Using Probability Distribution Functions to Correlate Breakthrough Curves of Water and air Contaminants

Khim Hoong Chu (✉ khimchu@gmail.com)

Honeychem Research <https://orcid.org/0000-0003-4012-0667>

Research Article

Keywords: asymmetric breakthrough curve, fixed bed adsorption, inflection point, probability distribution

Posted Date: May 4th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-441731/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Using probability distribution functions to correlate breakthrough curves of water and air contaminants

Khim Hoong Chu*

Honeychem Research, Auckland 1142, New Zealand

* Corresponding author. E-mail address: khimchu@gmail.com

Abstract

This paper reports the use of five probability cumulative distribution functions (normal, log-normal, logistic, Gompertz, and Weibull) to correlate published breakthrough data of water and air contaminants (ciprofloxacin, ammonium, hydrogen chloride, and hydrogen sulfide). Because the shape of the ciprofloxacin breakthrough curve is fairly symmetric, it is well correlated by all five functions ($R^2 > 0.99$). They also provide a good representation of the overall shape of the ammonium breakthrough curve ($R^2 > 0.99$). However, none can describe the leakage of ammonium during the initial period of column operation. The log-normal and Weibull functions give an excellent representation of the tailing HCl data while the normal, logistic, and Gompertz functions are quite poor. This difference in performance can be explained by the different characteristics of their inflection points. The log-normal and Weibull functions have a floating inflection point, which gives them flexibility in tracing the shape of the tailing data. The invariant inflection points of the normal, logistic, and Gompertz curves restrict their data fitting ability. Only the log-normal function can provide a reasonable fit to the H₂S data with strong tailing. It is shown that the invariant inflection point of a probability function can be converted to a floating one. A version of the Gompertz function so modified provides a good quantitative correlation of the tailing data of H₂S ($R^2 = 0.99$).

Keywords: asymmetric breakthrough curve; fixed bed adsorption; inflection point; probability distribution

Introduction

Fixed bed adsorption of water and air contaminants is the subject of this article. In this field of research, the bulk of experimental results collected in the form of breakthrough data are invariably processed using mathematical models. Mechanism-driven models are useful for exploring the fundamental aspects of fixed bed adsorption. The parameters of these models are often calibrated using information obtained from independent sources and/or engineering correlations. Breakthrough data are not used for parameter estimation; they are instead used to validate the predictive ability of the model. A variety of such models formulated at different levels of complexity have been tested in the environmental adsorption literature (Danish et al. 2021; Mirzaie et al. 2021; Roy et al. 2021). Because these models are cast in the form of coupled partial differential equations, specialized software and programming code are required for their numerical solution.

An alternative to mechanism-driven models is the so-called data-driven models. The terms “big data” and “data science” are becoming pervasive, affecting many aspects of water science and engineering (Hering 2019; Newhart et al. 2019). Large amounts of data are being used to develop optimization and process control strategies. In the context of fixed bed adsorption, “small data” sets are used to construct empirical models. Commonly used methods to build empirical models from breakthrough data include the design of experiments and neural network modeling approaches (Dalhat et al. 2021; Schio et al. 2021). These data-driven models may be used for process design or optimization, but they lack mechanistic merit and should not be used to predict breakthrough behavior outside the range of experimental conditions used for model calibration.

Another data-driven modeling approach involves simple phenomenological models of fixed bed adsorption, with the Bohart-Adams and Thomas models being the most popular ones (Bohart and Adams 1920; Thomas 1944). These two models consider adsorption equilibrium but their rate mechanisms are based on simple reaction kinetics. Although unrealistic, the assumption of reaction kinetics as the rate-limiting step allows analytical solutions to be derived. These relatively simple analytical expressions permit modeling work to be conducted using standard spreadsheet programs. We classify the Bohart-Adams and Thomas equations as data-driven models because breakthrough data are used to estimate their parameters. Another popular data-driven model is the Yoon-Nelson equation (Yoon and Nelson 1984), which often appears together with the Bohart-Adams and Thomas models in modeling studies. The Yoon-Nelson model is highly empirical since it is devoid of typical fixed bed variables such as flow rate and bed length.

The literature of environmental adsorption is replete with articles reporting the modeling of breakthrough data using the Bohart-Adams, Thomas, and Yoon-Nelson models. Unfortunately, the majority of these studies are not meaningful because the three models are analogous to one another and should not be treated separately (Chatterjee and Schiewer 2011; Chu 2020; Lee et al. 2015). Another shortcoming of the three models is that they cannot fit typical breakthrough curves of water and air contaminants which are mostly asymmetric. From a practical perspective, this is a serious limitation because a poorly fitted model will produce grossly inaccurate estimates of breakthrough and exhaustion times. Limited efforts have been made to enhance the data fitting ability of the three models, as summarized by Apiratikul and Chu (2021). To enlarge the pool of models capable of representing asymmetric breakthrough curves, this paper evaluates the feasibility of adopting probability distribution functions as breakthrough curve models.

Probability cumulative distribution functions are attractive because they are capable of producing sigmoid or S-shaped curves, which are similar to the shapes of contaminant breakthrough curves. Also, straightforward data analysis by spreadsheet calculation is possible because most of the probability functions have simple mathematical forms. As a preliminary investigation we have selected five probability functions for testing: normal, log-normal, logistic, Gompertz, and Weibull. Previous research in this area was restricted to evaluating some of these functions separately (Chu 2020b, 2021; Dima et al. 2020). This paper is the first systematic attempt to compare a set of probability functions using previously published breakthrough data of ciprofloxacin, ammonium, hydrogen chloride, and hydrogen sulfide.

Model development

Normal distribution

This distribution, sometimes referred to as Gaussian distribution, occurs throughout statistics. A particular normal distribution is completely determined by its mean and standard deviation. The formula for the normal cumulative distribution function, formulated to describe fixed bed adsorption, may be written

$$\frac{c}{c_0} = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{t - a_1}{b_1 \sqrt{2}} \right) \right] \quad (1)$$

where c is the effluent concentration at any time t , c_0 is the feed concentration, and a_1 and b_1 are respectively the mean and standard deviation of the normal function.

