Preparative chromatography is the most widely used purification method in the biotechnology and pharmaceutical industry. Here, one theoretical and one practical problem of preparative chromatography are investigated. The theoretical problem deals with the effect of experimental error on the precision of adsorption isotherm parameters estimated by numerical method. Numerical method has becoming popular in recent years as a quick alternative to the classical dynamic methods for adsorption isotherm determination. However, the fundamental problem of the numerical back-propagation of error has not been addressed in the literature. In this study, the variance in the final estimated parameters as a function of experimental error is quantified. The results indicate that it is theoretically possible to obtain sufficiently accurate parameters from multiple experimental replicates. The number of experimental replicates required to achieve isotherm parameters with less than 5% error is calculated explicitly for a variety of data sets. By using computer simulated experimental data, the effect of error in effluent
concentrations and error in retention times are addressed separately. The practical problem is about the optimization of preparative purification of chlorophyll a by countercurrent chromatography (CCC). CCC offers several advantages over high performance liquid chromatography (HPLC) and is a fundamentally preparative separation tool. The process reported here is capable of producing several hundreds of milligrams of chlorophyll a with a high purity in a short time. The CCC separation is optimized for good production rate and recovery yield with respect to loading conditions.
Studies in Preparative Chromatography

By
Ling Zhang

A DISSERTATION
Submitted to
Oregon State University

in partial fulfillment of
the requirements for the
degree of
Doctor of Philosophy

Presented September 26, 2003
Commencement June 2004
Doctor of Philosophy dissertation of Ling Zhang presented on September 26, 2003

APPROVED:

Frank Chapin, Oct 9, 2003
Co-Major Professor, representing Bioresource Engineering

Co-Major Professor, representing Bioresource Engineering

John P. Bolts, Oct 6, 2003
Head of the Department of Bioengineering

Sally Francis
Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Ling Zhang, Author

10-9-2003
CONTRIBUTION OF AUTHORS

Dr. John Selker of the Department of Bioengineering and Dr. Annie Qu of the Department of Statistics were involved with the development of the statistical analysis used in Chapter 2 and Chapter 3.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>1. General Introduction</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>7</td>
</tr>
<tr>
<td>Theory</td>
<td>8</td>
</tr>
<tr>
<td>Simulations</td>
<td>23</td>
</tr>
<tr>
<td>Results and discussion</td>
<td>29</td>
</tr>
<tr>
<td>Conclusions</td>
<td>53</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>54</td>
</tr>
<tr>
<td>References</td>
<td>55</td>
</tr>
<tr>
<td>3. Effect of Experimental Error on the Precision of Numerically Estimated Multicomponent Isotherms in Preparative Chromatography</td>
<td>60</td>
</tr>
<tr>
<td>Introduction</td>
<td>61</td>
</tr>
<tr>
<td>Methodology</td>
<td>67</td>
</tr>
<tr>
<td>Results and discussion</td>
<td>75</td>
</tr>
<tr>
<td>Conclusions</td>
<td>85</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>86</td>
</tr>
<tr>
<td>References</td>
<td>87</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS (Continued)

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Preparative Purification of Chlorophyll A from Spinach by High Speed Countercurrent Chromatography</td>
<td>90</td>
</tr>
<tr>
<td>Introduction</td>
<td>91</td>
</tr>
<tr>
<td>Experimental</td>
<td>93</td>
</tr>
<tr>
<td>Results and discussion</td>
<td>97</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>114</td>
</tr>
<tr>
<td>References</td>
<td>115</td>
</tr>
<tr>
<td>5. General Conclusion</td>
<td>119</td>
</tr>
<tr>
<td>Bibliography</td>
<td>121</td>
</tr>
<tr>
<td>Appendix: Fitting the Error in Retention Times by a Free Parameter</td>
<td>130</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1.</td>
<td>Flowchart describing the process of estimating isotherm parameters numerically</td>
<td>15</td>
</tr>
<tr>
<td>2-2.</td>
<td>Probability plot for estimated isotherm parameters</td>
<td>19</td>
</tr>
<tr>
<td>2-3.</td>
<td>Simulated chromatogram for exact data I ($V_{feed} = 0.3V_0$)</td>
<td>31</td>
</tr>
<tr>
<td>2-4.</td>
<td>Simulated chromatogram for exact data II ($V_{feed} = 0.6V_0$)</td>
<td>34</td>
</tr>
<tr>
<td>2-5.</td>
<td>RSS as a function of $n$, the number of experiments needed for results listed in Table 2-3 (solid line) and Table 2-4 (dashed line)</td>
<td>36</td>
</tr>
<tr>
<td>2-6.</td>
<td>Variations in simulated parameters (scaled with respect to the exact parameters) and RSS</td>
<td>40</td>
</tr>
<tr>
<td>2-7.</td>
<td>Simulated chromatogram for exact data III ($V_{feed} = 0.4V_0$)</td>
<td>42</td>
</tr>
<tr>
<td>2-8.</td>
<td>Simulated chromatogram for exact data IV ($V_{feed} = 0.6V_0$)</td>
<td>44</td>
</tr>
<tr>
<td>2-9.</td>
<td>Simulated chromatogram for exact data V ($V_{feed} = V_0$) with 24 non-zero points</td>
<td>46</td>
</tr>
<tr>
<td>2-10.</td>
<td>Simulated chromatograms for exact data II and VI ($V_{feed} = 0.6V_0$)</td>
<td>48</td>
</tr>
<tr>
<td>2-11.</td>
<td>Plot of the histogram obtained from the 70 replicates used in Table 2-6 (exact data II with 10% relative error), and the corresponding density of the Monte Carlo distribution with the same covariance with 50,000 replicates (solid line)</td>
<td>51</td>
</tr>
<tr>
<td>2-12.</td>
<td>Plot of the histogram obtained from the 30 replicates used in Table 2-7 (combination of exact data I and II both with 5% relative error), and the corresponding density of the Monte Carlo distribution with the same covariance with 50,000 replicates (solid line)</td>
<td>52</td>
</tr>
<tr>
<td>3-1.</td>
<td>Simulated chromatograms</td>
<td>68</td>
</tr>
<tr>
<td>3-2.</td>
<td>Probability plot for estimated isotherm parameters</td>
<td>72</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3-3.</td>
<td>RSS as a function of n, the number of experiments needed for data with 2% (solid line) or 3% (dotted line) relative error in retention times. ( V_{\text{feed}}=0.6 \ V_0 )</td>
<td>76</td>
</tr>
<tr>
<td>3-4.</td>
<td>Variations in simulated parameters (standardized with respect to the exact parameter values)</td>
<td>78</td>
</tr>
<tr>
<td>3-5.</td>
<td>RSS as a function of n, the number of experiments needed for data with 5% relative error in effluent concentrations and 3% relative error in retention times. ( V_{\text{feed}}=0.6 \ V_0 )</td>
<td>80</td>
</tr>
<tr>
<td>3-6.</td>
<td>Simulated chromatograms</td>
<td>82</td>
</tr>
<tr>
<td>4-1.</td>
<td>Chromatogram of HPLC analysis of crude extract</td>
<td>99</td>
</tr>
<tr>
<td>4-2.</td>
<td>Chromatogram of crude extract by glass column chromatography...</td>
<td>101</td>
</tr>
<tr>
<td>4-3.</td>
<td>Chromatogram of HPLC analysis of glass column effluent</td>
<td>102</td>
</tr>
<tr>
<td>4-4.</td>
<td>CCC Chromatogram of preparative purification of chlorophyll ( a ). Feed volume: 60 ml; feed concentration: 8.7 mg/ml of chlorophyll ( a )</td>
<td>103</td>
</tr>
<tr>
<td>4-5.</td>
<td>Chromatogram of HPLC analysis of chlorophyll ( a ) fraction from CCC elution</td>
<td>104</td>
</tr>
<tr>
<td>4-6.</td>
<td>CCC chromatogram of separation of chlorophyll ( a ) and chlorophyll ( b ). Feed concentration: 1 mg/ml of chlorophyll ( a ); feed volume: 5 ml; UV wavelength: 440 nm</td>
<td>106</td>
</tr>
<tr>
<td>4-7.</td>
<td>CCC chromatogram of preparative purification of chlorophyll ( a ). Feed concentration: 9.0 mg/ml of chlorophyll ( a ); feed volume: 60 ml</td>
<td>110</td>
</tr>
<tr>
<td>4-8.</td>
<td>CCC chromatogram of preparative purification of chlorophyll ( a ). Feed concentration: 8.2 mg/ml of chlorophyll ( a ); feed volume: 58.5 ml; injection method: dynamic injection</td>
<td>112</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4-9.</td>
<td>CCC chromatogram of preparative purification of chlorophyll $\alpha$. Feed concentration: 8.1 mg/ml of chlorophyll $\alpha$; feed volume: 60 ml.</td>
<td>113</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>2-1.</td>
<td>Isotherm parameter set.</td>
<td>21</td>
</tr>
<tr>
<td>2-2.</td>
<td>Numerically simulated experimental data set.</td>
<td>21</td>
</tr>
<tr>
<td>2-3.</td>
<td>Number of experiments required for exact data I with 5% relative error.</td>
<td>30</td>
</tr>
<tr>
<td>2-4.</td>
<td>Number of experiments required for exact data I with 10% relative error.</td>
<td>33</td>
</tr>
<tr>
<td>2-5.</td>
<td>Number of experiments required for exact data II with 5% relative error.</td>
<td>35</td>
</tr>
<tr>
<td>2-6.</td>
<td>Number of experiments required for exact data II with 10% relative error.</td>
<td>37</td>
</tr>
<tr>
<td>2-7.</td>
<td>Number of experiments required for combination of exact data I and II both with 5% relative error.</td>
<td>38</td>
</tr>
<tr>
<td>2-8.</td>
<td>Number of experiments required for combination of exact data I and II both with 10% relative error.</td>
<td>38</td>
</tr>
<tr>
<td>2-9.</td>
<td>Number of experiments required for exact data III with 5% relative error.</td>
<td>42</td>
</tr>
<tr>
<td>2-10.</td>
<td>Number of experiments required for exact data III with 10% relative error.</td>
<td>43</td>
</tr>
<tr>
<td>2-11.</td>
<td>Number of experiments required for exact data IV with 5% relative error and less data points than that of Table 2-5.</td>
<td>44</td>
</tr>
<tr>
<td>2-12.</td>
<td>Number of experiments required for exact data V with 5% relative error.</td>
<td>45</td>
</tr>
<tr>
<td>2-13.</td>
<td>Number of experiments required for exact data VI with 5% relative error.</td>
<td>47</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>2-14</td>
<td>Number of experiments required for exact data VI with 10% relative error.</td>
<td>47</td>
</tr>
<tr>
<td>3-1</td>
<td>Competitive Langmuir isotherm parameters.</td>
<td>73</td>
</tr>
<tr>
<td>3-2</td>
<td>Number of experiments required to obtain isotherm parameters with RSS=0.05 ($V_{\text{feed}}=0.6V_0$).</td>
<td>78</td>
</tr>
<tr>
<td>3-3</td>
<td>Number of experiments required for data with combined error to obtain isotherm parameters with RSS=0.05 ($V_{\text{feed}}=0.6V_0$).</td>
<td>81</td>
</tr>
<tr>
<td>3-4</td>
<td>Number of experiments required to obtain isotherm parameters with RSS=0.05 ($V_{\text{feed}}=0.3V_0$).</td>
<td>83</td>
</tr>
<tr>
<td>3-5</td>
<td>Number of experiments required for data with combined error to obtain isotherm parameters with RSS=0.05 ($V_{\text{feed}}=0.3V_0$).</td>
<td>83</td>
</tr>
<tr>
<td>4-1</td>
<td>Mass of chlorophyll $a$ in crude extract.</td>
<td>98</td>
</tr>
<tr>
<td>4-2</td>
<td>Loading conditions on glass columns.</td>
<td>100</td>
</tr>
<tr>
<td>4-3</td>
<td>Summary of individual CCC runs.</td>
<td>108</td>
</tr>
</tbody>
</table>
## LIST OF APPENDIX FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-1.</td>
<td>Objective function as a function of $t_0$</td>
<td>133</td>
</tr>
<tr>
<td>A-2.</td>
<td>Objective function as a function of $b_1$</td>
<td>133</td>
</tr>
</tbody>
</table>
To my parents,

in appreciation for your

unconditional love and support.
1. Preparative chromatography and nonlinear equilibrium isotherms

The distinction between analytical and preparative chromatography lies in the goal of separation, not in the size of the column used. Analytical chromatography aims at analysis and quantification of the components of a sample mixture whose composition and nature are little known. Preparative chromatography is used to purify from a well identified feed mixture one or more target components that are collected as final product. The different goals lead to different optimization strategies: while for analytical chromatography the resolutions of the components of interest are optimized in the shortest time, the objective function for preparative chromatography is the maximum throughput or production rate with recovery yield and purity constraints.

When a large sample is applied to a column as in preparative chromatography, it is usually observed that the band profiles and retention times are dependent on the sample size and concentration. The band profiles are no longer a series of Gaussian-shaped peaks and the shape of individual peaks varies as the sample composition does. It is the band profiles (the shape and relative position of peaks) that determine the amount of product can be recovered and its purity. Therefore, it is nearly impossible to optimize the operating conditions
without a clear understanding of how band profiles change in relation to loading conditions.

The complex peaks and their interdependence in preparative chromatography are resulted from nonlinear and competitive retention mechanisms. The concentration of individual component in the stationary phase at equilibrium is nonlinearly dependent on its concentration in the mobile phase. The amount of any component adsorbed on the stationary phase at equilibrium depends on not only its own concentration in the mobile phase but also the concentrations of all other components locally present in the mobile phase. All the components have to compete for the limited adsorption sites so that the amount of one component adsorbed on the stationary phase at equilibrium always decreases when the concentrations of other components increase. These relationships are accounted for by the nonlinear multicomponent isotherms. Accurate determination of multicomponent isotherms is prerequisite to optimization of preparative chromatography.

2. Countercurrent chromatography as a preparative separation tool

Countercurrent chromatography (CCC) is a form of liquid chromatographic technique that uses a liquid as the stationary phase. The principle of separation is liquid-liquid partitioning of solutes between two immiscible phases that are mixed and separated repeatedly in the column. The liquid stationary phase is retained inside the column by centrifugal force, and the mobile phase is passed through the
column by a pump. The CCC column is a continuous piece of tubing or a series of linked tubes that use no solid support.

Several types of CCC instruments are available on the market. The most popular one is planetary multilayer-coil CCC, which was first described and built by Yoichiro Ito, a pioneer in the development of CCC. The instrument consists of a spool of coiled tubing mounted on a column holder. The column is mounted around the central stationary shaft on a planetary gear. This arrangement produces synchronous planetary motion in a horizontal plane when the column is being rotated. The weight of the column is balanced by a counterweight mounted in a similar way.

