A Systematic Approach for Modeling Chromatographic Processes—Application to Protein Purification

Sharon Chan
Centre for Process Systems Engineering, Dept. of Chemical Engineering,
University College London (UCL), London, U.K.

Nigel Titchener-Hooker and Daniel G. Bracewell
Advanced Centre for Biochemical Engineering, Dept. of Biochemical Engineering,
University College London (UCL), London, U.K.

Eva Sørensen
Centre for Process Systems Engineering, Dept. of Chemical Engineering,
University College London (UCL), London, U.K.

DOI 10.1002/aic.11441
Published online February 26, 2008 in Wiley InterScience (www.interscience.wiley.com).

In this article, a systematic approach is presented which allows the estimation of model parameters including isotherm parameters for equilibrium-dispersive and general rate models including a procedure for how to subsequently select the most appropriate isotherm and model for a given application. The approach consists of three parts: determination of feed concentration, parameter estimation, and model selection. Accurate modeling of any separation process depends on the starting point, i.e. the feed and the feed concentration, however, for many chromatographic separations, this concentration is unknown. The first part of this work therefore presents a systematic procedure for feed concentration determination. The estimation of isotherm or other model parameters is done using the simulation and optimization tool gPROMS. Finally, the predictions from each model alternative are then compared using fractionation and maximum purification diagrams. The estimation approach is successfully applied to three case studies, including purification of alcohol dehydrogenase from a yeast homogenate supernatant and gradient elution ion-exchange of myoglobin, which illustrate the approach and the relative suitability and range of application of the models in describing the behavior of complex chromatographic separations. It is found that, while the equilibrium-dispersive model is a good approximation in most cases, there are cases when the more detailed general rate model is required. © 2008 American Institute of Chemical Engineers *AIChE J,* 54: 965–977, 2008
Keywords: chromatography, modeling, parameter estimation, equilibrium-dispersive model, general rate model, protein purification

Correspondence concerning this article should be addressed to E. Sørensen at e.sorensen@ucl.ac.uk.

Introduction

Chromatography is a key separation technique in the biological, pharmaceutical, and fine chemical industries. As

such, there is much interest in the modeling and optimization of process-scale chromatography as such studies allow for the investigation and screening of different design and operating alternatives with minimal experimental effort. However, to achieve this, accurate mathematical models of the chromatographic process are needed. Much of the work published on developing such models considers only analytical scale separations and further experimental work often needs to be performed to obtain all the required model parameters. Little research has been conducted on developing models for more complex separations. Thus, there is a need for a systematic approach to: (1) determine the model parameters in chromatography models from experimental data and (2) to select the most appropriate model for a given application.

In this work, a novel systematic approach is proposed for obtaining accurate model parameters from experimental and literature data based on commonly used chromatography models—demonstrated here using the equilibrium-dispersive (ED) model and the general rate (GR) model. A key feature in the approach is a procedure for approximating the feed concentration for mixtures where this is unknown, as is common for complex mixtures.

The objectives of this work are therefore to demonstrate: (1) a procedure for determining the feed concentration of an unknown mixture, (2) a procedure for determining model parameters including isotherm parameters, and (3) a graphical selection procedure for determining which model is most appropriate for a given process. The overall approach provides a tool for model development, as well as establishing guidelines for how to determine which model is appropriate for a given application.

Chromatography Models in Literature

Several mathematical models of chromatographic processes are available in the literature and have been applied successfully for the calculation of the band profiles obtained in chromatography.[1] Models studied in the literature include the ideal model, the ED model, and the GR model, with the latter two being most widely used.

The ED model accounts only for a finite extent of axial dispersion as it uses an apparent lumped dispersion coefficient to account for any band-broadening effects. It is accurate provided the mass transfer in the chromatographic column is controlled only by molecular diffusion across the mobile phase flowing around the packing particles and that the exchange between the stationary and mobile phases is very fast.[1] The model has been used extensively and verified against experimental data, and the success in employing the ED model in modeling and optimizing preparative chromatography has been clearly demonstrated.[2–5]

The GR model is widely acknowledged as being the most comprehensive among the chromatography models available in the literature[1,6,7] as it accounts for axial dispersion and all the mass transfer resistances, e.g. external mass transfer of solute molecules from bulk phase to the external surface of the adsorbent, diffusion of the solute molecules through the particle pores, and adsorption-desorption processes on the site of the particles. It has been successfully employed in many chromatography applications including complex processes such as simulated moving bed chromatography.[6,8–13] However, the GR model requires knowledge of a relatively large number of parameters to characterize the axial dispersion, external mass transfer and effective diffusion through the pores, and some of these parameters are difficult to measure accurately.[6] The GR model also involves two partial differential equations in its mass balances, and is therefore more computationally demanding than the simpler chromatography models.

Although the ED and GR models are the most popular in the literature, there are trade-offs associated with each. The ED model can be solved in a short time but with only a fair degree of accuracy, while the GR model captures the process dynamics more accurately, as it takes into account all the mass transfer resistances in the process, but it is computationally expensive. The application of either model, however, also requires the knowledge of how to obtain the model parameters from experimental and literature data as well as how to select the most appropriate model for a given application.

There has been some work in the literature which has compared the ED and GR models to other models such as the transport-dispersive model, which uses a lumped dispersion coefficient and a lumped mass transfer coefficient,[5,13] and the lumped pore diffusion model, which is a simplified form of the GR model employing a lumped overall mass transfer coefficient.[14] However, much of this work compares the models on a theoretical basis, or examines the conditions under which the models are applicable, using experimental results of simple systems on an analytical scale. There is thus a need for a systematic approach to discern which of the models are appropriate for complex separations, such as those found in biopharmaceutical applications, as well as how best to determine the parameters in the models. Note that, in this article, only the ED and GR models will be considered but the approach presented can be applied to any chromatography models irrespective of modeling complexities.

The approach can also be applied when studying isotherms, for the ED, GR, or any other model, and can be used to determine which isotherm model best describes a separation. This problem was also considered by Forseen et al.[15] although their ED model was simplified, as was the calculation of the Jacobian, to reduce the calculation time. Their approach was demonstrated on two systems, but both with only two components.

Systematic Approach for Model Parameter Determination

In this section, our approach for determining model parameters in a chromatography model is proposed. In the current literature, there are no methodical means by which unknown model parameters can be estimated for a mixture which consists of many unknown components. For example, for complex biomixtures containing several impurities, the actual composition of the feed is often unknown. The proposed approach details how the feed concentration may be approximated and demonstrates that detailed information of all the components in the mixtures is not necessary for successful model development. Nevertheless, some information is still needed for an accurate representation of the process, as will be discussed shortly.

The approach distinguishes between known and unknown model parameters, and these may again either be common to all the models considered or distinct to only one model.