The normal function predicts a sigmoid curve, whose shape is determined by its inflection point. An inflection point is where a curve changes concavity, that is, it is a point at which a curve goes from concave up to concave down, or vice versa. For the normal function, the location of its inflection point can be determined by finding at what time the second derivative of the function is zero. The second derivative of Eq. 1 is given by

$$\frac{\partial^2 (c/c_0)}{\partial t^2} = \frac{(a_1 - t) \exp\left[-(a_1 - t)^2 / 2b_1^2\right]}{b_1^3 \sqrt{2\pi}} \quad (2)$$

By equating the preceding expression to zero and solving the resulting equation for the t -term, we obtain $t = a_1$ at the inflection point. Substitution of this t value in Eq. 1 gives

$$\frac{c}{c_0} = \frac{1}{2} \quad (3)$$

The inflection point of the normal function is therefore located at the mid-point of the sigmoid curve.

Log-normal distribution

This function differs from the normal function in one aspect: the t -term in Eq. 1 is replaced by a $\ln(t)$ -term. The formula reads

$$\frac{c}{c_0} = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\ln(t) - a_2}{b_2 \sqrt{2}} \right) \right] \quad (4)$$

where a_2 and b_2 are fitting parameters.

The second derivative of Eq. 4 is given by

$$\frac{\partial^2 (c/c_0)}{\partial t^2} = \frac{[a_2 - b_2^2 - \ln(t)] \exp\left\{-[a_2 - \ln(t)]^2 / 2b_2^2\right\}}{b_2^3 t^2 \sqrt{2\pi}} \quad (5)$$

Setting the preceding equation to zero and solving for $\ln(t)$ leads to $\ln(t) = a_2 - b_2^2$.

Substitution of the last result in Eq. 4 gives

$$\frac{c}{c_0} = \frac{1}{2} \left[1 - \operatorname{erf} \left(\frac{b_2}{\sqrt{2}} \right) \right] \quad (6)$$

The preceding expression indicates that the location of the inflection point varies with the b_2 -term; the log-normal function therefore has a floating inflection point.

It should be mentioned that converting the t -term in the normal function to the $\ln(t)$ -term in the log-normal function requires some mathematical manipulation. Since one can take

the logarithm only of dimensionless numbers, the t -term, which is a dimensioned quantity, must first be made dimensionless. Here, we describe how this can be done. First, we assign a unit to the t -term in Eq. 1, say, min. Next, we multiply the t -term by t^*/t^* where $t^* = 1$ min. We now have $t^*(t/t^*)$ with the bracketed term being a dimensionless quantity. Finally, we take the logarithm of the bracketed term, obtaining $t^*\ln(t/t^*)$. So, Eq. 4 becomes

$$\frac{c}{c_0} = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{t^* \ln(t/t^*) - a_2}{b_2 \sqrt{2}} \right) \right] \quad (7)$$

The preceding expression is numerically analogous to the original Eq. 4 since the two t^* -terms are equal to unity. For simplicity, they may be omitted from Eq. 7. Then, Eq. 7 reverts to Eq. 4, but the logarithmic term is now dimensionless. Likewise, the a_2 -term should also be a dimensionless parameter.

Logistic distribution

This function has been used extensively for growth models. It is written as

$$\frac{c}{c_0} = \frac{1}{1 + \exp(a_3 - b_3 t)} \quad (8)$$

where a_3 and b_3 are fitting parameters.

To find its inflection point, we equate the second derivative of the logistic function to zero and solve the resulting equation for the t -term.

$$\frac{\partial^2 (c/c_0)}{\partial t^2} = \frac{2b_3^2 \exp(2a_3 - 2b_3 t)}{[1 + \exp(a_3 - b_3 t)]^3} - \frac{b_3^2 \exp(a_3 - b_3 t)}{[1 + \exp(a_3 - b_3 t)]^2} = 0 \quad (9)$$

From Eq. 9, $t = a_3/b_3$. Substitution of the last result in Eq. 8 yields

$$\frac{c}{c_0} = \frac{1}{2} \quad (10)$$

We see that the inflection point is located at the mid-point and is identical to that of the normal function.

Gompertz distribution

Whilst this function was originally developed to graduate human mortality rates, it is now used to model sigmoid curves across many areas of scientific research. A typical formula is written as

$$\frac{c}{c_0} = \exp\left[-\exp(a_4 - b_4 t)\right] \quad (11)$$

where a_4 and b_4 are parameters to be estimated.

As before, to find the location of its inflection point, the Gompertz function is differentiated twice with respect to the t -term.

$$\frac{\partial^2 (c/c_0)}{\partial t^2} = b_4^2 \left[\exp(a_4 - b_4 t) - 1 \right] \exp\left[-\exp(a_4 - b_4 t) + a_4 - b_4 t\right] \quad (12)$$

Setting the preceding result to zero and solving the resulting equation for the t -term, we get $t = a_4/b_4$. Substitution of the last result in Eq. 11 gives

$$\frac{c}{c_0} = \exp(-1) \quad (13)$$

The location of the inflection point is fixed at $c/c_0 = 0.37$.

Weibull distribution

This function is one of the most widely used lifetime distributions in reliability engineering. There is a large collection of probabilistic and stochastic models derived from the Weibull function. A commonly used two-parameter formula reads

$$\frac{c}{c_0} = 1 - \exp\left[-\left(\frac{t}{a_5}\right)^{b_5}\right] \quad (14)$$

where a_5 and b_5 are adjustable parameters.

The second derivative of the preceding expression is given by

$$\frac{\partial^2 (c/c_0)}{\partial t^2} = \frac{\exp\left[-(t/a_5)^{b_5}\right] b_5 (t/a_5)^{b_5} \left[b_5 - 1 - b_5 (t/a_5)^{b_5} \right]}{t^2} \quad (15)$$

Equating the second derivative to zero and rearranging the resulting equation gives

$$t = a_5 \left(\frac{b_5 - 1}{b_5} \right)^{1/b_5} \quad (16)$$

Substitution of the preceding result in Eq. 14 yields

$$\frac{c}{c_0} = 1 - \exp \left(- \frac{b_5 - 1}{b_5} \right) \quad (17)$$

The Weibull function has a floating inflection point because its location varies with the b_5 -term.