The use of liquid stationary phase is what distinguishes CCC from high performance liquid chromatography (HPLC), and the key that makes it a fundamentally preparative tool. The solutes dissolved in the stationary phase make use of the whole volume of the stationary phase, while in HPLC the solutes can only adsorb on the surface of the stationary phase. Thus, the capacity of a CCC column is much higher than that of a HPLC column with an equal volume of stationary phase. The scale-up of CCC separation is straightforward: just increase the injection volume and the volume of the two phases in proportion to the increase in the column size, while keep the rotation speed and mobile phase flow rate constant. Because the partition coefficients of the solutes do not change in the same solvent system, the quality of separation remains the same. The limit will
only be the solubility of the solutes and/or the physicochemical properties of the feed solution.

3. Brief overview of subsequent chapters

Chapter 2 and chapter 3 investigate the fundamental problem in numerical estimation of multicomponent isotherms in preparative chromatography. That is the effect of experimental error in measured chromatograms on the precision of the resulting isotherm parameters. Computer simulated chromatograms are used to represent experimental chromatograms in order to control, tightly and precisely, the source and magnitude of error. Chapter 2 focuses on the influence of error in the effluent concentrations caused by fraction collection and analysis. Chapter 3 focuses on the influence of error in the retention times due to slight changes in mobile phase composition or flow rate, and the combined effect of both concentration and time error. A relationship between the goodness of final isotherm parameters and the variance in the individual estimated parameters is established. The number of experimental replicates required to obtain a certain level of accuracy in the estimated isotherm parameters are calculated explicitly for a variety of experimental data sets.

Chapter 4 reports a preparative-scale purification process for rapid and economic production of chlorophyll a with extreme high purity using CCC as the purification technique. The efficacy of extraction and isolation steps is quantified. The production and recovery yield of several CCC operations are summarized. The
optimization of CCC separation with respect to loading conditions is discussed in
detail.
Chapter 2

Numerical Estimation of Multicomponent Adsorption Isotherms in Preparative Chromatography: Implications of Experimental Error

Ling Zhang¹, John Selker¹, Annie Qu², and Ajoy Velayudhan¹

¹Department of Bioengineering, Oregon State University, 116 Gilmore Hall, Corvallis, OR 97331-3906
²Department of Statistics, Oregon State University, 76 Kidder Hall, Corvallis, OR 97331-3906


Editorial Office
Journal of Chromatography
PO Box 681
NL-1000 AR Amsterdam
The Netherlands
Abstract

Since experimental methods for measuring multicomponent adsorption isotherms are extremely tedious, numerical approaches are an attractive alternative. Here, the variance in isotherm parameters as a function of experimental error in measured effluent concentrations is quantified. The number of experimental replicates needed to obtain isotherm parameters to a desired level of accuracy is calculated explicitly. After the covariance matrix of the parameters has been determined, Monte Carlo methods are found to be rapid and effective. The use of different kinds of experiments, the effect of resolution and loading, and the impact of the number of measured data points are described.

1. Introduction

Preparative chromatography is widely used in the isolation and purification of a variety of compounds from a range of feedstocks [1,2]. It has become clear that preparative chromatography is usually optimal with respect to production rate when the feed mass is high enough to cause nonlinear adsorption [2]. Thus, accurate knowledge of adsorption isotherms is an important prerequisite to optimization; clearly, these should be multicomponent isotherms, measured over the entire range of feeds and additive compositions that will be used in practical runs. In practice, however, multicomponent isotherms are rarely used to optimize a separation, principally because they require tedious and time-consuming
experiments. Recently, the problem of estimating multicomponent isotherms numerically (preferably using a small amount of easily obtained experimental data) is receiving considerable attention in the literature [3-9]. The goal of this paper is to investigate the effect of experimental error on the precision of the numerically estimated isotherm parameters. To the best of our knowledge, this important problem of estimating the numerical back-propagation of error has not been addressed in the literature.

2. Theory

2.1. Brief Review of Literature

For the determination of single-component isotherms, the most popular technique is probably elution on a characteristic point (ECP), since a single isocratic run gives many data points towards the isotherm. However, ECP has the drawback of not accounting for the contribution of non-equilibrium effects (pore diffusion, film mass-transfer) to the trailing edge of the chromatogram. The ensuing error in the extracted isotherm parameters has been characterized in the literature [9-15]. ECP cannot be used to measure multicomponent isotherms.

There are few experimental methods to measure multicomponent isotherms. Frontal analysis (FA) has been widely used to determine isotherms for binary systems [16-20], and recently this method has been applied to ternary systems [21,22]. Another experimental method that can measure multicomponent
adsorption is elution on a plateau. One approach involves radio-labeling [23,24], which is tedious and inconvenient. The simpler approach to elution on a plateau, which is called the concentration pulse or the step-and-pulse method, is to use unlabeled pulses. This is easier to run, and has been more widely used in liquid-solid adsorption [25-27]. There are also methods based on theoretical calculations, such as the hodograph method [28-30]. Jacobson et al. [16] proposed a method based on the h-transform of Helfferich and Klein [31] that measures binary isotherms for approximately Langmuirian behavior; results from this method agreed with those from the classical frontal method for their system. Jacobson and Frenz [32] also combined these two approaches to formulate a hybrid mass balance method.

The alternative approach to determine the isotherm is to approximate it numerically. One disadvantage of this approach is that the isotherm functional form must be chosen at the outset. Then, by minimizing the difference between calculated and experimental profiles, the parameter values are adjusted until a best fit to the experimental data is found. In practice, several isotherm forms should be tried in order to estimate the most suitable model.

Dose et al. [3] applied the simplex optimization algorithm to determine single-component and binary isotherms numerically. Langmuir and bi-Langmuir model parameters were identified successfully. This method was also used by Guan et al. [9] to determine single-component isotherm using a simulated "experimental" profile, and found more accurate than ECP. Another optimization
approach, the steepest descent algorithm, has been applied by James et al. [4] for
determination of single- or binary-competitive isotherms. For the two binary
systems studied (ketoprofen enantiomers in one case and benzyl alcohol and 2-
phenylethanol in the other), numerically identified model parameters were in good
agreement with those obtained from the ECP or FA methods. The authors
suggested that one drawback of their method was that the computation was long
and complex.

Choosing an isotherm model appropriate to a given experimental dataset is
an important part of numerical estimation, and has been studied by Guiochon's
group. Experimental data on the competitive adsorption of 2-phenylethanol and 3-
phenylpropanol in reversed-phase chromatography were fitted to a large number of
different isotherm models [18,33-35]. Results from this particular case show that
choosing model is a very complex and difficult task. Some of the models are
implicit, which increases the run time dramatically. For this data set, the implicit
models do not fit the data better than the explicit models, and are therefore
unattractive. Among the explicit models, the 11-parameter quadratic model
provides the most accurate fit to the binary data. Because no relation between the
number of parameters in a model and the accuracy of that model can be derived
from the results, simple models with fewer parameters would be preferable for the
calculation of band profiles in practice.

Since mobile phase modulators (e.g., organic modifiers in reversed phase
and salts in hydrophobic and electrostatic interaction chromatography) are widely
used in practice, it is important to know how they affect adsorption isotherms. Jandera et al. [36,37] investigated the effect of methanol concentration on single- and two-component solute isotherms in reverse phase chromatography. Langmuir, Jovanovic, competitive Langmuir, competitive Jovanovic and quadratic isotherms were used to fit the experimental data. A quadratic dependence of the logarithm of the isotherm parameters on modulator (methanol) concentration was found to apply for the Langmuir isotherms in the range 0-40% methanol. By contrast, in the Jovanovic forms, the parameters themselves (and not their logarithms) were described well by a quadratic expression in the modulator level. Jandera et al. [38,39] also investigated the mobile phase effects in normal phase chromatography. Here, the competitive Langmuir model failed to describe the data and it was found necessary to take into account the simultaneous effects of a competitive adsorption and of a possible multi-layer association of the already adsorbed molecules.

Antos et al. implemented the Marquardt algorithm for optimization to determine numerically the competitive isotherms of diolone acetate and benzophenone from a real post-reaction mixture at one mobile phase composition in a reversed phase system [5]. With a quadratic dependence of the logarithm of the isotherm parameters on modulator concentration, they were able to predict the separation for several different mobile phase compositions. Antos et al. also reported in a later paper [6] that using the same algorithm the numerical determination of the isotherm of methyl deoxycholate at various mobile phase compositions in a normal phase system from isocratic elution profiles. The
dependence of the isotherm parameters on the mobile phase composition was described by a three-parameter equation derived from the Snyder model [40]. The results were used to simulate overload gradient elution profiles. The predicted peak shapes are not in complete agreement with the experimental peak shapes. This is acceptable for single component runs, but for multicomponent runs, it may affect productivity or yield calculations appreciably.

2.2. Statement of Problem

Many questions remain regarding the numerical estimation of isotherms. In analytical chromatography, the feed concentrations are so low that the various mass-transfer and diffusion parameters (such as the film mass-transfer coefficient, the pore diffusivity, the axial dispersion coefficient) are all independent of feed concentration. And, by definition, the large number of binding sites relative to the number of adsorbing molecules implies that the equilibrium binding constants are independent of the rate constants. Neither of these convenient assumptions holds in nonlinear chromatography. At high feed concentrations, rate parameters can depend on concentration [41,42]. Further, once competitive binding supervenes, the resulting mass balances become nonlinear, making the decoupling between the thermodynamic isotherm parameters and the kinetic rate parameters problematic. This nonlinearity also implies that not all lumped models are equivalent, leading to the question of which kind of lumped model is the best. In the worst case, all rate
and equilibrium parameters may need to be determined simultaneously. Since we hope to use simple isocratic elution runs as the source of information for estimation, we must immediately be concerned with identifiability: can so many parameters in fact be uniquely and accurately extracted from a few simple runs? All these questions are beginning to be attacked in the literature as discussed earlier. In addition, such problems can be "ill-posed," i.e., the error in the results (the isotherm parameters) is extremely large even when the error in the input (the chromatogram) is very small. This question of ill posedness in chromatographic systems has been studied by James et al. [4,43].

Our approach here is aimed at an even more fundamental question: if we have a full understanding of the appropriate model, and identifiability is guaranteed, how does error in the experimental data translate into error in the resulting isotherm parameters? In other words, given a bound on the typical error associated with any point in the experimental chromatogram, can suitable bounds be found on the estimated parameters? This is an essential question in any parameter estimation or inverse problem. Here, we present an argument to bound the error in the parameters, and thereby establish a relationship between the experimental error and the number of experimental replicates needed to obtain isotherm parameters with a specified standard error.
2.3. Methodology

We use data simulated by a numerical code, called the solver, to represent the experimental chromatogram in order to ensure that the only source of error is in the chromatogram; the isotherm model and the exact values of its parameters must therefore be known exactly. Since the latter assumption is never possible when using real data (even if a given isotherm model fits real data well, there will always be some error involved), simulated chromatograms are the appropriate test bed for the present study. The rate parameters are also fixed, and the same numerical code that is used to generate the data is used in the estimation problem. Thus, error only arises from the chromatogram; we represent experimental error here by incorporating Gaussian noise. At each time, the exact chromatogram simulated by the solver is perturbed by adding a random number representing a certain relative error to produce an experimental chromatogram. We define relative error at a point in the chromatogram as the ratio of the error to the concentration at that point. The best-fit isotherm parameters are then generated iteratively using an optimizer that calls the solver repeatedly with changing isotherm parameter values as depicted schematically in Figure 2-1. For a given set of isotherm parameter values, the solver generates a simulated chromatogram, which the optimizer compares with the experimental chromatogram. The error between these chromatograms is minimized by the solver, and the corresponding isotherm parameter values are the best-fit values. The expression to be minimized is given explicitly in Section 3.
Figure 2-1. Flowchart describing the process of estimating isotherm parameters numerically.
A variety of multicomponent isotherms has been used to capture chromatographic retention behavior [2,44,45]. Here, we restrict attention to the simplest of these, the multicomponent Langmuirian isotherm:

\[
q_i = \frac{a_i c_i}{1 + \sum_{j=1}^{p} b_j c_j}, \quad i = 1, 2, 3, ..., p
\]  

where \( p \) is the total number of adsorbable components in the system. Each feed component contributes 2 parameters: \( a \) is the slope of the isotherm at the origin, and therefore represents the distribution coefficient at analytical concentrations (or Henry's-law constant); \( b \) is a measure of its affinity for the stationary phase; the ratio \( a/b \) is the corresponding saturation concentration. Although the Langmuirian isotherm is known to be thermodynamically inconsistent unless the saturation concentrations of all the species are identical [46,47], there are many situations in which it is a useful empirical description of realistic multicomponent adsorption. We use it here because it is a simple, explicit formalism, and allows us to concentrate on the question of estimation error. Ternary feeds are used in this work; consequently, 6 isotherm parameters must be obtained.

2.4. Statistical Analysis

We first introduce some convenient notation. The chromatogram generated by the solver for the exact isotherm parameters is called the exact chromatogram.
The variants generated by incorporating random errors into the exact chromatogram are called replicates, to distinguish them from real experimental runs, which will simply be called experiments. We will generate a large number of replicates for a given exact chromatogram, in order to be sure that a statistically representative distribution of the parameters is obtained. Later, we will see that a Monte Carlo distribution is found to be necessary for this purpose.