Table 1. Equilibrium-Dispersive Model: Distinct Model Parameters and Correlations

Parameters	Symbol	Correlation
Total porosity[13]	ε_T	$\varepsilon_T = \varepsilon_B + (1-\varepsilon_B)\varepsilon_P$
Plate number[16]	N_p	$N_p = 5.54\left(\frac{t_R}{w_{1/2}}\right)^2$
Apparent dispersion coefficient[1]	$D_{ap,i}$	$D_{ap,i} = \frac{uL}{2N_p}$
Dead time[9]	t_0	$t_0 = \frac{\pi d^2 L \varepsilon_B}{4Q}$

Table 2. Rate Model: Distinct Model Parameters and Correlations

Parameters	Symbol	Correlation
Particle porosity[1]	ε_P	0.5
Bed porosity[17]	ε_B	0.4
Axial dispersion coefficient[18]	$D_{ax,i}$	$\frac{vL}{D_{ax}} = \frac{L}{2R_p\varepsilon_B}(0.2 + 0.011Re^{0.48})$
Effective diffusivity coefficient[19]	$D_{e,i}$	$D_{e,i} = D_{m,i}(1 - 2.104\lambda + 2.09\lambda^3 - 0.95\lambda^5)/\tau_{tor}$
Mass transfer resistance coefficient[20]	$K_{pm,i}$	$K_{pm,i} = 0.687\nu^{1/3}\left(\frac{\varepsilon_B R_P}{D_{m,i}}\right)^{-2/3}$

Known model parameters are those that can be determined from the literature or the experimental set up. Some parameters are common in both the ED and the GR models and known common model parameters, such as column length L, column diameter D_C, particle radius R_P, and flow-rate of the mobile phase Q, are easily found from the experimental set-up. Other common model parameters, such as the feed concentration of the components and the isotherm parameters, may not be readily known, and are denoted unknown common model parameters and these must be estimated.

The distinct model parameters, which are the parameters characteristic of each model include the apparent axial dispersion coefficients in the ED model and the mass transfer resistance coefficients in the GR model. In this article, these parameters are determined using literature correlations (summarized in Tables 1 and 2), although as previously mentioned, they may also be estimated as part of the approach.

Procedure for feed concentration determination

In most industrial separations, the total concentration of the overall feed mixture is normally known. However, in complex separations where the feed contains impurities or many unknown components, the relative amount of each component or impurity is not generally known. To identify and model all of the components, and their associated model parameters, is difficult and time consuming. Thus a procedure to determine which are the necessary components for accurate modeling is needed.

The procedure developed in this article for feed concentration determination assumes that, for a mixture containing several unknown components in the feed, a set of chromatographic data in the form of chromatograms exists. The procedure also assumes that the identities of some of the components are known and chromatographic data on at least one component as well as that of the total mass of components eluted is required. The data itself, however, may provide little information as to the actual make-up of the feed mixture.

Figure 1 outlines the procedure for feed concentration determination from experimental chromatograms where some of the feed component concentrations are unknown. The procedure consists of a series of steps which are detailed below (each step is illustrated in detail in Case study 1 later):

Step 1: In the first step, input parameters are identified and defined: N_T is the assumed total number of components in the feed mixture and N_{NP} is the number of peaks on the chromatogram. R_C is a confidence ratio and is employed in the procedure to determine a reasonable number of components for which the estimated parameters are likely to be statistically significant. If too many components are fitted to too little data, the parameter values become meaningless as they are not statistically significant. The confidence ratio R_C ensures that this will not be the case. Care must be taken, however, in choosing R_C and some trial-and-error with larger or smaller R_C values may be necessary. (It is our experience that an R_C value of 2 ensures sufficient accuracy in the estimation.)

Step 2: In the next step, known components in the feed mixture which are likely to be retained on the column are identified. The characteristic property of the chromatographic separation will be known at this stage, e.g. affinity for stationary phase, molecular size, or hydrophobicity, as these characteristics determine if the components will be retained. The number of unretained components is identified as N_R.

Step 3: In the third step, the number of feed components appearing in negligible amounts, N_S, is identified. The total number of components which bind to the column and are subsequently eluted to form the chromatogram is then N_C, or $N_C = N_T - N_R - N_S$. Thus, components small in quantity (N_S) or not retained on the column (N_R) are not included in the model.

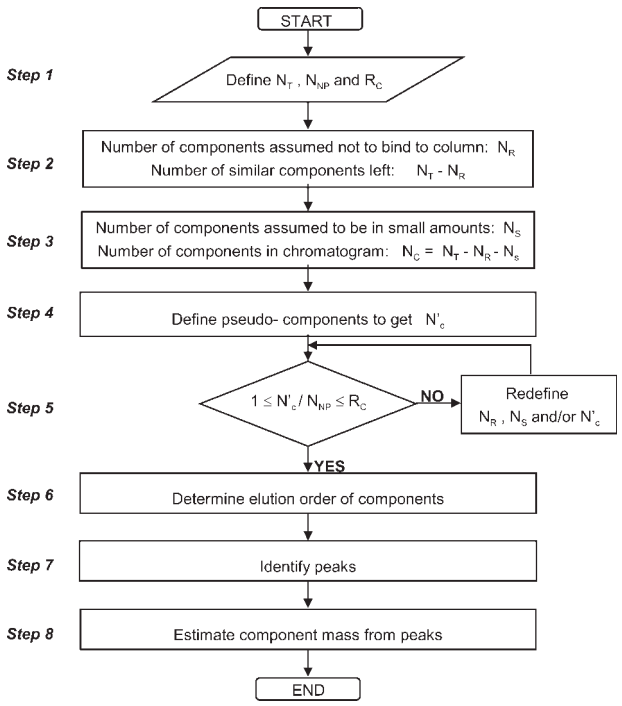

Figure 1. Procedure for determining feed concentration from an experimental chromatogram.

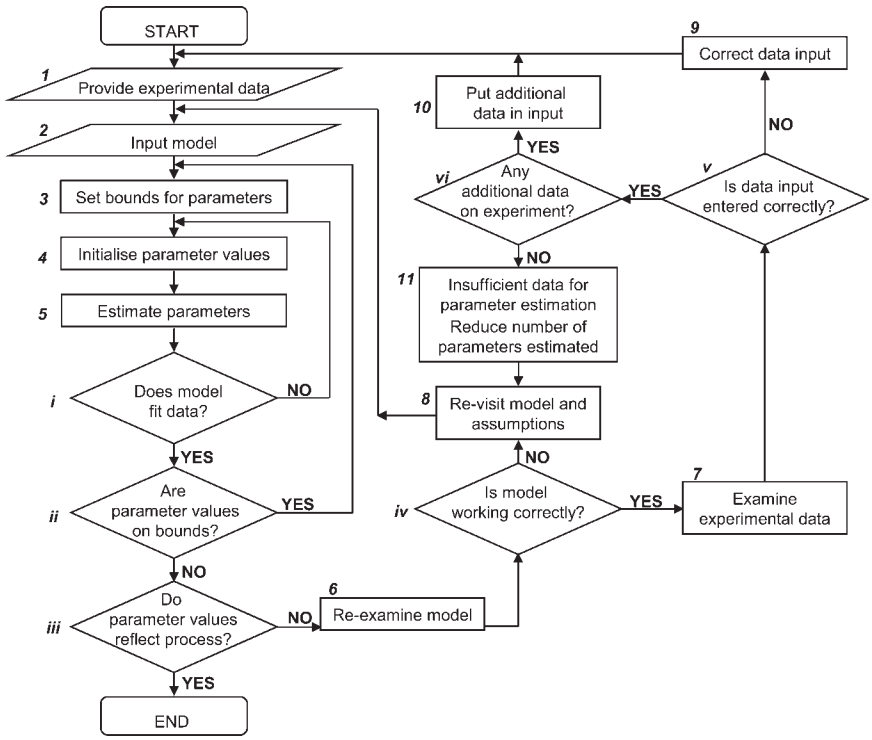

Figure 2. Process flowchart for estimating model parameters.

Step 4: In the fourth step, pseudocomponents are defined by lumping together components which have similar properties as they will elute in the same manner. This gives N'_C the number of components to be modeled. The lumping reduces the number of variables which need to be considered in the simulation and thereby reduces the complexity of the problem. This is required because, if too many parameters were to be estimated from limited data, the resulting predictions would not be accurate enough.