Nonlinear least-squares regression

Nonlinear least-squares regression was applied to the probability functions tested in this work to estimate their free parameters. The overall fit of a function was measured with the coefficient of determination R^2 and residual root mean square error (RRMSE). It is not necessary to use other statistical indicators such as the Akaike Information Criterion for model comparison because all the models examined here contain the same number of fitting parameters. The R^2 and RRMSE metrics are given by

$$R^2 = 1 - \frac{\sum_{j=1}^w (M_j - E_j)^2}{\sum_{j=1}^w (M_j - E_m)^2} \quad (18)$$

$$\text{RRMSE} = \sqrt{\frac{\sum_{j=1}^w (M_j - E_j)^2}{w - p}} \quad (19)$$

where w is the number of data points, M_j are model values of c/c_0 , E_j are experimental values of c/c_0 , E_m is the mean of experimental values, and p is the number of adjustable parameters. Note that the R^2 value is traditionally defined only for linear relationships. Nonetheless it is commonly used as an informal measure of the goodness of a nonlinear fit.

Result and discussion

Four sets of published breakthrough data of water and air contaminants are used to test and compare the data fitting abilities of the five probability functions described above. To provide a rigorous comparison, the data sets were carefully selected to reflect different curve characteristics. The first data set, reported by Sausen et al. (2018), exhibits a relatively symmetric curve shape. The second one, taken from the work of Nguyen et al. (2019), is characterized by a leakage of contaminant at the column exit. The third and fourth data sets, reported by Papurello et al. (2019), demonstrate different degrees of tailing behavior.

Case 1: breakthrough curve of ciprofloxacin

In this case study (Sausen et al. 2018), a series of column experiments were performed to measure the breakthrough curves of ciprofloxacin as functions of bed height and flow rate. Ciprofloxacin, a widely used antibiotic, has been detected in wastewater treatment plant effluents and surface waters. The fixed bed experiments were conducted in a glass column with an internal diameter of 1 cm and a length of 30 cm. The column was packed with a commercial cation exchanger with an average size of 0.653 mm. Figure 1 shows the selected data set, obtained with the following experimental conditions: feed concentration = 100 mg/L, flow rate = 5 cm³/min, and packed bed length = 7.6 cm.

Figure 1 plots the fitted curves of the normal, log-normal, logistic, Gompertz, and Weibull functions, computed using the best-fit parameters listed in Table 1, case 1. Visual inspection of Fig. 1 indicates that the measured data are well represented by the five functions. The values of R^2 (>0.99) and RRMSE (0.02-0.04) for the five fits, also shown in Fig. 1, support this observation. The Weibull, logistic, and normal functions are similar in performance and slightly better than the log-normal and Gompertz functions. Because of the symmetric nature of the measured curve, probability functions with an inflection point close to the mid-point are expected to do well. The inflection points of the logistic and normal functions are located at

the mid-point. The inflection points of the log-normal and Weibull functions, computed from Eqs. 6 and 17, are located at $c/c_0 = 0.41$ and $c/c_0 = 0.55$ respectively. The location of the Gompertz function's inflection point is given by $c/c_0 = 0.37$. It is thus not surprising that the normal, logistic, Weibull, and log-normal functions can handle symmetric data of this type rather well. The inferior performance of the Gompertz function can be explained by the fact that its inflection point is farthest from the mid-point. It is conclusive that the slightly asymmetric data set of Fig. 1 poses no great difficulty to the five functions.

Fig. 1 Probability function fits compared to observed ciprofloxacin breakthrough data reported by Sausen et al. (2018)

Table 1 Parameter estimates for the nonlinear fits in Figs. 1-4

Function	Parameter	Case 1	Case 2	Case 3	Case 4
Normal	a_1 (min); b_1	2071; 470.4	1437; 188.7	23.02; 18.6	78.39; 47.0
Log-normal	a_2 ; b_2	7.63; 0.23	7.27; 0.13	2.99; 0.95	4.28; 0.65
Logistic	a_3 ; b_3 (min ⁻¹)	7.39; 0.004	12.78; 0.01	2.03; 0.09	2.76; 0.04
Gompertz	a_4 ; b_4 (min ⁻¹)	4.57; 0.002	8.75; 0.01	0.99; 0.06	1.53; 0.03
Weibull	a_5 (min); b_5	2234; 5.19	1508; 8.69	29.32; 1.13	96.06; 1.65

Case 2: breakthrough curve of ammonium

This case reports the removal of ammonium by corncob biochar which was packed in a glass column with an internal diameter of 1.5 cm and a total column length of 50 cm (Nguyen et al. 2019). The effects of initial concentration, flow rate, and packed bed length on the breakthrough characteristics of ammonium were investigated. The selected breakthrough data set was obtained with a feed concentration of 10 mg/L, a flow rate of 1 cm³/min, and a packed length of 8 cm. Figure 2 presents the measured data and plots the fits of the five functions. The fit statistics and parameter estimates are presented respectively in Fig. 2 and Table 1, case 2. A visual comparison of the plots in Fig. 2 highlights the fact that there is little difference in the performances of the five functions. The second observation that jumps out is that the initial section of the breakthrough curve is poorly represented by the five functions. However, the R^2 scores for the five fits are very impressive (> 0.99 , see Fig. 2) because most of the data points above the initial stages of breakthrough are well tracked by the five functions. Overall fit indicators such as the R^2 metric do not always tell the whole picture. In assessing goodness of fit, it is inappropriate to trumpet high R^2 scores without first examining a graph of the data together with the model fit.