We have found that the variation in any estimated parameter is well approximated by a Gaussian distribution. A representative result for two parameters from different exact data sets is shown in Figure 2-2 (the corresponding isotherm parameter values and feed conditions are listed in Table 2-1 and 2-2). However, correlation among the parameters must also be accounted for, and this is done in this section. The individual estimates for $a_1$ are $a_{11}, a_{12}, ..., a_{1n}$, where $n$ is the number of replicate runs; similar notation applies for the other parameters. We scale the 6 average parameters before combining them into the parameter vector $\bar{X}$; thus its first component, $\bar{X}_1$, is $\frac{a_1 - a_{1,exact}}{a_{1,exact}}$ where $a_{1,exact}$ is the exact value of $a_1$, and $\bar{a}_1$ is the averaged value of $a_{11}, a_{12}, ..., a_{1n}$; the second component, $\bar{X}_2$, is $\frac{b_1 - b_{1,exact}}{b_{1,exact}}$; and the sixth component, $\bar{X}_6$, is $\frac{b_3 - b_{3,exact}}{b_{3,exact}}$. This has the advantage of making the average value of each component unity. The error in the parameters obtained from the optimization process can be represented through the residual sum of squares (RSS), defined as:
\[
\text{RSS} = \left[ \left( \overline{X_1} \right)^2 + \left( \overline{X_2} \right)^2 + \ldots + \left( \overline{X_6} \right)^2 \right]^{1/2}
\]

Since the standardized isotherm parameters were found to be Gaussian, we have:

\[
X_1 = \frac{a_{ij} - a_{1,\text{exact}}}{a_{1,\text{exact}}} \sim N\left(0, \sigma_1^2\right)
\]

where the \( N \) on the right-hand side represents a normal or Gaussian distribution

with a mean of 0 and a standard deviation of \( \sigma_1 \); similar results hold for all the other parameters. Then the \textit{averaged} parameters satisfy

\[
\overline{X_1} \sim N\left(0, \frac{\sigma_1^2}{n}\right) = \frac{\sigma_1}{\sqrt{n}} N(0,1)
\]

Similar results can readily be written for the other parameters. Taking the square of

\[
\frac{\sqrt{n}}{\sigma_1} \overline{X_1}
\]

gives a \( \chi^2 \) distribution with 1 degree of freedom [48]:

\[
\frac{n}{\sigma_1^2} \left( \overline{X_1} \right)^2 \sim \chi_1^2
\]

If the isotherm parameters were uncorrelated (i.e., they were independent of each other), then the RSS would be a weighted \( \chi^2 \) distribution with 6 degrees of freedom:

\[
\text{RSS} = \left[ \frac{\sigma_1^2}{n} X_1^2 + \frac{\sigma_2^2}{n} X_2^2 + \ldots + \frac{\sigma_6^2}{n} X_6^2 \right]^{1/2}
\]

\[
= \frac{1}{\sqrt{n}} \left[ \sigma_1^2 X_1^2 + \sigma_2^2 X_2^2 + \ldots + \sigma_6^2 X_6^2 \right]^{1/2}
\]
Figure 2-2 (a). Probability plot for estimated isotherm parameters. Estimated parameter $a_3$ from the exact data II with 10% relative error (70 replicates).
Figure 2-2 (b). Probability plot for estimated isotherm parameters. Estimated parameter $b_1$ from the exact data I with 5% relative error (30 replicates).
Table 2-1. Isotherm parameter set

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Exact value</th>
<th>Range</th>
<th>Parameter</th>
<th>Exact value</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>a₁ [-]</td>
<td>3.00</td>
<td>2.4 - 3.6</td>
<td>a₁ [-]</td>
<td>2.50</td>
<td>2.25 - 2.75</td>
</tr>
<tr>
<td>b₁ [ml/mg]</td>
<td>0.100</td>
<td>10³ - 10</td>
<td>b₁ [ml/mg]</td>
<td>0.0714</td>
<td>10⁻⁴ - 1</td>
</tr>
<tr>
<td>a₂ [-]</td>
<td>4.50</td>
<td>3.6 - 5.4</td>
<td>a₂ [-]</td>
<td>3.00</td>
<td>2.7 - 3.3</td>
</tr>
<tr>
<td>b₂ [ml/mg]</td>
<td>0.113</td>
<td>10³ - 10</td>
<td>b₂ [ml/mg]</td>
<td>0.075</td>
<td>10⁻⁴ - 1</td>
</tr>
<tr>
<td>a₃ [-]</td>
<td>6.75</td>
<td>5.4 - 8.1</td>
<td>a₃ [-]</td>
<td>3.50</td>
<td>3.15 - 3.85</td>
</tr>
<tr>
<td>b₃ [ml/mg]</td>
<td>0.135</td>
<td>10³ - 10</td>
<td>b₃ [ml/mg]</td>
<td>0.078</td>
<td>10⁻⁴ - 1</td>
</tr>
</tbody>
</table>

Table 2-2. Numerically simulated experimental data set

<table>
<thead>
<tr>
<th>Exact Data</th>
<th>Number of theoretical plates</th>
<th>Loading</th>
<th>Isotherm parameter set</th>
<th>Number of data points</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1000</td>
<td>c_{feed} = 2 mg/ml, V_{feed} = 0.3V₀</td>
<td>Set 1</td>
<td>179</td>
</tr>
<tr>
<td>II</td>
<td>1000</td>
<td>c_{feed} = 2 mg/ml, V_{feed} = 0.6V₀</td>
<td>Set 1</td>
<td>199</td>
</tr>
<tr>
<td>III</td>
<td>1000</td>
<td>c_{feed} = 5 mg/ml, V_{feed} = 0.4V₀</td>
<td>Set 2</td>
<td>84</td>
</tr>
<tr>
<td>IV</td>
<td>1000</td>
<td>c_{feed} = 2 mg/ml, V_{feed} = 0.6V₀</td>
<td>Set 1</td>
<td>24</td>
</tr>
<tr>
<td>V</td>
<td>1000</td>
<td>c_{feed} = 2 mg/ml, V_{feed} = V₀</td>
<td>Set 1</td>
<td>27</td>
</tr>
<tr>
<td>VI</td>
<td>2500</td>
<td>c_{feed} = 2 mg/ml, V_{feed} = 0.6V₀</td>
<td>Set 1</td>
<td>188</td>
</tr>
</tbody>
</table>

Note that the term in square brackets is independent of n. However, since the isotherm parameters are correlated in an overloaded run, we must replace the previous expression by
where the overbar on the chi-squared term indicates that the correlated structure of the parameters has been incorporated.

It is interesting to note that the mean and variance of $RSS^2$ can be calculated explicitly. The mean of $RSS^2$ is 6, which is exactly the mean of a chi-squared distribution with 6 degrees of freedom, $\chi^2_6$. The variance of $RSS^2$ is given by:

$$\sigma^2_{RSS^2} = 12 + 2 \sum_{i=1}^{6} \sum_{j=1}^{6} \rho_{i,j}^2$$

The first term on the right-hand-side of equation (3) is the variance of a chi-squared distribution with 6 degrees of freedom. The second term involves the sum of the correlatedness between the isotherm parameters taken pairwise, and is always positive. Therefore the correlatedness of the parameters always leads to an increase in the variance of $RSS^2$. However, the mean and variance of $RSS$ (which is what we’re interested in) cannot be explicitly calculated from those of $RSS^2$. We will therefore need to use simulations to determine the RSS distribution numerically.

Nevertheless, equation (3) is invaluable in pointing out that the variance of $\chi^2_6$ is independent of the number of replicates, $n$. Since this is also true of the mean of $\chi^2_6$, as pointed out above, we conclude that the mean and variance of $\chi^2_6$ also do
not depend on \( n \). It therefore follows from equation (2) that the RSS will decrease as the square root of the number of replicates. We can determine how many replicates will be needed to make the RSS sufficiently small.

Usually, requiring the RSS to be 0.05 should be sufficient for most subsequent applications of the isotherm parameters, and this value has been used in all the calculations below. It can easily be seen from the definition of RSS that this choice (RSS=0.05) guarantees that no estimated parameter is worse than 5% away from its exact value. Of course, RSS is also a random variable, with its own variance. So we will use the 95th quantile to be reasonably sure (with 95% confidence) that the calculated average value of RSS in a given optimization run lies below 0.05.

An important consequence of obtaining a sum of (correlated) \( \chi^2 \) distributions is that the variance in RSS will be comparable to the mean. This implies that, in some cases, a large number of replicates may be needed in order to determine the isotherm parameters to reasonable accuracy. We will see later that this is in fact the case (cf. Tables 2-10 ~ 2-12).

3. Simulations

Several simulation methods have been used in our group for preparative chromatography [49,50]. Typically, the lumped solid-phase mass-transfer model is used, which can be represented as:
Here, \( t \) and \( z \) are the independent variables time and distance into the column; the dependent variables are the mobile phase concentration \( c \) and the stationary phase concentration \( q \) for each adsorbable compound \( i \). The mobile phase velocity is \( u \), the volumetric phase ratio (the ratio of stationary phase to mobile phase volume) is \( \phi \), and the mass-transfer coefficient for the \( i^{th} \) component is \( k_i \). Note that \( q \) is not in equilibrium with \( c \), but simply represents the stationary phase concentration. The stationary phase concentration in equilibrium with \( c \) is \( q^* \), which is described through the multicomponent adsorption isotherm:

\[
q_i^* = f_i(c_1, c_2, \ldots, c_p)
\]  \hspace{1cm} (6)

where \( p \) adsorbable components are present in the column. This represents the solver. As mentioned earlier, only Langmuirian isotherms for ternary feeds (\( p = 3 \)) will be used here. Both Craig plate models and rate models based on the method of characteristics are used to solve the lumped equations above. It was found for these data sets that the Craig plate simulations were significantly faster than the rate model; since the optimization process involves many calls to the solver, the Craig process was preferred, and was used for all the results shown here.

\[
\frac{\partial c_i}{\partial t} + u \frac{\partial c_i}{\partial z} + \phi \frac{\partial q_i}{\partial t} = 0
\]  \hspace{1cm} (4)

\[
\frac{\partial q_i}{\partial t} = k_{M,i}(q_i^* - q_i)
\]  \hspace{1cm} (5)
In this paper, 5% and 10% relative error are widely used. These values are reasonable for many overloaded runs, for which direct detection by, e.g., a UV detector is usually impossible because the high feed concentrations saturate the detector. Thus sample collection, dilution, and separate analysis are necessary. This increases the overall error associated with the reconstructed chromatogram.

We expect that 5% relative error is fairly good for small molecules; proteins and other macromolecules often produce greater errors, and 10% might then be reasonable. The error is added into the exact chromatogram by using the random-number-generator FORTRAN code DRNOR, obtained from Netlib. This code generates normally distributed random iterates with mean 0 and variance 1; these are scaled to produce 5% relative error and added pointwise to the chromatogram to simulate experimental error. This is then used as the “experimental” chromatogram in the optimizer to find the best-fit isotherm parameters.

In order to scale the iterates, we must decide what fraction of the normal distribution lies within 5% relative error (since, with a Gaussian, there will always be some fraction of the values that lie outside of any specified bound). Here, we choose to specify that 90% of the iterates must lie within the specifications; thus, 10% of the distribution is allowed to lie outside of the ±5% limits. This choice was made in order to allow substantial likelihood for errant points (outliers). This implies that the estimate for the number of experiments obtained ultimately will be close to an upper bound; in other words, in most cases, fewer experiments may suffice. We therefore obtain a “worst-case” estimate with this choice.
The results from any run of the solver during the estimation process (as shown in Figure 2-1) are then compared against the experimental chromatogram, and an estimate of how well the simulation fits the experimental data is constructed. Here, an output least-squares method [51] is used; thus the objective function is the sum of the squares of the differences between the experimental and simulated concentration values at various instants. For example, for a single-component run where the isotherm is to be fitted to the Langmuir isotherm, the objective function $J$ would be given by:

$$J(c; a, b) = \sum_{m=1}^{M} \left[ c_{\text{sim}}(L, t_m; a, b) - c_{\text{exp}}(L, t_m) \right]^2$$  \hspace{1cm} (7)

where $M$ is the total number of data points at which the effluent history is known, $c_{\text{sim}}$ represents the result of the simulation at the column exit (at $z=L$) at those discrete times ($t=t_m$) for which experimental data is available. The isotherm parameters $a$ and $b$ are listed after the semicolon to emphasize that the simulated concentration depends on these parameters. The code that evaluates the objective function for given values of $a$ and $b$, and on this basis chooses new values of $a$ and $b$, is called the optimizer. The problem of estimating the “best” isotherm parameters, i.e. those that fit the experimental data best in the sense of minimizing the objective function $J$, is carried out iteratively. We start with a guess for $a$ and $b$. The optimizer feeds these values to the solver, and thus obtains the chromatogram corresponding to these values. It then evaluates $J$, and thus decides on a new pair $a, b$. The process is repeated until the fit is acceptably good, i.e., the
error lies below a specified tolerance. This description used a single-component isotherm for simplicity; as mentioned earlier, all the runs reported in this work are for ternary feeds. Such an iterative method is computationally convenient, allows for incorporation of regularization terms if the problem turns out to be sufficiently ill-posed to warrant it [52], and can be used for almost any kind of chromatogram.

Methods that try to invert the problem directly (i.e., working directly from the chromatogram and attempting to calculate the resulting isotherm parameters) are often more ill-posed, and are always limited by the nature of the chromatogram. Thus, a completely resolved chromatogram cannot be used in the direct method to obtain isotherm parameters that depend solely on interaction among the feeds. However, an iterative method simulates the entire chromatographic process, and therefore captures competition among the feeds as they pass down the column subsequent to feed introduction. Such an iterative method is likely to capture pure interaction parameters even when given a chromatogram that is completely resolved.

There are many effective optimization methods available. We use here a public-domain code written by Professor Tits' group at the University of Maryland [53]. This state-of-the-art FORTRAN code solves nonlinear optimization problems with nonlinear and linear equality and inequality constraints, and simple bounds on the variables. The code is based on Sequential Quadratic Programming (SQP) iterations [54,55], in which a sequence of quadratic sub-problems near the solution
of the original problem are solved to generate directions of search. Along these directions, better approximations to the solution are determined.

In the Langmuirian isotherm, the $a$ parameters are proportional to the linear (Henry's law) distribution coefficients, and can therefore be well estimated from separate analytical runs. The $b$ parameters, which relate to the saturation concentration or equivalently the isotherm curvature, cannot be determined easily from independent experiments. In fact, it is the difficulty of determining these nonlinear parameters experimentally that leads us to try numerical estimation in the first place! The two sets of isotherm parameters used in this study are listed in Table 2-1. Our boundaries on the $a$ parameters are fairly tight, at $\pm 20\%$ of the exact values for parameter set 1 and $\pm 10\%$ for parameter set 2. The bounds on the $b$ parameters are much wider, to minimize the loss of $b$'s appropriate to the data but excluded because of unduly narrow bounds. (If we had found some optimization runs failing with values of $b$'s close to or at a boundary value, we would have enlarged these bounds suitably. This was not a factor in our runs, and so we are fairly confident that our bounds were appropriate.)

In order to ensure that we reached the global minimum, for each experimental chromatogram we ran six widely different sets of initial guesses. In all cases, all six initial guesses converged to a numerically unique solution. (Slight differences in the final results are sometimes obtained; however, the differences are less than $1\%$, and we concluded that these represent the same minimum perturbed
by the different error in the chromatograms.) It is concluded that for this given bound on the parameters, convergence is guaranteed regardless the initial guesses.

4. Results and discussion

The first set of exact data used was produced by running the solver for the isotherm parameter set 1 and for a feed volume of \( V_{\text{feed}} = 0.3 \, V_0 \). All other conditions are listed in Table 2-2. The exact isotherm parameter values are: \( a_1 = 3.00 \); \( b_1 = 0.100 \, \text{ml/mg} \); \( a_2 = 4.50 \); \( b_2 = 0.113 \, \text{ml/mg} \); \( a_3 = 6.75 \); \( b_3 = 0.135 \, \text{ml/mg} \). These could be viewed as representative of moderately retained feeds (perhaps amino-acids or small peptides) on a reversed phase or hydrophobic interaction column under isocratic elution.

Both the chromatogram in the absence of error, and a representative one for 5% error, are shown in Figure 2-3 (a). The feed volume corresponds to bands that are just resolved at the column outlet. The number of theoretical plate is 1000, as listed in Table 2-2. We will see shortly that larger plate counts do not make much difference in these runs.