Step 5: In the fifth step, the ratio of the number of modeled components to the number of peaks on the chromatogram, N'_C/N_{NP}, is determined. N'_C cannot be less than N_{NP} as this would suggest that there are fewer components in the model than in the experimental data. However, at the same time, the estimated parameter values obtained may not be statistically significant if too many parameters are estimated simultaneously. Thus, a decision block is set up where the calculation continues to Step 6 only if the ratio N'_C/N_{NP} fulfils the criterion of a lower bound at 1 and an upper bound at R_C. Otherwise, the process is repeated from Step 2, and the number of components considered is re-evaluated. (It is our experience that an R_C value of 2 ensures sufficient accuracy in the estimation.)

Step 6: In this step, the components (both pseudocomponents and real) from Step 5 may now be regarded as the total number of components present in the separation for the purposes of modeling the process. The magnitude of the characteristic property of the components (e.g. affinity, molecular size, or hydrophobicity) is used to determine the interactions of the components with the stationary phase, and thus the order in which they will elute from the column.

Step 7: In the seventh step, the chromatogram peaks are identified by the order of elution of the components which was determined in Step 6. Determining which component(s) belongs to which peak when there are more components than peaks, is done by considering the relative amounts present in the feed and the size of the peaks. If this is ambiguous, then some trial-and-error may be necessary.

Step 8: In the final step, the feed concentration for each component is calculated using the trapezium rule on the area under each peak on the chromatogram.

The procedure is illustrated in three case studies shortly.

Procedure for parameter estimation of unknown parameters

Having obtained both the common and distinct model parameters from either the experimental setup or from literature correlations, what remains to be determined are normally the isotherm parameters and these are often determined experimentally. However, in cases where only limited chromatographic data is available and no further experimental work can be done, parameter estimation may be used to determine the unknown isotherm parameters. Parameter estimation refers to the process of obtaining the unknown values of the parameters by matching the model predictions to the available sets of experimental data.[15]

In the approach proposed in this article, it is assumed that the values of the isotherm parameters are not known, and that these need to be determined by parameter estimation. Should the isotherm parameters already be known, parameter estimation may be used to estimate other coefficients, e.g.

Table 3. Case Study 1: Dimensions of Column

Parameter	Symbol	Units	Value
Column length	L	cm	15
Column diameter	D_C	cm	60
Particle radius	R_P	μm	45
Volumetric flowrate	Q	ml/min	5,640 (about 120 cm^3/h)
Load volume		ml	148,000
Total component concentration		mg/ml	15

the effective diffusivities in the GR model. Such coefficients are usually not included because the increase in the number of estimated parameters based on limited experimental data renders the parameter values obtained statistically unreliable.

In this section, the proposed procedure for estimating unknown model parameters is outlined as shown in Figure 2. All the trapeziums and boxes are labeled numerically, while the decision blocks (diamond-shaped figures) are labeled using Roman numerals. The procedure is as follows:

Trapezium 1: Compile the experimental chromatographic data of the process.

Trapezium 2: Program the model of the process in the chosen modeling software (in this work, gPROMS is used[21]).

Box 3: Set the bounds for the parameters to be estimated. The parameters are estimated within the range imposed by the bounds.

Box 4: Initialize the values of the estimated parameters. During all the test runs, the parameter estimation will use these values as initial guesses.

Box 5: Carry out the parameter estimation (in this work, gPROMS is used[21]).

Decision i: Determine if the model prediction fits the shape of the experimental data given in Trapezium 1. If it does, proceed to the next step; otherwise return to Box 4 and change the values for initialization.

Decision ii: Check that the estimated parameter values do not lie on the bounds set in Box 3. If they do, return to Box 3 to increase the bounds; if not proceed to the next step. (If values lie on the bound, it may mean the parameter estimation has been constrained by the bounds on that parameter.)

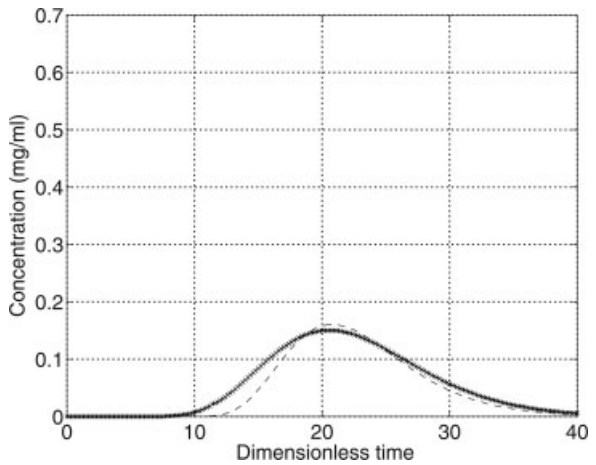

Figure 3. Component 6 chromatogram (x, experimental data; —, GR model; ---, ED model).

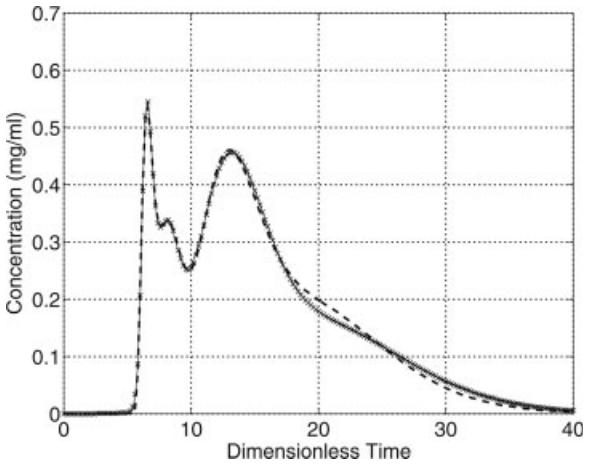

Figure 4. Total mixture chromatogram (x, experimental data; —, GR model; ---, ED model).

Decision iii: Check that the estimated parameter values obtained reflect the process accurately based on previous knowledge of the process. With the limited data provided for parameter estimation, the estimated parameters may not reflect certain aspects of the behavior correctly, e.g. the elution order or band width. If it does reflect the actual process, then the parameter estimation comes to an end. However, if it does not, then proceed to Box 6.

Box 6: Examine the model equations to determine why there is a discrepancy between the model prediction and the experimental data.

Decision iv: If the model is working correctly but the discrepancy is still present, proceed to Box 7. If it is determined not to be working correctly, then proceed to Box 8.

Box 7: Examine that the experimental data used in the estimation is correctly entered and proceed to Decision v.

Box 8: Revisit the model and the modeling assumptions. Return to Trapezium 2.

Decision v: If the experimental data input has been entered wrongly for the parameter estimation, proceed to Box 9. If the experimental data given is correct, proceed to Decision vi.

Box 9: Correct the experimental data input and return to Trapezium 1.

Table 4. Case Study 1: Known Components With Characteristic Property

Component	Molecular Weight (Da)	Amount Present (%)	Characteristic Property Value
A	50,000	5	7
B	150,000	5	7
C	55,000	11	8
D	15,000	8.3	9
E	40,000	9	10
F	10,000	23	11
G	30,000	7	12
H	6,000	5	12
I	150,000	10	1
J	62,000	15	2
K	23,000	0.5	7
L	16,000	1.2	9
		$\sum = 100$	

Table 5. Case Study 1: Pseudocomponents and Feed Component Concentrations

Component	Pseudocomponent	Average Molecular Weight	Amount Present (%)	Characteristic Property	Concentration (mg/ml)
A and B	1	100,000	10	7	1.23
C	2	55,000	11	8	1.50
D	3	15,000	8.3	9	2.13
E	4	40,000	9	10	1.14
F	5	10,000	23	11	3.20
G and H	6	20,000	12	12	1.80

Decision vi: Determine if there is any additional data available on the experiment. If there is additional data, proceed to Box 10. If there is no additional data on the experiment, proceed to Box 11.

