material in the model, and generating summary tables of the data (counts, quantiles, measure of spread and central tendency, etc.). These plots and tables can help determine if and how selective to be about location (e.g., site-specific only versus general for the material), type of analysis (e.g., grain-size analysis versus pumping test estimates), quality, etc.
One of the challenges in developing hydraulic conductivity distributions for the RDX project area is that there are few wells and therefore few pumping test hydraulic conductivity estimates. A greater density of K estimates in the Puye formation is found downgradient at the LANL Chromium project area. The Puye formation is continuous and large in extent (Figure 2), and although the wells are far away they may still be representative of the overall formation properties. The first example of using EDA to make filtering decisions is the comparison between RDX area and “other” nearby area K estimates (primarily the Chromium project area, but also other parts of the LANL site) in the Puye formation, shown in Figure 4a.

Figure 4. (a) RDX-area monitoring well K estimates compared to all other locations in the same geologic unit (Puye formation) and scale (intermediate) at the LANL site. (b) Small-scale (corehole and geophysics) data compared to intermediate-scale K estimates (“int”) across the LANL site.

Among the intermediate-scale Puye formation data, the median K estimate for RDX area samples is 6.0 ft/d (n = 74), while the “Other” area category median K estimate is 14.3 ft/d (n = 100). This may be explained by the depositional environment causing spatial trending in the material properties (the RDX area is closer to the mountain block source of the alluvial fan deposits of the Puye formation). However, there are few K estimates in the RDX area compared to the site as a whole, so although the medians are different, based on the distributional goal of identifying the plausible range of K values for the model, all LANL-area estimates were included in the distribution at this time. This can be re-evaluated in future iterations of additional estimates for hydraulic conductivity of the Puye formation are made within the RDX project.
area. There are no estimates in the database for the Tschicoma formation outside the RDX project area, so a similar comparison is unnecessary.

Figure 4b shows a comparison between small- (“corehole”, from the sonic coreholes in the Chromium project area) and intermediate-scale K estimates across the entire LANL site. Inclusion of the small-scale data in the database, despite the mismatch between corehole data and the model usage scale of the parameter defined in the distributional goal, allows for an additional check on the K estimates from other methods. Among the LANL area Puye formation data, the corehole K estimates have a lower median than the intermediate-scale values from pumping tests and similar methods, which is consistent with the expected scale-dependence of hydraulic conductivity (Carrera, 1993; Clauser, 1992; Neuman, 1990; Schulze-Makuch et al., 1999).

Other types of EDA may include plotting by location to determine spatial trends, comparing geologic units or subunits, interrogating metadata for measurement method assumptions, investigating correlations with other parameter values (e.g., porosity), and so on.

The EDA on the LANL area K data led to the following filtering and weighting decisions:

- All sitewide data in the Puye formation are used to inform the K distribution in Puye, as opposed to only Puye data near the RDX plume.
- At the LANL site, there are identified subunits within the Puye formation (Broxton & Vaniman, 2005), but K estimates from all subunits were included in the Puye distribution.
- Comparisons between the distributions of intermediate- and small-scale K estimates suggest that there are statistically significant differences between the medians of these populations. Intermediate scale data are value-weighted twice as high as small-scale data because of the representativeness to the scale of how K is used in the model. The weighting difference of a factor of 2 is arbitrary as it is not known how much more relevant intermediate-scale data are to the model scale than small-scale data.
2.5 Statistical Analysis

Once filtering and/or weighting of the data are determined and applied from the previous step, the final set of retained, weighted data in the database is processed to generate a distribution that is appropriate for use in the model (as defined by the distributitional goal, Section 2.2).

At anchor point locations, unique distributions are developed using only the data provided for each individual well. For pilot points, where the distribution depends on the geologic group (Puye or Tschicoma), records are first averaged within a well location before fitting the distribution. This is because multiple K estimates from the same well screen are not considered independent: in the case of pumping test analyses, for instance, multiple reported K estimates may come from several attempts to fit the same drawdown or recovery data using different approaches. At other well locations, only one “best estimate” might be reported, based on which of the methods is assumed to be most valid. Thus, averaging all data for a well screen is performed to reduce potential bias towards wells with greater numbers of reported estimates and ensure that independence assumptions are met.

Distributions for anchor points were fit using all available data for the well within the respective formation using the method of moments, with the exception of R-26, as discussed below (Figure 5a). The variance of the distribution estimated for an anchor point in this manner is representative of the variability in the applied measurement method(s) within an aquifer test analysis. The distributions are normal in log data space (log refers to log10 throughout this paper).