Since the estimation process is numerically intensive (recall that each simulated chromatogram was run for 6 initial guesses), we run a fair number of replicates (e.g., 30 or 40), and then try to extract the covariance matrix from these replicates using standard methods. A very large sample of replicates having this covariance is then generated by Monte Carlo simulations. The RSS distribution is
then generated, and equation (2) is used to determine the number of experiments needed to obtain isotherm parameters to a certain level of accuracy (here, the 95th quantile is used). The examples below will clarify exactly how the process is carried out. The package S-PLUS is used for statistical calculations [56]. Table 2-3 shows the number of experiments needed to 2 significant digits (in practice, we will round this to the nearest integer) as a function of the number of replicates and the size of the Monte Carlo simulation for the results in Figure 2-3(a). In this case, we did not go further because the number of experiments needed is clearly going to be less than 1, implying that 1 experiment would be sufficient to identify the isotherm parameters in this case. Of course, in practice, we would always do at least 2 experiments.

Table 2-3. Number of experiments required for exact data I with 5% relative error

<table>
<thead>
<tr>
<th>Size of Monte Carlo Matrix</th>
<th>Number of replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td>1000</td>
<td>0.82</td>
</tr>
<tr>
<td>10000</td>
<td>0.83</td>
</tr>
<tr>
<td>50000</td>
<td>0.83</td>
</tr>
</tbody>
</table>
Figure 2-3. Simulated chromatogram for exact data \( I (V_{\text{feed}} = 0.3V_0) \). Part (a) compares the exact case, i.e., no error added (symbols) with a representative run involving 5% relative error (solid line). Part (b) compares the "exact" or "no error" case (same symbols as in part (a)) with a representative run involving 10% relative error (solid line). In all cases, concentrations are reported approximately every second.
Table 2-4 shows the corresponding results for the case where the relative error in the chromatogram was increased to 10% (the feed volume remains at $V_{\text{feed}} = 0.3 V_0$). A representative chromatogram is shown in Figure 2-3 (b). For 10 replicates (the first column of the matrix), it can be seen that the number of required experiments varies somewhat as the size of the Monte Carlo sample increases. This variation decreases as the number of replicates increases, as can be seen from the second and third columns of the matrix. Thus the numerical process involves running as many replicates and choosing as large a Monte Carlo sample as necessary to achieve numerical convergence. It is clear from Table 2-4 that 30 replicates of the numerical estimation process (corresponding to 30 versions of the exact chromatogram to which error is added randomly as described above) are needed to obtain 2-digit accuracy in the result. On this basis, we conclude that 2 experiments should be done (a safe estimate would be 3 experiments). For each of these runs, the estimation process should be carried out to arrive at an estimate of the isotherm parameters. Taking the mean values of these parameters gives us the final result; the theory presented above indicates that the RSS associated with these mean isotherm parameters is below 0.05, with 95% confidence.
Table 2-4. Number of experiments required for exact data I with 10% relative error

<table>
<thead>
<tr>
<th>Size of Monte Carlo matrix</th>
<th>Number of replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td>1000</td>
<td>2.8</td>
</tr>
<tr>
<td>10000</td>
<td>2.6</td>
</tr>
<tr>
<td>50000</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Next, a simulation (exact data II in Table 2-2) using a larger feed volume ($V_{feed} = 0.6 V_0$) is used to generate the experimental chromatogram (see Figure 2-4). Now the bands are mixed on emerging from the column. Since mixed bands represent greater interaction among the feed components, it is interesting to compare the results of this case to the previous one. It should be remembered, however, that the entire chromatographic process is being simulated in the solver. Thus, even if the bands are resolved at the column outlet (as in Figure 2-3), the components were mixed over a considerable portion of the column, and interacted with each other over that length. Thus, feed-feed interactions may well be adequately described in this numerical approach even when the bands are resolved at the outlet. In this sense, this method is more general than methods based simply on the effluent chromatogram (such as all experimental methods). Again, we first generate 5% relative error using random iterates (a typical chromatogram is shown in Figure 2-4 (a)), and calculate the corresponding best-fit parameters. The results are shown in Table 2-5; now 40 replicates are needed for convergence to the RSS
Figure 2-4. Simulated chromatogram for exact data II ($V_{\text{feed}} = 0.6V_0$). Part (a) compares the exact case (symbols) with a representative run involving 5% relative error (solid line). Part (b) compares the exact case (same symbols as in part (a)) with a representative run involving 10% relative error (solid line). All other information as in Figure 2-3.
distribution, and we find 2 experiments should suffice to obtain isotherm parameters for which $RSS = 0.05$ with 95% confidence.

Table 2-5. Number of experiments required for exact data II with 5% relative error

<table>
<thead>
<tr>
<th>Size of Monte Carlo matrix</th>
<th>Number of replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td>1000</td>
<td>2.5</td>
</tr>
<tr>
<td>10000</td>
<td>2.5</td>
</tr>
<tr>
<td>50000</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Table 2-6 shows the corresponding case where the relative error is increased to 10% as the feed volume remains at $V_{\text{feed}} = 0.6 V_0$. A typical chromatogram is found in Figure 2-4 (b). Considerably more variation is found than that in the previous runs; now 70 replicates are needed to provide convergence to 2 significant digits. Notice that the number of experiments needed has gone up significantly relative to that in Table 2-5. The $RSS$ as a function of the number of experiments needed, $n$, is shown in Figure 2-5.

Consideration of the results in Tables 2-3 ~ 2-6 indicates that the chromatogram representing complete separation (Figure 2-3) needed fewer experiments for the same level of error than did the chromatogram with overlap (Figure 2-4). This is an interesting and somewhat unexpected result. This may be because of the iterative nature of the optimization method used here. As mentioned
Figure 2-5. RSS as a function of n, the number of experiments needed for results listed in Table 2-3 (solid line) and Table 2-4 (dashed line).
earlier if the bands are fully separated at the column outlet, the solver simulates the entire chromatographic process, including the period at the beginning of the column over which the bands underwent strong mutual interference. So even fully separated bands may be efficiently estimated by this method.

Table 2-6. Number of experiments required for exact data II with 10% relative error

<table>
<thead>
<tr>
<th>Size of Monte Carlo matrix</th>
<th>Number of replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td>1000</td>
<td>11</td>
</tr>
<tr>
<td>10000</td>
<td>10</td>
</tr>
<tr>
<td>50000</td>
<td>11</td>
</tr>
</tbody>
</table>

If another isotherm formalism were used that contained interaction terms, i.e., terms that vanished when only one component was present, then it is clear that mixed data of the kind shown in Figure 2-4 would be essential to identify the parameters contained in these interaction terms. Isotherms with such interaction terms have been proposed in the literature [18, 33-35].

We now examine the possibility of using both datasets above simultaneously in the optimizer, i.e., will fitting both small and large feed volume chromatograms at the same time provides better isotherm information? The results for 5% relative error in both datasets are shown in Table 2-7, and for 10% error in Table 2-8. The results seem better than those for either the small or large feeds
individually, since the number of experiments needed is smaller. But it should be kept in mind that a “single” experiment in Tables 2-7 and 2-8 corresponds to doing 1 small-feed and 1 large-feed run. Thus, for 10% error, using the small-feed run alone requires 2 runs, from Table 2-4; using the large run alone requires 6 (or 7) runs, from Table 2-6; using them both together requires 2 runs of each kind. However, this kind of combined run is often useful in that it is likely to be not much worse than the best case, and much better than the worst case. Since we do not know beforehand (for an arbitrary isotherm) which case is the best, the combined run may be quite attractive in practice.

Table 2-7. Number of experiments required for combination of exact data I and II both with 5% relative error

<table>
<thead>
<tr>
<th>Size of Monte Carlo Matrix</th>
<th>Number of replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td>1000</td>
<td>0.58</td>
</tr>
<tr>
<td>25000</td>
<td>0.53</td>
</tr>
<tr>
<td>50000</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Table 2-8. Number of experiments required for combination of exact data I and II both with 10% relative error

<table>
<thead>
<tr>
<th>Size of Monte Carlo matrix</th>
<th>Number of replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td>1000</td>
<td>2.1</td>
</tr>
<tr>
<td>25000</td>
<td>2.3</td>
</tr>
<tr>
<td>50000</td>
<td>2.3</td>
</tr>
</tbody>
</table>
As another way of examining the variation in the parameters, the individual results of the various replicates for the calculations above are shown in Figure 2-6. It is clear that the relative variation in the $a$ parameters is far less than in the $b$ parameters. This is to be expected in part because of the much tighter bounds on the former.

The runs above were based on simulated chromatograms for a ternary system with moderate selectivity (around 1.5 for each adjacent pair of components). While this is realistic for preparative runs, it is of interest to see how the results would change if the selectivity were decreased. Figure 2-7 is the exact chromatogram (exact data III in Table 2-2) for a ternary system in which the binary selectivities are 1.1, with $V_{\text{feed}} = 0.4 V_0$. The exact isotherm parameter values are listed in Table 2-1. Considerably more mixing is found than that in the earlier runs with comparable feed volumes for the system with higher selectivity (cf. Figure 2-3 and 4). When 5% relative error was added in the usual way, the results for up to 60 replicates are shown in Table 2-9. Now 7 experiments are needed, compared to 1 (Table 2-3) or 2 (Table 2-5). When 10% relative error is used, the number of experiments needed increases to 36 (Table 2-10). This is an indication that highly overloaded systems may have too much mixing among the feed bands to obtain the isotherm parameters effectively. Although this is a single result, we might cautiously suggest that there is a range of loading for which experimental estimation is facilitated, and that very low or very high loading complicates the estimation process.
Figure 2-6 (a). Variations in simulated parameters (scaled with respect to the exact parameters) and RSS. Exact data II with 10% relative (70 replicates).
Figure 2-6 (b). Variations in simulated parameters (scaled with respect to the exact parameters) and RSS. Combination of two exact data I and II both with 5% relative error (30 replicates).
Figure 2-7. Simulated chromatogram for exact data III ($V_{\text{feed}} = 0.4V_0$). Different isotherm parameters from those used in the earlier figures; this figure represents a more difficult separation.

Table 2-9. Number of experiments required for exact data III with 5% relative error

<table>
<thead>
<tr>
<th>Size of Monte Carlo matrix</th>
<th>Number of replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20</td>
</tr>
<tr>
<td>1000</td>
<td>5.7</td>
</tr>
<tr>
<td>10000</td>
<td>5.9</td>
</tr>
<tr>
<td>50000</td>
<td>5.8</td>
</tr>
<tr>
<td>100000</td>
<td>5.8</td>
</tr>
</tbody>
</table>
Table 2-10. Number of experiments required for exact data III with 10% relative error

<table>
<thead>
<tr>
<th>Size of Monte Carlo matrix</th>
<th>Number of replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20</td>
</tr>
<tr>
<td>1000</td>
<td>51</td>
</tr>
<tr>
<td>10000</td>
<td>46</td>
</tr>
<tr>
<td>50000</td>
<td>47</td>
</tr>
<tr>
<td>100000</td>
<td>51</td>
</tr>
</tbody>
</table>

Another factor that may affect the results is the number of data points in the experimental chromatogram. In the earlier runs (Figures 2-3 and 2-4), data points were generated approximately every second. While this is realistic for detectors in current use (in fact, typical detectors collect several points per second), in many cases the feed concentrations in preparative runs are high enough to saturate the detector. Then fractions must be collected, diluted if necessary, and re-analyzed in order to reconstruct the preparative chromatogram. In this case, the number of points in the chromatogram will be dramatically reduced. Figure 2-8 (exact data IV in Table 2-2) shows the same chromatogram as in Figure 2-4 (for \( V_{\text{feed}} = 0.6 \ V_0 \)), but with data points being generated every 10 seconds. The usual process of estimating the parameters was done, and the result for 5% relative error is shown in Table 2-11. It can be seen that the number of experiments required has increased significantly, to 11. This indicates that the large number of points in the earlier
runs (exact numbers are listed in Table 2-2) are in fact useful in minimizing the variance of the converged result.

Figure 2-8. Simulated chromatogram for exact data IV \( (V_{\text{feed}} = 0.6V_0) \). The same exact case as in Figure 2-4, but here the concentrations are reported approximately every 10 seconds (21 non-zero points).

Table 2-11. Number of experiments required for exact data IV with 5% relative error and less data points than that of Table 2-5

<table>
<thead>
<tr>
<th>Size of Monte Carlo matrix</th>
<th>Number of replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td>1000</td>
<td>8.2</td>
</tr>
<tr>
<td>10000</td>
<td>8.9</td>
</tr>
<tr>
<td>50000</td>
<td>8.8</td>
</tr>
</tbody>
</table>
One way to increase the number of data points when collecting fractions is to increase the loading; for isocratic elution, this should result in a wider band. The feed volume was increased to $V_{\text{feed}} = V_0$ in Figure 2-9 (exact data $V$ in Table 2-2), and the corresponding results shown in Table 2-12. It can be seen that the number of experimental data points did not increase appreciably in Figure 2-9, and the required number of experiments went to 10, which is a very small change from 11. Notice that the pure portion of the middle component is already quite narrow in Figure 2-8, and does not exist in Figure 2-9. Increasing the loading much further would in effect produce a case analogous to that of Figure 2-7, where the bands are highly mixed; the number of experiments required would therefore increase again.

The issue of increasing the number of experimental data points by using an appropriate combination of sampling (fraction collection) and detection is therefore one of great practical importance.

Table 2-12. Number of experiments required for exact data $V$ with 5% relative error

<table>
<thead>
<tr>
<th>Size of Monte Carlo Matrix</th>
<th>Number of replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td>1000</td>
<td>7.9</td>
</tr>
<tr>
<td>10000</td>
<td>8.5</td>
</tr>
<tr>
<td>50000</td>
<td>8.8</td>
</tr>
</tbody>
</table>
Figure 2-9. Simulated chromatogram for exact data $V (V_{\text{feed}} = V_0)$ with 24 non-zero points.

The effect of column efficiency on parameter estimation was investigated by repeating the runs above with higher plate counts. An exact data with 2500 plates was made (exact data VI in Table 2-2) with all other conditions the same as exact data II. Table 2-13 shows the number of experiments needed for this data set with 5% relative error. Convergence is achieved by 30 replicates, and 2 experiments should suffice. The results corresponding to 10% relative error are shown in Table 2-14, where 6 experiments are seen to be necessary. Comparing the chromatogram due to 2500 plates to that due to 1000 plates (Figure 2-10), a very slight increase in resolution is seen, due to the sharper peaks at $N=2500$. 
Slight differences can also be observed at the top and plateau portion of the second and third band. The effect on the estimation process is also small: the number of experiments needed decreased by 1 for 10% error, and remained the same for 5% error. Increasing the column efficiency further would make even less difference to the exact chromatogram, and it is likely that the number of experiments will be similarly insensitive.