Box 10: Add the additional experimental data to the input data and return to Trapezium 1.

Box 11: There is probably insufficient data for accurate parameter estimation. Since there is no additional data to be gathered, reduce the number of parameters to be estimated by making additional simplifying assumptions, e.g. by fixing some of the parameters, or redefining the pseudocomponents. This entails returning to Box 8 and revisiting the model.

The procedure ends successfully when the parameters estimated reflect the process at Decision iii to a predetermined level of tolerance, or unsuccessfully if there is insufficient data for parameter estimation.

Procedure for model selection

The procedures outlined in the previous sections are used to obtain the model parameters for all candidate models (in this work, the ED and GR models). The predictions from each model are then compared with the original experimental data to determine which of the models best predict the data. However, merely examining the goodness-of-fit of the model against the experimental data is not sufficient. Since a difference between predicted and experimental data is inherent, the crucial issue is which of the models will be the closest fit to the data and this may not be easily discernable by merely examining the chromatograms. What is needed is a means by which any deviation of the predicted data from the experimental data is highlighted.

Fractionation and maximum purification factor diagrams are recent graphical methods used to illustrate the trade-off between purity and recovery in chromatographic performance.[22] The diagrams are highly sensitive to how small changes in the chromatogram affect the purity and recovery of the process. Thus, by using these diagrams to compare the predicted and experimental data, the closeness between the predications and reality is made much more evident. This in turn will help the user to decide which of the models is best suited to describe the process as illustrated in the following case studies.

Case Studies

In the previous section, a systematic approach has been described where model parameters are estimated and the most appropriate model selected for a given separation process. In this section, the approach is illustrated for three case studies. The first case study is theoretical and is included to illustrate the steps in the approach, as well as to validate the parameter estimation technique. The second and third case studies employ real data using two different types of complex chromatography separations to illustrate the suitability of the approach.

Case study 1

The first illustration is a theoretical case study of modeling a step elution hydrophobic interaction chromatographic separation of 12 components, of which only the amount of one component (Component 6) is known. The dimensions of the column are shown in Table 3. The "experimental" data for this case study was generated using the GR model and are

Table 6. Case Study 1: Distinct Model Parameters

Equilibrium-Dispersive Model Parameters							
Parameter	Symbol	Value					
Total porosity	ε_T	0.7					
Dead time (s)	t_0	315					
Plate number	N_p	12.4					
Apparent dispersion coefficient $\times 10^3$ (cm^2/s)	$D_{ap,i}$	29					
General Rate Model Parameters							
Parameter	Symbol	Value					
Particle porosity	ε_P	0.5					
Bed porosity	ε_B	0.4					
Axial dispersion coefficient $\times 10^3$ (cm/s)	$D_{ax,i}$	1.48					
Components	i	1	2	3	4	5	6
Effective diff. coefficient $\times 10^7$ (cm^2/s)	$D_{e,i}$	7.64	1.08	2.04	1.27	2.43	1.78
Mass transfer coefficient $\times 10^3$ (cm/s)	$K_{pm,i}$	1.43	1.63	2.18	1.75	2.38	1.04

Table 7. Case Study 1: Comparison Between Model Parameters and Estimations for Isotherm Parameters*

	Parameter Estimation				Data Generation
Parameter	Initial Value	Lower Bound	Upper Bound	Estimation	GR Model
$K1_1$	90	65	250	87.2	100
$K1_2$	90	80	250	109	100
$K1_3$	90	80	250	104	100
$K1_4$	90	80	250	125	100
$K1_5$	90	80	250	90	100
$K1_6$*	90	80	250	100	100
$K2_1$	0.1	0.05	1	0.14	0.1
$K2_2$	0.1	0.05	1	0.26	0.3
$K2_3$	2	0.1	5	2.45	2.5
$K2_4$	2	0.1	30	6.45	8
$K2_5$	2	0.1	30	10.1	10
$K2_6$*	2	0.05	30	22.0	22

*Experimental Data Supplied.

given in Figure 3 for Component 6 and in Figure 4 for the total mixture. To reflect the limitations encountered in gathering industrial data, the only data used in the approach is of the one known component and of the total concentration. The primary purpose of this case study is to demonstrate the viability of the parameter estimation technique by comparing the estimated values obtained for the ED and GR models using the approach with the actual values employed in the GR model when generating the data.

Determining the Feed Concentration. Table 4 lists the components, and their known characteristic properties, which are present in the separation. The value of the property is used as an indication of the strength of a component's affinity for the stationary phase, with higher values having a stronger affinity for the stationary phase. Components with very low values do not bind to the stationary phase but rather go straight through the column. As they do not appear in the elution chromatogram, they are thus not considered in the simulation. Figure 1 showed the procedure for feed concentration determination which will now be applied to this case study:

Step 1: N_T is the total number of components in the feed mixture, $N_T = 12$. N_{NP}, the number of peaks on the chromatogram, can be observed from Figure 4 to be 4 distinguishable peaks. R_C, the confidence ratio, is defined as 2.

Step 2: From Table 4, Components *I* and *J* have separation properties which are much lower than the other components and are therefore considered to go straight through the column, i.e. $N_R = 2$. The number of components which will bind to the column is thus $N_T - N_R = 12 - 2 = 10$.

Step 3: In the next step, negligible amounts of components that are present in the mixture are grouped. From Table 4, the amounts of Components *K* and *L* are significantly lower compared to the rest of the component and these components are therefore regarded as negligible in the process, thus $N_S = 2$. The number of components eluting from the column is N_C, i.e. $N_C = N_T - N_R - N_S = 8$. Thus, components small in quantity (N_S), and not retained on the column (N_R), are not included in the model. (Note that if impurities are present which will have a significant impact on the separation in the column, they should be considered even if present in only very small quantities.)

Step 4: Components with similar chromatographic properties are lumped to form pseudocomponents. Components A and B are lumped to form Component 1, while G and H are lumped to form Component 6. (Table 5 lists the pseudocomponents.) N'_C, the number of components used in the model, is thus 6. (The molecular weights of the pseudocomponents are here assumed to be the arithmetic averages of the original components.)

Step 5: The ratio of the number of modeled components to the number of peaks on the chromatogram, N'_C/N_{NP}, is 6/4. This value is less than the confidence ratio R_C of 2, thus the six modeled components can be fitted to the four peaks, and the calculation continues to Step 6.

Step 6: The components (both pseudocomponents and real) from Step 5 may now be regarded as the only components present in the separation for the purpose of modeling the process, hence $N'_C = 6$. The magnitude of the characteristic property of the components (which the separation is based on) is used to obtain an estimate of the order in which the components elute from the column (see Table 5).

Table 8. Case Study 1: Comparison Between Model Parameters and Estimations for General Rate Model Diffusivity Coefficients*

	Parameter Estimation ($\times 10^{-7}$)			Data Generation	
Parameter	Initial Value	Lower Bound	Upper Bound	Estimation	GR Model
$D_{c,1}$	0.2	0.1	10	0.97	0.72
$D_{c,2}$	1.0	0.1	10	0.78	1.05
$D_{c,3}$	2.0	0.1	10	1.92	2.02
$D_{c,4}$	1.0	0.1	10	1.18	1.25
$D_{c,5}$	2.0	0.1	10	2.47	2.43
$D_{c,6}$*	2.0	0.1	10	1.78	1.77

*Experimental Data Supplied.

