Contract No:

This document was prepared in conjunction with work accomplished under Contract No. 89303321CEM000080 with the U.S. Department of Energy (DOE) Office of Environmental Management (EM).

Disclaimer:

SRNL
SAVANNAH RIVER NATIONAL LABORATORY
OPERATED BY SAVANNAH RIVER NUCLEAR SOLUTIONS

Chromatographic Modeling of 1-Foot Detritiation Test Column Results for the Optimization of Regeneration Time

From: R. G. Belliveau
To: L. Angelette
January 28, 2020

SRNL-L2110-2020-00012

Introduction

This report documents the methods and results of a modeling study on a series of data generated by 1-foot test columns for the Water Detritiation project. In order to compliment modeling work already in progress on this project, a purely empirical curve-fitting approach is used in this study to determine chromatographic performance indicators such as retention time, curve width, and the number of chromatographic stages.

The data used in this study comes from four sets of chromatographic test separations performed in March through May of 2017 using 1-foot columns. All are using the same flow rates and temperature (room temperature, nominally 25 C°). The first set of data (referred to in this report as the Column 1 data) is a series of twelve breakthrough curve measurements. Prior to each of these twelve runs, the column was regenerated by flowing hydrogen gas through the column, for increasing amounts of time, ranging from 2 hours to 25 hours. The second set of data (referred to as the Column 2 data) is comprised of five measurements of breakthrough curves measured in similar fashion as the Column 1 data, across a range of regeneration times from 2 to 21 hours. The third set of data (referred to as the Purge data) is of three breakthrough curves, measured after a regeneration time of 8 hours, which takes place after a nitrogen purge of 15, 30, or 45 minutes. The fourth data set is a series of five breakthrough curve measurements using a 2nd batch of stationary phase, taken after regeneration times ranging from 2 to 20 hours.

The overall modeling approach is as follows. First, the data is pre-processed for formatting, artifact removal, and is normalized. The data is then fit by integrated versions of four common chromatographic models. The best model is chosen based on the predictive performance of the model, and chromatographic performance values are calculated based on the best model's fit parameters. Each of these processes are detailed below.

Data Selection

Before the data can be subjected to the curve fitting processes, two issues with the data are addressed. First, the data are normalized to unit concentration (Equation 1). Assuming that the deuterium concentration adsorbed onto the stationary phase reaches saturation capacity at some point during each run, this scales the data to the maximum concentration of deuterium present in the data. For this work, each run is considered to have reached saturation capacity when the concentration of deuterium reported by the Picarro CRDS is approximately 142 ppm. This normalization (shown in Figure 1B) allows the models to need little adjustment to any scaling parameters in the fitting process and allows for the optimized parameters to be interpreted directly into chromatographic calculations.

$$C_{norm,i} = \frac{C_i}{C_{max}} \qquad \text{Eq. 1}$$

After normalization, the data are examined for artifacts. Frequently, the first 1 or 2 data points in a curve anomalously high. This is an artifact due to switching the column from regeneration mode to begin a separation run. The correction applied in this step (shown in Figure 1C) is to simply set the anomalous data points (again, only the first or first and second points) to be equal to the third data point.

