1   **Developing Hydraulic Conductivity Distributions for Use in Hydrologic Modeling**

2   **A.B. Jordan[1], D.S. Anderson[2], L. Gains-Germain[2], D.B. Boyle[2], L.M. Foster[2], P. Black[2]**

3   [1]Carbon Solutions LLC.

4   [2]Neptune and Company, Inc.

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

6

7   **Key Points:**

8   - Hydraulic conductivity distributions are needed to represent parameter uncertainty and
9   heterogeneity in stochastic groundwater models

10   - A method using pilot and anchor points is demonstrated for a groundwater model at a site
11   contaminated with high explosives

12   - Parameter distribution development includes setting a distributional goal, data collection,
13   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

43   for clays and well-sorted gravel typically range from 10^{-6} to 10^{-3} ft/d and 10 to 10^3 ft/d,

44   respectively (Fetter, 1994). Hydraulic conductivity of materials is a key input to most flow and

45   transport models, and often one of the most sensitive parameters.

46   In homogenous, isotropic materials, K is the same in all directions. In contrast, K within

47   geologic media is often directionally dependent due to the depositional history of the site. In this

48   paper, the directional dependence, or anisotropy, of K is not discussed. However, the process for

49   representing heterogeneity here is easily extendable to determining unique distributions for K_h

50   and K_v (K in the horizontal and vertical directions, respectively); K with axes that relate to the

51   direction of flow (e.g., longitudinal or transverse); or K_x, K_y, and K_z in a Cartesian coordinate

52   system.

53   Values of K in geologic media may show considerable spatial variability. Temporal

54   variability in K is usually not an issue for the timescales used in most groundwater models,

55   unless biotic or chemical reactions are present that contribute to significant modifications to the

56   solid matrix. In addition to spatial variability, estimates of hydraulic conductivity may be subject

57   to considerable uncertainty in estimation, depending on the methods used to determine them.

58   Finally, hydraulic conductivity is scale-dependent (Clauser, 1992; Neuman, 1990; Schulze-

59   Makuch et al., 1999). The scale at which heterogeneity is represented in a model can have a

60   significant impact on predicted contaminant transport (Carrera, 1993). These aspects of hydraulic

61   conductivity make it difficult to determine appropriate values to use in modeling (Foster &

62   Maxwell, 2019). Depending on the decision context for which the model will be used, the

63   model's structure must be designed to simulate appropriate hydraulic conductivities, and a

64   defensible statistical approach is required to determine realistic draws for the associated input

65   parameter(s) from a probabilistic distribution.

66   There are three typical approaches used to assign hydraulic conductivity in groundwater

67   models, with methods that may overlap between the three. All methods may be used in a

68   deterministic or stochastic modeling framework, and each have strengths and weaknesses (de

69   Marsily et al., 2005). An example of heterogeneous geologic materials is shown in Figure 1a,

70   followed by depictions of three common approaches to hydraulic conductivity representation

71   applied to this example in Figure 1b to 1d.

72   In the first approach, distributions of K values (isotropic or anisotropic) may be assigned

73 to geologic zones represented in the model (Figure 1b, a "layered cake" conceptual model),

74 where single draws for K are applied within each zone. At its extreme, a high-resolution, detailed

75 facies model may be developed under this paradigm (de Marsily et al., 2005). This method may

76 also be expanded to multiple interacting continua (Berkowitz, 2002) within a zone in order to

77 capture smaller-scale transport processes, if necessary for modeling the geologic media.

78   In the second approach (Figure 1c), geostatistical methods may be used to randomly

79 generate heterogeneous K fields, with or without spatial covariance, where parameters may be

80 used to differentiate between geologic media in a zoned model. The end result is a model domain

81 with K heterogeneity at scales smaller than the strata represented in the model, e.g., (Rubin et al.,

82 2010; Tompson & Gelhar, 1990).

83   A third approach, and the one that is described in this paper, uses values of K drawn from

84 distributions for pilot and anchor points to generate a K field, e.g., (Certes & de Marsily, 1991;

85 Doherty, 2003; LaVenue et al., 1995; LaVenue & Pickens, 1992; deMarsily, 1978; RamaRao et

86 al., 1995). Distributions may be specific to the geologic materials depicted in the model,

87 allowing for the representation of expected differences in properties between materials.

88 Interpolation is used to fill in the model nodes between pilot and anchor points (Figure 1d).

89 Optionally, boundaries between strata may also be enforced by regularization approaches that

90 link pilot and anchor points within a material but not across materials (Doherty, 2003).
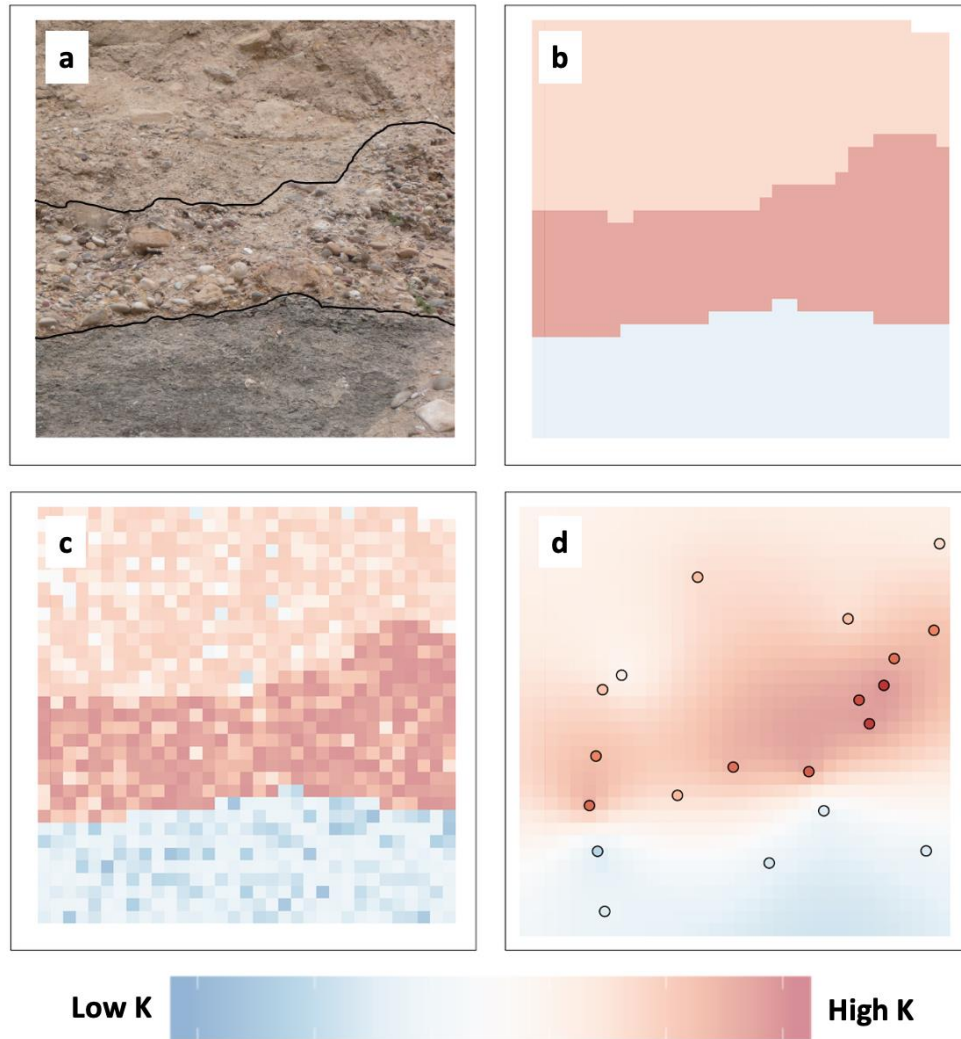
91

92

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.