Table 9. Case Study 2: Dimensions of Column

Parameter	Symbol	Units	Value
Column length	L	cm	12.5
Column diameter	D_C	cm	1.6
Particle radius	R_P	μ m	45
Volumetric flowrate	Q	ml/min	2.5
Load volume		ml	~250 (10CV)
Total component concentration		mg/ml	10

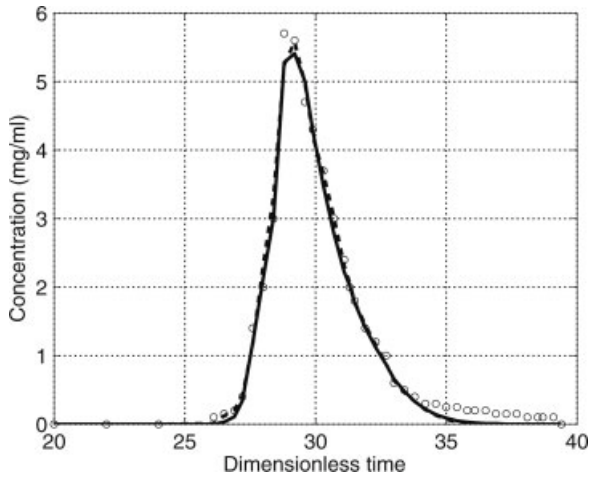

Figure 5. Case study 2: Total protein chromatogram (○, experimental data; —, GR model; ---, ED model).

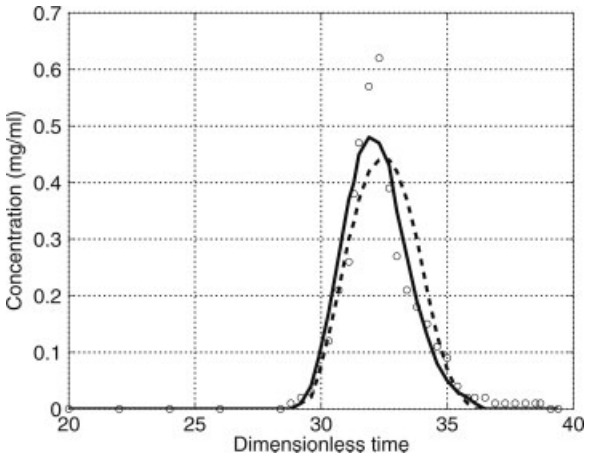

Figure 6. Case study 2: Alcohol dehydrogenase chromatogram (○, experimental data; —, GR model; ---, ED model).

Step 7: The chromatogram peaks are identified by the order of elution of the components—Peak 1 is made up of Components 1 and 2 (i.e. pseudocomponent of A and B and Component C); Peak 2 is Component 3 (Component D), Peak 3 is Components 4 (Component E) and 5 (Component F), while Peak 4 is Component 6 (pseudocomponent of G and H).

Step 8: The feed concentration for each component is calculated using the trapezium rule on the area under each peak on the total component mass and Component 6 chromatograms, and are given as the last column of Table 5.

Determining the Isotherm Parameters. The common and distinct model parameters can be determined from the experimental data and literature correlations. For this case study, the distinct parameters for both the ED and the GR models were obtained using the correlations in Tables 1 and 2 and the results are shown in Table 6.

The remaining parameters to be estimated are the isotherm parameters. Any isotherm model may be employed when using the approach presented in this article although a linear isotherm is used here to simplify this case study. The process uses step elution, where the mobile phase, and therefore also the isotherm, is changed during the process. Thus, for each component there are two isotherm parameters to be estimated, giving a total of 12 estimated parameters.

It is assumed that the final component eluted, Component 6, is the desired product. As data on all components are usually not available in a real process, only chromatograms for Component 6 and the total protein concentration are used for the model parameter estimation as explained earlier. The model predictions for both the ED and the GR model are given together with the experimental data in Figures 3 and 4. It can be seen that the ED model predicts the general trend for the total protein concentration (Figure 4) well. However, it is poorer at capturing the process dynamics for the individual elution profile of Component 6 (Figure 3), where the start of elution is predicted to be after the actual elution process.

Parameter Estimation Verification. The main purpose of this case study is to validate the parameter estimation technique used. Table 7 shows the original GR model parameters used to generate the experimental data, and the GR model parameters estimated using only the data for Component 6 and for the total protein (i.e. Figures 3 and 4). Different initial values for the estimated parameters were used in the estimation. By comparing the estimated parameters (second to the last column) with the values used when gen-

Table 10. Case Study 2: Pseudocomponents

Comp. No.	Feed Conc. (mg/ml)	Mol. Weight (Da)	Gene Name	Description	Charge Value*	Protein Abundance† (10^3 copies/cell)
1	0.03	5,500	PYK1	Pyruvate kinase	4.5	225.3
			YLR109W	Alkyl hydroperoxide reductase	−4.5	94.4
2	0.2	69,000	ENO2	Enolase II	4	775
			PGK	Phosphoglycerate kinase	−4	338.9
3	0.2	61,000	PDC1	Pyruvate decarboxylase	−3	326
4	0.2	35,800	TDH2	Glyceraldehyde-3-phosphate dehydrogenase 2	2	992.5
5	0.06	36,800	ADH1	Alcohol dehydrogenase	1	887.8
			PSA1	Mannose-1-phosphate guanyltransferase	−1	96.4

*Values from GeneED database.
†Values from Gygi et al.[24]

Table 11. Case Study 2: Distinct Model Parameters

Equilibrium-Dispersive Model Parameters						
Parameter	Symbol	Value				
Total porosity	ε_T	0.7				
Dead time (s)	t_0	243				
Plate number	N_p	108				
Apparent dispersion coefficient $\times 10^3$ (cm^2/s)	$D_{ap,i}$	3				
General Rate Model Parameters						
Parameter	Symbol	Value				
Particle porosity	ε_P	0.5				
Bed porosity	ε_B	0.4				
Axial dispersion coefficient $\times 10^3$ (cm/s)	$D_{ax,i}$	0.925				
Components	i	1	2	3	4	5
Effective diffusivity coefficient $\times 10^7$ (cm^2/s)	$D_{e,i}$	1.00	2.01	0.92	1.32	1.30
Mass transfer coefficient $\times 10^3$ (cm/s)	$K_{pm,i}$	1.02	1.32	1.00	1.12	1.12

erating the data (last column), it can be seen that the parameter estimation technique is very successful. It can also be seen that the best fit is obtained for Component 6, which is expected as this is the only component where individual data was available.

The parameter estimations for the ED model and the GR model take 1226 and 6827 CPU seconds, respectively, showing that the ED model is still useful for general predictive purposes, with a faster computational time, although the predictions are not as accurate.

Determining Other Parameters. In the previous section, parameter estimation was used to estimate the isotherm parameters, while other parameters were predicted using literature correlations. If, however, the isotherm parameters are already known, the procedure can instead be used to predict other model parameters. Estimating too many parameters numerically will render the values obtained statistically insignificant with low confidence levels. Therefore, predicting both isotherm parameters and diffusivity parameters from the same data is not recommended.

Table 8 shows a comparison between coefficients for the GR model used when generating the "experimental" data and those estimated using the parameter estimation procedure assuming that the isotherm parameters are known. It can be seen that again a good agreement is found.