240

241 Fig. 2 Probability function fits compared to observed ammonium breakthrough data reported
 242 by Nguyen et al. (2019)

243

244 Figure 2 reveals the presence of a phenomenon known as fronting or leakage, which
 245 refers to the appearance of a noticeable effluent concentration level soon after initiation of
 246 column operation. This nonzero effluent concentration level remains relatively stable for a
 247 period of time before the breakthrough curve rises rapidly. If the desired breakthrough
 248 concentration is comparable to the nonzero effluent concentration level, the leakage
 249 phenomenon will lead to a premature breakthrough time, and consequently resulting in severe
 250 underutilization of the column capacity. There are surely several reasons for the occurrence of
 251 leakage, with nonuniform flow being one. The problem of nonuniform flow is most likely due
 252 to poor column packing. The use of adsorbents with irregular particle shape or compressible

adsorbent material is rather common in this field of research. It is not easy to pack a column well with adsorbents of this type. Since fronting breakthrough curves are undesirable, there is little point in developing models that can track such curve shapes. Efforts should instead focus on eliminating fronting breakthrough curves by optimizing the experimental conditions and equipment.

Case 3: breakthrough curve of hydrogen chloride

This example, taken from the work of Papurello et al. (2019), reports the breakthrough characteristics of a simulated biogas stream containing hydrogen chloride and hydrogen sulfide in glass filters packed with two types of commercial activated carbon (Norit RST3 and Airdep Carbox) and a biochar material produced by gasification of wood. The adsorption experiments were conducted using adsorbent particles in the size range 0.1-0.18 mm and a gas hourly space velocity of 1636 h^{-1} . Feed concentrations of HCl in the range 10-377 ppm(v) and of H₂S in the range 10-200 ppm(v) and were used.

Figure 3 presents a set of HCl breakthrough data obtained with a biochar filter subjected to a feed concentration of 377 ppm(v). Also shown in Fig. 3 are the associated function fits, computed using the parameter estimates presented in Table 1, case 3. The observed breakthrough curve depicted in Fig. 3 is different from those of Figs. 1 and 2. It is asymmetric in shape, which is characterized by an immediate breakthrough of HCl, followed by a sharp rise in exit concentration, and finally a long period of filter saturation. Curve shape of this type is known as tailing. Note that the initial section of the experimental profile exhibits a very slight curve curvature, which makes the breakthrough curve look like a hyperbolic curve instead of a sigmoid curve.

By visual comparison of the plots in Fig. 3, we see that the five functions can be classified into two groups. The first group consists of the log-normal and Weibull functions while the remaining three functions, Gompertz, logistic, and normal, make up the second group.

As Fig. 3 shows, the asymmetric shape of the observed breakthrough curve is well represented by the first group, with reported $R^2 > 0.99$ and RRMSE < 0.03 . By contrast, the agreement between the observed data and the functions of the second group is unsatisfactory, with reported $R^2 < 0.99$.

Fig. 3 Probability function fits compared to observed HCl breakthrough data reported by Papurello et al. (2019)

The difference in performance between the two groups can be explained by the fact that the two functions in the first group have a floating inflection point while the other three in the second group have an invariant one. The inflection points of the log-normal and Weibull functions, computed from Eqs. 6 and 17 using the b -term values given in Table 1, are located respectively at $c/c_0 = 0.17$ and 0.11 . To fit the tailing breakthrough curve depicted in Fig. 3, it

is necessary to shift the inflection point away from the mid-point and toward the 0,0 origin. In the case of the log-normal and Weibull functions, this can be done by optimizing their b -term vales during the nonlinear regression process. The normal and logistic functions have an invariant inflection point located at the mid-point. As a consequence, they are confined to predicting perfectly symmetric curves, and therefore cannot describe the Fig. 3 data to a significant degree of precision. The Gompertz fit is slightly better than the normal and logistic fits because its inflection point is located below the mid-point at $c/c_0 = 0.37$.

In addition to giving poor fits to the tailing data of HCl, the normal, logistic, and Gompertz functions predict significant nonzero exit concentrations at $t = 0$, as may be seen in Fig. 3. At time zero, the dimensionless exit concentrations predicted by the normal (Eq. 1), logistic (Eq. 8), and Gompertz (Eq. 11) functions are respectively 10.8%, 11.6%, and 6.8%, which contradict the expected value of zero. The log-normal function is undefined at $t = 0$, but it predicts very small exit concentrations near time zero. The Weibull function is the only one that predicts zero outlet concentration at $t = 0$.

The log-normal fit with an R^2 score of 0.999 is remarkably accurate in tracking the tailing data of HCl. It is also slightly better than the Weibull fit. Since both functions have a floating inflection point, the difference in the two fits is likely due to the more flexible mathematical form of the log-normal function. Nevertheless, the results underscore the fact that a floating inflection point plays a critical role in determining the performance of a probability function.

Case 4: breakthrough curve of hydrogen sulfide

In the previous section, we tested the ability of the five probability functions to describe the tailing data of HCl taken from the work of Papurello et al. (2019). Here, another set of tailing breakthrough data reported in the same article is used to test the five functions. The selected data set was obtained with a carbon filter (Norit RST3) subjected to a feed stream of H_2S [$c_0 =$

76 ppm(v)]. The H₂S data set, presented in Fig. 4, differs from the HCl data set shown in Fig. 3 in two major aspects. First, there is no immediate breakthrough of H₂S. Second, the H₂S curve exhibits a significant degree of tailing. Figure 4 plots the associated function fits, computed using the parameter estimates listed in Table 1, case 4. In comparing the functions, Figure 4 shows that the log-normal provides the highest R^2 and the lowest RRMSE values. Although the log-normal function is the best performer, its R^2 score is less than 0.99. Evidently, the log-normal with a floating inflection point has some difficulty in representing this particular data set with strong tailing. The other four functions are clearly inferior to the log-normal, producing visually similar fits to the observed data which bypass most of the data points.

Fig. 4 Probability function fits compared to observed H₂S breakthrough data reported by Papurello et al. (2019)

As Fig. 4 shows, the log-normal function is vastly superior to the normal function. As demonstrated previously, the former can be easily derived from the latter via a logarithmic transformation. It is therefore of interest to convert the logistic and Gompertz functions to similar logarithmic forms. Applying the logarithmic transformation procedure described previously to the logistic and Gompertz functions yields

$$\frac{c}{c_0} = \frac{1}{1 + \exp[a_6 - b_6 \ln(t)]} \quad (20)$$

$$\frac{c}{c_0} = \exp\{-\exp[a_7 - b_7 \ln(t)]\} \quad (21)$$

Equation 20 is referred to as the log-logistic function while Eq. 21 is known as the log-Gompertz function (Chu 2020b). The t -term in Eqs. 20 and 21 is a dimensionless quantity defined by t/t^* where $t^* = 1$ min. Likewise, b_6 and b_7 are dimensionless parameters defined by the actual $b_6 \times t^*$ and the actual $b_7 \times t^*$ respectively.