99   There is some overlap in nomenclature between approaches; note that the term "anchor
100   point" is used in the random spatial fields method described in Rubin et al. (2010). In that case,
101   the scale of the data (local or nonlocal) determines how the anchor points are treated. The pilot
102   point method as outlined by Doherty (2003) does not use the term "anchor point," but its use is
103   promoted here as a way to differentiate between locations with hydraulic conductivity estimates

104 (where points are "anchored" to the data) and locations without them (pilot points). Pilot points

105 are typically assigned wider distributions based broadly on their geologic unit or other spatial

106 information. A discussion of the differences between the pilot point method described here and

107 other methods using similar nomenclature is found in Rubin et al. (2010). Overlap between

108 methods, such as using both pilot/anchor points as well as spatial random fields, is also done,

109 e.g., (Lavenue & de Marsily, 2001; Murakami et al., 2010).

110   The benefit of the pilot point/anchor point approach (Figure 1d) is that it provides a

111 particular flexibility for model calibration that may be absent from the other two common

112 approaches. If calibration target data include observations such as water levels, drawdown

113 response to pumping or injection (as in a pump-and-treat system), or contaminant concentrations,

114 including pilot points between anchor points allows the calibration to link observations by

115 varying spatially-explicit properties that drive hydraulic response, groundwater flow, and

116 contaminant migration. This allows the model to approach a desired level of parsimony which is

117 neither overly simplistic nor, at the other extreme, over-fitted based on the sparse field data

118 available (Hill, 2006). Although values estimated in a calibration using pilot points are unlikely

119 to be extremely accurate to "true" property values at those locations (Moore & Doherty, 2006),

120 due to upscaling, homogenization, etc., another strength of the pilot point approach is that it

121 provides the modeler with a broad picture of the hydraulic conductivity field at the site, which

122 may help inform the conceptual site model.

123   Hydraulic conductivity "data" all represent estimates, not direct measurements, of the

124 parameter. Estimates of hydraulic conductivity can be obtained in many ways, including: grain-

125 size analysis of sampled materials using methods such as Kozeny-Carman (Bear, 1972);

126 laboratory-scale fluid flow experiments on aquifer materials (Klute, 1965); in-situ pumping or

127 slug tests at well locations (Schwartz & Zhang, 2002); borehole geophysical tools (Maliva et al.,

128 2009); or estimates based on large-scale geophysical methods (Singha et al., 2007). Each type of

129 estimate has an associated spatial scale and uncertainty that affects its applicability at the model

130 scale.

131   The pilot point/anchor point method is presented here with an example from modeling a

132 plume of 1,3,5-trinitro-1,3,5-triazinane (known as Royal Demolition Explosive or RDX)

133 groundwater contamination at a Los Alamos National Laboratory (LANL) site. Distributions are

134 developed for the two primary geologic materials that affect the plume in the aquifer, the Puye

135 (Tpf) and Tschicoma (Tvt) formations. The older Tschicoma formation is comprised of dacite

136 lava flows from the eastern Jemez Mountains (N3B, 2019) (Figure 2). The Puye formation is

137 composed of alluvial-fan deposits sourced from rhyolitic dome complexes of the Tschicoma

138 formation. While there are other geologic materials in the model domain shown in Figure 2, the

139 plume is expected to remain within these two materials (N3B, 2020). Figure 2 also shows the

140 mesh used for the RDX groundwater model, with the plume area in a zoom inset.

141   After locating pilot and anchor points in the RDX groundwater model domain, the

142 distribution development process follows a protocol designed to rigorously assess realistic

143 parameter uncertainty (Brittingham et al., 2020; Gains-Germain et al., 2018; Higgs et al., 2017;

144 Jordan et al., 2017). These steps are outlined in Section 2.0: setting a distributional goal, data

145 collection and filtering, and statistical analysis. The results are summarized with discussion in
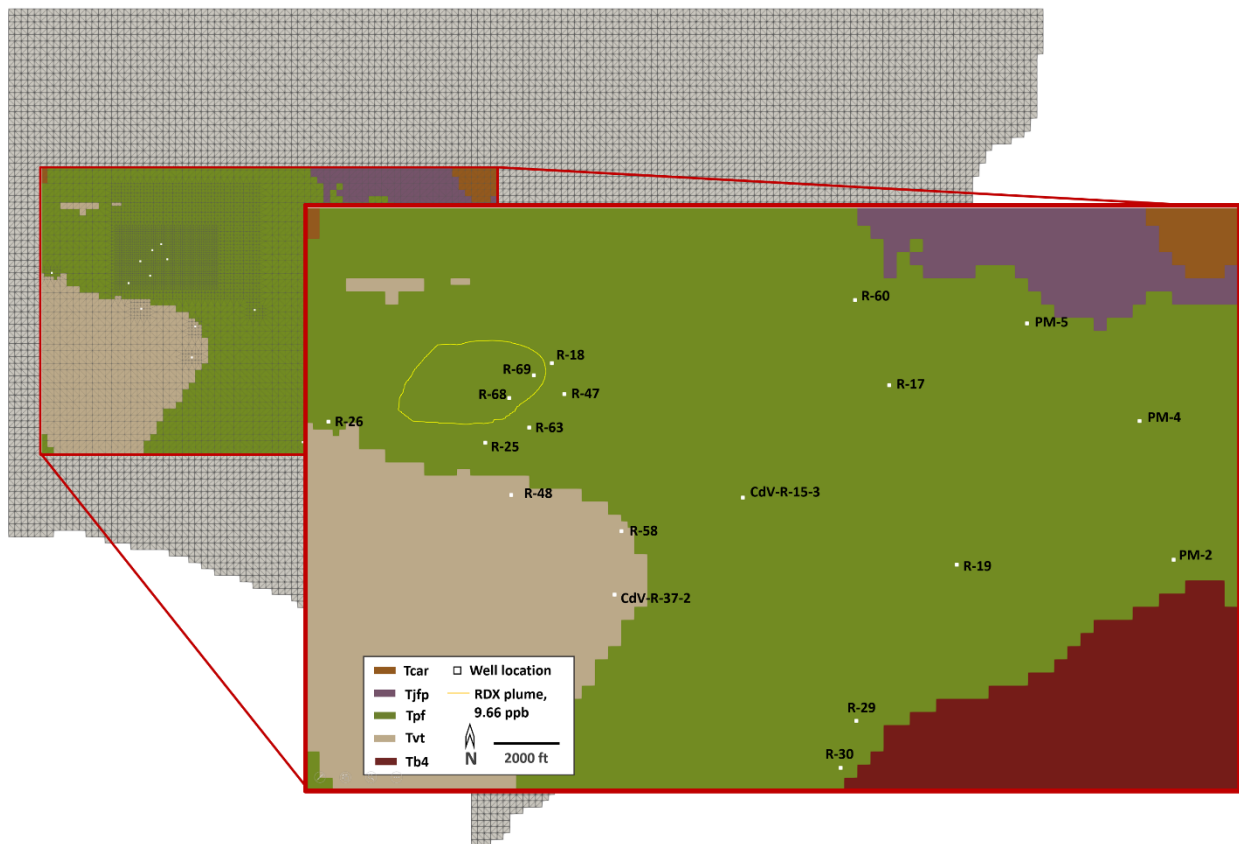
146 Section 3.0.

147

148 **Figure 2. RDX fate and transport model domain, mesh, geology, and approximate plume**

149 **extent at the water table based on a modeled RDX 9.66 ppb contour from N3B (2020).**

150 **Regional aquifer monitoring well locations are shown. The primary geologic units of**

151 **interest for the RDX plume model are the Puye (Tpf) and Tschicoma (Tvt) formations.**

152 **2 Methods**

153   2.1 Locating Pilot and Anchor Points

154   In the method presented here, anchor points are first determined by selecting monitoring

155 well locations with reliable hydraulic conductivity estimates. Not all monitoring well locations

156 are necessarily anchor points if hydraulic conductivity estimates are not available at these wells.