Table 12. Case Study 2: Values Used in the Parameter Estimation for the Equilibrium-Dispersive and the General Rate Models

	Parameter estimation			
	Initial Value	Lower Bound	Upper Bound	Estimation
ED Parameter				
$b1_1$	5	2	50	47.9
$b1_2$	10	8	25	16.3
$b1_3$	10	8	30	16.8
$b1_4$	10	8	50	15.6
$b1_5$*	30	15	50	19.3
$b2_1$	0.5	0.3	10	0.88
$b2_2$	0.5	0.3	10	1.67
$b2_3$	0.5	0.3	10	0.98
$b2_4$	0.7	0.5	10	1.29
$b2_5$*	1.6	1	10	2.27
C^∞	8.8	4	8.8	8.46
GR Parameter				
$b1_1$	45	10	100	19.58
$b1_2$	90	10	300	41.14
$b1_3$	155	20	300	280.31
$b1_4$	200	50	350	309.61
$b1_5$*	200	10	350	14.61
$b2_1$	0.071	0.03	0.2	0.042
$b2_2$	0.143	0.05	0.5	0.219
$b2_3$	0.200	0.05	0.5	0.418
$b2_4$	0.167	0.05	0.5	0.294
$b2_5$*	0.556	0.20	1.0	0.741
C^∞	8.8	4	8.8	7.79

Case study 2

The second case study considers the purification of a labile protein, alcohol dehydrogenase (ADH), from a yeast homogenate supernatant using hydrophobic interaction chromatography as described by Smith et al.[23] (data in Table 9). As experimental data, only the ADH and total protein chromatograms are available (Figures 5 and 6).

Determining the Feed Concentration. Figure 5 identifies the number of peaks N_{NP} as 3; one very small peak in the front, a tall one in the center of the chromatogram, and the third mostly made up of ADH. The feed mixture is yeast homogenate. The work of Gygi et al.[24] provides an extensive list of genes of proteins found in yeast *Saccharomyces cerevisiae*. Using the figures from their article, the total number of proteins, N_T is found to be 127. It is not known which of the yeast proteins (ADH aside) will bind to the column. Hence the characteristic property is deduced from the charge on the protein, listed on the GeneDB website (http://www.genedb.org/). The charge value provides an indication as to the hydrophobicity of the protein, and thus, whether or not it binds to the column. ADH has a charge of 1. Proteins with charge values close to this will be expected to bind to the column. Components with a charge greater than 4 are considered to be proteins which do not bind and these are grouped to form N_R, which is found to be 62. Fifty-seven components are present in negligible amounts (N_S) based on the protein abundance per cell given by Gygi et al.[24] Thus the number of components binding to the column, $N_C = N_T - N_R - N_S$

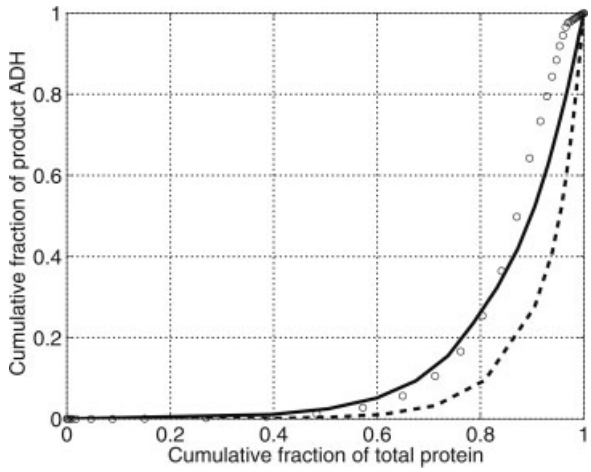

Figure 7. Case study 2: Fractionation diagram of ADH against total protein (○, experimental data; —, GR model; ---, ED model).

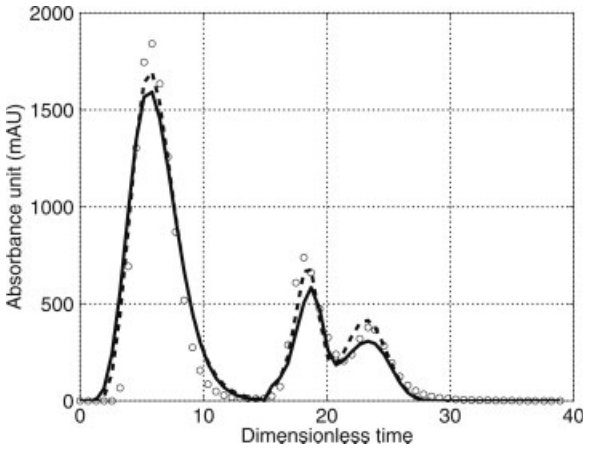

Figure 9. Case study 3: Egg white protein chromatogram (○, experimental data; —, GR model; ---, ED model).

= 8. Three pseudocomponents are formed, each consisting of two components, to give the number of components used in the model, $N'_C = 5$. The data for these components are given in Table 10. The ratio $N'_C/N_{NP} = 5/3$ and the criteria $1 \leq N'_C/N_{NP} < R_C$, where R_C is 2, is fulfilled.

Five components are thus to be modeled and the feed concentrations of these are determined from the chromatogram. The first peak contains Component 1, the second peak is made up of Components 2, 3, and 4, while Component 5 (ADH) is the third peak. The concentrations are determined from the area under each peak using the trapezium rule. It is assumed that the feed concentrations of Components 2, 3, and 4 are equally divided to form the concentration of the second peak as there is no means of ascertaining their individual concentrations.

Determining the Isotherm Parameters. The distinct parameters for both models were obtained using the correlations in Tables 1 and 2 and are shown in Table 11. The competitive Langmuir isotherm was used in this case study and the parameters estimated are shown in Table 12 together with the initial values and the lower and upper estimation bounds. Figures 5 and 6 show the comparison between the models predictions and the experimental chromatograms for ADH and total protein concentration. The figures show similar trends to those in Case study 1, with the ED model predicting the general trend well for the total protein concentration but showing a poorer fit for the ADH.

Comparing the performance of the ED and GR models, it is difficult to decide which model should be employed. Merely judging the "goodness-of-fit" against the experimen-

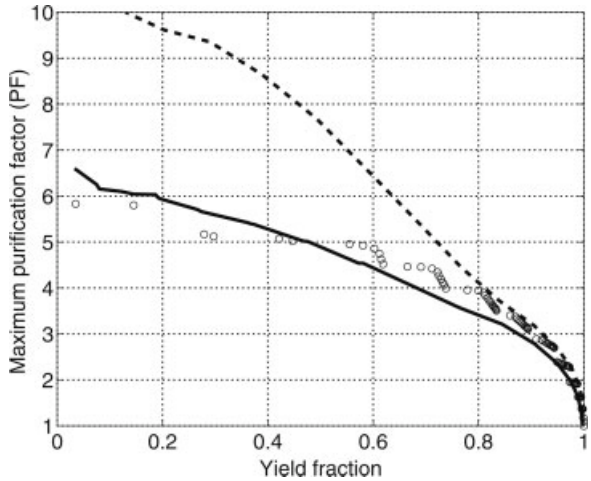

Figure 8. Case study 2: Maximum purification diagram for ADH (○, experimental data; —, GR model; ---, ED model).

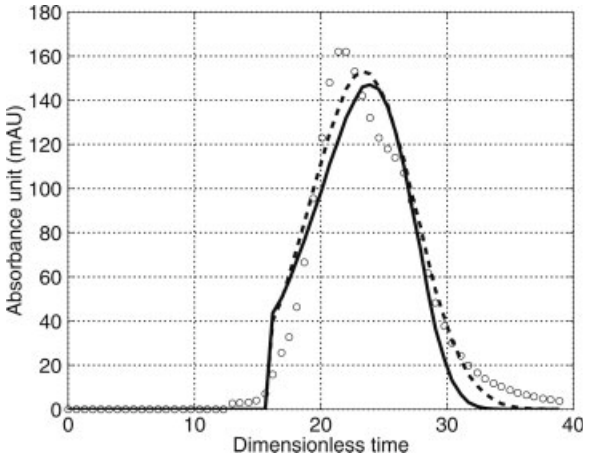

Figure 10. Case study 3: Myoglobin chromatogram (○, experimental data; —, GR model; ---, ED model).