Function	R^2	RRMSE
Log-Gompertz	0.989	0.038
Log-logistic	0.976	0.057

Fig. 5 Modified probability function fits compared to observed H₂S breakthrough data reported by Papurello et al. (2019)

The predictions of the log-logistic and log-Gompertz functions calculated according to Eqs. 20 and 21 respectively are compared with the H₂S data in Fig. 5. The parameter values used to compute the two curves are $a_6 = 11.07$, $b_6 = 2.60$, $a_7 = 7.73$, and $b_7 = 1.92$. It is apparent that the log-logistic fit depicted in Fig. 5 is considerably more accurate than the logistic fit depicted in Fig. 4. Likewise, the log-Gompertz fit illustrated in Fig. 5 is superior to the Gompertz fit shown in Fig. 4. Both the log-logistic and log-Gompertz fits do not display obvious nonzero effluent concentrations at $t = 0$.

The enhanced data fitting ability of the log-Gompertz and log-logistic functions can be explained by the fact that they have a floating inflection point. To find their inflection points, Eqs. 20 and 21 are first differentiated twice with respect to the t -term:

$$\frac{\partial^2 (c/c_0)}{\partial t^2} = \frac{\exp(a_6) b_6 t^{b_6-2} [\exp(a_6)(b_6 - 1) - t^{b_6}(b_6 + 1)]}{[\exp(a_6) + t^{b_6}]^3} \quad (22)$$

$$\frac{\partial^2 (c/c_0)}{\partial t^2} = b_7 t^{-2(b_7+1)} \exp[a_7 - \exp(a_7) t^{-b_7}] [\exp(a_7) b_7 - (b_7 + 1) t^{b_7}] \quad (23)$$

Next, equating Eqs. 22 and 23 to zero and solving the resulting equations for the t -term, we obtain

$$\ln(t) = \frac{1}{b_6} \left[a_6 - \ln \left(\frac{b_6 + 1}{b_6 - 1} \right) \right] \quad (24)$$

$$\ln(t) = \frac{a_7 + \ln[b_7/(b_7 + 1)]}{b_7} \quad (25)$$

Substituting Eqs. 24 and 25 in Eqs. 20 and 21 respectively gives

$$\frac{c}{c_0} = \frac{b_6 - 1}{2b_6} \quad (26)$$

$$\frac{c}{c_0} = \exp \left[- \left(\frac{b_7 + 1}{b_7} \right) \right] \quad (27)$$

The preceding equations confirm that the location of the inflection point is variable and depends on the value of the b -term. Given $b_6 = 2.60$ and $b_7 = 1.92$ as outlined above, the inflection points of the log-logistic and log-Gompertz computed from Eqs. 26 and 27 are located at $c/c_0 = 0.31$ and $c/c_0 = 0.22$ respectively. The inflection points of the original logistic and Gompertz functions are fixed at $c/c_0 = 0.5$ and $c/c_0 = 0.37$ respectively. These invariant inflection points restrict their ability to track the tailing H_2S data. For the four functions with a floating inflection point, the R^2 and RRMSE scores presented in Figs. 4 and 5 demonstrate that the log-Gompertz function is best supported by the H_2S data. The log-normal and log-logistic, which are similar in performance, are the next best supported by the data, while the Weibull function is the least supported. Papurello et al. (2019) used a mechanistic model to describe several sets of H_2S breakthrough data. It seems that the mechanistic model was only used to fit the initial stages of the data sets. It is not known whether the mechanistic model can track the entire profiles of the tailing breakthrough curves of H_2S .

Because of its improved data fitting ability, the log-Gompertz function can provide accurate breakthrough and exhaustion times corresponding to any predetermined effluent concentration level. However, it cannot be used for process design because it lacks typical fixed bed variables such as feed concentration, flow rate, and bed length. A simple way to fix this shortcoming is to let the two parameters of the log-Gompertz equation mimic the parameters of commonly used phenomenological models. For example, the Bohart-Adams model contains variables such as the feed concentration (c_0), superficial velocity (u), and bed length (L). So, the a_7 and b_7 terms of the log-Gompertz equation may be defined as follows: $a_7 = k_1 k_2 L / u$ and $b_7 = k_1 c_0$ where k_1 and k_2 are adjustable parameters. The log-Gompertz equation can now be used in the same way as the Bohart-Adams model for process design.

Linear forms of probability functions

The four functions with a variable inflection point—log-normal, Weibull, log-logistic, and log-Gompertz—can be transformed into linear forms and linear regression can then be used to estimate their parameters. For example, they can be expressed in the following linear forms:

$$\text{Log-normal: } \operatorname{erf}^{-1}\left(2\frac{c}{c_0}-1\right) = -\frac{a_2}{b_2\sqrt{2}} + \frac{1}{b_2\sqrt{2}}\ln(t) \quad (28)$$

$$\text{Weibull: } \ln\left[-\ln\left(1-\frac{c}{c_0}\right)\right] = -b_5\ln(a_5) + b_5\ln(t) \quad (29)$$

$$\text{Log-logistic: } \ln\left(\frac{c_0}{c}-1\right) = a_6 - b_6\ln(t) \quad (30)$$

$$\text{Log-Gompertz: } \ln\left(\ln\frac{c_0}{c}\right) = a_7 - b_7\ln(t) \quad (31)$$

These linear forms can be used to estimate the a - and b -terms by linear regression of the left-hand member of each equation vs. $\ln(t)$. In Eq. 28 $\operatorname{erf}^{-1}(y)$ designates the inverse error function. In Excel there is no inverse error function but it can be approximated using the inverse normal cumulative distribution [$\operatorname{erf}^{-1}(y) = \text{NORM.S.INV}((y+1)/2)/\text{SQRT}(2)$].