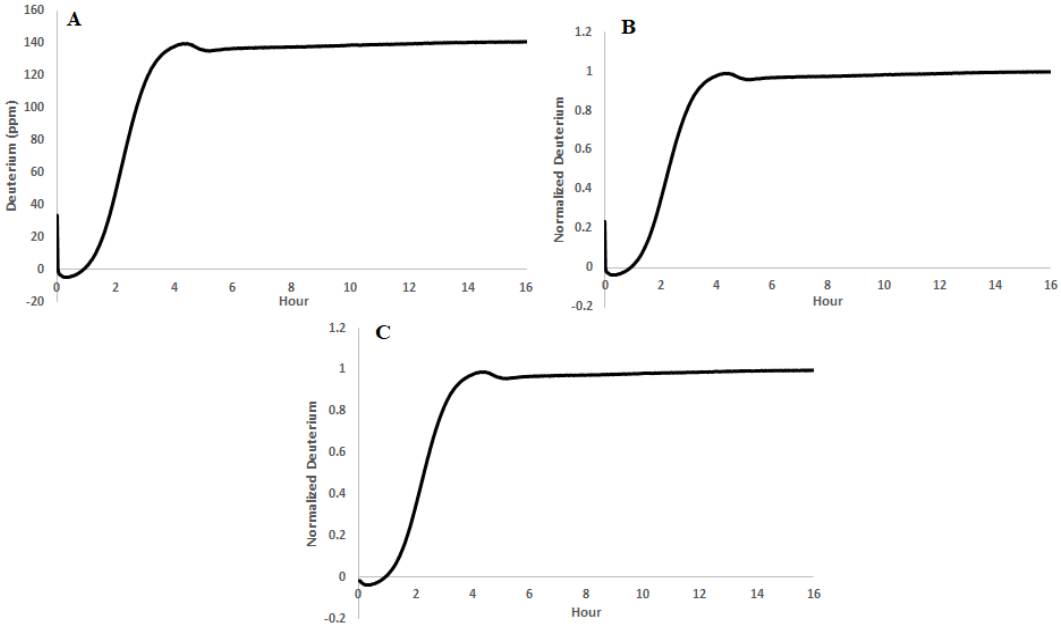

Figure 1. An example deuterium curve. A) The data unprocessed. B) The data after normalization. C) The data after artifact removal.

Curve Fitting

Traditionally, chromatographic modeling is performed on elution chromatography data, where analytes are represented in Gaussian-shaped peaks.[1,2] Frontal analysis (the examination of the breakthrough curve in data like what is examined in this report) attempts to achieve the same goals but

uses one of two different approaches. The derivative of a breakthrough curve is Gaussian-shaped, and so can be modeled by Gaussian and Gaussian-related functions in this form. Alternatively, the data can be natively modeled using integrated forms of the model functions.[2] This avoids errors in the parameter estimation related to numerical derivatization, and also avoids the need for the optimization of an extra derivatization-related parameter (the derivative gap size).

In this study, indefinite integrals of four common chromatographic peak models (Equations 1, 2, 8 and 9) are fit to the data from each run.[2] Chromatographic theory at its most basic assumes that the concentration profile of an analyte traveling through a column is represented as a normal distribution (a Gaussian function). The Gaussian function contains four parameters: **A -** the peak amplitude, $\mathbf{T_r}$ - the retention time of the peak, **S** - which is the Gaussian RMS width of the distribution, and **c** - the x-intercept. Of these parameters, $\mathbf{T_r}$ and **S** are used to calculate chromatographic values such as **N** (stage count), R_s (peak resolution), and α (separation factor). **A** and **c** are the previously-mentioned scaling parameters. These are normally 1 and 0 (respectively) for normalized data but are adjusted slightly during the optimization process. Values of **A** and **c** differing significantly from their nominal values are considered a sign of a poor curve fit. The integrated form of the Gaussian function is given in Equation 2.

$$\mathbf{G(x) = \frac{A}{2}\left[1 + erf\left(\frac{x - T_r}{S\sqrt{2}}\right)\right] + c} \qquad \textbf{Eq. 2}$$

A Gaussian function does not model tailing or fronting behavior in chromatographic data. A commonly-used model for systems with skewing behaviors is the exponentially-modified Gaussian (EMG). Shown in Equations 3-7, this model includes an additional parameter: **b** - the tailing factor. This parameter is another scaling factor, acting to stretch the distribution horizontally towards longer retention times.

$$\mathbf{G_{mod}(x) = \frac{A}{2} e^{\frac{-x}{b}}[F_1F_2 - F_3F_4] + c} \qquad \textbf{Eq. 3}$$