Table 13. Case study 3: Feed Concentrations

Component Number	Protein	pI Value	Amount of Total }Egg White Protein (%)	Average Molecular Weight (M_R)	Extinction Coefficient (ϵ)	Feed Conc. (mg/ml)
1	Ovalbumin	4.5	54	40,000	32,180	0.03
	Ovomuscoid	4.1	11			
2	Ovotransferrin	6.0	12–13	77,700	85,000	0.2
3	Lysozyme	10.7	3.4–3.5	14,300	36,000	0.2
4	Myoglobin	7	–	16,000	160,000	0.2

pI value and protein amounts taken from Awadé and Efstathiou,[27] M_R and ϵ values from Fasman.[28]

tal chromatograms is subjective, especially if both models look very similar. A graphical representation of the chromatograms proposed by Ngiam et al.[22] was therefore employed to highlight any significant differences between the performances of the candidate models.

The fractionation diagram of the ADH product against the total product for both models and the experimental data is given in Figure 7 and shows the change in the elution of product (ADH) against that of the total protein mixture eluted. The maximum purification diagram is given in Figure 8 and shows the trade-off between yield and purity of ADH for this separation. It can be seen from the figures that the GR model predicts the yield/purity relationship more accurately than the ED model for the ADH protein. The area of interest in most bioseparations lies in the region where the yield fraction is high, i.e. 0.8–1. Both models show similar levels of predictive performance in this region, although the GR model is slightly superior. Thus, it may be concluded that, for this case study, the GR model is the most suitable model.

Case study 3

The last case study considers ion exchange chromatography (IEX) which is one of the most widely used techniques for the purification of proteins in the biotechnology industry and is generally employed in the first steps of large scale purification processes.[25] The case study examines the more demanding case of gradient elution ion-exchange chromatography for a mixture of proteins from the work of Edwards-Parton as described by Chan.[26] With gradient elution, the isotherm parameters change dynamically.

Determining the Feed Concentration. Figures 9 and 10 confirm that the number of peaks N_{NP} is 4; three of the peaks being egg white proteins, while the fourth peak is the myoglobin. The feed mixture contains as many as 15 egg white proteins, from which one protein is determined as being nonbinding to the column. This is deduced from the isoelectric points, or pI values, of the egg white proteins, given by Awade and Efstathiou.[27] Of the remaining proteins, nine are present in negligible amounts. Thus the number of component binding to the column, N_C, is 6. Two pseudocomponents are formed, giving the number of components used in the model N'_C equal to 4. The ratio $N'_C/N_{NP} = 4/4$ and the criteria $1 \leq N'_C/N_{NP} < R_{C,}$ where R_C is 2, is fulfilled. Four components are to be modeled and the feed concentrations of these are determined from the chromatogram and the results are shown in Table 13. Individual extinction coefficients have been used to account for the fact that the components do not have the same detector response. The distinct model parameters for both models are given in Table 14.

Determining the Isotherm Parameters. Figures 9 and 10 compare the model predictions against the experimental data for the egg white proteins and the myoglobin. The fractionation and maximum purification diagrams are given in Figures 11 and 12, with myoglobin as the product. (It should be noted that due to difficulties in estimating the parameters for

Table 14. Case Study 3: Distinct Model Parameters

Equilibrium-Dispersive Model Parameters					
Parameter	Symbol	Value			
Total porosity	ε_T	0.7			
Dead time (s)	t_0	15.43			
Plate number	N_p	7			
Apparent dispersion coefficient $\times 10^3$ (cm^2/s)	$D_{ap,i}$	28.5			
General Rate Model Parameters					
Parameter	Symbol	Value			
Particle porosity	ε_P	0.5			
Bed porosity	ε_B	0.4			
Axial dispersion coefficient $\times 10^3$ (cm/s)	$D_{ax,i}$	2.86			
Components	i	1	2	3	4
Effective diffusivity coefficient $\times 10^7$ (cm^2/s)	$D_{e,i}$	2.62	1.83	4.26	4.05
Mass transfer coefficient $\times 10^3$ (cm/s)	$K_{pm,i}$	2.19	1.89	2.75	2.68

the GR model, some of the distinct parameters were modified from those found using the correlations in Table 2 to better fit the experimental data. The gradient elution implemented in this separation means the isotherm parameters change dynamically during the process, which increases the difficulty of parameter estimation.)

The fractionation diagrams and maximum purification factor diagrams (Figures 11 and 12) demonstrate that the performance of both models in predicting the process behavior improves for higher yields. When IEX is used in the first few processing steps, it is important to keep the yield high (even if not necessarily very pure) so that valuable product is not lost, and thus the region of interest in the maximum purification diagram for bioseparations is usually a yield fraction of 0.8–1. From Figure 12, it is clear that both models behave similarly in this yield fraction region. From the results, it is noted that the ED model is capable of predicting the process behavior well for this case study. Given similar predictions for ED and GR models, the ED model is preferable as it is less complex and computationally inexpensive. One way of improving the fit of the models for this case study would be to find a more suitable isotherm than the Langmuir isotherm which has been used here, although this has not been considered in this work.

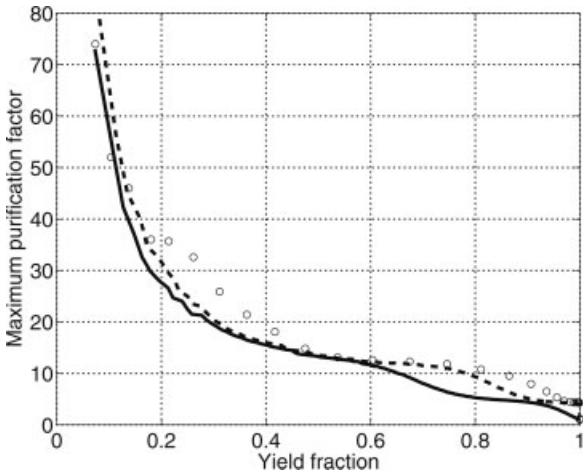

Figure 12. Case study 3: Maximum purification diagram for myoglobin (○, experimental data; —, GR model; ---, ED model).

Conclusions

This work has outlined an approach for modeling complex chromatographic separations based on a combination of experimental data, literature correlations, and parameter estimation. The approach allows the user to estimate model parameters and, subsequently, to evaluate the suitability of different models for a given process based on previously obtained experimental sets. The approach was applied to three case studies of increasing complexity in terms of experimental information and process dynamics. The approach was successful in determining the model parameters and model evaluation for all three processes. The GR model was shown to give slightly better predictions, but the difficulty in obtaining its model parameters means that, with similar model behavior, the ED model is preferable.

In the case studies, the approach was applied to evaluate the suitability of different chromatographic models, in particular, the ED and the GR models. The approach can also be applied to consider different isotherm models and can estimate the required isotherm parameter values as well as determine which isotherm best fits the experimental data.

This approach provides a useful tool to be used in conjunction with experimental work, thereby reducing the amount of experimentation required and thus saving both time and cost. Indeed, the software used for the parameter estimation in this work, gPROMS[21], has capabilities for design of experiments which can greatly assist in such work.