Figure 6 plots the linear fits, computed using the parameter estimates given in Table 2. All four linear fits are rather poor, yielding parameter estimates that are quite different from those generated by nonlinear regression (see Table 2). Thus, the results of linear regression are not statistically optimal. In nonlinear regression, initial values for each parameter in the model must be specified, usually based on intelligent guessing or previous experience. If the initial values are far from the final values, the nonlinear regression procedure can go in the wrong direction and never converge on a solution or converge on the wrong solution. To guard against this problem, one can use parameter estimates based on linear transformations as initial values for nonlinear regression. These estimates are usually fairly close to the final values. In a practical sense, Eqs. 28-31 are quite useful for preliminary data analysis.

Fig. 6 Linear fits of Eqs. 28-31 compared to observed H₂S breakthrough data reported by Papurello et al. (2019)

Table 2 Parameter estimates for the linear fits in Fig. 6 (values in parenthesis represent parameter estimates obtained from nonlinear regression)

Function	Parameter value
Log-normal	$a_2 = 4.35$ (4.28); $b_2 = 0.59$ (0.65)
Weibull	$a_5 = 106.92$ (96.06); $b_5 = 2.03$ (1.65)
Log-logistic	$a_6 = 13.42$ (11.07); $b_6 = 3.09$ (2.60)
Log-Gompertz	$a_7 = 8.98$ (7.73); $b_7 = 2.22$ (1.92)

Probability functions and fixed bed modeling

Here, we give an overview of previous research on the use of the five probability functions in breakthrough curve modeling. We also point out an interesting observation: some phenomenological models resemble the normal and logistic functions.

The model of Belter et al., published in a textbook in the late 1980's (Belter et al. 1988), reads

$$\frac{c}{c_0} = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{t - t_{0.5}}{\sigma_0 t_{0.5} \sqrt{2}} \right) \right] \quad (32)$$

where $t_{0.5}$ and σ_0 are adjustable parameters.

Although Belter et al. (1988) did not specify the origin of Eq. 32, it seems very likely that it is based on the normal function defined by Eq. 1. Its inflection point is thus located at $c/c_0 = 0.5$. Like the normal function, the Belter et al. model is confined to fitting symmetric breakthrough curves. It should not be used to represent tailing breakthrough data.

It is possible to rearrange the Belter et al. model to a linear form:

$$\operatorname{erf}^{-1} \left(2 \frac{c}{c_0} - 1 \right) = -\frac{1}{\sigma_0 \sqrt{2}} + \frac{1}{\sigma_0 t_{0.5} \sqrt{2}} t \quad (33)$$

A plot of the left-hand member of Eq. 33 vs. the t -term should yield a straight line if the data follow the linear model. Equation 33 can be used to generate initial values for a subsequent nonlinear fit of Eq. 32. As mentioned previously, to plot Eq. 33 in Excel, one may use the inverse normal cumulative distribution to approximate $\operatorname{erf}^{-1}(y)$, the inverse error function. As examples of recent applications of the Belter et al. model in environmental adsorption research one may cite the works of Saldaña-Robles et al. (2018) and Faisal et al. (2021). A modified form of the normal function incorporating some fixed bed variables for correlating breakthrough curves has been presented by Dima et al. (2020).

440 Interestingly, a handful of phenomenological models based on the assumption of linear
 441 adsorption bear a resemblance to the normal function. We present three such models here: (1)
 442 a simplified form of the model of Lapidus and Amundson (1952), (2) an asymptotic form of
 443 the model of Rosen (1952), and (3) an approximate form of the model of Anzelius (1926).
 444 These models may be written

$$445 \text{ Lapidus-Amundson: } \frac{c}{c_0} = \frac{1}{2} \left\{ 1 + \operatorname{erf} \left[\frac{\sqrt{Pe} (t/T - 1)}{2\sqrt{t/T}} \right] \right\} \quad (34)$$

446 where

$$447 Pe = \frac{vL}{D_a} \quad (35)$$

$$448 T = \frac{L}{v} \left[1 + \rho_p K \left(\frac{1 - \varepsilon}{\varepsilon} \right) \right] \quad (36)$$

$$449 \text{ Rosen: } \frac{c}{c_0} = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{1}{2} \sqrt{n_R} (\tau_0 - 1) \right) \right] \quad (37)$$

450 where

$$451 n_R = \frac{15D_s}{r_p^2} \frac{\rho_p KL}{v} \left(\frac{1 - \varepsilon}{\varepsilon} \right) \quad (38)$$

$$452 \tau_0 = \frac{1}{\rho_p K} \left(\frac{vt}{L} - 1 \right) \left(\frac{\varepsilon}{1 - \varepsilon} \right) \quad (39)$$

$$453 \text{ Anzelius: } \frac{c}{c_0} = \frac{1}{2} \left[1 + \operatorname{erf} \left(\sqrt{n_A \tau_0} - \sqrt{n_A} + \frac{1}{8\sqrt{n_A \tau_0}} + \frac{1}{8\sqrt{n_A}} \right) \right] \quad (40)$$

454 where

$$455 n_A = \frac{k_{LDF} \rho_p KL}{v} \left(\frac{1 - \varepsilon}{\varepsilon} \right) \quad (41)$$

456 In these equations, the symbols are defined as follows: $L \equiv$ the packed bed length, $v \equiv$
 457 the interstitial velocity, $\varepsilon \equiv$ the bed voidage, $\rho_p \equiv$ the particle density, $r_p \equiv$ the particle radius, K

458 $\equiv$ the Henry constant, $D_a \equiv$ the axial dispersion coefficient, $D_s \equiv$ the solid diffusion coefficient,
 459 and $k_{LDF} \equiv$ the linear driving force rate coefficient.