157 Next, pilot points are located between anchor points. If it is later found that the model calibration

158 struggles to find a good match to field data, additional pilot points may be added to increase

159 heterogeneity, but in this approach it is desired to represent the minimum amount of complexity

160 needed to achieve a successful calibration and make useful predictions given the decision context

161 (Hill, 2006). This method lends itself to encouraging parsimony, especially when combined with

162 regularization (Doherty, 2003), by setting up pilot points as direct linkages between locations

163 with observational data (e.g., water levels and drawdown due to pumping) that are calibrated in

164 the model. While in theory every node that is not an anchor point could be a pilot point (to

165 increase heterogeneity to the extreme), such an approach would vastly increase the number of

166 parameters required and could lead to spurious heterogeneity that is not supported by the

167 available information, related to model goals, or contributing to decision endpoints. More pilot

168 points should be placed in areas of (a) suspected heterogeneity and (b) where observational data

169 points are denser (Doherty, 2003). Moore & Doherty (2006) suggest the density of pilot points

170 should be commensurate with the density of observations. Additional strategic methods for

171 placing pilot points have been developed to improve the value of each pilot point location added

172 (LaVenue & Pickens, 1992; Moore & Doherty, 2006; Yang et al., 2012).

173   The pilot and anchor point locations for the RDX groundwater model are shown in Figure

174 3. In each iteration of the calibration, the likelihood of candidate K values for pilot and anchor

175 points are calculated using their respective distributions and the objective function.
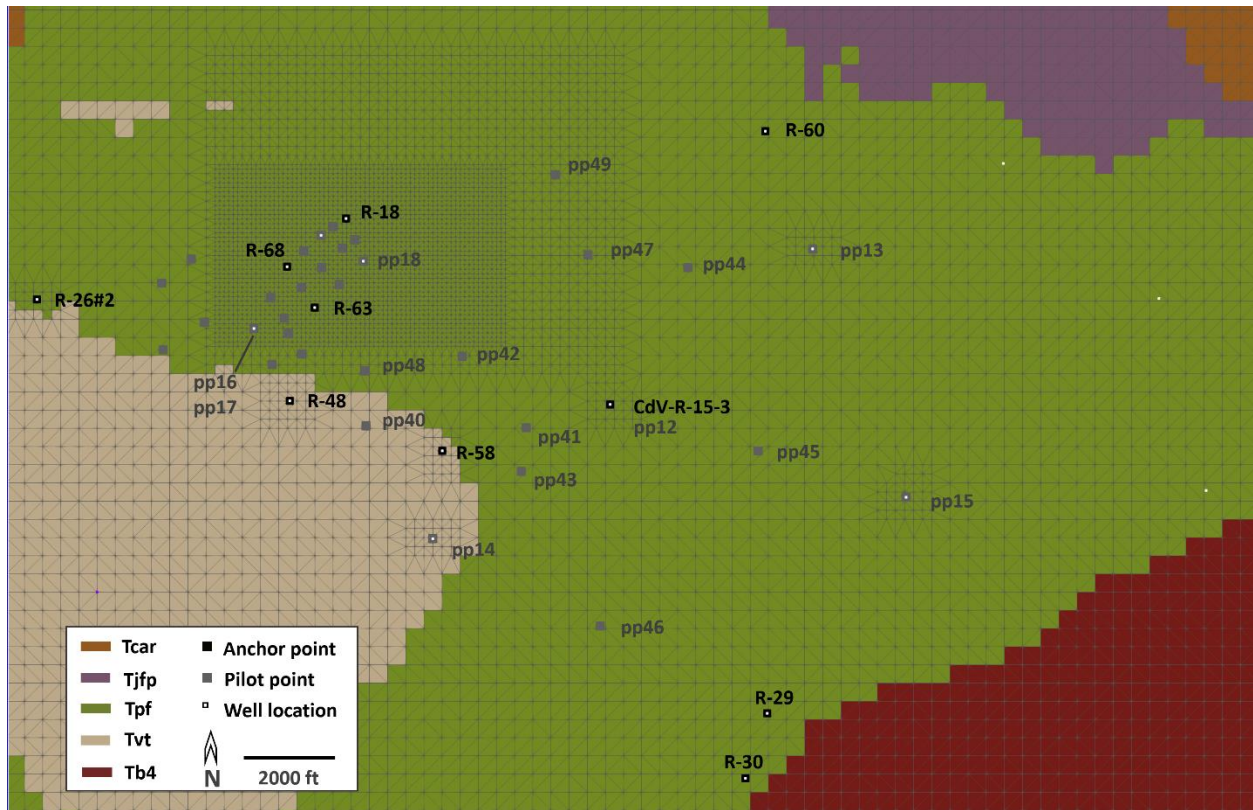
176

Figure 3. Pilot point (gray) and anchor point (black) locations in the 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 K_x, K_y, and K_z.

191   2.2 Distributional Goal

192   Setting a goal for the distributions is one of the most fundamental and often overlooked

193 steps of this process, because the distribution and methods must fit the model's specific needs

194 and spatial/temporal scales in order to serve the decision(s) that will be made based on results

195 (Gains-Germain et al., 2018; Higgs et al., 2017). In this step of the distribution development, the

196 modeling team must determine: (1) how the parameter uncertainty will be characterized in the

197 model; (2) the spatial and temporal scales over which each distribution is applied; and (3) the

198 sources of physical variability and uncertainty in the parameter K as they relate to the scales of

199 the data and the model application. This process leads to (4) an explicit statement of the goal of

200 statistical analysis, i.e., the outcome of the distribution development effort.

201   In this example, for the LANL RDX groundwater model, the distributional goal is written

202 as follows, given the steps above:

203   (1) A value will be drawn from K (horizontal) and K_z (vertical) distributions to represent an

204   average value of the parameter over a spatial extent (volume). Although a single K

205   distribution is used for both K_x and K_y, the model calibrates *x* and *y* directions

206   independently at all locations so final values may differ. For model calibration, an initial,

207   minimum, and maximum value are required. The draws within that range are determined

208   by the optimization algorithm. For uncertainty analysis and predictive modeling,

209   developed distributions are used as priors in a Bayesian modeling framework.

210   Distributions are needed for both pilot points and anchor points that represent an

211   appropriate state of uncertainty, at the appropriate scale.

212   (2) The spatial scale represented by a draw from the distribution is determined by pilot point

213   density and kriging parameters, but is expected to be at a scale of similar order of

214   magnitude to the spacing between monitoring well locations and pilot points (which itself

215   may vary considerably based on the distances between monitoring wells). Temporally,

216   draws from the distributions are used throughout the entire model run (tens to hundreds

217   of years) under the assumption that material properties will stay near constant over time.

218   (3) Sources of spatial variability in K are related to complex geologic depositional processes

219   that generate preferential flow paths (high K zones) and areas that resist flow (low K

220   zones) in subsurface sedimentary and volcanic materials at the site. It is expected that

221 heterogeneity in K exists at scales much smaller than the values as represented in the

222 model by draws from the distribution. The spatial scale used in the model generally

223 corresponds to that represented by pilot point density in concert with kriging parameters

224 (as discussed below). Sources of uncertainty in K are due to the methods used to estimate

225 their values, and the appropriateness of those estimates for the upscaling required for the

226 model.

227 (4) Probability distributions will be developed for pilot points in the model domain based on

228 the available data for the geologic materials they are located within. Distributions will be

229 developed for anchor points based on the available data at those locations. The values

230 must represent the plausible ranges for average values at the appropriate spatial scale

231 given the state of uncertainty in K values.

232 2.3 Data Collection

233 After the distributional goal is established, data gathering can begin. This may involve

234 screening (the process of determining which data are relevant to the model, and prioritizing that

235 information over all other possible sources of information). For example, if the model includes

236 zones of gravel and sand, then data collection takes place for those two materials, and not for

237 clays or other materials absent from the model. While this may seem obvious, sometimes the

238 screening process warrants additional documentation to address why certain data have not been

239 collected for completeness. For the LANL RDX groundwater model, screening meant seeking

240 out data in the Puye and Tschicoma formations specifically. Later, additional filtering may be

241 performed, as described in Section 2.4.