The log-transformed means of the distributions for the Puye and Tschicoma formations were estimated by calculating the mean among the well averages for all wells available, and the standard deviation of the distribution for each formation was estimated by calculating the standard deviation among the well averages and dividing by the square root of the number of wells. The variance in the distribution for each geologic formation represents the variability between wells. Figure 5b shows the resulting distributions for all pilot points in the model.

R-26 Screen 2 is assigned a very wide distribution because of considerable uncertainty reflected in the pumping test information at that location. While the pumping tests suggest an unusually low K estimate for that location (based on a recovery response analysis) compared to other Puye formation pumping tests, and it is also corroborated by the specific capacity lower-
bound K estimates, there is some uncertainty about whether this is reflective of the aquifer
system or if equipment malfunction is responsible for the low estimate. Additionally, estimates
were made at R-26 during drilling using borehole geophysical methods that suggested
considerably larger K estimates in some intervals in the borehole than the pumping test analyses
at screen 2 (Kleinfelder, 2005), although, if there was no equipment error, this could reflect true
vertical differences in hydraulic conductivity at this location. Therefore, the distribution for this
“anchor” point was unusually wide, even wider than the generic geology distribution for the
Puye formation used for pilot points, and is selected to be uniform rather than lognormal. While
this location has no detectable RDX contamination, because it is the furthest upgradient at the
RDX site, it plays a significant role in the model calibration by setting the hydraulic gradient
across the plume area. All of these factors contribute to its wide allowable range in a uniform
distribution. The upper screen at R-26 is not used as an anchor point in the model because it is
unknown if screen 1 is representative of the regional aquifer or was completed in perched
conditions (Kleinfelder, 2005).

The 1st and 99th percentiles of the distributions shown in Figure 5 were selected as the
allowable constraints for the calibration, as discussed below. For Puye, the range is 2.1 to 12 ft/d.
For Tschicoma, the range is 0.88 to 104 ft/d.
Figure 5. Distributions for (a) anchor points at well/screen locations and (b) pilot points in the Puye and Tschicoma formations. The lower panel of (a) shows $K$ estimates from the database.

3 Results and Discussion

The $K$ distributions developed above were implemented for use in the RDX groundwater model calibration. The calibration and model methods are described in N3B (2020). Calibration targets include field data such as water levels and water level gradients (flow direction), RDX yearly average concentrations, and concentration trends over time. In the non-linear least-squares calibration performed using the Levenberg-Marquardt (LM) algorithm implemented in MADS ("MADS: Model Analysis & Decision Support," n.d.), the $K$ distributions provide the constraints (minimum and maximum) that the calibration is allowed to test for $K$ at each pilot or anchor point location.
The mean of the distribution is used to initialize the pilot and anchor points in the model for the calibration, but any value throughout the range may be used depending on the history matching qualities of the parameter set. Since the calibration is sensitive to initialization parameters, the initial values determined by the distribution development are very important.

When the calibration process is completed, a set of parameters (including $K$ values at all pilot and anchor points) is found that minimizes the objective function. The kriged $K_x$ field for the LM calibration result is shown in Figure 6. As described above, $K_x$ and $K_y$ are calibrated independently, although the same $K$ distributions are used for both.

Figure 6. Model layer 1 (top layer) $K_x$ for the calibrated RDX groundwater model, in log10([ft/d]).

An unusually low $K_x$ value at R-26 PZ-2 is found in the model calibration and is consistent with the site data, as discussed above. R-58, on the high end of the anchor points based on site data, remains elevated after model calibration. In general, other than a few anchor points which represent outliers in the data set, the model calibration has achieved a fairly regular $K$ field in the horizontal dimension.
A check was performed to see if the model calibration moved any $K$ values to within 10% of their constraints (bounding ranges). A calibrated value that is pinned to the edge of the allowable range suggests an area where additional focused attention may be necessary, because the LM optimization may prefer a value outside the bounds developed from the data. Considerable deviations would suggest that scaling issues, model structure, or other factors would require model values to be further from the data values in order to match the calibration target data. In the RDX groundwater model calibration, for $K_x$ or $K_y$, 4% of pilot or anchor point locations were within 10% of the lower range limit, and 3% of locations were similarly near the upper range limit. None of the points were within 1% of the upper or lower limits.