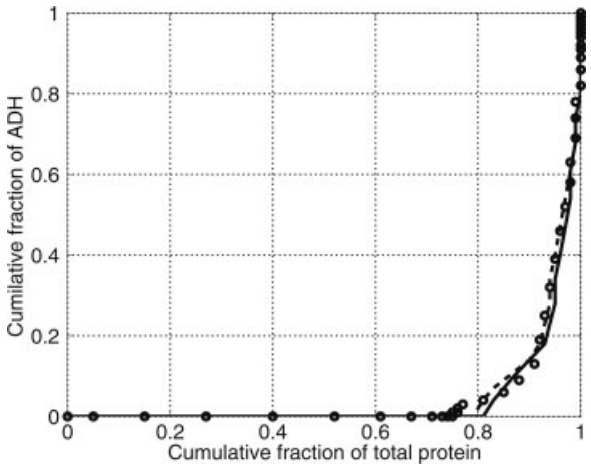

Figure 11. Case study 3: Fractionation diagram of myoblobin against total protein (○, experimental data; —, GR model; ---, ED model).

Acknowledgments

Grateful acknowledgments to Simon Edwards-Parton and R. Khanom of the Department of Biochemical Engineering for providing the experimental data used. Grateful acknowledgments to Simon Edwards-Parton for the use of his MATLAB program to generate the maximum purification diagrams, as well as discussions pertaining to the use of these diagrams. D.G. Bracewell and N. Titchener-Hooker acknowledge the support for the IMRC for Bioprocessing housed in The Advanced Centre for Biochemical Engineering by the Engineering and Physical Sciences Research Council (EPSRC) under the Innovative Manufacturing Research initiative.

Notation

D_C = diameter of the column (cm)
$D_{ap,i}$ = apparent dispersion coefficient (cm^2/s)
$D_{ax,i}$ = axial dispersion coefficient (cm/s)
$D_{e,i}$ = effective diffusivity coefficient of component i (cm^2/s)
$D_{m,i}$ = molecular diffusivity coefficient of component i (cm^2/s)
$K1_i$ = first isotherm coefficient in Case study 1

$K2_i$ = second isotherm coefficient in Case study 1
$K_{pm,i}$ = mass transfer resistance coefficient of component i
L = length of column (cm)
N_C = number of components retained
N'_C = number of components in model
N_{NP} = number of peaks on the chromatogram
N_p = plate number
N_R = number of components not retained on column
N_S = number of components with negligible amounts
N_T = total number of components
Q = flow-rate of mobile phase (mL/min)
R_C = confidence ratio
R_p = radius of the particle (cm)
Re = Reynolds number
t_0 = dead time (s)
t_R = retention time of component
u = interstitial velocity (cm/s)
v = interstitial velocity (cm/s)
$w_{1/2}$ = width of the peak at half peak height
ε_B = bed porosity
ε_P = particle porosity
ε_T = total porosity
λ = ratio of diameter of component to particle pore diameter
τ_{tor} = tortuosity of the component

Literature Cited

1. Guiochon G, Golshan Shirazi S, Katti AM. *Fundamentals of Preparative Nonlinear Chromatography*, 2nd ed. New York: Academic Press, 2006.
2. Katti AM, Diack M, Zoubair El Fallah M, Golshan-Shirazi S, Jacobson SC, Seidel-Morgenstern A, Guiochon G. Prediction of high concentration band profiles in liquid chromatography. *Acc Chem Res.* 1992;25:366–374.
3. Heuer C, Hugo P, Mann G, Seidel-Morgenstern A. Scale up in preparative chromatography. *J Chromatogr A*. 1996;752:19–29.
4. Teoh HK, Turner M, Titchener-Hooker N, Sørensen E. Experimental verification and optimisation of a detailed dynamic high performance liquid chromatography column model. *Comput Chem Eng*. 2001;25: 893–903.
5. Antos D, Kaczmarski K, Piątkowski W, Seidel-Morgenstern A. Concentration dependence of lumped mass transfer coefficients: linear versus non-linear chromatography and isocratic versus gradient operation. *J Chromatogr A*. 2003;1006:61–76.
6. Kaczmarski K, Antos D, Sajonz H, Sajonz P, Guiochon G. Comparative modeling of breakthrough curves of bovine serum albumin in anion-exchange chromatography. *J Chromatogr A*. 2001;925:1–17.
7. Teoh HK. Optimal design and operation of high performance liquid chromatographic processes, PhD Thesis. University College London, University of London, 2002.
8. Gu T, Tsai G-J, Tsao GT, Ladisch MR. Displacement effect in multicomponent chromatography. *J Am Inst Chem Eng*. 1990;36:1156–1162.
9. Li Z, Gu Y, Gu T. Mathematical modeling and scale-up of size-exclusion chromatography. *Biochem Eng J*. 1998;2:145–155.
10. Gu T. *Mathematical Modelling and Scale Up of Liquid Chromatography*. Berlin: Springer-Verlag, 1995.
11. Gu T, Zheng Y. A study of the scale-up of reversed-phase liquid chromatography. *Sep Purif Technol*. 1999;15:41–58.
12. Klatt K-U, Dünnebier G, Engell S. Modeling and computationally efficient simulation of chromatographic separation processes. *Math Comput Simul*. 2000;53:449–455.
13. Piątkowski W, Antos D, Kaczmarski K. Modeling of preparative chromatography processes with slow intraparticle mass transport kinetics. *J Chromatogr A*. 2003;988:219–231.
14. Kaczmarski K, Antos D. Modified Rouchon and Rouchon-like algorithms for solving different models of multicomponent preparative chromatography. *J Chromatogr A*. 1996;756:73–87.
15. Forseen P, Arnell R, Fornstedt T. An improved algorithm for solving inverse problems in liquid chromatography. *Comput Chem Eng*. 2006;30:1381–1391.
16. Synder LR, Kirkland JJ. *Introduction to Modern Liquid Chromatography*. New York: Wiley, 1974.
17. Unger KK. *Porous Silica: Its Properties and Use as Support in Column Liquid Chromatography (Journal of Chromatography Library S)*. Amsterdam: Elsevier, 1979.
18. Chung SF, Wen CY. Longitudinal dispersion of liquid flowing through fixed and fluidised beds. *J Am Inst Chem Eng*. 1968;14: 857–866.
19. Yau WW, Kirkland JJ, Bly DD. *Modern Size Exclusion Liquid Chromatography*. New York: Wiley, 1979.
20. Wilson EJ, Geankopolis CJ. Liquid mass transfer at very low Reynolds numbers in packed beds. *Ind Eng Chem Fundam*. 1966;5:9–14.
21. Process Systems Enterprise Ltd. *gPROMS User Manual*. London: Process Systems Enterprise Ltd., 2005.
22. Ngiam SH, Bracewell DG, Zhou Y, Titchener-Hooker NJ. Quantifying process tradeoffs in the operation of chromagraphic sequences. *Biotech Prog*. 2003;19:1315–1322.
23. Smith MP, Bulmer MA, Hjorth R, Titchener-Hooker NJ. Hydrophobic interaction ligand selection and scale-up of an expanded bed separation of an intracellular enzyme from Saccharomyces cerevisiae. *J Chromatogr*. 2002;968:121–128.
24. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mrna abundance in yeast. *Mol Cell Biol*. 1999;19:1720–1730.
25. Bonnerjea J, Oh S, Hoare M, Dunnill P. Protein purification: the right step at the right time. *Biotechnology*. 1986;4:954–958.
26. Chan S. Systematic approaches for modelling and optimisation of chromatographic processes, PhD Thesis, University College London, 2005.
27. Awadé AC, Efstathiou T. Comparison of three liquid chromatographic methods for egg-white protein analysis. *J Chromatogr B*. 1999;723:69–74.
28. Fasman DG, editor. *Practical Handbook of Biochemistry and Molecular Biology*. Boca Raton: CRC Press, 1992.

Manuscript received July 27, 2007, and revision received Dec. 29, 2007.