242 While peer-reviewed journal articles are the gold standard for references in the data

243 collection process, in many cases site-specific experimental data are found in reports or other

244 types of publications. "Values used" in previous model efforts may be included in the database

245 as a lower-quality source of information, as it may already include expert opinion, bias, or other

246 unknown modification based on the other model's goals. Nonetheless, values used in similar

247 models are occasionally helpful to include in the database for comparison.

248 The intention is to capture the most current state of knowledge about the parameter, with

249 site- or material-specific information as much as possible, or using general information or

250   literature review papers for the material if specific data are unavailable. The data are collected

251   into a database (e.g., a spreadsheet) that retains as much metadata as practical . The database

252   identifies whether the sources represent primary experimental data, literature review, expert

253   judgement, or values used in another model. The scale of measurement is included, where

254   "small" represents core or laboratory scales, "intermediate" represents portions of a site but not

255   the entire site, and "site" represents the entire site. For K estimates at LANL for the RDX site,

256   most of the information falls into the first two scales (Table 1).

257   The two primary small-scale measurement techniques used at the site both occur during

258   the drilling of boreholes. The first method is to retrieve core samples that have enough integrity

259   to perform particle-size analysis, which are then used to estimate hydraulic conductivities using

260   the Kozeny-Carman relationship (Bear, 1972).

261   This type of analysis was performed in the LANL Chromium project area, approximately

262   6 km downgradient of the RDX project area, with samples obtained from sonic coreholes CrCH-

263   1, CrCH-2, CrCH-3, CrCH-4, and CrCH-5 (LANL, 2018). These data are included in the

264   hydraulic database for completeness despite the small scale being potentially unrepresentative of

265   the intermediate-scale hydraulic conductivities needed to match the scale of the model.

266   The other small-scale technique of determining hydraulic conductivity is using borehole

267   geophysical methods. Results obtained from Combinable Magnetic Resonance (CMR) analysis

268   for the Puye in well R-26 in the RDX area (Kleinfelder, 2005) is included in the database. Slug

269   testing would also be considered small-scale (Gh de Marsily et al., 2005), but slug test estimates

270   are not commonly available at LANL wells.

271   The "intermediate" scale data are obtained from standard aquifer testing practices at the

272   LANL site. Pumping test analyses can be performed in a single-well format (the pumped well is

273   also monitored for drawdown) or in a multi-well format (a nearby monitoring well is used for

274   drawdown rather than, or in addition to, the pumped well itself). Both single- and multi-well tests

275   are considered intermediate scale for the purposes of this analysis, although the volume of

276   aquifer interrogated by any pumping test depends on the time frame and rate of pumping. Most

277   of the LANL pumping test analyses are of the single-well format. In either case, drawdown

278   versus time is plotted over the course of a pumping test and the recovery period is fit by one of

279   many empirical or semi-analytical solutions for aquifer testing that exist in the literature. The

280 appropriateness of the selected method depends on assumptions in that method relative to the

281 true behavior of the aquifer (e.g., homogeneity, isotropy, confinement, etc.) and the well

282 characteristics (e.g., fully or partially penetrating). Not all well completion reports include

283 hydraulic conductivity estimates; some only present transmissivity, which is related to hydraulic

284 conductivity by $K = T/b$, where T is transmissivity and b is effective aquifer thickness (Schwartz

285 & Zhang, 2002). Transmissivity estimates were not included in the database directly, but

286 hydraulic conductivity estimated from transmissivities in the literature are included.

287 Methods of fitting the time-series drawdown data that appear commonly in LANL well

288 completion reports include Theis and the related Cooper-Jacob method for confined aquifers, the

289 Hantush equation for partially-penetrating wells, and the Neuman method for unconfined

290 aquifers (Schwartz & Zhang, 2002). The exact methods used vary between analysts and pumping

291 test configurations in the well completion report collection at LANL. These analyses may

292 introduce subjectivity if the fitting is performed by eye, or even when using semi-automated

293 methods to minimize residuals. The time selected for curve-fitting along the drawdown or

294 recovery process is also a factor in the appropriateness of the hydraulic conductivity estimate and

295 its scale; in the case of single well tests, early-time drawdown data interrogates a smaller portion

296 of the aquifer located closer to the well screen, and late-time data represents a larger volume of

297 the aquifer system as pumping stresses reach further away from the well screen .

298 Another intermediate-scale single well pumping test method is referred to as the specific

299 capacity method. The specific capacity method described in McLin (2005), which estimates

300 lower bounds for hydraulic conductivities, is typically used in this dataset.

301 A reason that well K estimates may also be uncertain for appropriate use in the model

302 that is unrelated to upscaling has to do with potential bias in location of the well screens. In some

303 cases where K values are estimated from pumping tests, monitoring or infrastructure (pump and

304 treat) wells may have had higher K strata targeted for their well screen. Therefore, the K

305 estimates from pumping tests could overestimate an appropriate average value for the model

306 volume over which they are applied.

307

308   Table 1.

309   *Summary Data for K Estimates in the Puye and Tschicoma Formations*

310

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

311   **Table 1. Summary of** K **estimates for Puye and Tschicoma formation well screens in the**

312   **database collected for this analysis across the LANL site, including the RDX area. The**

313   **number of independent estimates (*N*) represents the number of unique monitoring**

314   **well/screen locations for each type of analysis.**

315   2.4 Filtering

316   The data collection should aim to be comprehensive, although it may stay within the

317   bounds of the screening identified earlier; the filtering step is used after evaluating all of the

318   collected data for relevance, scale, and quality. Filtering may be thought of as applying a weight

319   of zero to certain data in the database, while fractional "value" weights (between 0 and 1) may be

320   used to address various aspects of data quality and relevance (Edwards, 1977).

321   Exploratory data analysis (EDA) is a tool used to investigate any patterns present among

322   the data. It is an essential process that helps guide the filtering and/or data weighting procedures

323   to be performed. EDA can include making visualizations of the data for each material in the

324   model, and generating summary tables of the data (counts, quantiles, measure of spread and

325   central tendency, etc.). These plots and tables can help determine if and how selective to be

326   about location (e.g., site-specific only versus general for the material), type of analysis (e.g.,

327   grain-size analysis versus pumping test estimates), quality, etc.

328   One of the challenges in developing hydraulic conductivity distributions for the RDX

329   project area is that there are few wells and therefore few pumping test hydraulic conductivity

330   estimates. A greater density of K estimates in the Puye formation is found downgradient at the

331   LANL Chromium project area. The Puye formation is continuous and large in extent (Figure 2),

332   and although the wells are far away they may still be representeative of the overall formation

333   properties. The first example of using EDA to make filtering decisions is the comparison

334   between RDX area and "other" nearby area K estimates (primarily the Chromium project area,

