

# Reproducing tailing in breakthrough curves: are statistical models equally representative and predictive?

Daniele Pedretti<sup>1,\*</sup>, Marco Bianchi<sup>2</sup>

<sup>1</sup>*Geological Survey of Finland (GTK), Espoo, FI-02151, Finland (\*corr auth: daniele.pedretti@gtk.fi)*

<sup>2</sup>*British Geological Survey (BGS), Keyworth, Nottingham, NG12 5GG, United Kingdom*

## Abstract

Breakthrough curves (BTCs) observed during tracer tests in highly heterogeneous aquifers display strong tailing. Power laws are popular models for both the empirical fitting of these curves, and the prediction of transport using upscaling models based on best-fitted estimated parameters (e.g. the power law slope or exponent). The predictive capacity of power law based upscaling models can be however questioned due to the difficulties to link model parameters with the aquifers' physical properties. This work analyzes two aspects that can limit the use of power laws as effective predictive tools: (a) the implication of statistical subsampling, which often render power law undistinguishable from other heavily tailed distributions, such as the logarithmic (LOG); (b) the difficulties to reconcile fitting parameters obtained from models with different formulations, such as the presence of a late-time cutoff in the power law model. Two rigorous and systematic stochastic analyses, one based on benchmark distributions and the other on BTCs obtained from transport simulations, are analyzed. It is found that a power law model without cutoff (PL) results in best-fitted exponents ( $\alpha_{PL}$ ) falling in the range of typical experimental values reported in the literature ( $1.5 < \alpha_{PL} < 4$ ). The PL exponent tends to lower values as the tailing becomes heavier. Strong fluctuations occur when the number of samples is limited, due to the effects of subsampling. On the other hand, when the power law model embeds a cutoff (PLCO), the best-fitted exponent ( $\alpha_{CO}$ ) is insensitive to the degree of tailing and to the effects of subsampling and tends to a constant  $\alpha_{CO} \approx 1$ . In the PLCO model, the cutoff rate ( $\lambda$ ) is the parameter that fully reproduces the persistence of the tailing and is shown to be inversely correlated to the LOG scale parameter (i.e. with the skewness of the distribution). The theoretical results are consistent with the fitting analysis of a tracer test performed during the MADE-5

experiment. It is shown that a simple mechanistic upscaling model based on the PLCO formulation is able to predict the ensemble of BTCs from the stochastic transport simulations without the need of any fitted parameters. The model embeds the constant  $\alpha_{CO} = 1$  and relies on a stratified description of the transport mechanisms to estimate  $\lambda$ . The PL fails to reproduce the ensemble of BTCs at late time, while the LOG model provides consistent results as the PLCO model, however without a clear mechanistic link between physical properties and model parameters. It is concluded that, while all parametric models may work equally well (or equally wrong) for the empirical fitting of the experimental BTCs tails due to the effects of subsampling, for predictive purposes this is not true. A careful selection of the proper heavily tailed models and corresponding parameters is required to ensure physically-based transport predictions.

*Keywords: upscaling; power law; tracer tests; tailing; anomalous transport; MADE site.*

## **1. Introduction**

Solute transport in advection-dominated highly heterogeneous aquifers typically results in a strongly non-symmetric shape of the breakthrough curves (BTCs). Strong late-time tailing is the result of large contrasts in flow velocity and of solute channeling along preferential paths (e.g. Bianchi and Pedretti, 2017; Fiori, 2014; Le Borgne et al., 2008; Willmann et al., 2008). Even though the BTC tails may account for only a few % of the total initial contaminant mass, the corresponding concentrations can still exceed an identified limit of water toxicity, generating a risk for humans and other sensible receptors exposed to such polluted groundwater. Therefore, modeling of solute transport in heterogeneous aquifer must be able to adequately represent and predict the persistence of concentrations in time.

Because the non-symmetric shape of the BTCs complicates the interpretation of tracer tests by means of the classic Fickian interpretation of the transport processes, alternative non-Fickian approaches have been proposed in recent years to reproduce tailing (e.g. Benson et al., 2000; Berkowitz et al., 2006; Haggerty et al., 2000). These upscaling or “proxy” models (Fiori et al., 2015) embed effective functions, such as memory functions (Carrera et al., 1998), whose parametric forms resemble that of the experimental curves (e.g. Haggerty et al., 2000), allowing the models to mimic the BTC tails.

Power law distributions seem to be the model of choice to describe BTC tailing for upscaling purposes (e.g. Becker and Shapiro, 2003; Dentz and Berkowitz, 2003; Dreuzy and Carrera, 2016; Edery et al., 2014; Farrell and Reinhard, 1994; Fiori and Becker, 2015; Gouze et al., 2008; Luo et al., 2007; Sanchez-Vila and Carrera, 2004; Willmann et al., 2008; Zhang et al., 2013). Although a variety of other non-symmetric parametric statistical functions can be adopted for the same purpose (Haggerty et al. 2000), e.g. the logarithmic model (e.g. McKenna et al., 2001; Pedit and Miller, 1994), the popularity of power law models can be ascribed to the apparent linear behavior formed by the BTC tails when plotted in double-log scales (Figure 1). From model fitting of experimental tracer tests, the power law exponent or slope ( $\alpha$ ) has been observed to vary between  $\alpha \approx 1$  and  $\alpha \approx 5$  when fitting experimental BTCs observed during tracer tests conducted in a variety of flow regimes and transport conditions in heterogeneous aquifers (e.g. Bianchi et al., 2011; BRGM, 1990; Hadermann and Heer, 1996; Haggerty et al., 2000; Pedretti et al., 2013; Sanchez-Vila and Carrera, 2004; Willmann et al., 2008; Zhang et al., 2013).

The predictive ability of effective models has been questioned by several authors (e.g. Fiori et al., 2015; Neuman and Tartakovsky, 2009). A key problem relies in the lack of a solid link between mathematical parameters such as the power law exponents and the physical properties of the aquifers (e.g. Flach, 2012; Willmann et al., 2008; Zhang et al., 2013), for instance the spatial distribution of the hydraulic conductivity ( $K$ ). Indeed, the mechanisms leading to power law tailing in the BTCs have been identified only in very limited circumstances. For instance,  $\alpha = 3/2$  is expected in the case of matrix diffusion (Hadermann and Heer, 1996), while Pedretti et al., (2013) found  $\alpha = 1$  for radially convergent transport. Fiori et al. (2007) used a power-law based approach to show that, for  $K$  fields with univariate power-law distributions of  $\ln K \rightarrow 0$ , the expected scaling of a travel time distribution at late time is also a power-law function, with  $\alpha$  linked to the slope of the  $\ln K$  distribution. Zhang et al. (2014) showed that  $\alpha$  can be related to the statistical distribution of volumetric fractions of low permeable facies in alluvial aquifer systems consisting of series of mobile and immobile zones. For many other types of aquifers and transport conditions, however, a *universal* mechanistic model for the description of late time tailing of the BTC has still not been identified. As such, the exponents of the power law models used in upscaled transport models are generally empirical since their estimation is based on fitting or

calibration of the experimental data (i.e. *ex post* evaluation) rather than on a predictive analysis (i.e. *ex ante* evaluation).

Using power laws as empirical fitting tools is, however, not trivial and uncertain especially when the number of observations is limited, as the effects of subsampling can potentially confuse the interpretation of the results. Indeed, for small datasets, other skewed statistical distributions such as the exponential, Weibull, gamma, Zipf, or the lognormal distributions may resemble power laws when plotted in double log scales (Clauset et al., 2009; Goldstein et al., 2004; Mitzenmacher, 2004). Sparseness of data especially toward the late times is not uncommon for BTCs observed in field conditions, due to the interruption of the monitoring after a certain experimental time (e.g. Haggerty et al., 2004) or because of limitations (e.g. detection limits) of the methods used to measure the concentrations. Some of the difficulties of finding a representative model for the description of the tailing in the BTC are illustrated in Figure 1. The formulation of all models used in this example are reported in Table 1, and presented in detail in next sections.

In Figure 1a, a reference BTC is generated using the power law model without cutoff (PL) model as the memory function of the multi-rate mass transfer transport code STAMMT-L (Haggerty and Reeves, 2002), and then fitted using the lognormal model (LOG). In Figure 1b, the two models are inverted, i.e. the reference BTC is generated with a LOG memory function, while the best fitted curve is a PL model. The entry parameters for the reference distributions in the both left and right examples are physically based and in line with the compilation of results by Haggerty et al. (2000). Note how easily PL and LOG models can be misinterpreted as the correct models to describe the BTC tailing. The identification is further complicated by the fact that the tails of the LOG model follow apparent straight lines in log-log plots similar to the PL model. Indeed, it has been already demonstrated that a visual inspection to double log plots is insufficient to determine if a curve is actually power law distributed (e.g. Clauset et al., 2009). Among the existing mathematical tools to estimate the best fitting parameters, such as maximum likelihood estimators (MLEs) or regressive methods (e.g. Benson et al., 2001; Haggerty and Reeves, 2002; Kelly et al., 2017; Mehdinejadi, 2017), it has been shown that MLEs provide more accurate and reliable results in presence of limited datasets (Clauset et al., 2009) and should be therefore preferred.

Another potential difficulty of using power laws as empirical fitting tools stems from the misinterpretation of best fitting parameters when different power law formulations are adopted and compared. This is illustrated in Figure 1c, which shows the concentrations measured during a dipole flow tracer test (MADE-5 test) performed in the alluvial aquifer in Columbus, USA (Zheng et al., 2011). Details of the experiment are presented by Bianchi et al. (2011). Here, the experimental BTC is normalized using concentration peak and corresponding time, and plotted in double log scales to emphasize the power-law-like behavior of the late-time tail. Two power law models have been fitted to the experimental data to describe BTC tails, the power law model without cutoff (PL) and the power law model with exponential late-time cutoff (PLCO). While the PL model is probably the most adopted formulation for fitting purposes (e.g. Willmann et al., 2008; Zhang et al., 2013), the PLCO model has also been introduced to describe the very late arrival of extremely diluted concentrations and the sudden drop in the BTC tail at large time scales (e.g. (Bijeljic et al., 2011; Dentz et al., 2004; Wang and Cardenas, 2017; Zhang et al., 2013). Fewer studies have applied the PLCO model compared to the PL model, probably because the PLCO requires calibration of two parameters (i.e. slope and cut-off rate) while the PL model needs just one (i.e. the slope). A notable difference in the estimated exponents for the two power law models is observed. When the PL is fitted to the data, the estimated exponent is  $\alpha_{PL} = 1.54$ , while the corresponding value for the PLCO model is  $\alpha_{CO} = 1.26$ , i.e. being significantly smaller than the one estimated without embedding the cutoff in the power law based formulations.

It is therefore reasonable to ask the following questions. How different power laws are from other similar statistical models (e.g. the LOG model) and what is the consequence of choosing one or the other on the predictability of transport behavior? Is the difference in estimated slopes between PL and PLCO models in Figure 1c a specific result of the considered case study or a more general outcome? More importantly, does it matter to select a specific heavily tailed model for what concerns both representativeness (i.e. capacity of fitting a BTC) and predictability of transport? This work aims to answer to these questions by means of two stochastic analyses, one based on a combination of benchmark simulations and the other based on stochastic transport simulations in highly heterogeneous aquifers. The analysis is supported by a Supplementary Information (SI) online document, which reports additional analyses and details of the study not reported in the main manuscript..

## 2. Theoretical analysis

A theoretical analysis based on benchmark distributions is performed to assess in detail (a) the effect of subsampling on the accuracy and representativeness of power law models and on the robustness of their best-fitted parameters, and (b) the departure of PL and PLCO models to describe tailed distributions when used as fitting models. The testing methodology follows the well-established approaches described in Clauset et al (2009) and adopts a scenario-based Monte Carlo framework that follows three main steps.