Where:

$$\mathbf{F_1 = e^{\left(\frac{S\sqrt{2}+x}{b}\right)}} \qquad \textbf{Eq. 4}$$

$$\mathbf{F_2 = erf\left(\frac{-T_r+S\sqrt{2}+x}{S\sqrt{2}}\right)} \qquad \textbf{Eq. 5}$$

$$\mathbf{F_3 = e^{\left(\frac{S^2+\frac{T_r}{b}}{2b^2}\right)}} \qquad \textbf{Eq. 6}$$

$$\mathbf{F_4 = erf\left(\frac{\frac{x-T_r}{S}-\frac{S}{b}}{\sqrt{2}}+1\right)} \qquad \textbf{Eq. 7}$$

Two other models used are the Lognormal and Poisson distributions (Equations 8 and 9, respectively).[2] The Lognormal model uses factors related to the x-dimension (time (x) and $\mathbf{T_r}$) in natural-log-space. It otherwise assumes the form of a normal distribution. The Poisson distribution is used in a form containing the gamma (Γ) function, which allows it to be calculated as a continuous function, instead of containing a factorial, and being calculated as a discreet function. In the integrated forms, some of the models contain the error function, and the Poisson model contains the (incomplete) gamma function. Both of these functions are computed using native Python functions ('math.erf', and 'gammainc', respectively).

$$\mathbf{L(x)} = \frac{\mathbf{A}}{\mathbf{2}}\left[\mathbf{1} + \mathbf{erf}\left(\frac{\mathbf{\ln(x) - \ln(T_r)}}{\mathbf{S\sqrt{2}}}\right)\right] + \mathbf{c} \qquad \textbf{Eq. 8}$$

$$\mathbf{P(x)} = \mathbf{AT_r}\left(-\mathbf{e^{(S-1)}}\right)\left(\frac{\mathbf{x}}{\mathbf{T_r}}\right)^{\mathbf{S}}\left(\frac{\mathbf{x(S-1)}}{\mathbf{T_r}}\right)^{\mathbf{-S}}\mathbf{\Gamma}\left(\mathbf{S}, \frac{\mathbf{x(S-1)}}{\mathbf{T_r}}\right) + \mathbf{c} \qquad \textbf{Eq. 9}$$

Examples of the Gaussian and lognormal model fits to experimental data are shown in Figure 2.