Table 2-13. Number of experiments required for exact data VI with 5% relative error

<table>
<thead>
<tr>
<th>Size of Monte Carlo Matrix</th>
<th>Number of replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td>1000</td>
<td>1.7</td>
</tr>
<tr>
<td>10000</td>
<td>1.6</td>
</tr>
<tr>
<td>50000</td>
<td>1.6</td>
</tr>
</tbody>
</table>

Table 2-14. Number of experiments required for exact data VI with 10% relative error

<table>
<thead>
<tr>
<th>Size of Monte Carlo Matrix</th>
<th>Number of replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td>1000</td>
<td>6.5</td>
</tr>
<tr>
<td>10000</td>
<td>6.7</td>
</tr>
<tr>
<td>50000</td>
<td>6.6</td>
</tr>
</tbody>
</table>
Figure 2-10. Simulated chromatograms for exact data II and VI ($V_{\text{feed}} = 0.6V_0$). Solid line for exact dataset VI (2500 plates) and symbols for exact data II (1000 plates).

Finally, we compare the distributions generated by the relatively small number of replicates and those produced by the large Monte Carlo simulations. Figure 2-11 shows the histogram from the replicates and the density distribution generated by the Monte Carlo simulations for the case described in Table 2-6. It is clear that the original replicates would have provided a poor estimate of the 95th quantile, from which we calculate the number of experiments required to obtain 5% precision in the final isotherm parameters. The density distribution is seen to be asymmetric, with a significant tail. It should be noted that the number of experiments calculated from the Monte Carlo estimate of the 95th quantile is itself a
variable with some error. The process of generating Monte Carlo distributions is therefore repeated several times (typically 20) for the bottom right entry in each of the tables, in order to determine a standard deviation for the reported number of experiments. In all cases we have studied, these relative standard deviations are less than 1% of the mean, which indicates that the results of the tables can be regarded with confidence.

Figure 2-12 shows the same comparison of the histogram from the replicates with the density generated from the Monte Carlo simulations for the case described in Table 2-7. Again, a tailing density distribution is found. The single entry in the histogram with an RSS between 0.07 and 0.08 is an event of extremely low probability, since the corresponding Monte Carlo simulation with 50,000 replicates does not have a significant density value in this range.

The work presented here could be expanded in several ways. Firstly, other perturbations on the experimental data could be considered. For example, slight changes in flow rate or mobile phase composition or gradient timing and slope would result in changes in the times at which the bands emerged from the column. Since we have so far only considered perturbations in the effluent concentrations, but kept the times at which they emerge fixed to their exact values, this would provide an additional source of data regarding error propagation in the estimation of isotherms. Secondly, and more importantly, this work must be tested on real experimental data, for which the “exact” isotherm model is unknown; then the combined effect of model error and experimental error on the required number of
experiments must be assessed. While this is a vital problem, it can only be studied once the role of each kind of error has been estimated. The current work provides an explicit basis for quantifying the effect of experimental error on isotherm parameter estimation.
Figure 2-11. Plot of the histogram obtained from the 70 replicates used in Table 2-6 (exact data II with 10% relative error), and the corresponding density of the Monte Carlo distribution with the same covariance with 50,000 replicates (solid line).
Figure 2-12. Plot of the histogram obtained from the 30 replicates used in Table 2-7 (combination of exact data I and II both with 5% relative error), and the corresponding density of the Monte Carlo distribution with the same covariance with 50,000 replicates (solid line).
5. Conclusions

Numerical estimation of isotherm parameters is rapidly becoming an attractive and viable method for determining multicomponent adsorption data. In this paper, the influence of experimental error on the variance of the resulting parameters is addressed quantitatively. It is shown that the variation in the RSS, which is a measure of the goodness of fit, varies inversely as the square root of the number of replicates. From this result, we have shown how many experiments of a certain kind will be needed to identify parameters in a ternary mixture to 5% accuracy. Using experimental data of different kinds is discussed; in many practical separations, this may be a simple way to explore a larger fraction of the parameter space effectively. An attractive feature of the numerical approach used here is its effectiveness even when the effluent chromatogram consists of fully resolved peaks. The extent of loading, resolution and the number of data points in the experimental chromatogram are shown to be significant parameters in the estimation process; however, estimation is relatively insensitive to plate count for $N$ larger than 1000. Comparison of the discrete histograms with the Monte Carlo density distributions shows that the process is statistically sound.

An important feature of the results is the wide variation in the number of experimental replicates needed to obtain tightly bounded isotherm parameters. In some cases (e.g., Table 2-10), the estimation method described here would require an unrealistically large number of experimental replicates. As against that, the marked decrease in the number of experimental replicates required when more than
one kind of experiment was run (e.g., Table 2-8) is very promising, and may provide a way to achieve good estimates from relatively few experimental runs.

In the present work, we have focused on experimental errors that changed the effluent concentration values. As mentioned in the text, other sources of error can lead to changes in the effluent times. We hope to present results incorporating these additional sources of error in a future publication.

Acknowledgements

The code FFSQP was used for the optimization runs; this code was created by Professor André Tits of the University of Maryland, and was used here with his permission.
References


Chapter 3

Effect of Experimental Error on the Precision of Numerically Estimated Multicomponent Isotherms in Preparative Chromatography

Ling Zhang¹ and Ajoy Velayudhan²

¹Department of Bioengineering, Oregon State University, 116 Gilmore Hall, Corvallis, OR 97331-3906
²Bistol-Myers Squibb Company, P.O. Box 4755, Syracuse, NY 13057

Manuscript in preparation
Abstract

The classical chromatographic methods of isotherm determination are extremely tedious and time-consuming. Their application to multicomponent isotherms is limited. Recently, numerical method has drawn more attention and been proven promising to generate good results. Earlier, we addressed the question of how the precision of the resulting isotherm parameters is affected by experimental error for the first time in the literature. By studying the effect of error in effluent concentrations, we developed an approach to bound the error in final estimated parameters given certain amount of error in experimental chromatograms. Here, we demonstrate that the same approach is applicable to cases in which error is in retention times or both types of error (concentration as well as time) are present. The number of experimental replicates required to achieve a desired level of accuracy in estimated isotherm parameters is calculated. The extent of effect from different error source is compared.

1. Introduction

Preparative chromatography is recognized by the pharmaceutical industry as the only purification method capable of producing extremely high purity of therapeutics with a high production rate and a good recovery yield. The optimal production rate in preparative chromatography is usually achieved when the column is overloaded [1]. Since under overloaded conditions the equilibrium isotherm becomes nonlinear and band profiles are interdependent on each other, the
determination of the optimum conditions for a separation is very complex. In the practice of optimization of preparative chromatography, determination of isotherm is necessary, and the accuracy of the isotherm parameters is of great importance in assuring the quality of design of the optimum experimental conditions.

The methods for isotherm determination basically fall into three groups. They are static methods, dynamic methods and numerical methods. Static methods are rarely used nowadays because they involve tedious laboratory work that is extremely time-consuming and requires a large amount of material. A few dynamic methods are available to determine multicomponent isotherms. Frontal analysis (FA) is probably the most popular one and has been widely applied to measure competitive isotherms for binary mixtures [2-5]. It has also been used to determine isotherms for ternary mixtures [6, 7]. Another dynamic method is the perturbation method or elution on a plateau method. Perturbation method has been applied to generate competitive isotherms prerequisite for the design and optimization of simulated moving bed (SMB) [8, 9].

Numerical method derives isotherms from eluted band profiles by minimizing the differences between experimental chromatograms and simulated chromatograms using chosen isotherm model and parameters. Dose et al. [10] reported numerically determined the single-component isotherms of N-benzoyl-(L,D)-alanine and N-benzoyl-(L,D)-phenylalanine. They used a modified simplex algorithm to search for parameter values of bi-Langmuir isotherms and found a good agreement with those determined by frontal analysis. James et al. [11]
applied a conjugate gradient algorithm for optimization to determine single-component and competitive isotherms for two binary mixtures. In one case, bi-Langmuir isotherm was used for the ketoprofen enantiomers; in the other Langmuir isotherm was used for benzyl alcohol and 2-phenylethanol. Good agreement was observed, in both cases, between the isotherms obtained by numerical method and those by elution by characteristic points and frontal analysis. Recently, Felinger et al. [5] implemented two search algorithms — the super modified downhill simplex and the Levenberg-Marquardt steepest descent, to determine the competitive isotherms of the mixture of 1-phenyl-1-propanol enantiomers. Four isotherms were tested; they are Langmuir, bi-Langmuir, Tóth and Langmuir-Freundlich.

Antos et al. [12] applied the least-squares Marquardt algorithm in determining the competitive Langmuir isotherms of two components (benzophenone and diolone acetate) and the modulator (ethyl acetate) in the mobile phase simultaneously. They simulated the chromatograms for different sample concentrations and mobile phase flow rates using the numerical estimated isotherm parameters, and found a good agreement with experimental results. Using the same isotherm parameter values for one of the components and the modulator, combined with an equation describing the dependence of isotherm parameters on modulator concentration for another component, they were also able to predict separation processes for different mobile phase compositions. Antos et al. used the same Marquardt algorithm in another paper [13] to determine Langmuir isotherm of methyl deoxycholate at various mobile phase compositions from isocratic elution
band profiles. The effect of mobile phase composition on the isotherm parameters was described by a three-parameter equation. The results were used to simulate overload gradient elution and satisfactory agreement with experimental results was obtained.

The application of the numerical method to determine adsorption isotherms is receiving more attention in the design of simulated moving bed separations (SMB) in the past several years. Ching et al. [14] identified the single-component Langmuir isotherm for both Fenoprofen enantiomers respectively, then applied the competitive Langmuir isotherms with the single-component parameter values to simulate racemic separations. The results were satisfactory so that the numerical determined isotherms were further used in the study of the operation region of a SMB system. The parameters of a modified competitive Langmuir isotherm were determined by Juza [15], and applied in a SMB performance test for a binary mixture consisting of cyclopentanone and cycloheptanone. The competitive Langmuir isotherms of the same two compounds and their dependence on mobile phase compositions were determined by Antos and Seidel-Morgenstern [16] in their study of a two-step gradient SMB process.

One of the advantages of numerical method over dynamic method in isotherm determination is that numerical method is readily applicable to mixtures of more than three components. The application of dynamic methods involves injection of mixtures with known compositions (known concentrations of all the components). For each composition, samples at various concentration levels are
injected to cover a large range of isotherm curve. That means a considerable amount of pure substances must be available to make up the sample mixtures in each experiment. However, pure substances are usually available in limited quantities during the early stage of purification process development. Numerical method does not require sample mixtures with known compositions or systematic change of sample concentrations, i.e. any preparative band profiles for which the individual component concentrations are known can be used as experimental data. Thus no separate preparation of sample mixtures is needed as the number of components increases. Moreover, if the number of components is above two, the difficulties of signal interpretation [1] and the amount of time required by dynamic methods make them unattractive for practical applications. The fact of the matter is that desired products are often present in fermentation broth, cell culture or natural sources from which they must be purified. A mixture from these sources usually contains several impurities as well as desired products. The ideal method ought to have the characteristics that make it easy to be extended to mixtures of such nature. The cost associated with numerical method as the number of components grows is, of course, increased computation time. With the relatively easy access to high-speed computers and fast-developing computational technology, it is very promising to achieve simultaneous parameter estimation for complex mixtures in an acceptable amount of time.

To determine isotherms by numerical method, the isotherm model must be chosen beforehand. With some initial values of isotherm parameters, band profiles
can be calculated. The isotherm parameter values can be adjusted by comparing the calculated band profiles to experimental profiles so that the deference between the two gets smaller. The best-fit isotherm parameters are obtained when the deference between the calculated and experimental band profiles is minimized. A fundamental question arises from the numerical method is how error in the experimental band profiles translate into error in the estimated isotherm parameters, provided that the isotherm model is appropriately chosen. When this numerical parameter determination process is repeated on several experimental profiles (experimental profiles reproduced under the same conditions), how different the isotherm parameters could be resulted from individual profiles with experimental error caused by variations in flow rates, mobile phase composition, detection or fraction collection, etc? Is there any relationship between the amount of error in the experimental data and the precision of estimated isotherm parameters?

We reported [17] for the first time the study of the effect of experimental error on the precision of numerically estimated isotherm parameters for a ternary system. We quantified the variance in isotherm parameters as a function of experimental error in effluent concentrations and developed a method to calculate the number of experimental replicates needed to obtain a certain level of precision in estimated isotherm parameters explicitly. Here we present studies of the influence of experimental error in effluent times (one of the two types of error in a two-dimensional band profile) and the combination of both error sources for the same ternary system.
2. Methodology

2.1. Simulation and optimization

In order to ensure the only source of error is in the chromatograms, we have used the computer simulated chromatograms as the “experimental” chromatograms. The isotherm model and its parameters are chosen so that an exact chromatogram can be simulated. Experimental error is added to the exact chromatogram by incorporated Gaussian noise. A representative of such computer simulated “experimental” chromatogram is shown in Fig. 3-1. We define relative error in effluent concentrations at each data point as the ration of the error to the concentration at that point. We use a random number generator to generate a series of normally distributed random numbers with mean 0 and variance 1. Then those random numbers are scaled to produce the desired amount of relative error and added point-wise into the exact chromatogram. As shown in Fig. 3-1, each data point has a relative error of no more than 5%. The relative error in retention times is defined as the ration of error to the dead time (the breakthrough time of the mobile phase). Unlike the way error is added into effluent concentrations, only one random number from a normal distribution is generated by the random number generator. We use that number to scale the dead time for a given chromatogram to produce the desired amount of error in time. So a chromatogram with 3% relative error in retention times means there is no more than 3% error in the dead time, but the exact amount could be anywhere between -3% to 3%. 
Craig plate model has been applied successfully to simulate non-linear chromatographic separations [18-23] and is used in our study. The mass balance equation of solutes is given as

\[ c(i, j, k) + \phi q(i, j, k) = c(i, j - 1, k - 1) + \phi q(i, j, k - 1) \]  

(1)

where \( c(i, j, k) \) and \( q(i, j, k) \) are the mobile phase and stationary phase concentration respectively for solute \( i \) in the \( j \)-th plate at the \( k \)-th instant (\( c \) and \( q \) are assumed to be in equilibrium with each other); \( \phi \) is the volumetric phase ratio.
The multicomponent Langmuir isotherm model is used to describe the competitive adsorption of the solutes:

\[ q_i = \frac{a_i c_i}{1 + \sum_{j=1}^{p} b_j c_j}, \quad i=1, 2, 3, ..., p \]  

(2)

where \( p \) is the total number of solutes. There are two isotherm parameters for each solute: \( a \) is the distribution coefficient at low concentrations and \( b \) is the equilibrium constant. As we mentioned in the previous paper [17], we choose Langmuir isotherm model because it is simple and explicit. There have been comparison studies of different isotherm models [24-27] on how well they fit the experimental measurements. More complicated models (with relatively more parameters) did not necessarily fit data better than simple models. Simple isotherm model is preferable in numerical estimation because the difficulty of optimization search and the associated time cost increase as the number of parameters does. By using a relatively simple isotherm model, we are able to focus on answering the fundamental question of estimation error. For the same reason, implicit models are not among the first choices.

