Advances in Water Resources 113 (2018) 236–248

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

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

Daniele Pedretti[a,*], Marco Bianchi[b]

[a] Geological Survey of Finland (GTK), Espoo FI-02151, Finland
[b] British Geological Survey (BGS), Keyworth, Nottingham NG12 5GG, United Kingdom

ARTICLE INFO

Keywords:
Upscaling
Power law
Tracer tests
Tailing
Anomalous transport
MADE site

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 renders power laws 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 considered. It is found that a power law model without cutoff (PL) results in best-fitted exponents (α_{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 (α_{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 (λ) 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 λ. 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.

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.,

* Corresponding author.
E-mail address: daniele.pedretti@gtk.fi (D. Pedretti).

https://doi.org/10.1016/j.advwatres.2018.01.023
Received 18 November 2017; Received in revised form 23 January 2018; Accepted 23 January 2018
Available online 31 January 2018

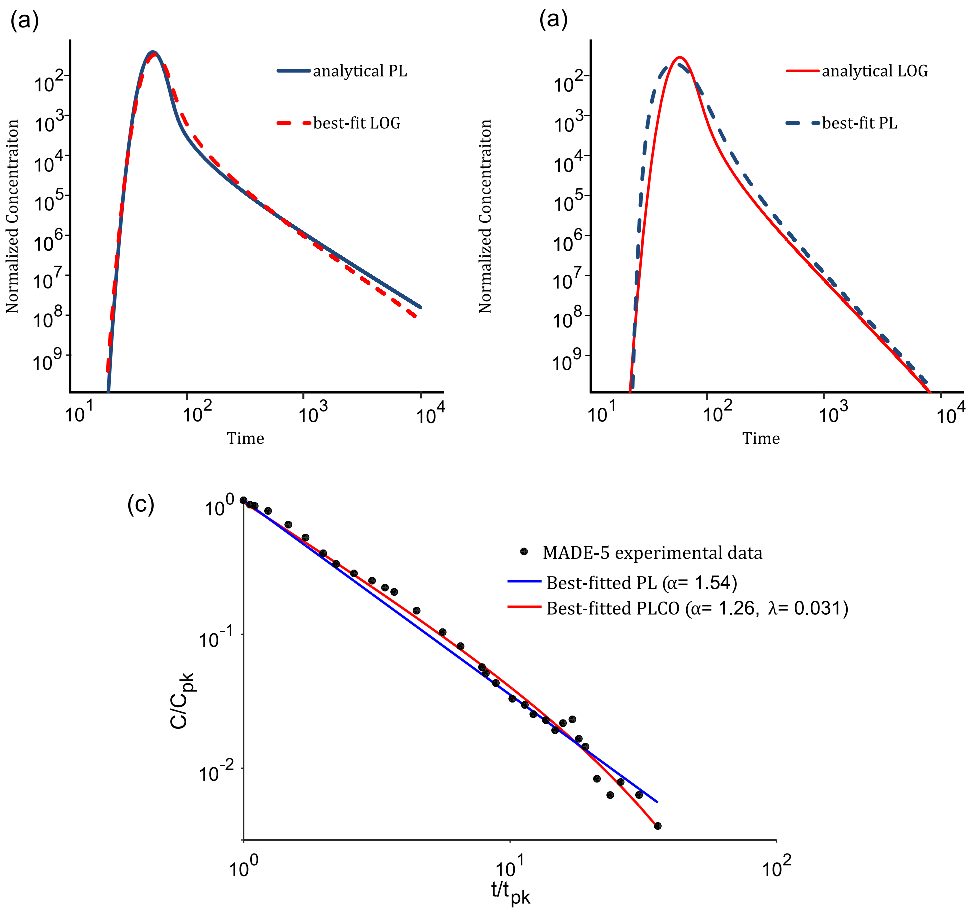

Fig. 1. (a, b) Examples of heavily tailed BTCs analytically generated and numerically fitted using the code STAMMT-L (Haggerty and Reeves, 2002). In (a), the analytical curve adopts a PL distribution of mass transfer times with $\alpha_{PL} = 2$, while the fitted curve is obtained using a LOG distribution of mass transfer rates. In (b), the reference curve adopts a LOG model with scaling factor σ = 3, while the fitted curve is obtained using a PL model. (c) Experimental results from a dipole 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of thisarticle.)

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 (Fig. 1). The power law exponent or slope (α) has been observed to vary between α ≈ 1 and α ≈ 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 α = 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 $\ln K \to 0$, the expected scaling of a travel time distribution at late time is also a power-law function, with α linked to the slope of the ln*K* distribution. Zhang et al., (2014) showed that α 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 Fig. 1. The formulation of all models used in this example are reported in Table 1, and presented in detail in next sections.

In Fig. 1a, a reference BTC is generated using the power law model