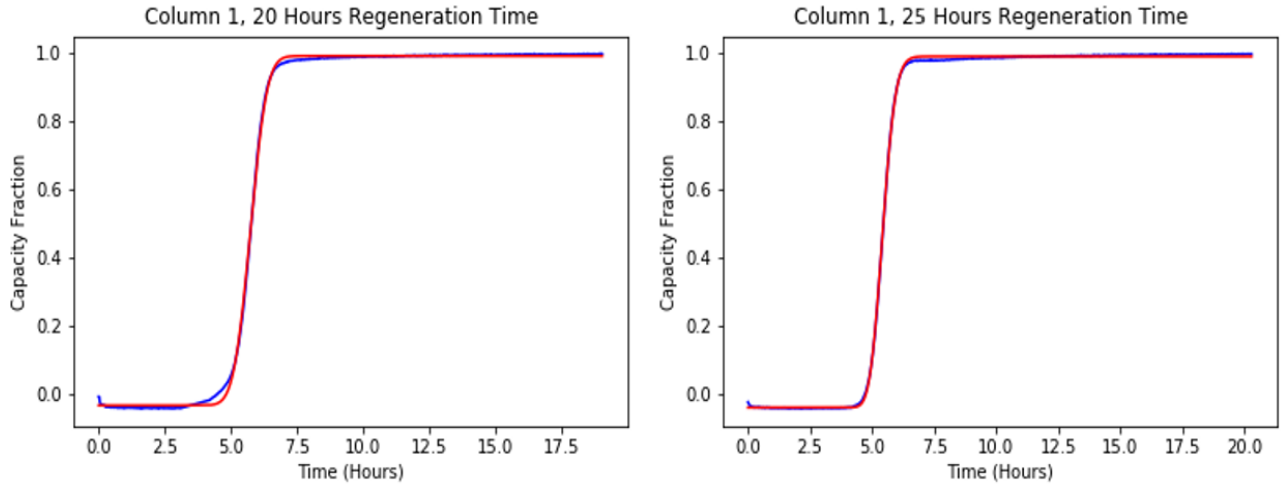

Figure 2. Curve fits of (left) the Gaussian model to the Column 1, 20-hour regeneration curve, and of (right) the Lognormal model to the Column 2, 25-hour regeneration curve. The experimental data is shown in blue and the curve fit is shown in red.

Model Selection

The best-performing model is determined by calculating the root-mean-square error of estimate (RMSEE) for each model that succeeds in fitting the data.[3] This calculation is given in Equation 10. As the data is normalized to 1, this means that the RMSEE represents the average percent deviation (in decimal form) between the model and the data. Acceptable values are on the order of 2-3% deviation.

$$\mathbf{RMSEE} = \sqrt{\frac{\sum(\mathbf{y} - \mathbf{y_{est}})^2}{\mathbf{n}}} \qquad \textbf{Eq. 10}$$

Where y are the normalized deuterium curve data, y_{est} is the modeled normalized deuterium curve data, and n is the number of data points in the curve.

Table 1 shows the regeneration time for each run, the model that best fits the deuterium curve in that data set, and the RMSEE for that model. The Gaussian or lognormal models best fit each of the curves. The lognormal models perform better than the Gaussian models in some cases where the full sigmoidal shape of the curve is not present in the data, such as because the run begins part-ways into the sigmoid, as in the 2 hour runs. For each run, both the Gaussian and lognormal models result in acceptable RMSEE values.

Sample	Regen. Time (Hours)	Model	RMSEE
Column 1	1.9	lognormal	0.0148
Column 1	3.9	lognormal	0.0152
Column 1	5.9	gaussian	0.0100
Column 1	6.3	gaussian	0.0058
Column 1	7.6	gaussian	0.0064
Column 1	9.8	gaussian	0.0059
Column 1	12.0	gaussian	0.0058
Column 1	14.3	gaussian	0.0140
Column 1	16.2	gaussian	0.0240
Column 1	16.0	gaussian	0.0220
Column 1	19.9	gaussian	0.0086
Column 1	25.2	lognormal	0.0052
Column 2	1.9	gaussian	0.0054
Column 2	1.9	gaussian	0.0029
Column 2	13.9	gaussian	0.0117
Column 2	19.7	gaussian	0.0119
Column 2	21.0	gaussian	0.0067
Purge 15	7.5	gaussian	0.0055
Purge 30	7.8	gaussian	0.0032
Purge 45	8.1	gaussian	0.0045
Batch 2	2.0	lognormal	0.0109
Batch 2	8.0	gaussian	0.0089
Batch 2	7.2	gaussian	0.0051

Batch 2	14.0	gaussian	0.0059
Batch 2	20.0	gaussian	0.0125

Table 1. Regeneration times, best performing models, and root-mean-square error of estimation of the best model for each run.

Chromatographic Figures of Merit

Chromatographic performance is normally estimated by the number of stages (**N**, Equation 11) or by how well two analytes are separated (resolution or separation factor). In the case of this data, the second analyte, water, is at or near saturation capacity at the start of the separation run and the goal is simply to slow the deuterium from making its way through the column.

$$\mathbf{N} = \left(\frac{\mathbf{T_r}}{\mathbf{S}}\right)^{\mathbf{2}} \qquad \textbf{Eq. 11}$$