To calculate chromatographic band profiles, the system of nonlinear equations, Eq. (1), is solved numerically using an algorithm based on a modified Brent’s method [28-30]. The isotherm parameters are estimated by the following procedure: firstly, guess initial values for the selected isotherm model (competitive Langmuir model in this study); secondly, simulate the band profiles at the column
outlet using algorithms mentioned above; finally, compare the simulated and experimental band profiles and minimize their differences by adjusting isotherm parameter values. The third step of the procedure is performed by an optimization program [31] which implements sequential quadratic programming (SQP) algorithm to minimize objective functions. The objective function \( J \) in our problem is defined as:

\[
J(c; a, b) = \sum_{i=1}^{p} \sum_{m=1}^{M} \left[ c_i^{\text{sim}}(L, t_m; a, b) - c_i^{\text{exp}}(L, t_m) \right]^2
\]

where \( c_i^{\text{sim}} \) and \( c_i^{\text{exp}} \) represent the simulated and experimental effluent concentrations for component \( i \) at time \( t_m \) at which concentrations are measured in experimental chromatograms. \( M \) is the total number of data points that cover the entire effluent history; \( p \) is the total number of components. \( L \) is the length of the column. When a minimum of \( J \) is found, the optimization program returns the best-fit isotherm parameters. To maximize the chances of finding the global minimum, we always try several widely different initial guesses of isotherm parameters for every experimental chromatogram.

### 2.2. Statistical analysis

We have presented in the previous paper [17] the relationship between the experimental error in effluent concentrations and the error in the final estimated isotherm parameters. The same statistical analysis is applied here to the study the effect of experimental error in retention times on the precision of isotherm
parameters. We define the error in the estimated isotherm parameters as the residual sum of squares (RSS):

$$RSS = \left[ (\overline{X}_1)^2 + (\overline{X}_2)^2 + \ldots + (\overline{X}_6)^2 \right]^{1/2}$$

where \(\overline{X}_i = \frac{a_i - a_i,\text{exact}}{a_i,\text{exact}}, \overline{X}_i = \frac{b_i - b_i,\text{exact}}{b_i,\text{exact}}, \ldots, \overline{X}_i = \frac{b_3 - b_3,\text{exact}}{b_3,\text{exact}}\). The isotherm parameters of the first component are called \(a_1\) and \(b_1\). Similarly, \(a_2\) and \(b_2\) are the isotherm parameters for the second component; \(a_3\) and \(b_3\) are for the third component. One set of best-fit parameters is returned when the optimization process is performed on one experimental chromatogram. If we repeat the optimization process on \(n\) experimental chromatograms, we obtain \(n\) sets of best-fit parameters. The average best-fit parameters of \(n\) replicates are \(\overline{a}_1, \overline{a}_1, \ldots, \overline{b}_3\). The exact values of parameters are \(a_1,\text{exact}, b_1,\text{exact}, \ldots, b_3,\text{exact}\). \(\overline{X}_i\) is the average parameter vector scaled to make the average value of each parameter unity.

We have found that the individual estimates of each parameter from \(n\) replicates follow a Gaussian distribution. Two examples of such distribution for two parameters obtained from data with different types of experimental error are shown in Fig. 3-2. The exact isotherm parameter values are listed in Table 3-1. The elution band profiles are shown in Fig. 3-1.
Figure 3-2(a). Probability plot for estimated isotherm parameter $a_3$ from data with 3% relative error in retention times (100 replicates).
Figure 3-2(b). Probability plot for estimated isotherm parameter $b_2$ from data with 3\% relative error in retention times and 5\% relative error in effluent concentrations (100 replicates).

Table 3-1. Competitive Langmuir isotherm parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$a_1$ [-]</th>
<th>$b_1$ [ml/mg]</th>
<th>$a_2$ [-]</th>
<th>$b_2$ [ml/mg]</th>
<th>$a_3$ [-]</th>
<th>$b_3$ [ml/mg]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact value</td>
<td>3.00</td>
<td>0.100</td>
<td>4.50</td>
<td>0.113</td>
<td>6.75</td>
<td>0.135</td>
</tr>
</tbody>
</table>
Since the distribution of each standardized isotherm parameter can be approximated by a normal distribution, that is:

\[ X_1 = \frac{a_{1,j} - a_{1,exact}}{a_{1,exact}} \sim N(0, \sigma_i^2) \]  

(5)

where \( a_{1,j} \) is any individual estimate from \( n \) replicates. Similar results are found for all the other parameters. Then the average parameter follows a weighted normal distribution:

\[ \overline{X}_1 \sim N\left(0, \frac{\sigma_1^2}{n}\right) = \frac{\sigma_1}{\sqrt{n}} N(0,1) \]  

(6)

Thus, the square of \( \overline{X}_1 \) has a weighted \( \chi^2 \) distribution with 1 degree of freedom [32]:

\[ \left(\overline{X}_1\right)^2 \sim \frac{\sigma_1^2}{n} \chi_i^2 \]  

(7)

Similar notions apply to all the other parameters. Consider the definition of RSS, its dependence on the average parameter vector can be represented as:

\[
RSS = \left[ \frac{\sigma_1^2}{n} \chi_1^2 + \frac{\sigma_2^2}{n} \chi_1^2 + \ldots + \frac{\sigma_6^2}{n} \chi_1^2 \right]^{1/2}
\]

(8)

The term inside of square brackets is independent of the number of replicates \( n \).

We have reported [17] that the mean and variance of the term in the square bracket do not depend on \( n \) either. Therefore, taking into account of the correlations among
all isotherm parameters, the distribution of RSS can be simulated numerically. From equation (8), we can calculate the number of replicates \( n \) required to obtain any value of RSS.

3. Results and discussion

Since the RSS of the final estimated isotherm parameters is inversely proportional to the square root of \( n \), it is possible to make RSS sufficiently small by doing multiple experiments. RSS value of 0.05 implies, in a statistical sense, that there is no more than 5% error in any estimated isotherm parameters. This is a reasonable requirement for the estimated isotherm parameters to be used in other applications. It is interesting to find out how many experimental replicates are needed to obtain RSS=0.05 given certain amount of experimental error in the chromatograms. As mentioned earlier, the effect of concentration error has been presented in the previous paper, here we present the results obtained from data with error in retention times and with both types of error.

The first exact data set used was simulated using isotherm parameters listed in Table 3-1. The feed sample concentration is 2.0 mg/ml, feed volume is 0.6\( V_0 \) and flow rate is 1.0 ml/min. The exact chromatogram and a representative chromatogram with experimental error are shown in Fig. 3-1. The number of theoretical plates used in all the simulations presented in this paper is 1000. The error in retention times, in dead time more precisely, causes bind profiles to shift back or forward in time depending on the individual case. It should be noted that it
is not a simple shift without any other changes in band profiles. The curvature of the elution peaks has changed slightly although the basic shape of band profiles remains the same.

The dependence of RSS on the number of experiments is shown in Fig. 3-3 for data with 2% or 3% relative error in retention times. It can be read from the curve in Fig. 3-3 that at least 3 or 4 experiments are required to obtain RSS=0.05 if there is 2% error in retention times. That number will go up to 7 if the amount of error increases to 3% in order to achieve the same level of accuracy in estimated parameters. When there is 5% relative error in retention times (figure not shown), it would need 20 experiments to get estimates of isotherm parameters with RSS=0.05.

![Figure 3-3. RSS as a function of n, the number of experiments needed for data with 2% (solid line) or 3% (dotted line) relative error in retention times. $V_{\text{feed}}=0.6 \ V_0$.](image-url)
Compare the effect of retention time error to that of concentration error (Table 3-2); it is obvious that significantly more experimental replicates are required to obtain the same level of accuracy in estimated isotherm parameters for data with error in retention times than those with the same amount of error in effluent concentrations. One way of examining the difference in the extent of effect by two error sources is to look at the variance of individual parameters. Fig. 3-4 shows the variance of estimated parameters for data with 10% concentration error (Fig. 3-4(a)) or 3% retention time error (Fig. 3-4(b)). The number of experiments needed to obtain RSS=0.05 for the two cases is very close, 6 for one and 7 for the other. Since half experiment makes no sense in practice, we always round it to the nearest integer. With error only present in concentrations, the spread of all a parameters is far less than that of b parameters as shown in Fig.3-4(a). In the case of error in retention times (Fig. 3-4(b)), the variation in a parameters are no longer tightly bounded. This is because changes in a parameters alone can shift the chromatogram back or forward in time while keeping the peak shapes roughly consistent. But changes in b parameters alone can not make the chromatogram back or forward in time without changing peak shapes dramatically. The “shifting” of chromatogram caused by error in retention times contributes more variation to the estimated a parameters. Since a and b parameters are correlated and being estimated simultaneously, b parameters do not necessarily vary less in order to compensate error in a parameters.
Table 3-2. Number of experiments required to obtain isotherm parameters with \( \text{RSS}=0.05 \) \( (V_{\text{feed}}=0.6V_0) \)

<table>
<thead>
<tr>
<th>Relative error</th>
<th>2%</th>
<th>3%</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>In effluent concentrations</td>
<td>-</td>
<td>-</td>
<td>1.8</td>
<td>6.3</td>
</tr>
<tr>
<td>In retention times</td>
<td>3.2</td>
<td>6.7</td>
<td>20</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 3-4(a). Variations in simulated parameters (standardized with respect to the exact parameter values). Data with 10% relative error in effluent concentrations (70 replicates).
After the effect of each type of error has been studied, we now move on to examine the case where both types of error are present. This is a more realistic representation of real experimental data. Fig. 3-5 shows how RSS decreases as a function of $n$, the number of experiments needed. Table 3-3 gives the $n$ values to obtain estimated isotherm parameters with RSS=0.05. Each one of the four $n$ values in the table is resulted from data with the combination of error in effluent concentrations (in columns, either 5% or 10%) and error in retention times (in rows, either 2% or 3%). It is good to know that the number of experiments needed
for combined error case is less, in general, than the sum of two separate cases. For example, it needs approximately 4 experiments for data with 5% concentration error and 2% retention time error. For data with 5% relative error in concentrations alone (Table 3-2), it needs about 2 experiments; for data with 2% relative error in retention times alone, it needs at least 3 experiments.

Figure 3-5. RSS as a function of $n$, the number of experiments needed for data with 5% relative error in effluent concentrations and 3% relative error in retention times. $V_{\text{feed}}=0.6 \; V_0$. 

![Graph showing RSS as a function of n]
Table 3-3. Number of experiments required for data with combined error to obtain isotherm parameters with RSS=0.05 ($V_{\text{feed}}=0.6V_0$)

<table>
<thead>
<tr>
<th>Combined error</th>
<th>In effluent concentrations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td>In retention times</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>3%</td>
</tr>
</tbody>
</table>

Another exact data set with smaller feed volume was simulated using the same isotherm parameters listed in Table 3-1. The feed volume is $0.3V_0$ and all other conditions are the same as the first data set. The exact chromatogram and a representative “experimental” chromatogram can be found in Fig. 3-6. All three components are fully separated now. In the larger loading run (Fig. 3-1), there is considerable mixing between any two adjacent two components and the middle component has a narrow portion of pure band.

The number of experiments needed to obtain isotherm parameters with RSS=0.05 for the smaller loading run are listed in Table 3-4 and Table 3-5 respectively for data with only one type of error and data with combined error. Again, it takes more experimental replicates to achieve the same level of precision in estimated isotherm parameters for data with error in retention times than in effluent concentrations. With 5% relative error in concentrations, it needs single experiment; but with 5% relative error in retention times, the number of
Figure 3-6. Simulated chromatograms. Solid lines represent the exact chromatogram and dotted lines with symbols represent chromatogram with 5% relative error in effluent concentrations and 2% relative error in retention times. ($c_{\text{feed}} = 2 \text{ mg/ml}, V_{\text{feed}} = 0.3V_0$)

Experiments required goes up significantly to 11. When both types of error are present in the data, it requires more experiments than each separate case to obtain $RSS=0.05$ in final isotherm parameters. It takes about 4 experiments for 2% time error and 10% concentration error combined, for instance, and it takes about 2 experiments in each separate case.
Table 3-4. Number of experiments required to obtain isotherm parameters with 
\( \text{RSS}=0.05 \) \( (V_{\text{feed}}=0.3V_0) \)

<table>
<thead>
<tr>
<th>Relative error</th>
<th>2%</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>In Effluent concentrations</td>
<td>-</td>
<td>1.0</td>
<td>2.1</td>
</tr>
<tr>
<td>In retention times</td>
<td>1.8</td>
<td>11</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 3-5. Number of experiments required for data with combined error to obtain 
Isotherm parameters with \( \text{RSS}=0.05 \) \( (V_{\text{feed}}=0.3V_0) \)

<table>
<thead>
<tr>
<th>Combined error</th>
<th>In effluent concentrations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td>In retention times</td>
<td>2%</td>
</tr>
</tbody>
</table>

We reported in the previous paper that for data with concentration error 
alone it needed fewer experimental replicates to obtain the same level of precision 
in estimated parameters using chromatograms of lower loading run than those of 
larger loading run. By comparing results in Table 3-2 and 3-3 to those in Table 3-4 
and 3-5, we find out that the lower loading run needs fewer experiments for error in 
retention times as well as the combined error case. This confirms our conclusion 
that the interaction among sample components are well captured to result in 
efficiently estimated isotherm parameters even though all components are fully 
resolved at column outlet because numerical method simulates the entire separation
process from sample injection where all components are mixed and continue to be mixed and interact over a portion of the column while they travel along. It should also be noticed that, in the special case of the multicomponent Langmuir isotherm, the same isotherm parameters are found in both single-component and multicomponent isotherms. Thus, for component $j$, the parameters $a_j$ and $b_j$ are identical in the single and multicomponent forms. It is therefore not necessary to have highly overlapping peak profiles in order to assess competitive behavior. In other, more complex, isotherms, there are parameters that are found only in the multicomponent case, e.g., a term such as $a_j c_j$ will obviously not be found in a single-component isotherm. It would be worthwhile to repeat the current work on such isotherms to see if the present conclusions continue to hold.