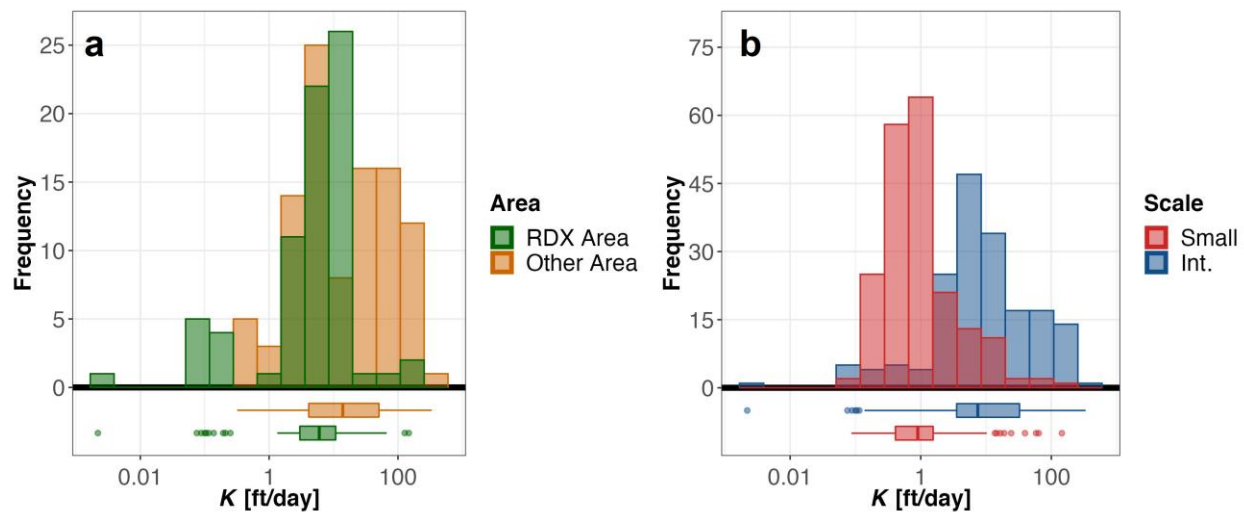335   but also other parts of the LANL site) in the Puye formation, shown in Figure 4a.

336
337   **Figure 4. (a) RDX-area monitoring well K estimates compared to all other locations in the**

338   **same geologic unit (Puye formation) and scale (intermediate) at the LANL site. (b) Small-**

339   **scale (corehole and geophysics) data compared to intermediate-scale K estimates ("int")**

340   **across the LANL site.**

341   Among the intermediate-scale Puye formation data, the median K estimate for RDX area

342   samples is 6.0 ft/d (n = 74), while the "Other" area category median K estimate is 14.3 ft/d (n =

343   100). This may be explained by the depositional environment causing spatial trending in the

344   material properties (the RDX area is closer to the mountain block source of the alluvial fan

345   deposits of the Puye formation). However, there are few K estimates in the RDX area compared

346   to the site as a whole, so although the medians are different, based on the distributional goal of

347   identifying the plausible *range* of K values for the model, all LANL-area estimates were

348   included in the distribution at this time. This can be re-evaluated in future iterations of additional

349   estimates for hydraulic conductivity of the Puye formation are made within the RDX project

350   area. There are no estimates in the database for the Tschicoma formation outside the RDX

351   project area, so a similar comparison is unnecessary.

352   Figure 4b shows a comparison between small- ("corehole", from the sonic coreholes in

353   the Chromium project area) and intermediate-scale K estimates across the entire LANL site.

354   Inclusion of the small-scale data in the database, despite the mismatch between corehole data and

355   the model usage scale of the parameter defined in the distributional goal, allows for an additional

356   check on the K estimates from other methods. Among the LANL area Puye formation data, the

357   corehole K estimates have a lower median than the intermediate-scale values from pumping tests

358   and similar methods, which is consistent with the expected scale-dependence of hydraulic

359   conductivity (Carrera, 1993; Clauser, 1992; Neuman, 1990; Schulze-Makuch et al., 1999).

360   Other types of EDA may include plotting by location to determine spatial trends,

361   comparing geologic units or subunits, interrogating metadata for measurement method

362   assumptions, investigating correlations with other parameter values (e.g., porosity), and so on.

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

364   decisions:

365   •   All sitewide data in the Puye formation are used to inform the K distribution in Puye, as
366   opposed to only Puye data near the RDX plume.

367   •   At the LANL site, there are identified subunits within the Puye formation (Broxton &
368   Vaniman, 2005), but K estimates from all subunits were included in the Puye distribution.

369   •   Comparisons between the distributions of intermediate- and small-scale K estimates
370   suggest that there are statistically significant differences between the medians of these
371   populations. Intermediate scale data are value-weighted twice as high as small-scale data
372   because of the representativeness to the scale of how K is used in the model. The
373   weighting difference of a factor of 2 is arbitrary as it is not known how much more
374   relevant intermediate-scale data are to the model scale than small-scale data.

375   2.5 Statistical Analysis

376   Once filtering and/or weighting of the data are determined and applied from the previous

377   step, the final set of retained, weighted data in the database is processed to generate a distribution

378   that is appropriate for use in the model (as defined by the distributional goal, Section 2.2).

379   At anchor point locations, unique distributions are developed using only the data

380   provided for each individual well. For pilot points, where the distribution depends on the

381   geologic group (Puye or Tschicoma), records are first averaged within a well location before

382   fitting the distribution. This is because multiple K estimates from the same well screen are not

383   considered independent: in the case of pumping test analyses, for instance, multiple reported K

384   estimates may come from several attempts to fit the same drawdown or recovery data using

385   different approaches. At other well locations, only one "best estimate" might be reported, based

386   on which of the methods is assumed to be most valid. Thus, averaging all data for a well screen

387   is performed to reduce potential bias towards wells with greater numbers of reported estimates

388   and ensure that independence assumptions are met.

389   Distributions for anchor points were fit using all available data for the well within the

390   respective formation using the method of moments, with the exception of R-26, as discussed

391   below (Figure 5a). The variance of the distribution estimated for an anchor point in this manner

392   is representative of the variability in the applied measurement method(s) within an aquifer test

393   analysis. The distributions are normal in log data space (log refers to log10 throughout this

394   paper).

395   The log-transformed means of the distributions for the Puye and Tschicoma formations

396   were estimated by calculating the mean among the well averages for all wells available, and the

397   standard deviation of the distribution for each formation was estimated by calculating the

398   standard deviation among the well averages and dividing by the square root of the number of

399   wells. The variance in the distribution for each geologic formation represents the variability

400   between wells. Figure 5b shows the resulting distributions for all pilot points in the model.

401   R-26 Screen 2 is assigned a very wide distribution because of considerable uncertainty

402   reflected in the pumping test information at that location. While the pumping tests suggest an

403   unusually low K estimate for that location (based on a recovery response analysis) compared to

404   other Puye formation pumping tests, and it is also corroborated by the specific capacity lower-

405   bound K estimates, there is some uncertainty about whether this is reflective of the aquifer

406   system or if equipment malfunction is responsible for the low estimate. Additionally, estimates

407   were made at R-26 during drilling using borehole geophysical methods that suggested

408   considerably larger K estimates in some intervals in the borehole than the pumping test analyses

409   at screen 2 (Kleinfelder, 2005), although, if there was no equipment error, this could reflect true

410   vertical differences in hydraulic conductivity at this location. Therefore, the distribution for this

411   "anchor" point was unusually wide, even wider than the generic geology distribution for the

412   Puye formation used for pilot points, and is selected to be uniform rather than lognormal. While

413   this location has no detectable RDX contamination, because it is the furthest upgradient at the

414   RDX site, it plays a significant role in the model calibration by setting the hydraulic gradient

415   across the plume area. All of these factors contribute to its wide allowable range in a uniform

416   distribution. The upper screen at R-26 is not used as an anchor point in the model because it is

417   unknown if screen 1 is representative of the regional aquifer or was completed in perched

418   conditions (Kleinfelder, 2005).

419   The 1st and 99th percentiles of the distributions shown in Figure 5 were selected as the

420   allowable constraints for the calibration, as discussed below. For Puye, the range is 2.1 to 12 ft/d.