The more time an analyte spends traveling through the stationary phase of a column, generally speaking, the more diffusion it will undergo, and the broader shape the peak or front will have. The number of stages, **N,** indicates how well a column manages to prevent that unwanted diffusion. It describes how 'sharply' an analyte elutes from the column for the amount of time it spent traveling through the stationary phase. Because the goal of this project is to elute the most amount of clean water before deuterium-contaminated water is eluted (not necessarily the same thing as achieving a high **N**), a more contextual figure of merit is also used. As a figure of merit for this project, column performance is based on the molar ratio of H_2 gas used in the column regeneration process to the amount of deuterium-free water eluted by the column during the separation run of the column. Water which is 'deuterium-free' is defined for this study as a concentration lower than 4 ppm. Empirically, this is estimated as 1.5 RMS width units (**S**) before the retention time.[1] To illustrate the behavior in column performance when considering at what point to consider the run 'complete' in the sense that the run has stopped eluting clean water, three cut-off points for the run are considered. First, the model accounts for only water eluted 1.5 RMS width units before the modeled retention time. The model also calculates the amount of water eluted up to the modeled retention time, and then 1.5 RMS widths after the retention time. This last value is on the inflection of the breakthrough curve as it begins to plateau at the stationary phase saturation capacity, and is considered to nominally be the point at which the deuterium front reaches saturation.

[1] A calculation of errors between the described method of determining the point to consider as a cut-off for deuterium-free water and a method in which the time where the closest measured value to 4 ppm deuterium shows only small residuals in all of the data except for the 16-hour regeneration data sets, in which the error is around 10%. It appears that there is a degree of 'fronting' in these runs, during which the 4 ppm mark is reached some minutes earlier than a Gaussian-behaving chromatographic front would normally.

Results & Discussion

Figures 3, 4, and 5 show the model-determined retention time (**R_t**), distribution RMS width (**S**), and number of stages (**N**), respectively. Deuterium is retained in-column longer with increasing regeneration times, up to around 16 hours spent in regeneration (Figure 3). After which, the retention times do not change appreciably.

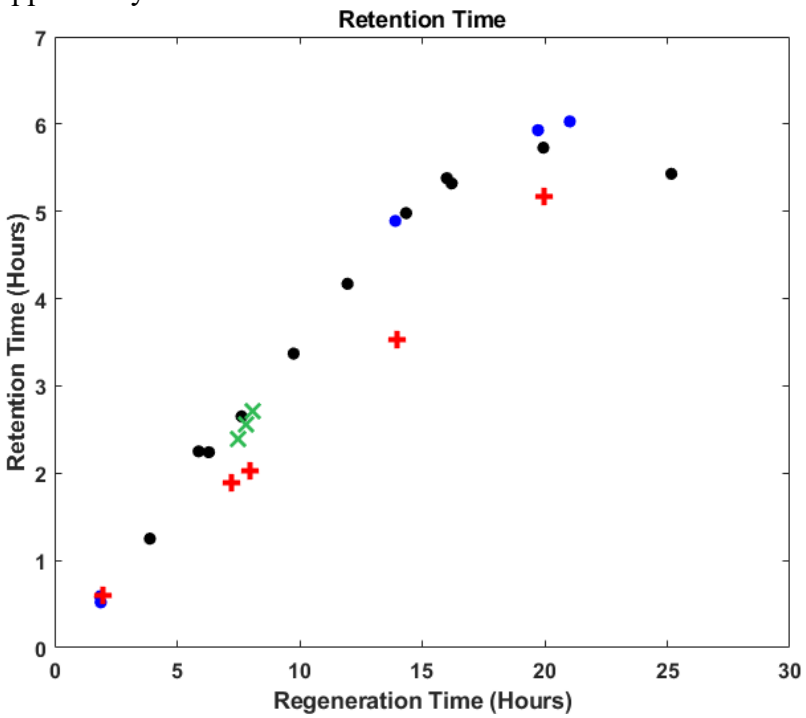

Figure 3. Deuterium retention time for each data set. Black is the Column 1 data. Blue is the Column 2 data. Green is the Purge data. Red is the Batch 2 data. Retention time plateaus around 5-6 hours.