1. For each analyzed model (PL, PLCO, LOG), a set of specific reference parameters is selected and ensembles of 1000 vectors containing  $n$  independently and identically distributed pseudo-random values are generated with different sizes from  $n = 10^2$  to  $n = 10^6$ .
2. Maximum likelihood estimation is performed to obtain best fitting parameters for each statistical model in order to explore the impact of subsampling on the estimations and the similarity among heavily tailed models.
3. The ensemble of estimated parameters for each tested scenarios is analyzed and discussed.

It is noted that, while Clauset et al (2009) already addressed some aspects analyzed in this work (in particular, the departure of a PL model from a LOG model), we extend that work by adding a full evaluation of the behavior of the PLCO model, which to our best knowledge has not been reported yet in the literature.

### 2.1 Methodology

Table 1 summarizes the key equations (Equations 1-11) defining the models adopted in this work and corresponding best estimators and log-likelihoods ( $L$ ). Equation 1 describes the probability density function (pdf) of the PL model (e.g. Bauke, 2007; Clauset et al., 2009; Goldstein et al., 2004; Mitzenmacher, 2004; Reed, 2001), where  $\alpha_{PL}$  is the slope or exponent,  $t$  is the random variable,  $p$  is the probability density and  $t_m$  is the “early time truncation” setting the beginning of the tail (not to be confused with the late time truncation of the power law model with exponential cutoff, defined hereafter). Equation 2 describes the pdf of the power law model with exponential cutoff (PLCO) (e.g. Clauset et al., 2009), where  $\Gamma$  is the Gamma function,  $\alpha_{CO}$  is the

slope or exponent and  $\lambda$  is the parameter setting the “late time truncation” or end of the PL regime, such that  $\lambda^{-1} > t_m$ . Equation 3 describes the pdf of the LOG model (e.g. Mitzenmacher, 2004), where  $\mu$  is the scale parameter and  $\sigma$  is the shape parameter. Similar to the PL, this model requires only two fitting parameters. It is important to recall that in the LOG model  $\sigma$  is the only parameter controlling the skewness, or degree of tailing, of the distribution ( $\mu$  does not control the shape of the distribution). The greater  $\sigma$ , the more skewed the distribution and heavier the tail (e.g. Blackwood, 1992), while the distributions tends to a more symmetric distribution as  $\sigma \rightarrow 0$  (e.g. Crow and Shimizu, 1988, p. 3). A LOG model can also embed an early time truncation  $t_m$ , such that the pdf takes the form of Equation 4 (Clauset et al., 2009). In all models, in the context of BTCs, the density  $p$  corresponds to the solute concentration [ $\text{ML}^{-3}$ ],  $t$  is the experimental time [T], and  $t_m$  [T] can coincide with the concentration peak ( $t_{pk}$ ). The late-time cutoff  $\lambda$  has units [ $\text{T}^{-1}$ ].

Ensembles of vectors containing pseudo-random values  $t_i$  (where  $i = 1, \dots, n$ ) are generated with a standard and well-adopted transformation approach (Clauset et al., 2009 - App. D; Press et al., 1992). In the context of solute transport, each vector can represent the distribution of first-passage times of solute particles at a control plane in the aquifer, from which a BTC can be estimated (e.g. Pedretti and Fernández-García, 2013). Best-fitted parameters of the models fitting each vector are estimated using MLE methods. From Table 1, note that sometimes the best estimator of a parameter can be obtained analytically; when it is not possible, a numerical minimization of  $L$  is needed (in this work we used a least-square minimization approach). For the PL, the best estimator of the slope,  $\alpha_{PL}'$ , can be obtained analytically minimizing Equation 1, which is obtained by taking the derivative with respect to  $\alpha_{PL}$  and setting it to zero. On the other hand, for the PLCO, the estimated parameters ( $\alpha'_{CO}, \lambda'$ ) need to be found numerically. For LOG model, the best estimated parameters  $\mu', \sigma'$  can be fully found analytically. For the truncated version of the LOG model, on the other hand, a numerical model is required to minimize the corresponding log-likelihood function (Equation 11), where  $P_{LOG}(t_m)$  is the value of the cumulative probability function of the non-truncated LOG model at  $t_m$ .

The MLEs for the PL, PLCO and truncated LOG models require defining the start of the tail of the distribution ( $t_m$ ). For BTCs, it is reasonable to assume that  $t_m$  corresponds to the time of the concentration peak. Therefore,  $t_m$  does not need to be fitted. If  $t_m$  is unknown, however, one

option would to estimate it using a Kolmogorov-Smirnov (KS) test, minimizing the distance ( $D$ ) between the cumulative density function (CDF) of the empirical data,  $F$ , and the CDF of the estimated data,  $F'$ .

For a pair of statistical models, the normalized log-likelihood ratio,  $R$ , can be adopted to identify which between the two models best represents the data. The log-likelihood ratio  $R'$  is defined as

$$R' = \frac{L_1}{L_2} \quad (12)$$

where  $L_1$  and  $L_2$  are the calculated log-likelihoods for the pair of models to be compared. A normalization is required to compare  $R$  values from distributions of different sizes, such that (Clauset et al. 2009)

$$R = \frac{R'}{n\sigma_R} \quad (13)$$

where  $\sigma_R$  is the standard deviation of  $R'$ . Two distributions are considered similar in terms of representativeness of the reference data when  $R \rightarrow 0$ , whereas  $R$  tending to large (positive or negative) numbers indicates a stronger affinity of the data towards one distribution.

The goodness-of-fit (*gof*) test is performed to determine the statistical significance of  $R$ . The *gof* adopts the KS test, which compares the empirical CDF and an ensemble of  $N_v =$  random vectors following the same distribution. The *gof* is the probability ( $p$ ) that the empirical CDF is larger than the CDF of the hypothesized distributions with a set of estimated parameters. As in work of Clauset et al. (2009), we set  $N_v = 1000$  and a probability of 0.1 to define the threshold, below which the hypothesis that “*the empirical data are PL distributed*” is rejected. The more  $p \rightarrow 1$ , the more the differences between empirical data and model can be attributed to statistical fluctuation only.

## 2.2 Analysis

### 2.2.1 PL distributed random numbers

A first test was performed using the PL model as the reference model to generate random numbers. The distributions consider  $t_m=1$  and  $\alpha_{PL} = 2$ , which is similar to the minimum slopes



for BTCs reported by Willmann et al. (2008). The ensemble of estimated parameters and statistics are shown in Figure 2 in the form of boxplots for different sample sizes. The top panel (a) shows the boxplots of the PL estimated  $\alpha'_{PL}$ , the mid panels (b,c) show the PLCO estimated  $\alpha'_{CO}$  and  $\lambda'$ , while the bottom panels (d,e) show the truncated LOG estimated  $\mu', \sigma'$ . The median of the distributions is hereafter indicated by a superscript  $X^{0.5}$  for a generic variable  $X$ , and corresponds to the red line in the box plots in all figures reported in this work.

From Figure 2a, it can be observed that the estimated  $\alpha'_{PL}$  converges to the reference value  $\alpha_{PL} = 2$  for an increasing sample size. The MLE slopes become accurate for  $n \geq 10^3$ , as the fluctuations of the first interquartile range are of the order of  $\alpha'_{PL} \approx 2 \pm 0.25$  when  $n = 10^2$ , while dropping to  $\alpha'_{PL} \approx 2 \pm 0.1$  when  $n = 10^3$ . Observing Figure 2b, the PLCO model provides very similar results for the estimated slope  $\alpha'_{CO}$ . This behavior can be expected considering that the PLCO is a generalization of the PL model. Notably, for increasing  $n$ , the best estimates of the PLCO model are obtained for constant  $\alpha'_{CO}$  and increasing  $\lambda'$ . In other words, for increasing sample size  $n$ , the late-time cutoff is estimated to be further away from the early time cutoff,  $t_m$ . The median of the exponential cutoff value increases at a rate roughly close to  $\log_{10} \lambda'^{0.5} \propto n$ . This can be explained considering that, for the same set of parameters and due to the effects of subsampling, a distribution with larger  $n$  has higher chances to embed extreme values. As such, the tails tend to be more persistent and extended over a larger number of log cycles, as  $n$  increases. Therefore,  $\lambda$  can provide information regarding to the temporal extension (or *persistence*) of the tail of the distribution, which in the context of solute transport can be interpreted as the persistence of the late-time concentrations on the BTC.

Observing Figure 2d and Figure 2e, the LOG model generates a very erratic distribution of estimated parameters. This is particularly true for small  $n$ , where large fluctuations around the median values are noted. As  $n$  increases, the median values stabilize around strongly negative mean values ( $\mu'^{0.5} \approx -500$ ) and high standard deviation ( $\sigma'^{0.5} \approx 22$ ). Although it is not straightforward to provide a quantitative explanation for these specific estimated values, we note that a LOG model with very high standard deviation tends to converge towards a power law model (this is formally proved in the next section, in the context of LOG distributed random numbers). Another important aspect to note is that  $\sigma'^{0.5}$  increases as  $n$  increases. This behavior is inversely

proportional to that of  $\lambda^{0.5}$  estimated using PLCO (Figure 2c). As in the case of  $\lambda$ , this is explained considering that the probability of occurrence of large random number increases as  $n$  increases. This means that not only the skewness of distribution (controlled by  $\sigma$  in the LOG model) is a function on  $n$ , but also that  $\lambda$  can provide information about both persistence and skewness of the tails.

For each pair of tested distributions, the corresponding ensembles of resulting statistical parameters  $R$  and  $p$ -values are provided in Figure 3. At the top, Figure 3a and Figure 3b compare the PL versus LOG normalized ratio. It can be noted that the median of the log-likelihood ratio is always close to  $R^{0.5} \rightarrow 0$ , independently from  $n$ . This result confirms (e.g. Bauke, 2007; Clauset et al., 2009; Fiori et al., 2007; Goldstein et al., 2004; Haggerty et al., 2000; Mitzenmacher, 2004) that, even if the random numbers are generated using a PL model, it is very difficult to distinguish a PL model from a LOG model when fitting empirical datasets. In the context of solute transport modeling, it means that even if the mechanisms controlling the late-time concentrations are PL distributed, it is virtually impossible to discern a PL or LOG based mechanism building up the BTC tails only on the basis of an empirical curve-fitting exercise. This issue is further corroborated from the analysis of the  $p$ -values, which measure the statistical significance of  $R$ . For the pair PL-LOG models, the threshold  $p=0.1$  is often exceeded, while the median value has an inconsistent behavior with  $n$ . This suggests that  $R$  cannot statistically discriminate which model is more representative of the reference data. Similar observations can be drawn observing the other pairs of distributions PLCO-LOG (Figure 3c and Figure 3d) and PLCO-PL (Figure 3e and Figure 3f). In summary, this analysis suggests that, when the reference data are PL distributed, any fitting model can equally well describe the data.

### 2.2.2 PLCO distributed random numbers

Two tests using PLCO as reference model were performed. The first test, described hereafter, is inspired on the PLCO best-fitted parameters from the MADE-5 dataset (Figure 1c) and embeds  $t_m=1$ ,  $\alpha_{CO} = 1.2$  and  $\lambda = 10^{-3}$ . The corresponding estimated parameters using the PL, PLCO and LOG models are reported in Figure 4. The second test embeds a more persistent tail ( $\lambda = 10^{-5}$ ) and is reported in the SI (Text S1), as it provides qualitatively similar conclusions as the first test.