It is very interesting to notice that the number of experiments required for lower loading data with combined error seems to be exactly the sum of the two data with only one type of error. Remember that, for the larger loading data set, the number of experiments required to obtain RSS=0.05 in final parameters for the combined error case is generally less than the sum for the two separate cases. It is certainly expected that more experiments need to be done when more error is present in the experimental data. But it is not clear how much each type of error contributes to the final results. When real experimental data are used where no “exact” chromatogram exists, it always helps to minimize variations in retention times since a relatively small variation in retention times causes more variations in estimated parameters compared to error in concentrations.
4. Conclusions

As numerical method gains popularity in recent years as a viable alternative to the most commonly used dynamic methods for adsorption isotherm estimation, some important problems remain unsolved. Among them is the fundamental question of the effect of experimental error on the precision of the estimated parameters. The concentration range in which multicomponent isotherms are estimated is usually too high to give direct readings on a detector. Fractions must be collected and analyzed offline to reconstruct the chromatogram. This inevitably introduces error in the reconstructed chromatograms. Another type of error occurs due to variations in flow rate or mobile phase composition that results in shifts of elution peaks in time. Our investigation attempts to quantify the influence of each type of experimental error on the variance of final estimated isotherm parameters.

The focus of the present work is the effect of error in retention times as well as combined effect of both time error and concentration error. We have found that the RSS of averaged parameters resulted from estimation on $n$ experimental replicates is inversely proportional to the square root of $n$. The number of experimental replicates required to obtain any value of RSS can be determined theoretically. We have calculated the $n$ values to make RSS=0.05 for data with various amount of error in time and for data with different combinations of error in both concentrations and time. Compared to the effect of concentration error, error in retention times gives rise to more variation in $a$ parameters. More replicates are required to achieve the same level of accuracy in estimated parameters for data
with error in retention times than in effluent concentrations. The combined effect of both types of error results in even larger $n$ to obtain RSS=0.05, but it not clear how the two are correlated. For the two sets of exact data tested in this study, the effect of combined error is at most the sum of the effect from separate error sources.

**Acknowledgements**

We thank Professor André Tits of University of Maryland for his permission to use the optimization code FFSQP.
References


Chapter 4

Preparative Purification of Chlorophyll A from Spinach by High Speed Countercurrent Chromatography

Ling Zhang\textsuperscript{1} and Ajoy Velayudhan\textsuperscript{2}

\textsuperscript{1}Department of Bioengineering, Oregon State University, 116 Gilmore Hall, Corvallis, OR 97331-3906
\textsuperscript{2}Bristol-Myers Squibb Company, P.O. Box 4755, Syracuse, NY 13221

Manuscript in preparation
Abstract

High speed countercurrent chromatography (CCC) was used for the first time to purify chlorophyll a in preparative scale. Fresh spinach leaves were blended in acetone and chlorophylls were extracted to petroleum ether by liquid-liquid extraction. Glass columns packed with polymeric reversed phase were used to separate chlorophylls from the more retentive carotenoids and to partially remove the less retained impurities. The fraction containing chlorophylls from column chromatography was injected into CCC column and chlorophyll a was purified. The highest CCC production achieved was to produce 448 mg of chlorophyll a with a purity of 99% and a recovery yield of 86% in two hours. CCC separation was optimized and evaluated in terms of loading conditions.

1. Introduction

Chlorophylls are a group of pigments widely present in photosynthetic organisms including higher plants, algae and some bacteria. Chlorophylls are often used as a quantitative reference in physiological and ecological studies. They are the only natural green plant pigment found in abundance [1]. Recent studies [2-7] have proven that chlorophylls and their water-soluble derivatives known as chlorophyllins show antigenotoxic and anticarcinogenic properties. For detailed studies of the role of chlorophylls in cancer prevention, large quantities of pure materials must be purified in an efficient and economic way.
The most essential method for preparative purification of chlorophylls has been the conventional column chromatography packed with diethylaminocellulose [8], agar [9, 10] or powdered sugar [11] as absorbents. Recently, more reproducible and faster preparative HPLC have been used in either normal [12] or reversed-phase [13] mode. Application of high speed countercurrent chromatography (HSCCC) to purification of chlorophylls has not been reported in the literature.

Countercurrent chromatography is recognized as a preparative separation technique [14, 15, 16] that offers several advantages over HPLC. CCC uses liquid as stationary phase and has no solid matrices so that the loss of material due to irreversible adsorption on solid matrices is avoided. The CCC column is basically a piece of empty plastic tubing that only filled with the liquid of choice when in use. Thus, no cleaning or regeneration of column is needed, and rather crude samples possibly containing some particulates can be injected. The separation principle is based on different partition of solutes between phases. So the nature of the column can be altered by changing the solvent system, or changing the direction of column rotation for the same solvent system to run in either normal or reverse mode. Most importantly, solvents are relatively inexpensive compared to HPLC column packing materials. Because the stationary phase in CCC is a liquid, the loading capacity of a CCC column is much higher that of an HPLC column with an equal internal volume [16].
The viability of CCC as a preparation separation tool is evident from the numerous applications published in the literature. Many of these applications deal with preparative purification of natural products [15, 17-31]. The aim of our study is to develop a process that can produce hundreds of milligrams of chlorophyll $a$ with a purity of at least 99% by using HSCCC as the purification method. The CCC separation is optimized in terms of loading conditions for high production and recovery yield.

2. Experimental

2.1 Materials

Spinach (Spinacia oleracea) was bought from local supermarkets. ACS grade acetone and petroleum ether were purchased from Fisher Scientific (Pittsburgh, PA, USA). HPLC grade ethyl acetate and acetonitrile were also purchased from Fisher Scientific. Non UV petroleum ether (OmniSolv®) was purchased from EM Science (Gibson, NJ, USA). Food grade anhydrous ethanol was purchased from AAPER Alcohol and Chemical Co. (Shelbyville, KY, USA). Deionized-distilled water was obtained from a Mill-Q ultra-pure water system (Millpore, Bedford, MA, USA) in our lab. Chlorophyll $a$ and chlorophyll $b$ standards were purchased from Sigma (St. Louis, MO, USA).
2.2 Apparatus

The multilayer-coil countercurrent chromatography (CCC) device is made by P.C. Inc. (Potomac, MD, USA). The column is a right-handed coil of 1.68 mm I.D. polytetrafluoroethylene (PTFE) tubing with a capacity of 350 ml. A piston pump (Eldex Laboratories, Inc., San Carlos, CA) was used to deliver liquid phases. A V4 UV detector (ISCO, Inc., Lincoln, NE) was attached to the CCC column outlet to detect elution peaks.

Waters (Milford, MA, USA) HPLC system consists of a Model 600 quaternary pump, Model 717 plus auto-sampler and Model 486 UV detector. The HPLC system is controlled by a personal computer with Waters Millennium software. Alltima C18 HPLC column (0.46 cm I.D. × 25 cm, 5 μm) and guard column (0.46 cm I.D. × 0.75 cm) were purchased from Alltech (Deerfield, IL, USA).

Flex-column (2.5 cm I.D. × 50 cm) and Chromaflex (1 cm I.D. × 60 cm) standard glass columns were purchased from Kimble/Kontes (Vineland, NJ, USA). Amberchrom reversed-phase resin (75 μm) was purchased from TosoHaas (Montgomeryville, PA, USA).

2.3 Procedures

2.3.1 Batch Extraction

All the experiments were carried out at temperature between 20 °C and 25 °C in a light-reduced room. Only fresh and clean spinach leaves were used in the
experiments. The leaves were blended in acetone (ACS grade) at 15,000 rpm for two minutes. Then petroleum ether (ACS grade) was added into the slurry and mixed thoroughly by stirring vigorously. The mixture was settled on a shaker for at least 10 minutes at 80 rpm. The dark green petroleum ether layer formed on top of the mixture was collected and saved. The extraction was repeated three or four times depending on the color of the slurry by adding fresh petroleum ether until the spinach residual looked pale green. The dark green top layers from all batches were pooled and evaporated by rotary evaporation at 30 °C until a sticky paste formed in the evaporation container, i.e., there was no more free-flowing liquid. Since heat would cause chlorophylls to degrade quickly, spinach leaves were frozen in advance in order to keep the temperature low during blending. Without freezing, the temperature of the slurry could reach 35-37 °C after two minutes. In the current method, the final temperature remained below room temperature.

2.3.2 Isolation by column chromatography

The extract paste formed after rotary evaporation was re-suspended in ethyl acetate/ethanol (25%/75%, v/v). The sample solution was centrifuged at 7,800g for 15 minutes. The supernatant was eluted through a glass column packed with Amberchrom reversed-phase resin by ethyl acetate/ethanol (25%/75%, v/v) as mobile phase. Two sizes of columns were tried. One (1 cm I.D. × 60 cm) was packed with 47 ml of stationary phase and the other (2.5 cm I.D. × 50 cm) with 200 ml. The smaller column was eluted at 4 ml/min and the bigger column at 9 ml/min
by mobile phase driven by a peristaltic pump. The fraction containing chlorophylls was collected and evaporated by rotary evaporation at 30 °C. The sample paste was re-suspended in CCC mobile phase. This was the sample ready to be injected into the CCC column.

2.3.3 Purification by CCC

The solvent system used in CCC is petroleum ether (HPLC grade) and 90% ethanol (90% food grade anhydrous ethanol and 10% DI water) at equal volume [32]. The two solvents were mixed in a separation funnel by shaking vigorously to equilibrate. The two phases were separated shortly before use. The upper phase (petroleum ether rich) was used as stationary phase and lower phase (ethanol rich) as mobile phase. The empty CCC column was first filled up with the stationary phase. Then the column inlet was disconnected from the pump and the feed sample was injected into the column by a syringe. After sample injection, the column was reconnected to the pump. Elution began when start pumping the mobile phase after the column was brought up to speed. The effluent was detected by a UV detector at 500 nm unless mentioned otherwise. The column was rotated at 780 rpm. Mobile phase flow rate was 3.5 ml/min. The flow and rotation were stopped after chlorophyll a was completely eluted. The phases left in the column and the un-eluted compounds dissolved in them were blown out by nitrogen.
2.3.4 HPLC analysis

The crude extract from spinach leaves, feed samples and fractions from each chromatographic separation were analyzed on an Alltima C18 column eluted by acetonitrile/ethanol (55%/45%, v/v). Proper dilutions were made to make sure the effluent concentrations fall into the calibration curve constructed by analyzing standard chlorophyll samples of known concentrations. The injection volume was 10 µl and the flow rate was 1 ml/min. The effluent was monitored at 440 nm by a UV detector.

3. Results and discussion

3.1 Batch Extraction

We found that two minutes was sufficient to achieve complete homogenization without heating up the slurry during the blending step. The volume/mass ratio of acetone/spinach was kept roughly between 1 and 2 ml/g. The volume of petroleum ether added at each extraction step was one third of acetone volume used for blending. These two ratios combined with three or four times of extraction were appropriate to achieve efficient chlorophyll extraction from spinach leaves of various qualities all year round. The amount of chlorophyll a extracted from different batches is listed in Table 4-1. We were able to extract on average 47 mg of chlorophyll a from 100g fresh spinach leaves.
Table 4-1. Mass of chlorophyll \( a \) in crude extract

<table>
<thead>
<tr>
<th>Batch #</th>
<th>Spinach (g)</th>
<th>Acetone (ml)</th>
<th>Number of extractions</th>
<th>Chl ( a ) extracted (mg) / 100g spinach</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1800</td>
<td>2000</td>
<td>3</td>
<td>22.1</td>
</tr>
<tr>
<td>2</td>
<td>1000</td>
<td>1500</td>
<td>3</td>
<td>49.0</td>
</tr>
<tr>
<td>3</td>
<td>970</td>
<td>1500</td>
<td>3</td>
<td>59.4</td>
</tr>
<tr>
<td>4</td>
<td>705</td>
<td>1500</td>
<td>3</td>
<td>69.5</td>
</tr>
<tr>
<td>5</td>
<td>1750</td>
<td>1500</td>
<td>3</td>
<td>33.8</td>
</tr>
<tr>
<td>6</td>
<td>1300</td>
<td>1500</td>
<td>3</td>
<td>62.3</td>
</tr>
<tr>
<td>7</td>
<td>1500</td>
<td>1500</td>
<td>4</td>
<td>35.6</td>
</tr>
<tr>
<td>8</td>
<td>1170</td>
<td>1500</td>
<td>4</td>
<td>53.7</td>
</tr>
<tr>
<td>9</td>
<td>790</td>
<td>1500</td>
<td>4</td>
<td>38.3</td>
</tr>
</tbody>
</table>

The HPLC analysis of crude extract is shown in Fig. 4-1. Compared to the HPLC analysis of authentic samples, the peak at 12 min was identified as chlorophyll \( a \) and the peak at around 9 min was identified as chlorophyll \( b \). The major impurities other than chlorophyll \( b \) from spinach according to their elution order on a C-18 reversed-phase column [33] are believed to be: (i) xanthophylls (a group of peaks before chlorophyll \( b \) in Fig. 4-1); (ii) hydrocarbonaceous carotenoids (unresolved peaks after chlorophyll \( a \) in Fig. 4-1).
3.2 Isolation by column chromatography

During elution on the glass column, three color bands were observed on both columns. The first band was yellow (xanthophylls) that was followed by a green band (chlorophylls), and an orange band (carotenoids) followed the green band. The loading for the smaller column was about 3 mg of chlorophyll a per ml of stationary phase and 1.5 mg/ml for the bigger column. Typical loading volumes and mass are listed in Table 4-2. Recovery of chlorophyll a on both columns was more than 92%. Only the green band, which will be further separated by CCC later, was collected. After that the column was washed using 100% ethyl acetate.
Table 4-2. Loading conditions on glass columns

<table>
<thead>
<tr>
<th>Run #</th>
<th>Stationary phase volume (ml)</th>
<th>Mass of chl a in feed sample (mg)</th>
<th>Feed volume (ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>47</td>
<td>134</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>47</td>
<td>144</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>47</td>
<td>147</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>47</td>
<td>137</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>47</td>
<td>135</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>47</td>
<td>133</td>
<td>15</td>
</tr>
<tr>
<td>7</td>
<td>200</td>
<td>314</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>200</td>
<td>302</td>
<td>20</td>
</tr>
</tbody>
</table>

A typical chromatogram from glass column elution is shown in Figure 4-2.

The yellow and green bands were mixed in the first peak with a plateau. The vertical line at approximately 26 min shows the cutting point, at which collection of the fraction ceased and washing began. The orange band was washed out and shown as several mixed peaks on the chromatogram. The green band containing chlorophylls was collected and analysis on a reversed-phase HPLC column. The chromatogram of HPLC analysis is given in Fig. 4-3. The impurities eluted before chlorophyll b were significantly reduced and the impurities eluted after chlorophyll a were almost totally eliminated.
Figure 4-2. Chromatogram of crude extract by glass column chromatography.
(column size 1 cm I.D. × 60 cm, flow rate 4 ml/min)
3.3 Purification by CCC

The fraction containing chlorophylls eluted from glass column was evaporated and re-suspended in CCC mobile phase. This sample was injected into CCC column and chlorophyll a was separated from other compounds. The highest production from single run we achieved was 448 mg chlorophyll a with a purity of 99%. The CCC chromatogram is given in Fig. 4-4 and loading conditions are listed in Table 4-3 (entry 1). Chlorophyll a started coming out from the column outlet at approximately 80 min and was completely eluted in about two hours. The front of chlorophyll a peak is not totally resolved from the previous band. So fractions were collected and analyzed by HPLC to check their purity. The HPLC
Figure 4-4. CCC Chromatogram of preparative purification of chlorophyll \( a \). Feed volume: 60 ml; feed concentration: 8.7 mg/ml of chlorophyll \( a \).
chromatogram of high purity chlorophyll $a$ fraction is shown in Fig. 4-5. The recovery yield of this run is 86%.