The RMS width (**S**) of each curve reaches a peak near 14 hours of regeneration time, and then sharply drops (Figure 4). This can be seen in the data for Column 1: the experimental data for the 25-hour regeneration run overlaid with the experimental data for the 2-hour regeneration run (but shifted in time) would show two curves that are nearly identical in shape. However, the experimental data for the 14-hour regeneration runs (from Column 1, Column 2, or Batch 2) show a much broader curve than either the 2 or 25-hour data sets.

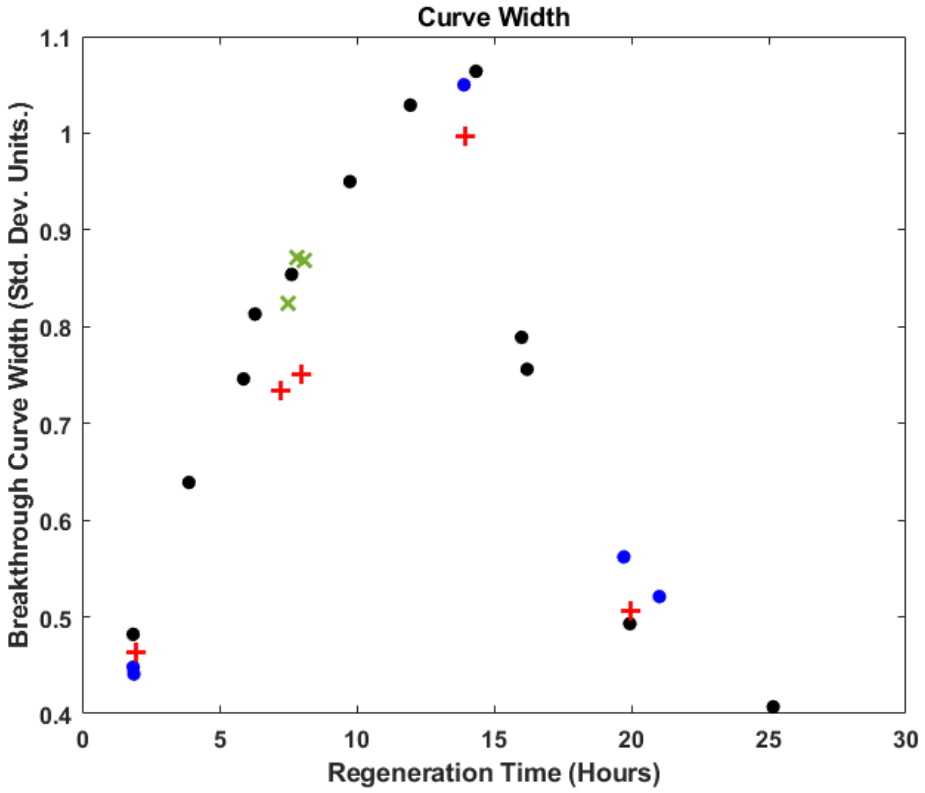

Figure 4. Deuterium breakthrough curve RMS width for each data set, measured in distribution standard deviations. Black is the Column 1 data. Blue is the Column 2 data. Green is the Purge data. Red is the Batch 2 data. Breakthrough curves reach a maximum in width after a 14-hour regeneration.

The number of stages (Figure 5), being positively correlated to the ratio of the retention time and curve width, increases rapidly with decreasing curve RMS width. This inflection in the number of stages coincides with the inflection in curve RMS width at regeneration times higher than 14 hours. The large increase in the number of stages may be exaggerated due to inaccuracies in estimating the width of such sharp curves as seen in the high regeneration time runs. In Heung et al, a similar (but opposite) issue is seen, where models using high numbers of stages lead to breakthrough curves that are increasingly sharper, but negligibly so.[4] This implies that at high stage numbers, there are inaccuracies inherent to empirical determination by frontal analysis.