In both tests it was found that the ensemble median PLCO parameters have been correctly identified for large datasets, as the fluctuations of the estimated slope  $\alpha'_{CO}$  around the median reduce as  $n$  increases consistently with the behavior of  $\alpha'_{PL}$  estimated for the PL distributed random numbers (Section 2.2.1; Figure 2a). For  $\lambda'$ , the fluctuations tend to become small (statistical errors of the order of  $\pm 0.1 \log_{10} \lambda'$ ) for number of samples  $n > 10^3$ . Recalling that  $n$  represents the data forming the curves tailing,  $n > 10^3$  means that a large amount of values is needed at late time in order to reduce the statistical errors during the fitting analysis. This may become an issue for practical applications, given that experimental data are typically limited in time, while for numerical simulations more computationally demanding simulations (e.g. more particles in Lagrangian simulations) may be needed to make accurate estimations of the PLCO values.

When the PL model is used to fit the PLCO dataset (Figure 4c), the estimated  $\alpha'_{PL}$  also tends to converge to a unique value as for  $\alpha'_{CO}$ . However, note that  $\alpha'_{PL}$  converges to a wrong value close to  $\alpha'_{PL} \approx 1.5$ , i.e. to a higher slope than the reference one. This departure in estimated slopes between fitted PL and reference PLCO models was also found when these models were fitted to the MADE-5 site data (Figure 1c). The LOG-estimated parameters (Figure 4e and Figure 4f) also tend to converge to a unique set of parameters as  $n$  increases, with  $\mu' \approx 1$  and  $\sigma' \approx 3$ . The latter value is consistent with the expected shape of heavily tailed BTCs, as reported for instance by Haggerty et al. (2000). Very low fluctuations around the median are also observed for the LOG best-fitted parameters as  $n$  increases.

In Figure 5a, the log-likelihood ratio between the PLCO and LOG models tends to large positive values as  $n$  increases, rightly suggesting that the PLCO model is a more capable than the LOG model to describe the PLCO distributed reference data. This might seem a foregone conclusion, but it is recalled that when the reference data were generated according to a PL distribution, all models were found to be equally valid descriptions according to their respective  $R$  values (end of Section 2.2.1). In Figure 5b, it can also be observed that for increasing  $n$  the  $p=0.1$  threshold is never exceeded, reflecting the fact that  $R$  is a statistically reliable metric to distinguish between the two distributions. The comparison between PLCO and PL models (Figure 5c and Figure 5d) also shows that the PL model is not particularly effective in representing PLCO data as shown by the behavior of  $R$  (Figure 5e), which tends to very negative values as  $n$  increases. Note also that

the results further suggest that a LOG model outperforms the PL model to statistically describe the distribution of the data. The validity of this result is corroborated by the  $p$  value (Figure 5f), which does not exceed the  $p=0.1$  threshold, except for small  $n$  values.

The results from the PLCO distributed data highlights two important points worth discussing. The first point is that, while the PLCO model is a good approximation of PL distributed data (as discussed in Section 2.3.1), the opposite is not true. The PL model does not reproduce the reference slope when the data are PLCO distributed. The reason is associated to the effects of the truncation in the PLCO model, which limits the extension of the tails to a finite value, while the tails of the PL model extends endlessly. A cutoff is always expected during solute transport tracer tests due to the finite nature of the injected tracer mass. Therefore, one should opt to a PLCO formulation rather than a PL formulation when using power laws to reproduce BTC tailing for the interpretation of tracer tests. , although the effects of subsampling may create a significant uncertainty in the statistical assessment. To support the importance of the cutoff on the power law model selection, it is also recalled that the moments of the PL model are by definition infinite, while the truncation generates finite moments to the PLCO model. The LOG model is also characterized by finite moments, explaining why the LOG model outperforms the PL model when describing PLCO random datasets. The second point is that a suite of models need to be considered and tested to identify which one best describes the reference dataset (i.e. the BTC). Comparing only two models is not enough for this purpose, even if the  $R$  and  $p$  values indicate that one model dominates over the other (as in the case of LOG over the PL).

### 2.2.3 LOG distributed data

Three final tests were run to evaluate the theoretical behavior of MLE best-fitted PL and PLCO parameters from random data generated using the LOG model. Two tests were performed with an identical methodology as for the analysis of PL and PLCO distributed data. Using different reference LOG parameters, the results (SI, Text S2-S3) show consistency with the previous analyses in what concerns the effects of the statistical subsampling and estimated best fitted parameters. In particular, these tests remark once again the existence of a correlation between the PLCO exponential cutoff rate  $\lambda$  and the LOG shape parameter  $\sigma$ , and that the power law slopes is observed to be less clearly correlated to the distribution tailing.

The third test was performed with a slightly different approach than the previous tests. Rather than focusing on the effect of subsampling, the third test focuses on the behavior of PL and PLCO best fitting parameters as a response to a systematic increment in the degree of tailing in the LOG distribution. To this end, all individual random distributions are obtained imposing  $\mu = 0$  and setting  $n = 10^5$ , while increasing the shape parameters from  $\sigma = 1$  to  $\sigma = 6$ . This range is in line with the variability of LOG shape parameters estimated from the analysis of experimental tracer tests reported by Haggerty et al. (2000).

The results of the third test are reported in Figure 6. The results plotted in panel (a) confirm that the effects of subsampling are negligible, while panel (b) shows that as the tailing (i.e.,  $\sigma$ ) increases, the estimated  $\alpha'_{PL}$  tend to drop from  $\alpha'_{PL} \approx 3 \pm 1$  to a minimum asymptotic value  $\alpha'_{PL} \approx 1.5$ . The range of estimated PL exponents is remarkably close to the range of variability of experimental slopes reported in the literature (e.g. Haggerty et al., 2000; Pedretti and Fiori, 2013; Willmann et al., 2008; Zhang et al., 2013), and similar to the slope found from the interpretation of the MADE-5 tracer test using the PL (Figure 1c). The asymptotic behavior of the PL slopes observed in our theoretical analysis is also consistent with the behavior of the PL slopes estimated from the analysis of transport in heterogeneous permeable fields with different connectivity by Willmann et al. (2008). Another significant result is that the minimum PL slope value  $\alpha'_{PL} \approx 1.5$  is not dissimilar from the minimum values estimated for the BTCs simulated by Zhang et al. (2013).

The PLCO model generates a constant best fitted  $\alpha'_{CO} \approx 1.2$  independently from degree of tailing of the reference distribution (Figure 6c). This value is similar to the one found for the best fitted PLCO model of the MADE-5 data (Figure 1c), and lower than the minimum slopes reported in the literature by Willmann et al. (2008) and Zhang et al. (2013). The fact that the slope is closer to the unit is more consistent with the expected power law scaling of a LOG model for increasing higher skewness. Indeed, from the log-transformation of the terms of Equation (3) one obtains (e.g. Mitzenmacher, 2004)

$$\ln p_{LOG}(t) = \ln \left\{ p_{LOG}(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{\ln(t)-\mu}{\sigma} \right)^2 \right] \right\} = -\ln t - \ln \sqrt{2\pi}\sigma - \frac{(\ln(t)-\mu)^2}{2\sigma^2}. \quad (14)$$

Rearranging (14), it turns out that

$$\ln p_{LOG}(t) = -\frac{1}{2\sigma^2}(\ln(t))^2 + \left(\frac{\mu}{\sigma^2} - 1\right)\ln(t) - \ln\sqrt{2\pi}\sigma - \frac{\mu^2}{2\sigma^2} \quad (15)$$

such that, for large  $\sigma$ , the solution can be approximated as

$$\ln p_{LOG}(t) \approx -\ln(t) + k \quad (16)$$

where  $k$  is a constant. Equation (16) shows that LOG distribution with increasingly higher skewness can be approximated by a power law model with exponent  $\alpha = 1$ . Thus, the fact that the estimated PLCO slopes ( $\alpha_{CO} \approx 1.2$ ) are closer to the unit than the estimated PL slopes ( $\alpha_{PL} \approx 1.5$ ) supports the higher validity of PLCO models as fitting tools compared to the PL model.

The results in Figure 6d remark the existence of a clear inverse correlation between LOG  $\sigma$  and PLCO  $\lambda$ . This last result is a further proof for the use of  $\lambda$  as a metric to describe the persistence of tailing and the skewness of the distribution. This important property is adopted in the following section to provide an alternative approach for using power-law based upscaling models as predictive tools, relying on  $\lambda$  as the key predictive variable instead of  $\alpha$ .

### 3. Fitting transport in highly heterogeneous media

The analyses presented in this section focus on assessing the use of the PL, PLCO, and LOG models as empirical tools to describe BTCs tailing. A stochastic numerical analysis simulating transport in a highly heterogeneous alluvial systems is adopted to obtain reference BTCs with heavy tailing. Since for this type of aquifers and transport conditions it is not possible to assume *a priori* a parametric function form of the BTC tails, the use of MLE-based statistics can provide insights regarding (a) which among the tested parametric models provide the most representative description of the data, (b) the implication of the subsampling on the parametric inference, and (c) the resulting power law slopes estimated using PL and PLCO models.

#### 3.1 Reference BTCs

The reference BTCs are obtained from new numerical simulations of flow and transport in one of the aquifer scenarios previously considered by Bianchi and Pedretti (2017). This scenario is based on the heterogeneity of the Lawrence Livermore National Laboratory alluvial aquifer described in several publications (e.g. LaBolle and Fogg, 2001; Zhang et al., 2013). Transport simulations

in this work differ from those reported in Bianchi and Pedretti (2017) exclusively in terms of number of particles ( $n_p$ ) used to run the Lagrangian transport simulations and of the distance between injection and observation plane. However, the numerical setup is left unchanged and the reader is referred to the previous work for a detailed description.

The simulations reproduce 3D conservative tracer tests in  $n_r=100$  realizations of the distribution of 5 hydrofacies generated using the Markov-chain / transition probability algorithm T-PROGS (Carle and Fogg, 1997). The hydrofacies with highest hydraulic conductivity ( $K$ ) is characterized by  $K=5.184$  m/day and account for 18% of the aquifer volume. The total variance of the system is  $\sigma_{\ln(K)}^2 = 24.9$ , i.e. a much higher heterogeneity for instance than the MADE site where  $\sigma_{\ln(K)}^2 < 6$  (Bohling et al., 2016), since it is representative of an aquifer system with multimodal  $K$  distribution (e.g. Zhang et al., 2013). The high  $K$  hydrofacies have mean lengths equal to 10m, 50m and 1.3m in x, y, and z directions, respectively; the mean flow is parallel to y in these simulations. In each simulation, flow is resolved using a finite difference code (Harbaugh et al., 2000), while transport is resolved using random walk particle tracking (Salamon et al., 2006). For all the hydrofacies, the model consider local isotropic dispersivity of 0.01 m, local effective diffusion of  $5.2 \times 10^{-5}$  m<sup>2</sup>/d and constant effective porosity of 0.1. The injection of the tracer mass, discretized into  $n_p$  particles of equal mass, takes place as a pulse along a fully penetrating passive borehole using a flux-averaged injection scheme. The simulations are run using different sets of  $n_p$  (up to  $10^6$ ) in order to test the effects of statistical subsampling (analogously to  $n$  used in the theoretical tests). The resulting BTCs are measured from the ensemble of first passage times crossing a vertical plane located at a distance  $L = 100$ m downgradient of the injection line. The cumulative distribution functions (cdfs) of particles arrival times are then transformed into area-normalized BTCs using the approach described in Pedretti and Fernández-García (2013).

For illustrative purposes, the ensemble of BTCs obtained from the numerical simulations using the maximum number of particles is reported in the SI (Text S4). It can be observed that the BTCs are significantly right skewed, with peaks located between  $t=10^2$  d and  $t=10^3$  d from the beginning of the simulations and tailing persistent over a few log cycles, with cutoffs between  $t=10^4$  d and  $t=10^6$  d. This strongly non-symmetric behavior is due to the combination of an extremely high variance ( $\sigma_{\ln(K)}^2 \approx 25$ ) and the spatial organization of the hydrofacies resulting in a low relative

geoentropy,  $H_R = 0.32$  (Bianchi and Pedretti, 2017). As shown in Bianchi and Pedretti (2017), the resulting BTCs are more skewed than for instance those simulated, for similar transport conditions, assuming that the aquifer heterogeneity is similar to that of MADE site according to the analysis of Bianchi and Zheng (2016).