460 Each of the three phenomenological models assumes a single curve broadening
 461 mechanism. The Lapidus-Amundson model assumes that axial dispersion is the dominant rate
 462 mechanism. In the Rosen model, the rate-limiting step is solid diffusion. The Anzelius model
 463 assumes that intraparticle diffusion can be represented by a linear driving force rate expression.
 464 Note that the expression selected for the Anzelius model is known as the Klinkenberg
 465 approximation (Klinkenberg 1948). We see that the normal function, Lapidus-Amundson,
 466 Rosen, and Anzelius equations all share the same mathematical form defined by Eq. 42, with
 467 different error function arguments. This similarity is most likely a sheer coincidence.

$$468 \quad \frac{c}{c_0} = \frac{1}{2} [1 + \operatorname{erf}(y)] \quad (42)$$

469 Next, a quasi-log-normal distribution function has been used to derive an approximate
 470 solution to a fixed bed model which includes linear adsorption, the axial dispersion of mass as
 471 well as external film and intraparticle diffusional resistances to mass transfer (Xiu and Li 1999;
 472 Xiu et al. 1997). The approximate solution reads

$$473 \quad \frac{c}{c_0} = \mu_0 \int_{-\infty}^{(\ln \tau - \mu)/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\beta^2}{2}\right) d\beta \quad (42)$$

474 where β is an integration parameter and μ_0 , μ , σ , and τ are parameters which include all the
 475 relevant fixed bed variables and equilibrium and rate parameters. For a complete definition of
 476 these parameters, the reader is referred to Xiu et al. (1997). The semi-infinite integral of Eq. 42
 477 is somewhat cumbersome; a more tractable alternative given in a simple error function form is
 478 available in the literature (Sigrist et al. 2011):

$$479 \quad \frac{c}{c_0} = \frac{1}{2} \mu_0 \left[1 + \operatorname{erf}\left(\frac{\ln(\tau) - \mu}{\sigma\sqrt{2}}\right) \right] \quad (43)$$

The last result is analogous to the log-normal function given by Eq. 4. Given $\mu_0 = 1$ (Xiu et al. 1997), if we let $\mu = a_2$, $\sigma = b_2$, and $\tau = t$, Eq. 43 reduces to Eq. 4. The two equations differ in the definitions of their parameters: the μ , σ , and τ terms comprise physically significant parameters while the a_2 and b_2 terms are empirical parameters. It seems that under the assumption of linear adsorption a fixed bed breakthrough curve can be considered as a probability cumulative distribution curve, and it is possible to link the fixed bed's variables to the probability function's parameters. We have shown that the log-normal function with a floating inflection point is quite effective in tracking the tailing breakthrough data of HCl and H₂S. Consequently, Eq. 43 should have no problem in handling the same type of breakthrough data.

We now turn our attention to the logistic function. Oulman (1980) pointed out that a simplified form of the Bohart-Adams model is the same as the logistic function. It is now known that the Thomas and Yoon-Nelson models are also analogous to the logistic function (Chatterjee and Schiewer 2011; Chu 2020; Lee et al. 2015). Because there is no intrinsic relation between the three models and the logistic function, this mathematical equivalence is sheer coincidence. Similar to the logistic function which is confined to predicting perfectly symmetric breakthrough curves, the widely used Bohart-Adams, Thomas, and Yoon-Nelson models cannot describe tailing breakthrough curves.

Few studies have investigated the ability of the remaining two probability functions, Gompertz and Weibull, to represent breakthrough data. Two recent studies (Chu 2020b, 2021) used asymmetric breakthrough data to test the two functions, but no attempts were made to compare their data fitting ability.

Conclusions

Probability cumulative distribution functions provide a useful method for correlating breakthrough data of water and air contaminants. In this work the data fitting abilities of the normal, log-normal, logistic, Gompertz, and Weibull functions have been studied using literature breakthrough data of ciprofloxacin, ammonium, HCl, and H₂S. The five functions are shown to provide excellent correlations of the relatively symmetric breakthrough data of ciprofloxacin. For the fronting breakthrough data of ammonium, all five functions provide satisfactory overall fits but none can track the leakage of ammonium during the initial period of column operation. When challenged with the tailing breakthrough data of HCl, the probability functions show apparent differences in performance. The normal, logistic, and Gompertz functions, which have an invariant inflection point, provide unsatisfactory correlations of the tailing data. The log-normal and Weibull functions, which have a floating inflection point, are more accurate in describing the HCl data. Only the log-normal function provides a reasonably accurate representation of the H₂S data with strong tailing. A modified version of the Gompertz function, which has a floating inflection point, provides the best result for the H₂S data. It is concluded that probability functions with a variable inflection point can be used quite reliably to correlate asymmetric breakthrough data. Further studies extending to other probability functions with a floating inflection point are needed to test the validity of the present results.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

530 **Availability of data and materials**

531 All data generated or analyzed during this study are included in this published article.

532

533 **Competing interests**

534 The author declares he has no competing interests.

535

536 **Funding**

537 The author received no specific funding for writing this article.

538

539 **Authors' contributions**

540 This was a single-author article.

541

542 **References**

543 Anzelius A (1926) Über Erwärmung vermittelt durchströmender Medien. Z Angew Math
544 Mech 6:291–294

545 Apiratikul R, Chu KH (2021) Improved fixed bed models for correlating asymmetric
546 adsorption breakthrough curves. J Water Process Eng 40:101810

547 Belter PA, Cussler EL, Hu, W-S (1988) Bioseparations: downstream processing for
548 biotechnology. Wiley, New York.

549 Bohart GS, Adams EQ (1920) Some aspects of the behavior of charcoal with respect to chlorine.
550 J Am Chem Soc 42:523–544

551 Chatterjee A, Schiewer S (2011) Biosorption of cadmium(II) ions by citrus peels in a packed
552 bed column: effect of process parameters and comparison of different breakthrough curve
553 models. Clean Soil Air Water 39:874–881