The next step in the model workflow after calibration is uncertainty analysis, which is used to expand the deterministic result from the calibration into a set of input parameters to run the model probabilistically (N3B, 2020). The uncertainty analysis is performed using the Markov Chain Monte Carlo (MCMC) sampler as implemented in MADS (“MADS: Model Analysis & Decision Support,” n.d.). This analysis produces 1018 model runs which are intended to capture the state of uncertainty in the parameters (including $K$) while still achieving acceptable history-matching to the field data.

Figure 7 shows a comparison of $K_x$ and $K_y$ posterior distributions (represented as histograms of MCMC-generated samples) for pilot points against the prior distributions (black line) for horizontal $K$ for the Puye and Tschicoma formations. For the Puye formation pilot points, the modeled values after uncertainty analysis follow the distribution developed from the site data. In essence, the maximum a’posteriori estimates from the MCMC calibration are consistent with maximum likelihood estimates from the LM calibration. For the Tschicoma formation, there are only two pilot points, where each $K_x$ and $K_y$ cluster in a particular location in parameter space after the uncertainty analysis. One pilot point (pp40 in Figure 3) moved to the edge of the distribution during the LM calibration (98 ft/d) and remained near there during the MCMC. Taken together, the results for pilot points in the Puye and Tschicoma did not drive any re-evaluation of the model structure, but the pilot points near the extremes of the distribution will be re-analyzed in future iterations.
4 Conclusions

The pilot point/anchor point method, coupled with statistical analysis for distribution development, is a rigorous approach to grounding a parsimonious numerical model in plausible site data for hydraulic conductivity while allowing the model calibration to explore the full range of uncertainty. The distribution development step is critical for the initialization of the parameter in the calibration, as well as defining plausible ranges that would trigger an investigation into why a model-calibrated value strays far from the site data. The approach is iterative, with each iteration based on the currently available state of knowledge about the parameter (K, in this case). Additional iterations may be performed if model calibration results or sensitivity analyses suggest that there would be additional benefit to further analysis of the underlying data set, or if new data become available.

The pilot and anchor point method to spatial distribution of K values provides considerable flexibility that may be lost when using rigidly defined stratigraphic zones, which may have considerable overlap in hydraulic properties once uncertainty is included. The connection this method provides between site-specific knowledge (at anchor points), as well as
flexibility for calibration at pilot points (within the realm of plausible values for the geologic formation), is both defensible and practical for numerical modeling purposes. It optimizes the representation of upscaled physical processes within a parsimonious model framework. The method is easily extendable to other heterogeneous parameters and modeling applications.

In the example presented here for a model of groundwater contamination with RDX, two geologic units are present in the plume area and their pilot and anchor point distributions reflect the K estimates collected in a database of LANL-area observations. The data collection spans a wider range of locations, geologic units, and scales than are likely to be representative, so a filtering/weighting step is performed. Based on the explicitly defined distributional goal, statistical analysis is performed to identify the distributions and $1^{\text{st}}/99^{\text{th}}$ percentiles are used as constraints in model calibration. The resulting calibration found excellent matches to field data while staying within the bounds provided by the statistical analysis (N3B, 2020). This suggests that the K estimates from the site provided a plausible and defensible starting location for the model calibration (relative to other parameter values) with some adjustment allowed during calibration to account for model grid resolution, scale, simplifications, assumptions, and uncertainty.

Future work on the use of the method for LANL-area groundwater models could include a second iteration on the RDX site data to evaluate which pilot and anchor points are particularly sensitive. Additional focused analysis can be performed on those literature data for hydraulic conductivity estimates to see if uncertainty may be narrowed.

Another useful improvement in the application of the method to the RDX groundwater model would be making the distributions multivariate in $K_x$, $K_y$, and $K_z$. At present, the distributions are drawn independently, although the $K_z$ distribution covers a significantly lower range of K values based on observed and estimated anisotropy at the site. Therefore, the model has a built-in mechanism for calibrating with $K_z < K_x$, $K_y$. However, the statistical analysis and model implementation could be updated by adjusting the distributional goal to include anisotropy correlation explicitly.

Acknowledgments
This work was funded by the Environmental Management Los Alamos Field office of the Department of Energy. The authors would like to thank colleagues who have provided critical reviews and discussions related to these materials, including D. Broxton and B. Robinson, as well as the original developers of the pilot point approach within the LANL Chromium Groundwater Model, V. Vesselinov and D. O’Malley.

Open Research


References


https://doi.org/10.1029/171GM16


https://doi.org/10.1029/WR026i010p02541


https://doi.org/10.1029/2012WR011864