The temporal scale of the early times and of the exponential cutoff in the simulated BTCs is quite variable among the realizations, as illustrated by the small inset in Figure S7 (SI) which depicts two BTCs randomly drawn from the ensemble. A normalization is done such that the cdfs of arrival times are normalized by the concentration peak time ( $t_{pk}$ ). The corresponding pdfs recalculated after this normalization are shown in Figure S8 (SI). It can be observed that the curves now collapse and, on a visual inspection, a “power-law-like” behavior with a clearly identified exponential cutoff is now observed at a normalized time ( $\tau = t/t_{pk}$ ) between  $\tau = 10^3$  and  $\tau = 10^4$ . The adopted normalization intrinsically implies that the beginning of the tails,  $t_m$ , corresponds to  $t_{pk}$ .

### 3.2 Parametric estimation of BTC tailing and analysis

MLE-based estimators are applied to obtain the best fitting parameters from the normalized BTCs. The results are shown in Figure 7. The best-fitted PL slope  $\alpha'_{PL}$  (Figure 7a) was found to fluctuate between  $\alpha'_{PL} \approx 1.3$  and  $\alpha'_{PL} \approx 2$ , being quite insensitive to the number of particles. For the largest dataset ( $n_p = 10^6$ ) the median value results in  $\alpha'^{0.5}_{PL} = 1.51$ , while the median values for the estimated slopes of the PLCO model (Figure 7b) are equal to a value  $\alpha'^{0.5}_{CO} = 1.16$ . The range of the cutoff rates varies between  $\lambda' = 10^{-2}$  and  $\lambda' = 10^{-4}$  (Figure 7c), which indicates that the cutoff lies somewhere between 2 and 4 time-normalized log cycles after the BTC peak ( $\tau = 1$ ), as one can also see in Figure S8 (SI). The range of estimated PLCO parameters is insensitive to the number of particles used in the simulations. The behavior of the estimated LOG parameters (Figure 7d and Figure 7e) is also consistent with the results of the theoretical analysis. In particular, the median value for the shape parameter is  $\sigma'^{0.5} = 2.07$ , in line with the results illustrated in Figure 6 for the corresponding best fitted power law slopes. The analysis of  $R$  and  $p$ -values suggests that both LOG and PLCO models largely outperform PL (results not reported). Comparing the PLCO/ LOG log-likelihood ratio (Figure 7f and Figure 7g) it was found that  $R$  tends towards positive values, suggesting that the PLCO model is statistically more representative



than the LOG model. This result gains relevance as  $n_p$  increases and it is well supported by the  $p$ -values, which are always below the threshold for  $n_p \geq 10^4$ .

Analyzing the behavior of estimated parameters of the reference BTCs from the numerical simulations, several analogies emerge with the results presented in Section 2. The  $\alpha'_{PL}{}^{0.5} = 1.51$  is in line with the slopes reported by Zhang et al. (2013), who analyzed similar transport conditions, with the minimum  $\alpha'_{PL}$  obtained from the analysis of heavily-tailed distributed benchmark random numbers (Section 2), and with the estimated PL slope for the BTC observed during MADE-5 experiment (Figure 1c). The  $\alpha'_{CO}{}^{0.5} = 1.16$  found for the PLCO model also in line the theoretical results presented in Section 2 and the analysis of the MADE-5 data, and again tending to  $\alpha = 1$  expected from skewed distribution represented by LOG models. The results cast some doubts however on the validity of the PL formulation to describe the tails, primarily because the  $R$  tests (supported the  $p$ -values) suggest that the PLCO is much more valid model than the PL. Moreover, the fact that estimated PL slope is substantially different than the slope of the PLCO model generates also doubt on the actual reliability of  $\alpha'_{PL}{}^{0.5} \approx 1.5$ . The fact that the PLCO model outperforms the LOG model suggests that the transport mechanisms forming the BTC tails can indeed generate a power law like behavior on the tail. However, for *predictive* upscaling purposes, the cutoff associated with the finite injected mass should not be ignored when using power laws, as shown in the following Section.

#### 4. Predicting transport in highly heterogeneous media

From the analyses presented so far, it was found that:

- Transport simulations showed that BTC tailing may be better described using a power law model with cutoff (PLCO) than using the power law without cutoff (PL) or a LOG model.
- When fitting distributions with increasing tailing (Section 2.2.3), the best-fitted power law exponent of the PLCO model tends to a constant unit value insensitive to the degree of tailing. In particular, a power law behavior with  $\alpha = 1$  is expected for highly skewed distributions, as such those described by LOG distributions with large  $\sigma$ . Experimental results from tracer tests in highly heterogeneous systems (e.g. Haggerty et al., 2000) confirm the empirical estimation of large  $\sigma$  in these systems.

- A power law value close to the unit has been found from both the stochastic simulations performed in extremely heterogeneous system and from the analysis of a tracer test at the MADE-site, which is characterized by a relative milder degree of heterogeneity.
- The lateral persistence of the tailing is fully simulated by the PLCO cutoff rates  $\lambda$ , which determine the characteristic time ( $\lambda^{-1}$ ) at which the power law regime drops. Being correlated to  $\sigma$ ,  $\lambda$  is also a metric for the skewness of the distribution.

Based on these evidences, it is hereby explored if solute transport can be predicted by invoking a PLCO model embedding a *universally valid*  $\alpha_{CO} = 1$  and assuming the cutoff rate ( $\lambda$ ) as the critical parameter controlling transport. To support this hypothesis, a simple mechanistic PLCO-based model is formulated and used to make a prediction of the ensemble of BTCs obtained from the stochastic transport simulation presented in Section 3. The prediction is free of fitting parameters, as  $\lambda$  can be estimated assuming purely advective stratified transport, as follows.

In the highly heterogeneous simulations presented in this work, the distance of plume migration is comparable to the characteristic scales of the aquifer heterogeneity. Under these conditions, a stratified model is a reasonable approximation to describe the solute plume behavior (e.g. Dentz and Carrera, 2007; Gelhar et al., 1979; Pedretti and Fiori, 2013). Using a similar conceptual model as the one adopted by Pedretti et al. (2013) to describe the formation of tailing under radial convergent transport, a BTC can be seen as the convolution of multiple “sub BTCs” formed by pools of particles characterized by similar travel times. The pool of fastest particles forms the early time segment of the BTC, while the pool of slowest particles form the late time tail of the BTC. Figure S9 (SI) describes the evolution in space (parallel to the mean aquifer gradient) of the Lagrangian velocities of two particles. One particle (in blue) is representative of the Lagrangian velocities of the fastest particles ( $v_f$ ), while the other (in red) is representative of the slowest particles ( $v_s$ ). For short travel distances ( $L < 100\text{m}$ , i.e. less than two times the mean facies lengths along the mean aquifer gradient), these velocities are very different and fluctuate over several order of magnitude. The control plane used to calculate the BTC in this work is located exactly at  $L = 100\text{m}$ . For larger travel distances ( $L > 100\text{m}$ ), the velocities become comparable, as the particles transition from slow- to fast-flow paths and subsequently tend to remain within these. In a perfectly stratified media, the time delay ( $\Delta\tau$ ) between the fastest and the slowest pools of particles arriving at the control plane can be calculated as

$$\Delta\tau = \frac{L}{v_s} - \frac{L}{v_f} \quad (17)$$

where  $v_s$  and  $v_f$  are respectively the maximum and minimum Lagrangian velocities at the injection line. Recalling that in the PLCO model the cutoff rate ( $\lambda$ ) was found to be proportional to the persistence of the tail, it is possible to express the cutoff rate as the inverse of the time delay, i.e.

$$\lambda_{est} = \frac{t_{pk}}{\Delta\tau}. \quad (18)$$

Since the BTCs in our numerical experiment are obtained from stochastic simulations,  $\Delta\tau$  is a random variable, as the maximum difference in transport rates randomly varies depending on the realization-specific transport conditions. An ensemble average of  $\lambda_{est}$  can be then calculated from the harmonic mean ( $\lambda_H$ ) of the  $\Delta\tau$  values from the each realization of the  $K$  field. For the numerical experiment in this study, this calculation results in  $\lambda_H=0.0017$ .

The resulting BTCs generated using PLCO models embedding  $\lambda_H=0.0017$  and different  $\alpha_{CO}$  are shown in Figure 8a, along which the ensemble mean of the numerical BTCs is reported as a thick black line. The PLCO based model embedding  $\alpha_{CO} = 1$  (red dotted line), hypothesized to be predictive of the ensemble, agrees reasonably well with the ensemble and when compared with the model embedding the  $\alpha_{CO} = 1.2$  (close to the best fitting  $\alpha_{CO}^{0.5} = 1.16$ ). Not only the initial part of the tail is reproduced fairly well, but the cutoff rate is also nicely matched. A sensitivity around the slope indicates that the models embedding increasingly higher  $\alpha_{CO}$  tends to diverge from the reference dataset. In particular, the model embedding  $\alpha_{CO} = 1.5$  (close to the best fitted slope from the PL model) is able to reproduce the initial part of the tail but underestimates the concentrations as time elapses. Thus, the PLCO-based stratified model works well exclusively when the power law slope tends to the unit, as hypothesized. A different slope is not predictive, including that estimated from the PL model, even when the cutoff rate has been correctly identified. It is remarked that the applicability of this simple model is limited to transport scales for which  $\lambda$  can be directly related to the conditions at the source. The approximation may no longer hold for larger transport distances, i.e. when the plumes diverges from a stratified behavior.

For comparison purposes, the results from the predictions using the PL (Figure 8b) and LOG (Figure 8c) models are also reported. The PL model fits the reference curves only when it embeds the best-fitted  $\alpha_{PL} = 1.5$  and until  $\lambda_H^{-1}$ . All models embedding  $\alpha_{PL} < 1.5$  overestimate the reference curves, while all models embedding  $\alpha_{PL} > 1.5$  largely underestimate the reference curve. This result confirms that the proper use of power laws for predictive purposes is strongly model dependent, and in particular that the PL model is suboptimal for predictive purposes compared to the PLCO model. The only parameter feeding the PL model ( $\alpha_{PL}$ ) does not have a direct link with the physical properties of the aquifer, contrarily for instance to the cutoff rates in the PLCO model, which can be linked to the different velocities at the source (provided the stratified approximation holds).

The ensemble of BTCs tend to be reproduced by the LOG model when the scale parameters are comprised between  $\sigma = 2 - 3$ . These values are in line with the best fitted values described above and also consistent with the relationship between  $\sigma$  and  $\lambda$  identified in Figure 6, as  $\lambda \approx 10^{-2} - 10^{-3}$  when  $\sigma = 2 - 3$ . This also means that  $\sigma$  does not necessarily has to be very large to predict transport in highly heterogeneous systems. Notice also that for a few log cycles (up to  $\tau = 10^2$ ) all curves embedding  $\sigma > 3$  match the power-law like portion of the ensemble BTC. This agrees with the observations by Haggerty et al. (2000). They acknowledged, however, that “the lognormal density function with larger  $\sigma$  cannot provide a true power law BTC but hold the PL slope relatively constant over a long time”. Contrarily to  $\lambda$ , however, it seems more difficult to provide a mechanistic justification for a specific value of  $\sigma$  to be used to predict the curves. Finding a physical explanation for  $\sigma$  is currently under investigation.