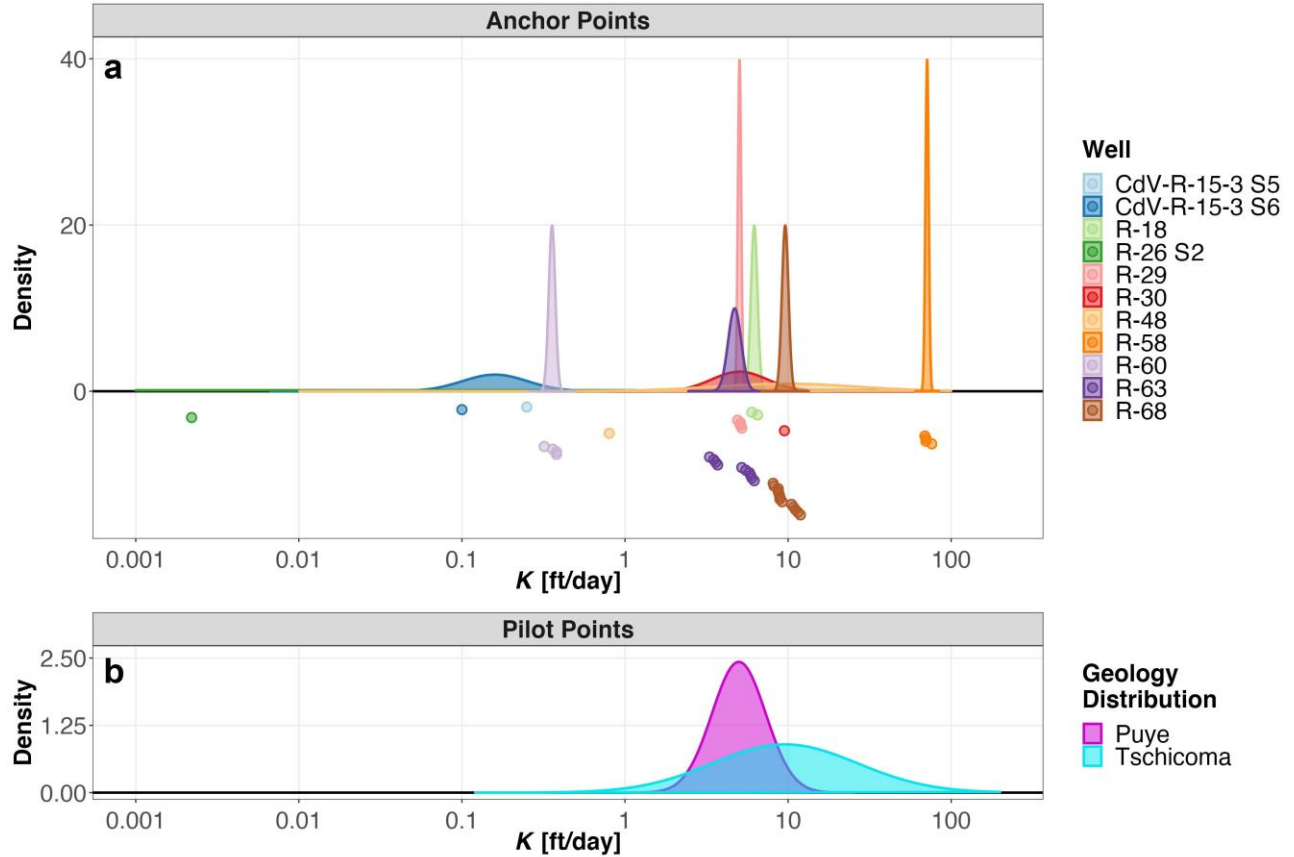421   For Tschicoma, the range is 0.88 to 104 ft/d.

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.

435   The mean of the distribution is used to initialize the pilot and anchor points in the model

436   for the calibration, but any value throughout the range may be used depending on the history

437   matching qualities of the parameter set. Since the calibration is sensitive to initialization

438   parameters, the initial values determined by the distribution development are very important.

439   When the calibration process is completed, a set of parameters (including K values at all

440   pilot and anchor points) is found that minimizes the objective function. The kriged K_x field for

441   the LM calibration result is shown in Figure 6. As described above, K_x and K_y are calibrated

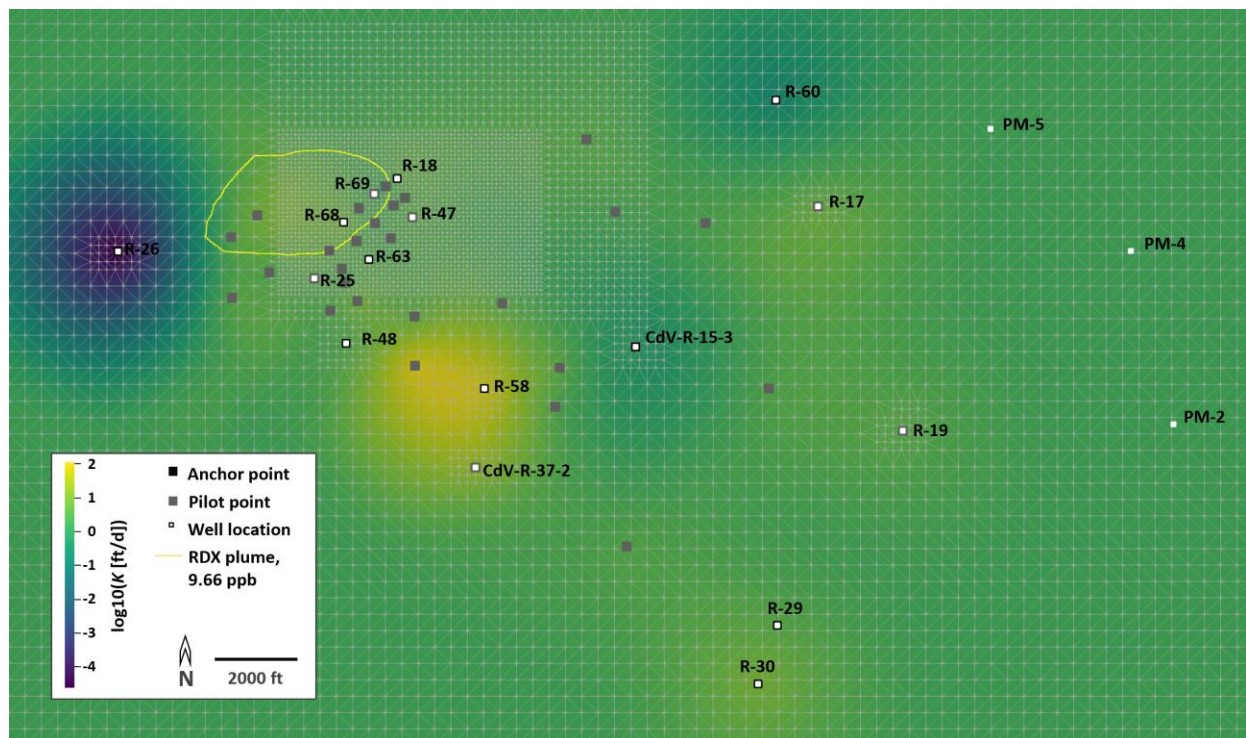442   independently, although the same K distributions are used for both.

444   **Figure 6. Model layer 1 (top layer) K_x for the calibrated RDX groundwater model, in**

445   **log10([ft/d]).**

446   An unusually low K_x value at R-26 PZ-2 is found in the model calibration and is

447   consistent with the site data, as discussed above. R-58, on the high end of the anchor points

448   based on site data, remains elevated after model calibration. In general, other than a few anchor

449   points which represent outliers in the data set, the model calibration has achieved a fairly regular

450   K field in the horizontal dimension.

451   A check was performed to see if the model calibration moved any K values to within

452   10% of their constraints (bounding ranges). A calibrated value that is pinned to the edge of the

453   allowable range suggests an area where additional focused attention may be necessary, because

454   the LM optimization may prefer a value outside the bounds developed from the data.

455   Considerable deviations would suggest that scaling issues, model structure, or other factors

456   would require model values to be further from the data values in order to match the calibration

457   target data. In the RDX groundwater model calibration, for K_x or K_y, 4% of pilot or anchor point

458   locations were within 10% of the lower range limit, and 3% of locations were similarly near the

459   upper range limit. None of the points were within 1% of the upper or lower limits.