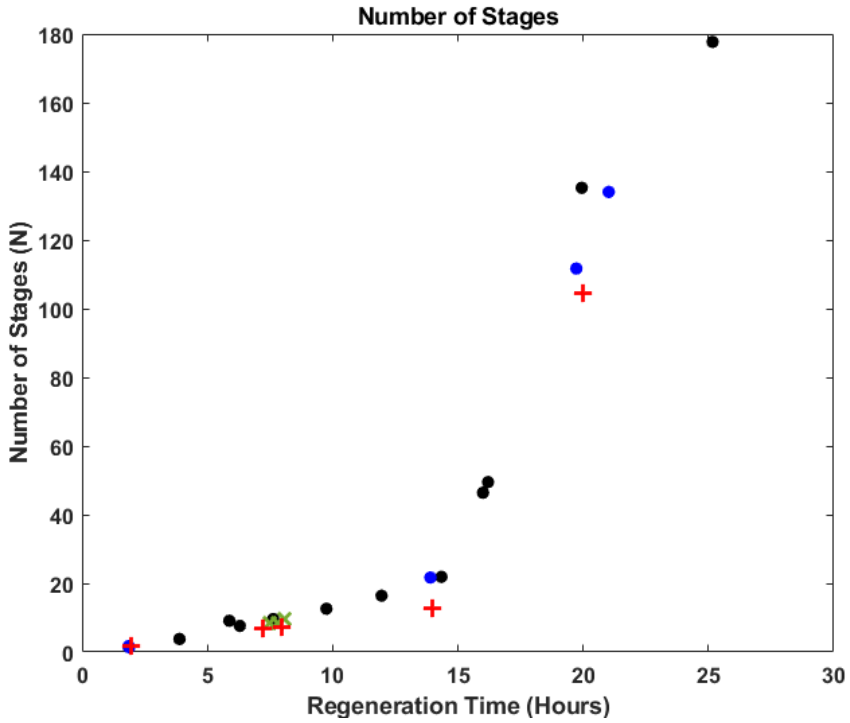

Figure 5. Number of stages (as determined by the deuterium breakthrough curve). Black is the Column 1 data. Blue is the Column 2 data. Green is the Purge data. Red is the Batch 2 data. The inflection at regeneration times higher than 14 hours coincides with the inflection to lower curve widths at higher regeneration times.

The Purge data shown in Figures 3, 4, and 5 show increasing performance as the pre-regeneration nitrogen purge time increases from 15 to 45 minutes. The changes are modest, but a small increase in the number of stages is seen (8.4 to 9.7) with increasing purge time. The Column 2 data set does not show any distinction from the Column 1 data set. However, the Batch 2 data shows decreased performance, particularly in terms of retention times. For a given regeneration time, the first batch of stationary phase (used in the Column 1 and 2 data, as well as the purge test data) retains the deuterium longer than the second batch, allowing more clean water to be eluted before deuterium breakthrough.

Figure 6 shows the calculated H_2/H_2O ratios when considering only water eluted prior to reaching 4 ppm deuterium in the eluent (black dots), water eluted up to the modeled retention time (blue), and then water eluted up to saturation of the column with deuterium (red). At low regeneration times, the amount of water eluted before the deuterium begins to break through is small, which means including in the ratio calculation more of the water eluting during the deuterium breakthrough has a relatively larger impact than is seen at high regeneration times. When considering only water eluted prior to reaching 4 ppm deuterium, there is a minimum in the ratio at 17.6 hours, shown in Figure 7. This was determined by fitting a polynomial to the calculated ratios, given in Equation 12.

$$\mathbf{Ratio} = \mathbf{0.0327(Ret.\,Time)^2 - 1.152(Ret.\,Time) + 15.12} \qquad \textbf{Eq. 12}$$