## Summary and conclusion

Despite the development of effective models able to reproduce transport at late time (e.g. fitting BTCs tailing), the predictive capability of these tools is limited by the use of empirically-based model parameterizations and the lack of a link between physical and mathematical parameters. The limitations become particularly important when these models embed formulations (e.g. memory functions) with functional forms following power law distributions. This study systematically addressed the issues affecting the use of power-law like as empirical models for

solute transport modeling, contending that these difficulties are intrinsic in the specific formulation of the adopted model. We found the following key results:

- The best-fitted power-law slope ( $\alpha$ ) is strongly model dependent, requiring attention in its selection for predictive purposes. The use of a power law model without cutoff (PL) as fitting model generates best-fitted slopes decreasing from  $\alpha_{PL} \approx 3-4$  for more symmetric reference distributions to an asymptotic minimum value close to  $\alpha_{PL} \approx 1.5$  for heavily tailed reference distributions. The minimum estimated values are in line with the minimum values reported in the literature for transport in strongly heterogeneous systems generating preferential flow and solute transport channeling (e.g. Willmann et al., 2008; Zhang et al., 2013).
- Using the power law model with cutoff (PLCO) to fit the same datasets results in a constant slope  $\alpha_{CO} \approx 1$  independently from the degree of tailing of the reference distributions. A unit slope is consistent with the expected scaling of LOG models when  $\sigma > 1$ , which have been frequently reported from the interpretation of tracer tests (e.g. Haggerty et al., 2000). A large  $\sigma$  means very highly skewed distributions, which is expected for transport in highly heterogeneous systems with strong connectivity, organized structures and low geentropy (Bianchi and Pedretti, 2017). In the PLCO model, the degree of tailing and more precisely the tailing persistence (i.e. the continuity of late-time concentrations in the BTCs) is well reproduced by the cutoff rate  $\lambda$ , which sets the inverse of the characteristic time defining the end of PL regime on the distributions. More symmetric distributions are characterized by  $\lambda \rightarrow 1$ , while more skewed distributions are characterized by  $\lambda \rightarrow 0$ .
- Subsampling may limit dramatically the ability to differentiate a power law model from another heavily-tailed distribution when fitting experimental BTCs. It was shown quantitatively that, for instance, even if the mechanisms controlling the late-time concentrations are PL distributed, it is virtually impossible to discern a power law based mechanism building up the BTC tails from LOG based mechanism. Moreover, the range of estimated parameter may fluctuate significantly, suggesting that in presence of limited datasets (e.g. those coming from field experiments), care must be taken when fitting empirical datasets.

- It has been shown and discussed that a simple PLCO model embedding non-fitted parameters was able to reproduce reasonably well the ensemble of BTCs generated via Lagrangian transport in highly heterogeneous stochastic alluvial aquifers. The model relies on a stratified-based approximation, such that  $\lambda$  can be calculated from the harmonic mean of the maximum differences in Lagrangian velocities at the source for each realization. The inverse of the estimated value agrees with the expected scaling of the tails at late time. The PLCO embeds also a non-fitted  $\alpha_{CO} = 1$ , which is consistent with the expected power law like behavior of a LOG model with  $\sigma > 1$ . On the other hand, PL models deviated significantly from the reference curves, while it was difficult to link LOG parameters to any physical parameter describing the aquifer heterogeneity.

In summary, the following two general conclusions were achieved:

1. For what concerns *representing* tailing, any model may work equally well (or equally wrong) for empirical curve-fitting purposes. Unless a large number of samples is available, such as in the case of numerical simulations, the effect of statistical subsampling may strongly affect the experimental datasets, invalidating the identification of a specific model from a statistical perspective.
2. For what concerns *predicting* tailing, on the other hand, not all models work equally well. The selection of a specific heavily-tailed parametric model becomes fundamental to make right predictions. In particular, the correct model needs to be invoked in order to link physical and mathematical parameters and generate reliable predictions.

## **Acknowledgements**

The authors would like to acknowledge Aldo Fiori and another anonymous reviewer for the constructive comments that helped improving the quality of our manuscript.

## Figures

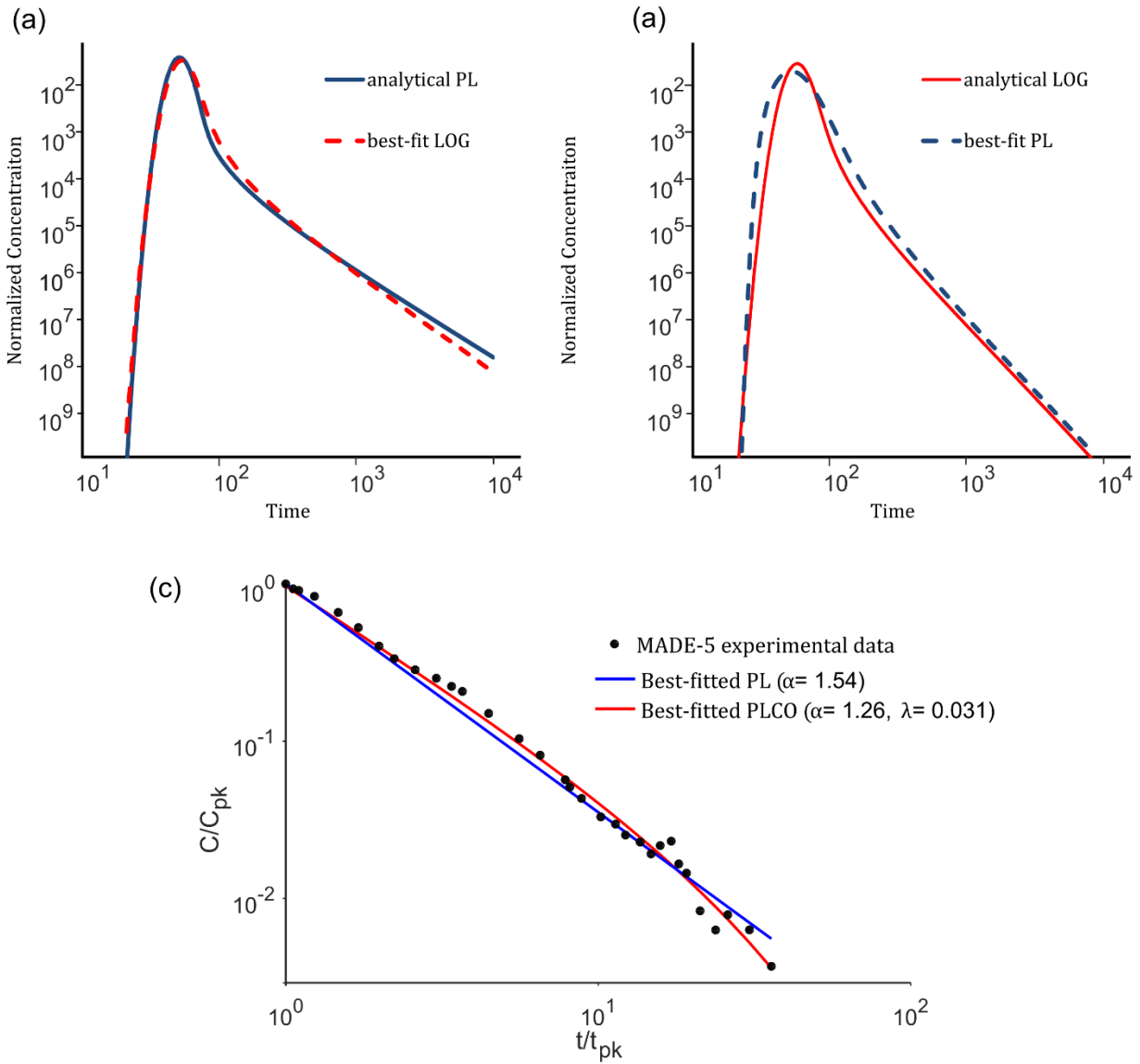


Figure 1 (a-b) Examples of heavily tailed BTCs analytically generated and numerically fitted using the code STAMMT-L (Haggerty and Reeves, 2002). In (a), the analytical curve adopts a PL distribution of mass transfer times with  $\alpha_{PL} = 2$ , while the fitted curve is obtained using a LOG distribution of mass transfer rates. In (b), the reference curve adopts a LOG model with scaling factor  $\sigma = 3$ , while the fitted curve is obtained using a PL model. (c) Experimental results from a convergent flow tracer test (MADE-5), and best-fitted power law models with or without exponential cutoff (respectively, PL and PLCO). Concentration peak and corresponding time are used as normalization variables.



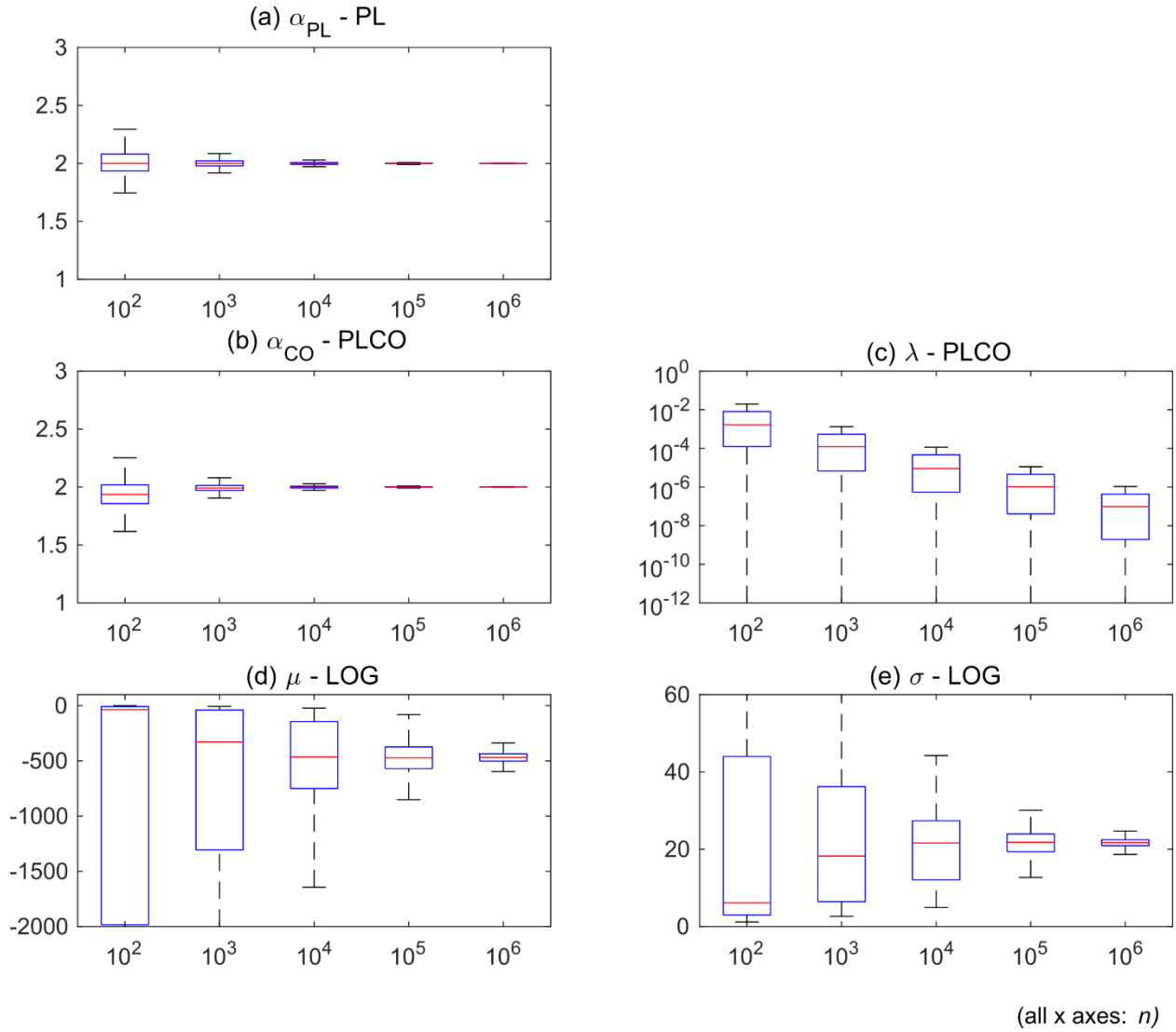


Figure 2 Boxplots representing the statistical distributions of MLE-fitted parameters from 1000 realizations of  $n$  random numbers generated using a PL model with  $\alpha_{PL} = 2$  and  $t_m = 1$ .