460   The next step in the model workflow after calibration is uncertainty analysis, which is

461   used to expand the deterministic result from the calibration into a set of input parameters to run

462   the model probabilistically (N3B, 2020). The uncertainty analysis is performed using the Markov

463   Chain Monte Carlo (MCMC) sampler as implemented in MADS ("MADS: Model Analysis &

464   Decision Support," n.d.). This analysis produces 1018 model runs which are intended to capture

465   the state of uncertainty in the parameters (including K) while still achieving acceptable history-

466   matching to the field data.

467   Figure 7 shows a comparison of K_x and K_y posterior distributions (represented as

468   histograms of MCMC-generated samples) for pilot points against the prior distributions (black

469   line) for horizontal K for the Puye and Tschicoma formations. For the Puye formation pilot

470   points, the modeled values after uncertainty analysis follow the distribution developed from the

471   site data. In essence, the maximum a'posteriori estimates from the MCMC calibration are

472   consistent with maximum likelihood estimates from the LM calibration. For the Tschicoma

473   formation, there are only two pilot points, where each K_x and K_y cluster in a particular location

474   in parameter space after the uncertainty analysis. One pilot point (pp40 in Figure 3) moved to the

475   edge of the distribution during the LM calibration (98 ft/d) and remained near there during the

476   MCMC. Taken together, the results for pilot points in the Puye and Tschicoma did not drive any

477   re-evaluation of the model structure, but the pilot points near the extremes of the distribution will

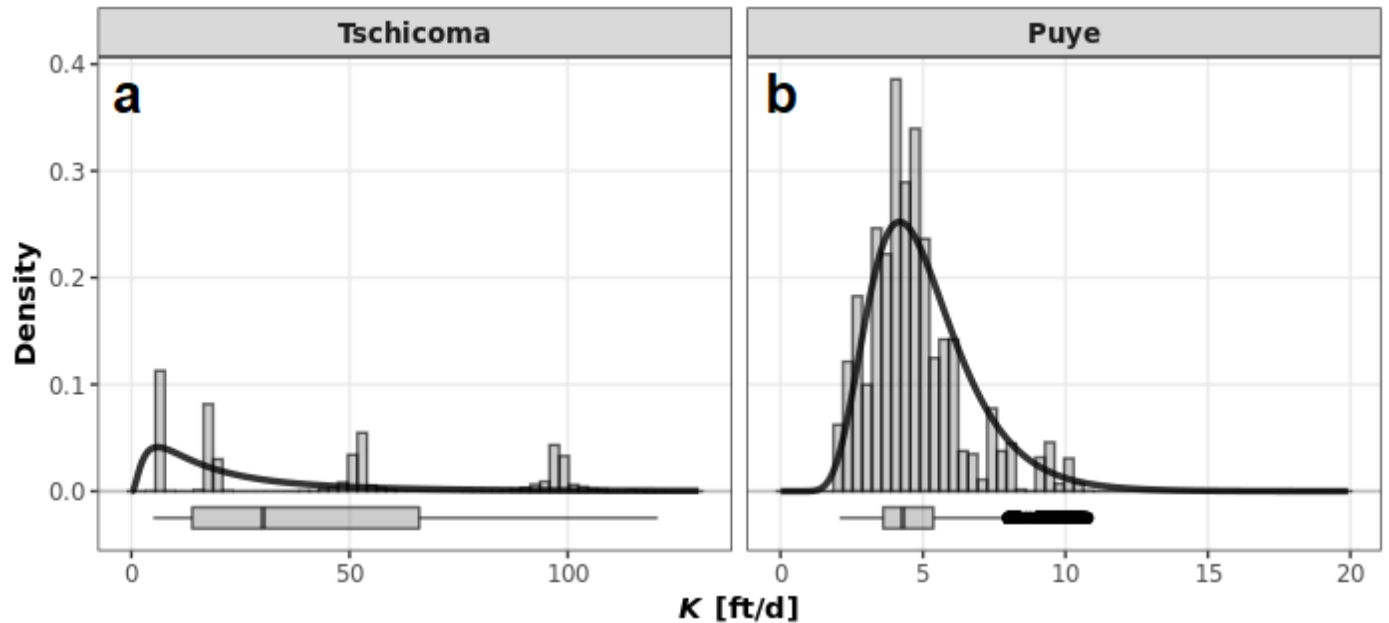478   be re-analyzed in future iterations.

479

Figure 7: K values generated from the uncertainty analysis for pilot points in the model, for the Tschicoma (a) and Puye (b) formations.

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

497 flexibility for calibration at pilot points (within the realm of plausible values for the geologic

498 formation), is both defensible and practical for numerical modeling purposes. It optimizes the

499 representation of upscaled physical processes within a parsimonious model framework. The

500 method is easily extendable to other heterogeneous parameters and modeling applications.

501 In the example presented here for a model of groundwater contamination with RDX, two

502 geologic units are present in the plume area and their pilot and anchor point distributions reflect

503 the K estimates collected in a database of LANL-area observations. The data collection spans a

504 wider range of locations, geologic units, and scales than are likely to be representative, so a

505 filtering/weighting step is performed. Based on the explicitly defined distributional goal,

506 statistical analysis is performed to identify the distributions and $1^{st}/99^{th}$ percentiles are used as

507 constraints in model calibration. The resulting calibration found excellent matches to field data

508 while staying within the bounds provided by the statistical analysis (N3B, 2020). This suggests

509 that the K estimates from the site provided a plausible and defensible starting location for the

510 model calibration (relative to other parameter values) with some adjustment allowed during

511 calibration to account for model grid resolution, scale, simplifications, assumptions, and

512 uncertainty.

513 Future work on the use of the method for LANL-area groundwater models could include

514 a second iteration on the RDX site data to evaluate which pilot and anchor points are particularly

515 sensitive. Additional focused analysis can be performed on those literature data for hydraulic

516 conductivity estimates to see if uncertainty may be narrowed.

517 Another useful improvement in the application of the method to the RDX groundwater

518 model would be making the distributions multivariate in K_x, K_y, and K_z. At present, the

519 distributions are drawn independently, although the K_z distribution covers a significantly lower

520 range of K values based on observed and estimated anisotropy at the site. Therefore, the model

521 has a built-in mechanism for calibrating with $K_z < K_x$, K_y. However, the statistical analysis and

522 model implementation could be updated by adjusting the distributional goal to include anisotropy

523 correlation explicitly.

524 **Acknowledgments**

530   **Open Research**

531   The data presented here can be found in N3B (2020) (https://ext.em-

532   la.doe.gov/GovFTPFiles/api/GetFiles/GetFile?fileName=EMID-700925-01_EMLA-2020-1452-

533   02-001_RDX_Deep_GW_Risk_Rpt_052820.pdf).

534   **References**

535   Bear, J. (1972). *Dynamics of fluids in porous media*. New York, London, Amsterdam.: America Elsevier Publishing

536   Company, Inc. .

537   Berkowitz, B. (2002). Characterizing flow and transport in fractured geological media: A review. *Advances in Water

538   Resources*, *25*(8–12). https://doi.org/10.1016/S0309-1708(02)00042-8

539   Brittingham, H., Gains-Germain, L., Anderson, D., Jordan, A., Foster, L., Rice, A., et al. (2020). Distribution

540   Development for the RDX Regional Model at Los Alamos National Laboratory-20385. In *Waste Management

541   Symposia, March 8–12, Phoenix, AZ.*

542   Broxton, D. E., & Vaniman, D. T. (2005). Geologic Framework of a Groundwater System on the Margin of a Rift

543   Basin, Pajarito Plateau, North-Central New Mexico. *Vadose Zone Journal*, *4*(3).

544   https://doi.org/10.2136/vzj2004.0073

545   Carrera, J. (1993). *An overview of uncertainties in modelling groundwater solute transport*. *Journal of Contaminant

546   Hydrology* (Vol. 13).

547   Certes, C., & de Marsily, G. (1991). Application of the pilot point method to the identification of aquifer

548   transmissivities. *Advances in Water Resources*, *14*(5), 284–300.