![Chromatogram of HPLC analysis of chlorophyll $a$ fraction from CCC elution.](image)

In this run, the mobile phase broke through at 15 minute at flow rate of 3.5 ml/min. There was 175 ml of stationary phase retained in the column at the end of the run. Since 60 ml of stationary phase was displaced during sample injection, the volume of stationary phase initially in the column at the beginning of elution was 285 ml. The loss of 100 ml of stationary phase after mobile phase breakthrough brings the final retention of stationary phase down to 61% from its initial 82%. We did a CCC run with only two phases, i.e., we started pumping the mobile phase through the column filled with the stationary phase without sample injection at a
slightly higher flow rate to rest leaching of stationary phase. The retention of stationary phase was 81% when mobile phase broke through at 18 min at flow rate of 4 ml/min. The equilibrium of two phases was reached soon after with only 3 ml of stationary phase being lost in the next 20 minutes. In another run with very small loading (5 ml of 1 mg/ml of chlorophyll a), the retention of stationary phase at mobile breakthrough was 85% at flow rate 3.5 ml/min. About 82% stationary phase remained in the column at the end of two-hour run. Baseline separation of chlorophyll a from chlorophyll b was obtained as shown in Fig. 4-6. The leaching of stationary phase in preparative CCC separation can be explained by changes due to the increase of sample viscosity and density at high concentrations [16]. When the viscosity and density of the sample solution is very different from those of the liquid phases, both factors affect the mixing between the sample solution and the phases. Also the partitioning of the solutes lowers the density difference of the two phases. Consequently, the equilibrium of the two phases is not as stable as it could be without a very viscous and dense injection plug at the column inlet.

The instability of phase equilibrium caused by large loading of concentrated sample and its physical properties imposes a maximum injection concentration and volume. The feed concentration of the highest production run was 8.7 mg/ml of chlorophyll a and feed volume was 60 ml. We increased the feed concentrations in steps while keeping feed volume constant (entry 5-9 in Table 4-3) and found that 10 mg/ml was a practical upper bound. For instance, one run with 10.3 mg/ml feed concentration (entry 7 in Table 4-3) generated 242 mg of product while another run
Figure 4-6. CCC Chromatogram of separation of chlorophyll $a$ and chlorophyll $b$. Feed concentration: 1 mg/ml of chlorophyll $a$; feed volume: 5 ml; UV wavelength: 440 nm.
with comparable feed concentration (entry 8 in Table 4-3) completely failed. What happened in the failed run was that the sample solution injected into the column inlet formed a liquid piston that pushed the stationary phase ahead of it during the run. No mixing of the stationary phase with the feed and the mobile phase ever occurred, so the feed sample came out following stationary phase as a plug. Complete displacement by injection piston occurred again when the feed concentration was increased to 11.8 mg/ml (entry 9 of Table 4-3). When the feed volume was increased and feed concentration remained in the safe range (entry 10 of Table 4-3), the separation was so poor that the recovery yield was only 17%. The equilibrium of the phases was extremely unstable as 75% of the stationary phase had been lost when mobile phases broke though and the final retention is only 5%.

If the feed concentration was kept between 8 mg/ml to 9 mg/ml and maximum 60 ml feed solution was injected, one CCC operation could usually produce at least 400 mg chlorophyll a with better than 98% purity and recovery yield above 80% as listed in Table 3-4, entry 1-4. Small variation in production or recovery yield may be expected due to variation in feed samples (mass ratio of chlorophyll a to impurities). The run shown as entry 4 has a recovery yield slightly lower than 80% because its feed sample has a smaller product/impurities ratio compared to the other three runs. Its production (392 mg) is not far different from runs with similar loading conditions. Recovery yield falls below 70 when feed
Table 4-3. Summary of individual CCC runs

<table>
<thead>
<tr>
<th>Entry</th>
<th>Chl $a$ con. in feed (mg/ml)</th>
<th>Feed volume (ml)</th>
<th>Chl $a$ purity in feed (%)</th>
<th>Chl $a$ recovered (mg)</th>
<th>Chl $a$ purity in product (%)</th>
<th>Recovery yield (%)</th>
<th>Retention of s.p. (%) $(1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.7</td>
<td>60</td>
<td>51</td>
<td>448</td>
<td>99</td>
<td>86</td>
<td>82</td>
</tr>
<tr>
<td>2</td>
<td>8.1</td>
<td>60</td>
<td>50</td>
<td>414</td>
<td>99</td>
<td>85</td>
<td>78</td>
</tr>
<tr>
<td>3</td>
<td>8.5</td>
<td>61</td>
<td>49</td>
<td>416</td>
<td>97</td>
<td>80</td>
<td>76</td>
</tr>
<tr>
<td>4</td>
<td>8.5</td>
<td>60</td>
<td>47</td>
<td>392</td>
<td>98</td>
<td>77</td>
<td>67</td>
</tr>
<tr>
<td>5</td>
<td>9.0</td>
<td>60</td>
<td>52</td>
<td>378</td>
<td>96</td>
<td>70</td>
<td>66</td>
</tr>
<tr>
<td>6</td>
<td>9.5</td>
<td>60</td>
<td>47</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>10.3</td>
<td>60</td>
<td>47</td>
<td>242</td>
<td>96</td>
<td>39</td>
<td>63</td>
</tr>
<tr>
<td>8</td>
<td>10.1</td>
<td>59.5</td>
<td>48</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>11.8</td>
<td>65.5</td>
<td>46</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>8.2</td>
<td>70</td>
<td>50</td>
<td>101</td>
<td>94</td>
<td>17</td>
<td>25</td>
</tr>
<tr>
<td>11</td>
<td>8.2</td>
<td>58.5</td>
<td>50</td>
<td>388</td>
<td>98</td>
<td>81</td>
<td>82</td>
</tr>
</tbody>
</table>

(1) The first column is the retention of stationary phase when mobile phase breaks through; the second column is the retention of stationary phase at the end of the run.

(2) Retention of stationary phase at 95 min.

* Complete displacement of stationary phase by injection plug.
concentration is above 9 mg/ml (entry 5 and 6 in Table 4-3). Very low production and recovery yield are expected for run listed as entry 6 although no fractions were analyzed because the loss of stationary phase was significant and only 7% left before chlorophyll a was completely eluted. Fig. 4-7 shows the chromatogram of the run with 70% recovery yield (entry 5 in Table 4-3). A comparison of Fig. 4-7 and Fig. 4-4 indicates how much the resolution between the chlorophyll a peak and the band eluted before it has worsened as the retention of stationary phase drops from 66% to only 38% at the end of the run as opposed to from 82% to 60%.

In all the runs mentioned so far, we applied a stationary injection method to introduce the feed into the column. That means the feed is injected into a column filled with stationary phase without rotating the column. Rotation begins after feed injection and the mobile phase is then pumped through while the column spins at a selected speed. Another classical injection method is dynamic injection [34], in which the feed is introduced into an equilibrated CCC column with both stationary and mobile phases through an injection valve with a sample loop. We did a CCC run following the dynamic injection method as follows: first fill up the column with the stationary phase; then pump the mobile phase into the column after bring the column up to speed at 780 rpm; wait for the mobile phase to exit the column outlet and allow the column to equilibrate for a few minutes; inject the feed by switching the injection valve to direct the mobile phase into sample loop; keep pumping the mobile phase following feed injection. The loading conditions (entry 11) are similar to those of run entry 2 as listed in Table 4-3. This run was finished in about
Figure 4-7. CCC Chromatogram of preparative purification of chlorophyll $a$. Feed concentration: 9.0 mg/ml of chlorophyll $a$; feed volume: 60 ml.
2 hours and 50 minutes, which is 50 minutes longer than runs done with stationary injection because it took extra time to equilibrate column before sample injection. The resolution is much worse as shown in Fig. 4-8 compared to that of run entry 2 as shown in Fig. 4-9. These two runs started with comparable amount of stationary phase (82% retention at mobile phase breakthrough for one and 78% for another), but the run that took longer to finish ended up with only 38% stationary phase retained as opposed to 60% in the other run. Thus, dynamic injection is not preferable for preparative CCC operation where loss of stationary phase continues after mobile phase breaks through.
Figure 4-8. CCC Chromatogram of preparative purification of chlorophyll $a$. Feed concentration: 8.2 mg/ml of chlorophyll $a$; feed volume: 58.5 ml; injection method: dynamic injection.
Figure 4-9. CCC Chromatogram of preparative purification of chlorophyll $a$. Feed concentration: 8.1 mg/ml of chlorophyll $a$; feed volume: 60 ml.
Acknowledgements

This research was partially supported by a grant from Marine/Freshwater Biomedical Sciences Center, Oregon State University. We thank Dr. William H. Gerwick for providing the CCC instrument used in this study.
References


Chapter 5

General Conclusion

Computer simulated chromatograms were proven appropriate test bed for the study of the effect of experimental error on the numerically estimated isotherm parameters for the only source of error is in the chromatograms. Individual isotherm parameters resulted from a set of experimental replicates were found to follow a Gaussian distribution. The RSS of average parameters from $n$ replicates was shown to be inversely proportional to the square root of $n$. This relationship indicates that it is theoretically possible to achieve numerically estimated isotherm parameters to any degree of accuracy by doing multiple experiments.

The number of experimental replicates needed to obtain RSS=0.05 calculated for several data sets turned out to be smaller than 10 in most cases provided that no more than 10% relative error in effluent concentrations or 3% relative error in retention times were present. These results indicate that it is promising to obtain isotherm parameters with a satisfactory level of accuracy by numerical method with a few experiments if the experiments are carefully conducted and proper experimental chromatograms are used. However, there are certain cases that require a fairly large number of replicates which are not practical to do. Results obtained in this study provide several suggestions that may make life easier when real experimental data are used. First, always try to reduce error in the retention times because a relatively small variation in time causes relatively big
variations in estimated parameters. Second, get more data points whenever possible if fractions have to be collected. Space out the time points at which fractions are collected in different replicates if the interval is rather big so that more time points are covered. This is to make sure no important curvature of band profiles is missed. Third, use data of different loading simultaneously, but avoid those with too much overlapping between peaks.

As for application of CCC to purification of chlorophyll a, CCC was shown a superior technique for preparative separations. Loading was limited by the stability of the phase equilibrium in the column. Up to 17% of CCC column volume sample at maximum concentration was successfully loaded and separated. Although the stationary phase kept leaching out of the column after the mobile phase broke through when loading was high, significant amount of stationary phase was retained to offer good resolution. The isolation step by column chromatography effectively removed the carotenoids that are more soluble in the CCC stationary phase, therefore enhanced the loading capability for the product. The whole process is fast and reproducible. It is capable of producing 400 mg of chlorophyll a with a purity of better than 98% and a recovery yield of 80%.
Bibliography


Chavent, G., in Proc. 5th IFAC Symp., Darmstadt, Germany, R. Isermann (Ed.), Vol. 1 (1979) 85.


McAlpine, J., in R. J. P. Cannell (Editor), Natural Products Isolation, Humana Press, Totowa, New Jersey, 1998, Ch. 8, p. 247.


Ramanan, S. Laboratory Notebook, unpublished work done in this laboratory, 2000.


Zhang, T.-Y., in Y. Ito and W. D. Conway (Editors), High-Speed Countercurrent Chromatography, John Wiley & Sons, New York, 1996, Ch. 8, p. 225.


Appendix
Appendix

Fitting the Error in Retention Times by a Free Parameter

In the study of the effect of experimental error on the precision of numerically estimated isotherm parameters, the effect of two types of error was investigated: error in effluent concentrations and error in retention times. As the results in chapter 3 have pointed out that the contribution of error in times to the variance of final estimated parameters is more significant than that of the same magnitude of error in concentrations. In the effort of reducing the variation in the final estimated parameters, one possible solution is to fit the error in retention times by a free parameter so that the effect of time error may be separately reflected by a parameter other than the isotherm parameters. Such a parameter can be found for error in retention times but not for error in concentration because of the way each type of error is present in the experimental chromatograms. Error in retention times is usually caused by slight changes in the mobile phase composition or flow rate. Each data point of across the entire chromatogram experiences the same magnitude of influence. In other words, all the components respond to the changes in exactly the same way and no discrimination among data points can be found. That influence is represented in the numerical simulations by a single parameter ($t_0$) – the column dead time, which is the retention time of the non-retained component. While the error in concentrations is caused by fraction collection and off-line analysis, so the amount of error in individual fractions (corresponding to individual
data points in the reconstructed chromatograms) is random. Those error in
concentrations are added point-wise in the computer simulated chromatograms,
therefore can not be represented by one or a few parameters.

Fitting the error in retention times by parameter $t_0$ makes the total number
of parameters to be optimized increase to seven for three component system. The
objective function to be minimized remains similarly defined. The data set with a
feed volume of $0.6 V_0$ was tested on fitting both $t_0$ and isotherm parameters for
chromatograms with both concentration and time error. The results indicate that
the optimization program has difficulty varying $t_0$. Most of the time, the initial
value of $t_0$ remained unchanged while all the isotherm parameters had been varied
to optimize the objective function. But the minimum value of the objective
function had not been reached yet because the best fit of $t_0$ was not found. To make
sure that the global minimum is the objective function value corresponding to the
exact values of all isotherm parameters as well as $t_0$, the response surface of the
objective function was mapped by changing one parameter at a time while keeping
the other parameters constant. It was found that the response surface of the
objective function in the $t_0$ dimension has a feature different from those in the
isotherm parameters. The plot of the objective function as a function of $t_0$ is shown
in Fig. A-1 and one representative plot of the objective function as a function of
isotherm parameter is shown in Fig. A-2. Only a small segment of the curve is
plotted in both figures in order to show the details clearly. It is obvious that the
difference between the two curves is that the objective function is not smooth in the $t_0$ dimension.

![Figure A-1. Objective function as a function of $t_0$.](image1)

![Figure A-2. Objective function as a function of $b_1$.](image2)
Indeed, the optimization program's failure to vary $t_0$ values confirms that the objective function as a function of both $t_0$ and isotherm parameters is nonsmooth mathematically. The optimization program used in this study implements algorithms based on sequential quadratic programming (SQP), which deals with exclusively smooth functions. Because of the improper optimization method for a nonsmooth problem, the question of whether or not the effect of error in retention times can be lumped into one single parameter other than isotherm parameters remains unanswered in the current study.