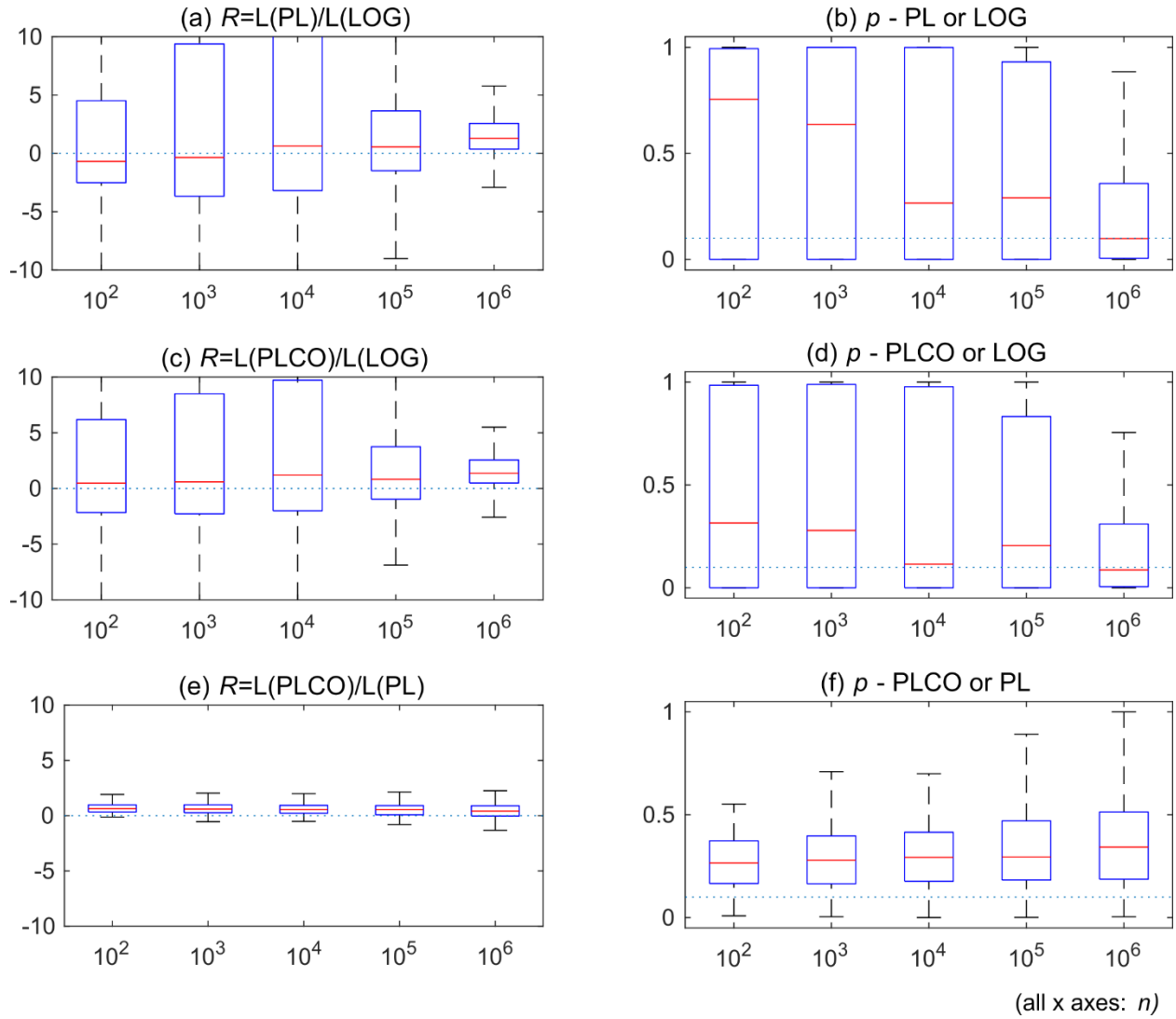


Figure 3 Boxplots representing the statistical distributions of log-likelihood ratio (R) and p-values from 1000 realizations of  $n$  random numbers generated using a PL model with  $\alpha_{PL} = 2$  and  $t_m = 1$ .

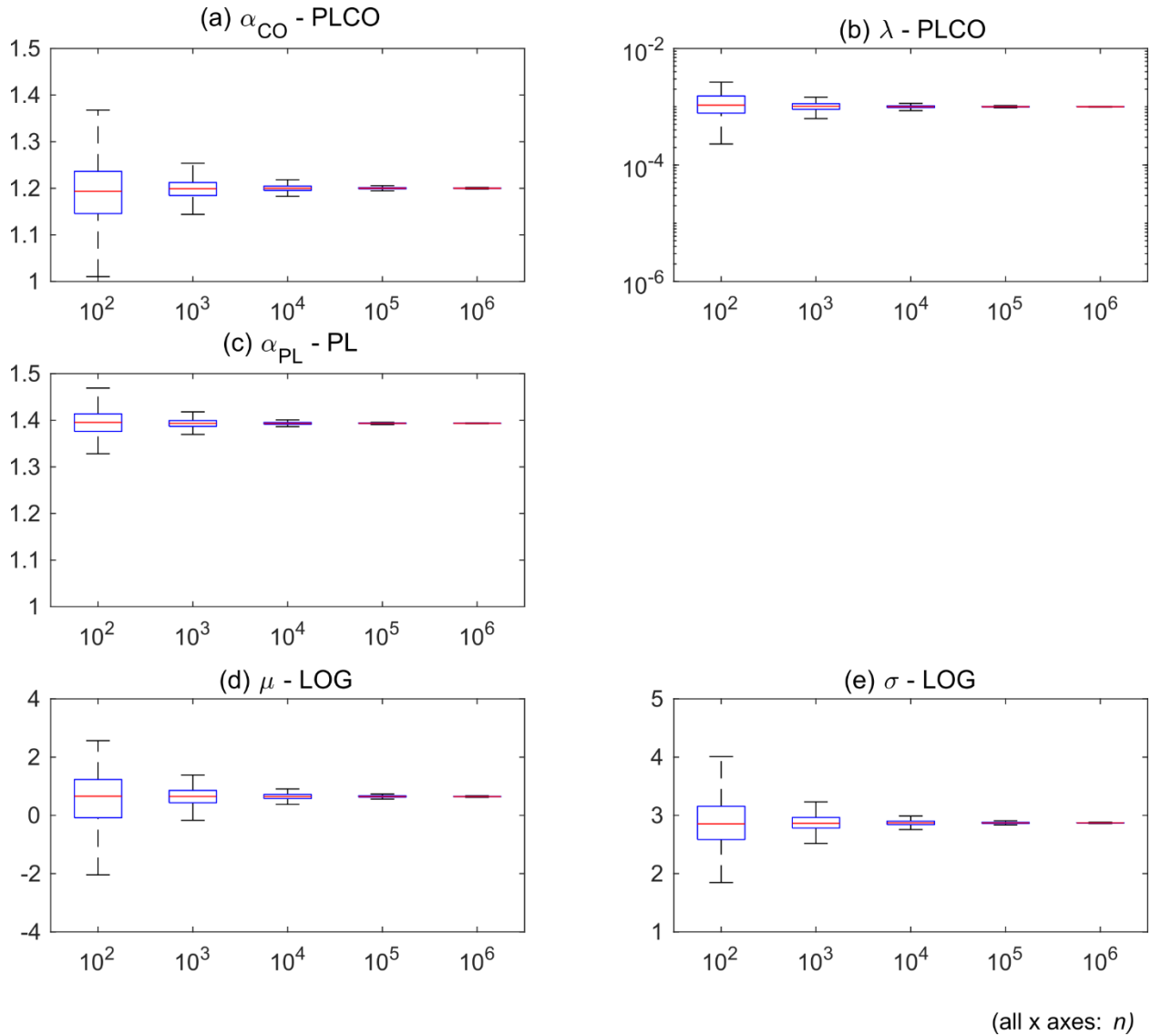


Figure 4 Boxplots representing the statistical distributions of MLE-fitted parameters from 1000 realizations of  $n$  random numbers generated using a PLCO model with  $\alpha_{CO} = 1.2$ ,  $\lambda = 10^{-3}$  and  $t_m = 1$ .

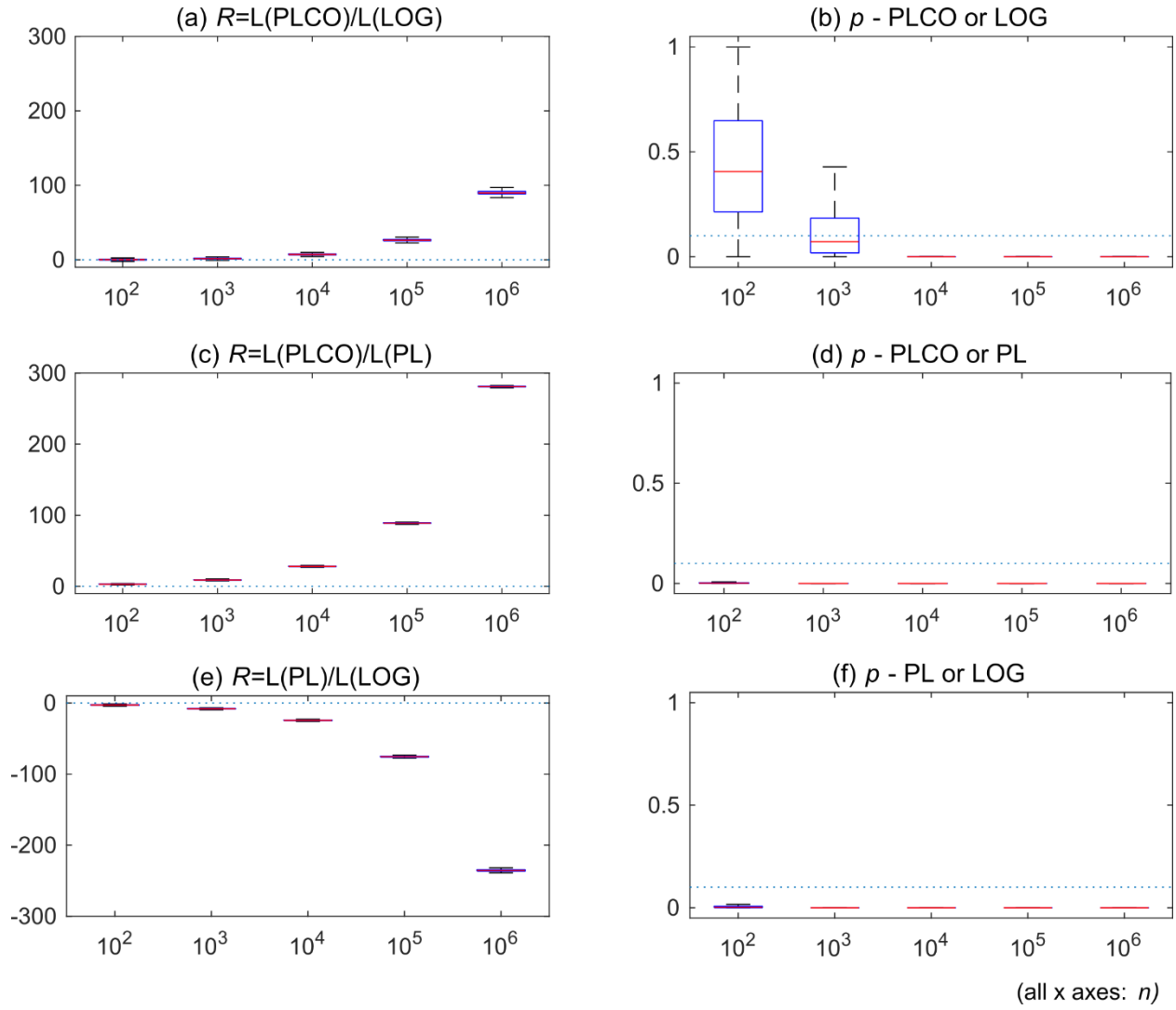


Figure 5 Boxplots representing the statistical distributions of log-likelihood ratio (R) and p-values from 1000 realizations of  $n$  random numbers generated using a PLCO model with  $\alpha_{CO} = 1.2$ ,  $\lambda=10^{-3}$  and  $t_m = 1$ .