554 Chu KH (2020a) Breakthrough curve analysis by simplistic models of fixed bed adsorption: in
 555 defense of the century-old Bohart-Adams model. *Chem Eng J* 380:122513
 556 Chu KH (2020b) Fitting the Gompertz equation to asymmetric breakthrough curves. *J Environ*
 557 *Chem Eng* 8:103713
 558 Chu KH (2021) Fixed bed adsorption of chromium and the Weibull function. *J Hazard Mater*
 559 *Lett* 2:100022
 560 Dalhat MA, Mu'azu ND, Essa MH (2021) Generalized decay and artificial neural network
 561 models for fixed-Bed phenolic compounds adsorption onto activated date palm biochar. *J*
 562 *Environ Chem Eng* 9:104711
 563 Danish M, Ansari KB, Aftab RA, Danish M, Zaidi, S, Trinh QT (2021) gPROMS-driven
 564 modeling and simulation of fixed bed adsorption of heavy metals on a biosorbent:
 565 benchmarking and case study. *Environ Sci Pollut Res* (in press).
 566 <https://doi.org/10.1007/s11356-021-13207-y>
 567 Dima JB, Ferrari M, Zaritzky N (2020) Mathematical modeling of fixed-bed columns
 568 adsorption: hexavalent chromium onto chitosan flakes. *Ind Eng Chem Res* 59:15378–
 569 15386
 570 Faisal AAH, Jasim HK, Naji LA, Naushad Mu., Ahamad T (2021) Cement kiln dust-sand
 571 permeable reactive barrier for remediation of groundwater contaminated with dissolved
 572 benzene. *Sep Sci Technol* 56:870–883
 573 Hering JG (2019) From slide rule to big data: how data science is changing water science and
 574 engineering. *J Environ Eng* 145:02519001
 575 Klinkenberg A (1948) Numerical evaluation of equations describing transient heat and mass
 576 transfer in packed solids. *Ind Eng Chem* 40:1992–1994

577 Lapidus L, Amundson NR (1952) Mathematics of adsorption in beds. VI. The effect of
 578 longitudinal diffusion in ion exchange and chromatographic columns. *J Phys Chem*
 579 56:984–988

580 Lee C-G, Kim J-H, Kang J-K, Kim S-B, Park S-J, Lee S-H, Choi J-W (2015) Comparative
 581 analysis of fixed-bed sorption models using phosphate breakthrough curves in slag filter
 582 media. *Desalin Water Treat* 55:1795–1805

583 Mirzaie M, Talebizadeh AR, Hashemipour H (2021) Mathematical modeling and experimental
 584 study of VOC adsorption by Pistachio shell–based activated carbon. *Environ Sci Pollut*
 585 *Res* 28:3737–3747

586 Newhart KB, Holloway RW, Hering AS, Cath TY (2019) Data-driven performance analyses
 587 of wastewater treatment plants: a review. *Water Res* 157:498–513

588 Nguyen LH, Vu TM, Le TT, Trinh VT, Tran TP, Van HT (2019) Ammonium removal from
 589 aqueous solutions by fixed-bed column using corncob-based modified biochar. *Environ*
 590 *Technol* 40:683–692

591 Oulman CS (1980) The logistic curve as a model for carbon bed design. *J Am Water Works*
 592 *Assn* 72:50–53

593 Papurello D, Silvestri S, Lanzini A (2019) Biogas cleaning: trace compounds removal with
 594 model validation. *Sep Purif Technol* 210:80–92

595 Rosen JB (1952) Kinetics of a fixed bed system for solid diffusion into spherical particles. *J*
 596 *Chem Phys* 20:387–394

597 Roy D, Neogi S, De S (2021) Multicomponent transport model-based scaling up of long-term
 598 fixed bed adsorption of reactive dyes from textile effluent using aminated PAN beads.
 599 *Environ Sci Pollut Res* (in press). <https://doi.org/10.1007/s11356-021-13661-8>

600 Saldaña-Robles A, Damian-Ascencio CE, Guerra-Sanchez RJ, Saldaña-Robles AL, Saldaña-
601 Robles N, Gallegos-Muñoz A, Cano-Andrade S (2018) Effects of the presence of organic
602 matter on the removal of arsenic from groundwater. *J Clean Prod* 183:720–728

603 Sausen MG, Scheufele FB, Alves HJ, Vieira MGA, da Silva MGC, Borba FH, Borba CE (2018)
604 Efficiency maximization of fixed-bed adsorption by applying hybrid statistical-
605 phenomenological modeling. *Sep Purif Technol* 207:477–488

606 Schio RR, Salau NPG, Mallmann, ES, Dotto GL (2021) Modeling of fixed-bed dye adsorption
607 using response surface methodology and artificial neural network. *Chem Eng Commun* (in
608 press). <https://doi.org/10.1080/00986445.2020.1746655>

609 Sigrist ME, Beldomenico HR, Tarifa EE, Pieck CL, Vera CR (2011) Modelling diffusion and
610 adsorption of As species in Fe/GAC adsorbent beds. *J Chem Technol Biotechnol* 86:1256–
611 1264

612 Thomas HC (1944) Heterogeneous ion exchange in a flowing system. *J Am Chem Soc*
613 66:1664–1666

614 Xiu G-H, Li P (1999) Linear chromatographic elution and breakthrough curves: quasi-
615 lognormal distribution function and its derived equations. *Chem Eng Sci* 54:377–387

616 Xiu G-H, Nitta T, Li P, Jin G (1997) Breakthrough curves for fixed-bed adsorbers: quasi-
617 lognormal distribution approximation. *AIChE J* 43:979–985

618 Yoon YH, Nelson JH (1984) Application of gas adsorption kinetics. I. A theoretical model for
619 respirator cartridge service life. *Am Ind Hyg Assoc J* 45:509–516

620

Figures

Figure 1

Probability function fits compared to observed ciprofloxacin breakthrough data reported by Sausen et al. (2018)

Figure 2

Probability function fits compared to observed ammonium breakthrough data reported by Nguyen et al. (2019)

Figure 3

Probability function fits compared to observed HCl breakthrough data reported by Papurello et al. (2019)

Figure 4

Probability function fits compared to observed H₂S breakthrough data reported by Papurello et al. (2019)

Function	R^2	RRMSE
Log-Gompertz	0.989	0.038
Log-logistic	0.976	0.057

Figure 5

Modified probability function fits compared to observed H₂S breakthrough data reported by Papurello et al. (2019)

Figure 6

Linear fits of Eqs. 28-31 compared to observed H₂S breakthrough data reported by Papurello et al. (2019)