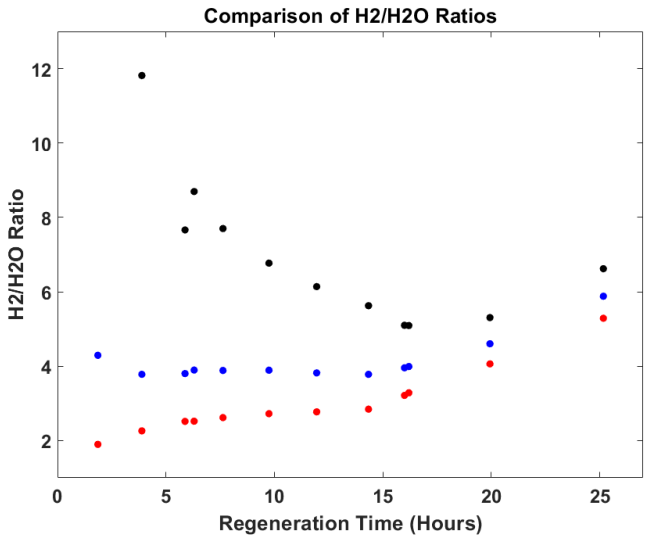

Figure 6. H_2/H_2O molar ratios for the three cases described above. Each set is of the column 1 data, but considering different times throughout the elution to stop considering further eluted water in the ratio calculation. Black includes water eluted up to 1.5 times the RMS width prior to the deuterium front (nominally 4 ppm deuterium). Blue includes water eluted up to the modeled retention time. Red considers water eluted up to 1.5 RMS widths after the modeled retention time (nominally when column becomes saturated with deuterium).

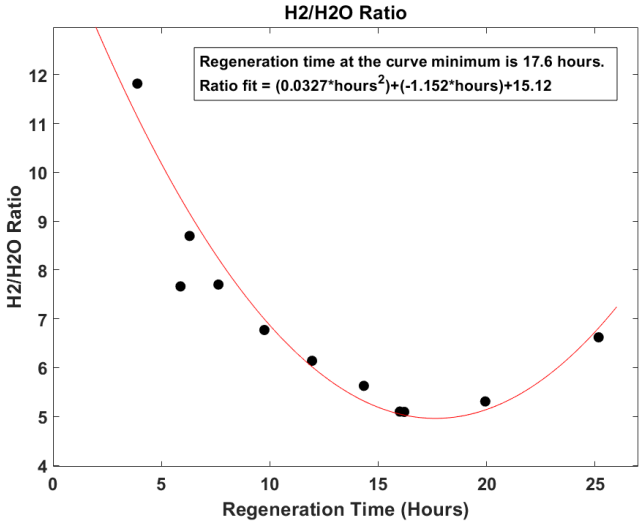

Figure 7. H_2/H_2O molar ratio when considering only water recovered prior to eluting 4 ppm deuterium. Black dots show the Column 1 data, the red curve is a second order polynomial fit to the data. The fit shows a minimum at 17.6 hours of regeneration prior to the separation run.

Conclusions

Chromatographic performance of 1-foot columns used to test varying regeneration times is modeled based on four functions commonly used in chromatographic modeling. Most data sets were successfully modeled as Gaussian, with the remainder modeled slightly better as Log-Normal functions. These models were used to determine retention times and widths for each deuterium front in the data sets. These values were used to estimate the number of stages and the ratio of moles of hydrogen gas used to regenerate the column to the moles of 'clean' water recovered by the separation run. The modeling shows a marked improvement in chromatographic performance after 14 hours of column regeneration, with a minimum H_2/H_2O ratio found at 17.6 hours of regeneration time.

References

1. C. Helling, T. Dams, B. Gerwat, et al. Physical characterization of column chromatography: Stringent control over equipment performance in biopharmaceutical production. *Trends in Chrom.* **2013**, 8: 55-71.
2. V. B. Di Marco, G. G. Bombi. Mathematical functions for the representation of chromatographic peaks. *J. Chrom. A* **2001**, 931: 1-30.
3. R. Kramer. Chemometric Techniques for Quantitative Analysis. Marcel Dekker, 1998, New York, New York. ISBN: 0-8247-0198-4
4. L. K. Heung, G. C. Staak, J. E. Klein, et al. Method to test isotopic separation efficiency of palladium packed columns. WSRC-STI-2007-00119.