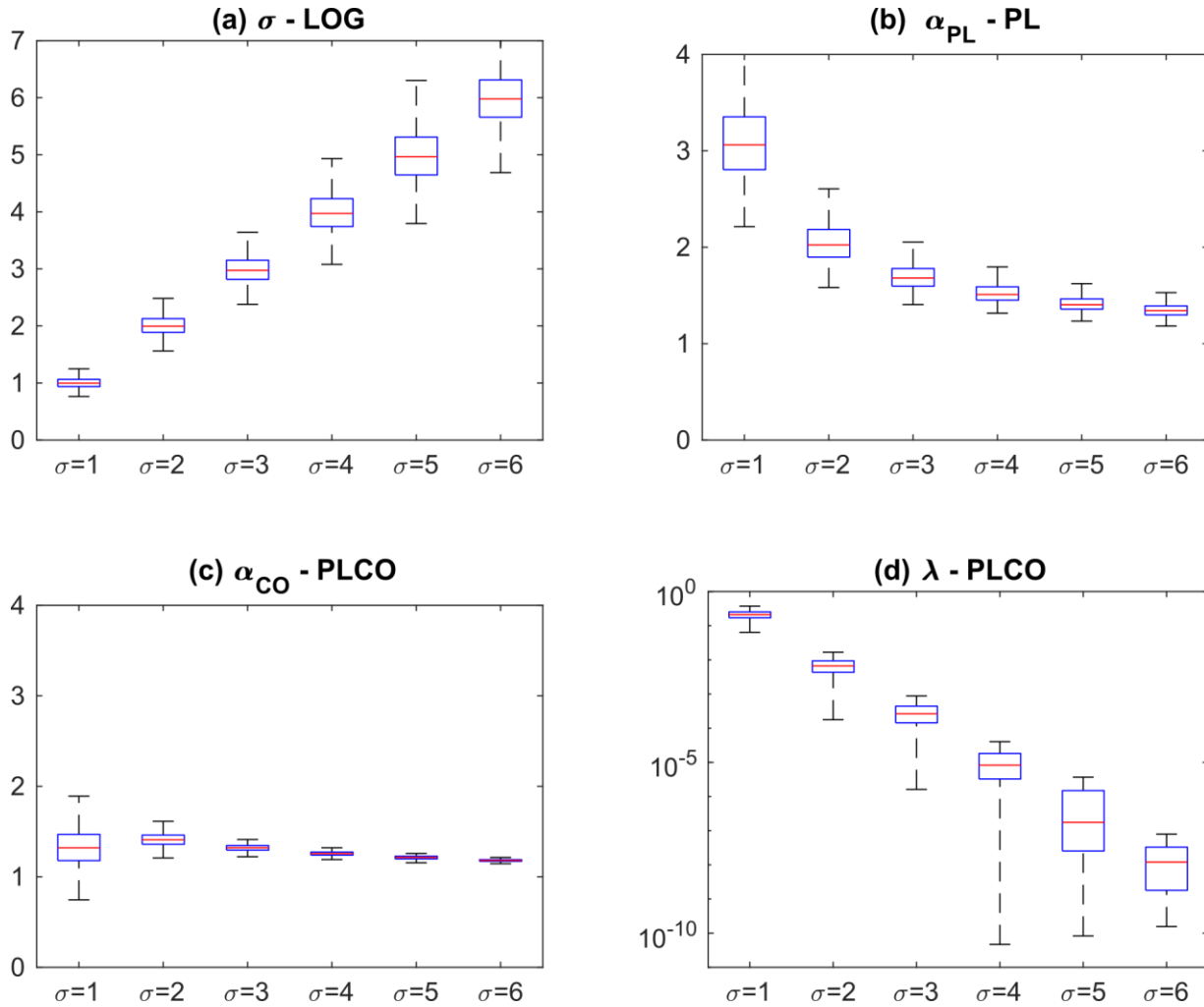


Figure 6 Comparison of estimated parameters fitting LOG distributed random numbers with increasing  $\sigma$ . Note that the slope of PL, the power-law model without cutoff ( $\alpha'_{PL}$ ) converges to a minimum value close to  $\alpha_{PL} \approx 1.5$ , while the slope of PLCO, the power-law model with cutoff ( $\alpha'_{CO}$ ) is a constant. In the PLCO, the tailing is measured by the cutoff rate  $\lambda$ .

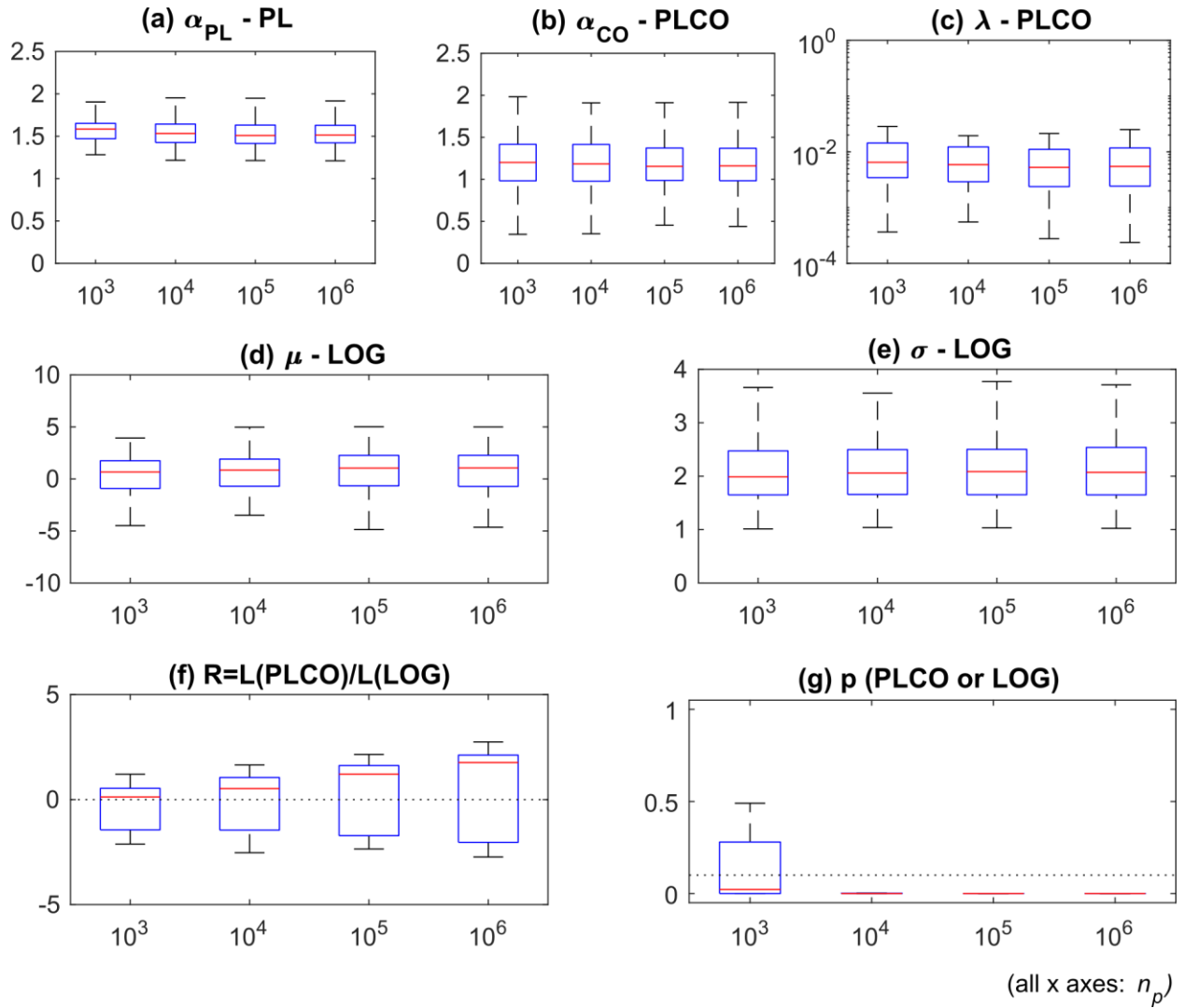


Figure 7 (a-e) Estimated parameters for the different tested models; (f,g) R and p-values comparing the PLCO and LOG models.

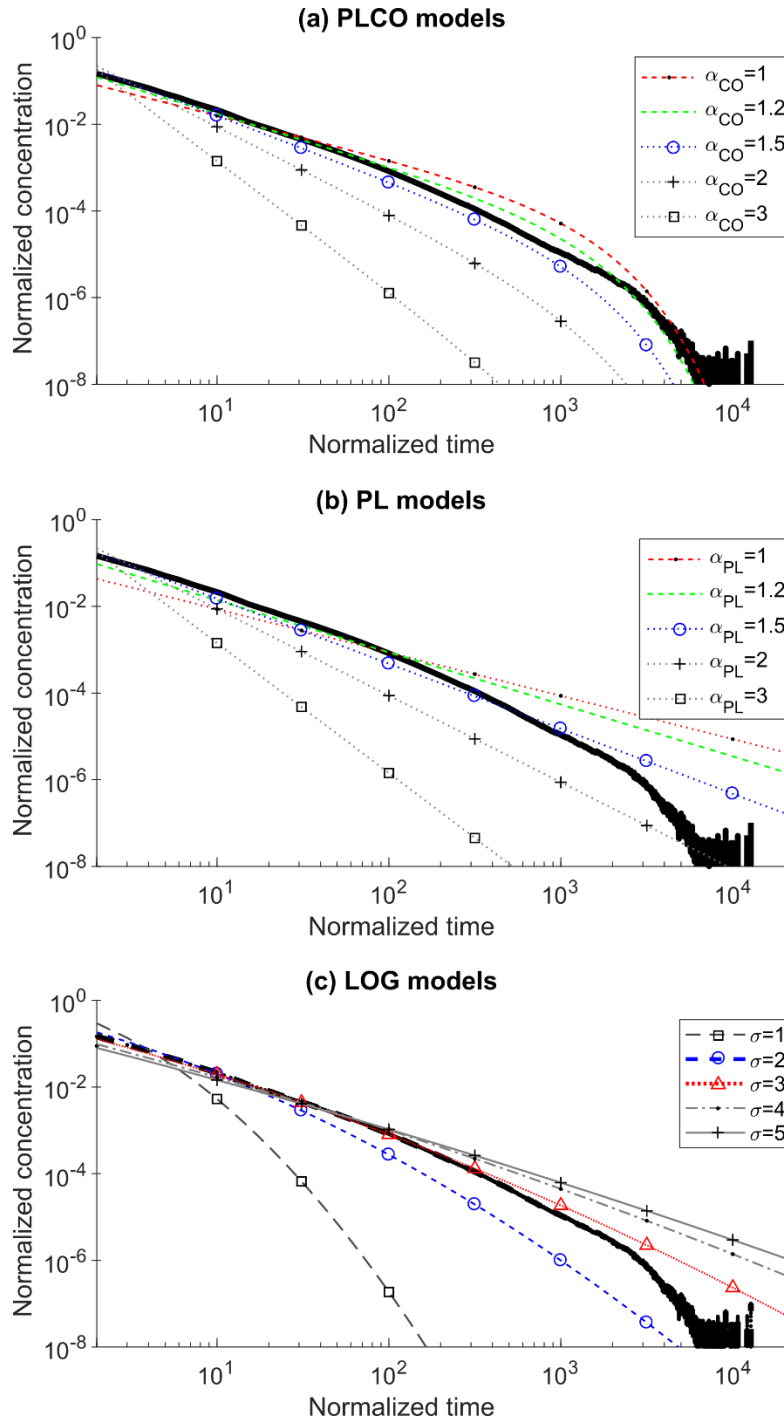


Figure 8 (a) Predictions made by PLCO model embedding variable slopes  $\alpha_{CO}$  and a cutoff rate  $\lambda_H$  estimated from the ensemble of maximum difference in Lagrangian velocities at the source. (b) Predictions made by a PL embedding variable slopes  $\alpha_{PL}$ . (c) Predictions made by a LOG model with  $\mu = 0$  and variable  $\sigma$ . The black thick line represents the reference ensemble curve.