549   Clauser, C. (1992). Permeability of crystalline rocks. *Eos, Transactions American Geophysical Union*, *73*(21).

550   https://doi.org/10.1029/91EO00190

551   Cressie, N. (1988). Spatial prediction and ordinary kriging. *Mathematical Geology*, *20*(4).

552   https://doi.org/10.1007/BF00892986

553   Doherty, J. (2003). Ground water model calibration using pilot points and regularization. *Ground Water*, *41*(2).

554   https://doi.org/10.1111/j.1745-6584.2003.tb02580.x

555   Edwards, W. (1977). How to Use Multiattribute Utility Measurement for Social Decisionmaking. *IEEE

556   Transactions on Systems, Man and Cybernetics*, *7*(5). https://doi.org/10.1109/TSMC.1977.4309720

557   Fetter, C. (1994). *Applied hydrogeology, Third Edition*. New York, NY, Macmillan Publishing Co., 1988.

558   Foster, L. & Maxwell, R. (2019). Sensitivity analysis of hydraulic conductivity and Manning's n parameters lead to
559   new method to scale effective hydraulic conductivity across model resolutions. *Hydrological Processes*, *33*(3),
560   332–349. https://doi.org/10.1002/hyp.13327.

561   Gains-Germain, L., Mccandless, S., Anderson, D., Brittingham, H., Stockton, T., & Black, P. (2018). Development
562   of Waste Inventory for Probabilistic Performance Assessment Modeling-19361. In *Waste Management
563   Symposia, March 18–22, Phoenix, AZ.*

564   Higgs, M., Black, P., Jordan, A., Duffy, P., Stockton, T., & Tauxe, J. (2017). Statistical Methods for Effective
565   Spatial and Temporal Scaling in Support of Probabilistic Performance Assessments-17340. In *Waste
566   Management Symposia, March 5–9, Phoenix, AZ.* WM Symposia, Inc., PO Box 27646, 85285-7646 Tempe,
567   AZ (United States).

568   Hill, M. C. (2006). The practical use of simplicity in developing ground water models. In *Ground Water* (Vol. 44).
569   https://doi.org/10.1111/j.1745-6584.2006.00227.x

570   Jordan, A., Fitchett, S., Catlett, K., Levitt, D., Occhiogrosso, G., Tauxe, J., et al. (2017). Realistic Geochemical
571   Parameter Uncertainty for Performance Assessment Modeling-17066. In *Waste Management Symposia,
572   March 5–9, Phoenix, AZ.*

573   Kleinfelder. (2005). *Final Well R-26 Completion Report, Los Alamos National Laboratory, Los Alamos, New
574   Mexico, Project No. 37151, Rev. 1.* Albuquerque.

575   Klute, A. (1965). Laboratory measurement of hydraulic conductivity of saturated soil. In C. A. Black (Ed.), *Methods
576   of Soil Analysis, Part 1: Physical and Mineralogical Properties, Including Statistics of Measurement and
577   Sampling* (pp. 210–221). https://doi.org/10.2134/agronmonogr9.1.c13

578   LANL (2018). *Compendium of Technical Reports Conducted Under the Work Plan for Chromium Plume Center
579   Characterization.* Los Alamos, NM.

580   LaVenue, A. M., & Pickens, J. F. (1992). Application of a coupled adjoint sensitivity and kriging approach to
581   calibrate a groundwater flow model. *Water Resources Research*, *28*(6), 1543–1569.
582   https://doi.org/10.1029/92WR00208

583   LaVenue, A. M., RamaRao, B. S., de Marsily, G., & Marietta, M. G. (1995). Pilot Point Methodology for
584   Automated Calibration of an Ensemble of Conditionally Simulated Transmissivity Fields: 2. Application.
585   *Water Resources Research*, *31*(3), 495–516. https://doi.org/10.1029/94WR02259

586   LaVenue, M., & de Marsily, G. (2001). Three-dimensional interference test interpretation in a fractured aquifer
587   using the pilot point inverse method. *Water Resources Research*, *37*(11), 2659–2675.
588   https://doi.org/10.1029/2000WR000289

589   MADS: Model Analysis & Decision Support. (2021). Retrieved October 27, 2021, from http://madsjulia.github.io/

590   Maliva, R. G., Clayton, E. A., & Missimer, T. M. (2009). Application of advanced borehole geophysical logging to
591   managed aquifer recharge investigations. *Hydrogeology Journal*, *17*(6), 1547–1556.
592   https://doi.org/10.1007/s10040-009-0437-z

593   de Marsily, Gh, Delay, F., Gonçalvès, J., Renard, P., Teles, V., & Violette, S. (2005, March). Dealing with spatial
594   heterogeneity. *Hydrogeology Journal*. https://doi.org/10.1007/s10040-004-0432-3

595   de Marsily, Ghislain. (1978). *De l'identification des systemes hydrologiques [On the calibration of hydrologic*

596   *systems]*. Doctoral thesis.

597   McLin, S. G. (2005). *Analyses of the PM-2 Aquifer Test Using Multiple Observation Wells*. Los Alamos, NM.

598   Moore, C., & Doherty, J. (2006). The cost of uniqueness in groundwater model calibration. *Advances in Water*

599   *Resources*, *29*(4). https://doi.org/10.1016/j.advwatres.2005.07.003

600   Murakami, H., Chen, X., Hahn, M. S., Liu, Y., Rockhold, M. L., Vermeul, V. R., et al. (2010). Bayesian approach

601   for three-dimensional aquifer characterization at the Hanford 300 Area. *Hydrology and Earth System Sciences*,

602   *14*(10), 1989–2001. https://doi.org/10.5194/hess-14-1989-2010

603   N3B (2019). *Investigation Report for Royal Demolition Explosive in Deep Groundwater*. Los Alamos, NM.

604   N3B (2020). *Fate and Transport Modeling and Risk Assessment Report for RDX Contamination in Deep*

605   *Groundwater*. Los Alamos, NM.

606   Neuman, S. P. (1990). Universal scaling of hydraulic conductivities and dispersivities in geologic media. *Water*

607   *Resources Research*, *26*(8). https://doi.org/10.1029/WR026i008p01749

608   RamaRao, B. S., LaVenue, A. M., de Marsily, G., & Marietta, M. G. (1995). Pilot Point Methodology for

609   Automated Calibration of an Ensemble of conditionally Simulated Transmissivity Fields: 1. Theory and

610   Computational Experiments. *Water Resources Research*, *31*(3), 475–493. https://doi.org/10.1029/94WR02258

611   Rubin, Y., Chen, X., Murakami, H., & Hahn, M. (2010). A Bayesian approach for inverse modeling, data

612   assimilation, and conditional simulation of spatial random fields. *Water Resources Research*, *46*(9).

613   https://doi.org/10.1029/2009WR008799

614   Schulze-Makuch, D., Carlson, D. A., Cherkauer, D. S., & Malik, P. (1999). Scale Dependency of Hydraulic

615   Conductivity in Heterogeneous Media. *Ground Water*, *37*(6). https://doi.org/10.1111/j.1745-

616   6584.1999.tb01190.x

617   Schwartz, F. W., & Zhang, H. (2002). *Fundamentals of Ground Water*. New York, NY: John Wiley & Sons.

618   Singha, K., Day-Lewis, F. D., & Moysey, S. (2007). Accounting for tomographic resolution in estimating

619   hydrologic properties from geophysical data. In *Geophysical Monograph Series* (Vol. 171).

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

621   Tompson, A. F. B., & Gelhar, L. W. (1990). Numerical simulation of solute transport in three-dimensional,

622   randomly heterogeneous porous media. *Water Resources Research*, *26*(10).

623   https://doi.org/10.1029/WR026i010p02541

624   Yang, Y., Over, M., & Rubin, Y. (2012). Strategic placement of localization devices (such as pilot points and

625   anchors) in inverse modeling schemes. *Water Resources Research*, *48*(8).

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

627