## References

- Bauke, H. (2007). Parameter estimation for power-law distributions by maximum likelihood methods. *Eur. Phys. J. B* 58, 167–173.
- Becker, M., and Shapiro, A.M. (2003). Interpreting tracer breakthrough tailing from different forced-gradient tracer experiment configurations in fractured bedrock. *Water Resour. Res.* 39, W1014.
- Benson, D.A., Schumer, R., Meerschaert, M.M., and Wheatcraft, S.W. (2001). Fractional Dispersion, Lévy Motion, and the MADE Tracer Tests. In *Dispersion in Heterogeneous Geological Formations*, (Springer, Dordrecht), pp. 211–240.
- Bianchi, M., and Pedretti, D. (2017). Geological entropy and solute transport in heterogeneous porous media. *Water Resour. Res.* 53.
- Bianchi, M., and Zheng, C. (2016). A lithofacies approach for modeling non-Fickian solute transport in a heterogeneous alluvial aquifer. *Water Resour. Res.* 52, 552–565.
- Bianchi, M., Zheng, C., Tick, G.R., and Gorelick, S.M. (2011). Investigation of Small-Scale Preferential Flow with a Forced-Gradient Tracer Test. *Ground Water* 49, 503–514.
- Bijeljic, B., Mostaghimi, P., and Blunt, M.J. (2011). Signature of Non-Fickian Solute Transport in Complex Heterogeneous Porous Media. *Phys. Rev. Lett.* 107, 204502.
- Blackwood, L.G. (1992). The lognormal distribution, environmental data, and radiological monitoring. *Environ. Monit. Assess.* 21, 193–210.
- Bohling, G.C., Liu, G., Dietrich, P., and Butler, J.J. (2016). Reassessing the MADE direct-push hydraulic conductivity data using a revised calibration procedure. *Water Resour. Res.* 52, 8970–8985.
- BRGM (1990). *Operations de multitracages en écoulement influence par pompage pour la détermination des paramètres hydrodynamiques de l'aquifère superficiel du site d'El Cabril (Espagne)*. Univ. Politecnica de Catalunya, Barcelona.
- Carle, S.F., and Fogg, G.E. (1997). Modeling Spatial Variability with One and Multidimensional Continuous-Lag Markov Chains. *Math. Geol.* 29, 891–918.
- Carrera, J., Sanchez-Vila, X., Benet, I., Medina, A., Galarza, G., and Guimera, J. (1998). On matrix diffusion: formulations, solution methods and qualitative effects. *Hydrogeol J* 6, 178–190.
- Clauset, A., Shalizi, C.R., and Newman, M.E.J. (2009). Power-law distribution in empirical data. *SIAM Rev.* 51, 661–703.
- Crow, E., and Shimizu, K. (1988). *Lognormal Distributions: Theory and Applications* (Dekker, New York).
- Dentz, M., and Berkowitz, B. (2003). Transport behavior of a passive solute in continuous time random walks and multirate mass transfer. *Water Resour. Res.* 39, 1111–1131.
- Dentz, M., and Carrera, J. (2007). Mixing and spreading in stratified flow. *Phys. Fluids* 19, 017107.



- Dentz, M., Cortis, A., Scher, H., and Berkowitz, B. (2004). Time behavior of solute transport in heterogeneous media: transition from anomalous to normal transport. *Adv. Water Resour.* 27, 155–173.
- Dreuzy, J.-R. de, and Carrera, J. (2016). On the validity of effective formulations for transport through heterogeneous porous media. *Hydrol. Earth Syst. Sci.* 20, 1319–1330.
- Ederly, Y., Guadagnini, A., Scher, H., and Berkowitz, B. (2014). Origins of anomalous transport in heterogeneous media: Structural and dynamic controls. *Water Resour. Res.* 50, 1490–1505.
- Farrell, J., and Reinhard, M. (1994). Desorption of halogenated organics from model solids, sediments, and soil under unsaturated conditions. 2. Kinetics. *Environ. Sci. Technol.* 28, 63–72.
- Fiori, A. (2014). Channeling, channel density and mass recovery in aquifer transport, with application to the MADE experiment. *Water Resour. Res.* n/a-n/a.
- Fiori, A., and Becker, M.W. (2015). Power law breakthrough curve tailing in a fracture: The role of advection. *J. Hydrol.* 525, 706–710.
- Fiori, A., Janković, I., Dagan, G., and Cvetković, V. (2007). Ergodic transport through aquifers of non-Gaussian log conductivity distribution and occurrence of anomalous behavior. *Water Resour. Res.* 43.
- Fiori, A., Zarlenga, A., Gotovac, H., Jankovic, I., Volpi, E., Cvetkovic, V., and Dagan, G. (2015). Advective transport in heterogeneous aquifers: Are proxy models predictive? *Water Resour. Res.* 51, 9577–9594.
- Flach, G. (2012). Relationship between dual-domain parameters and practical characterization data. *Ground Water* 50, 216–229.
- Gelhar, L.W., Gutjahr, A.L., and Naff, R.L. (1979). Stochastic analysis of macrodispersion in a stratified aquifer. *Water Resour. Res.* 15, 1387–1397.
- Goldstein, M.L., Morris, S.A., and Yen, G.G. (2004). Problems with fitting to the power-law distribution. *Eur. Phys. J. B - Condens. Matter Complex Syst.* 41, 255–258.
- Gouze, P., Le Borgne, T., Leprovost, R., Lods, G., Poidras, T., and Pezard, P. (2008). Non-Fickian dispersion in porous media: 1. Multiscale measurements using single-well injection withdrawal tracer tests. *Water Resour. Res.* 44, W06426.
- Hadermann, J., and Heer, W. (1996). The Grimsel (Switzerland) migration experiment: integrating field experiments, laboratory investigations and modelling. *J. Contam. Hydrol.* 21, 87–100.
- Haggerty, R., and Reeves, P.C. (2002). STAMMT-L 1.0, Formulation and User's Guide, Technical Report ERMS 520308 (Sandia National Laboratories, Albuquerque, NM, USA).
- Haggerty, R., McKenna, S.A., and Meigs, L.C. (2000). On the late-time behavior of tracer test breakthrough curves. *Water Resour Res* 36, 3467–3479.
- Haggerty, R., Harvey, C.F., Freiherr von Schwerin, C., and Meigs, L.C. (2004). What controls the apparent timescale of solute mass transfer in aquifers and soils? A comparison of experimental results. *Water Resour. Res.* 40, W01510.

Harbaugh, A.W., Banta, E.R., Hill, M.C., and McDonald, M.G. (2000). MODFLOW-2000, the U.S. Geological Survey modular ground-water model – User guide to modularization concepts and the Ground-Water Flow Process (U.S. Geological Survey Open-File Report 00-92, 121 p).

Kelly, J.F., Bolster, D., Meerschaert, M.M., Drummond, J.D., and Packman, A.I. (2017). FracFit: A robust parameter estimation tool for fractional calculus models. *Water Resour. Res.* 53, 2559–2567.

LaBolle, E.M., and Fogg, G.E. (2001). Role of Molecular Diffusion in Contaminant Migration and Recovery in an Alluvial Aquifer System. *Transp. Porous Media* 42, 155–179.

Le Borgne, T., Dentz, M., and Carrera, J. (2008). Spatial Markov processes for modeling Lagrangian particle dynamics in heterogeneous porous media. *Phys. Rev. E* 78, 026308.

Luo, J., Dentz, M., Cirpka, O.A., and Kitanidis, P.K. (2007). Breakthrough curve tailing in a dipole flow field. *Water Resour. Res.* 43.

McKenna, S., Meigs, L., and Haggerty, R. (2001). Tracer tests in a fractured dolomite. 3. Double-porosity, multiple-rate mass transfer processes in convergent flow tracer tests. *Water Resour Res* 37, 1143–1154.

Mehdinejadi, B. (2017). Estimating the solute transport parameters of the spatial fractional advection-dispersion equation using Bees Algorithm. *J. Contam. Hydrol.*

Mitzenmacher, M. (2004). A Brief History of Generative Models for Power Law and Lognormal Distributions. *Internet Math.* 1, 226–251.

Neuman, S.P., and Tartakovsky, D.M. (2009). Perspective on theories of non-Fickian transport in heterogeneous media. *Adv. Water Resour.* 32, 670–680.

Pedit, J.A., and Miller, C.T. (1994). Heterogeneous sorption processes in subsurface systems. 1. Model formulations and applications. *Environ. Sci. Technol.* 28, 2094–2104.

Pedretti, D., and Fernández-García, D. (2013). An automatic locally-adaptive method to estimate heavily-tailed breakthrough curves from particle distributions. *Adv. Water Resour.* 59, 52–65.

Pedretti, D., and Fiori, A. (2013). Travel time distributions under convergent radial flow in heterogeneous formations: Insight from the analytical solution of a stratified model. *Adv. Water Resour.* 60, 100–109.

Pedretti, D., Fernández-García, D., Bolster, D., and Sanchez-Vila, X. (2013). On the formation of breakthrough curves tailing during convergent flow tracer tests in three-dimensional heterogeneous aquifers. *Water Resour. Res.* 49, 4157–4173.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (1992). *Numerical Recipes in C (2Nd Ed.): The Art of Scientific Computing* (New York, NY, USA: Cambridge University Press).

Reed, W.J. (2001). The Pareto, Zipf and other power laws. *Econ. Lett.* 74, 15–19.

Salamon, P., Fernández-García, D., and Gomez-Hernandez, J. (2006). Modeling mass transfer processes using random walk particle tracking. *Water Resour Res* 42, W11417.

Sanchez-Vila, X., and Carrera, J. (2004). On the striking similarity between the moments of breakthrough curves for a heterogeneous medium and a homogeneous medium with a matrix diffusion term. *J. Hydrol.* *294*, 164/175.

Wang, L., and Cardenas, M.B. (2017). Transition from non-Fickian to Fickian longitudinal transport through 3-D rough fractures: Scale-(in)sensitivity and roughness dependence. *J. Contam. Hydrol.* *198*, 1–10.

Willmann, M., Carrera, J., and Sanchez-Vila, X. (2008). Transport upscaling in heterogeneous aquifers: What physical parameters control memory functions? *Water Resour Res* *44*, W12437.

Zhang, Y., Green, C.T., and Fogg, G.E. (2013). The impact of medium architecture of alluvial settings on non-Fickian transport. *Adv. Water Resour.* *54*, 78–99.

Zhang, Y., Green, C.T., and Baeumer, B. (2014). Linking aquifer spatial properties and non-Fickian transport in mobile-immobile like alluvial settings. *J. Hydrol.* *512*, 315–331.

Zheng, C., Bianchi, M., and Gorelick, S.M. (2011). Lessons Learned from 25 Years of Research at the MADE Site. *Ground Water* *49*, 649–662.