From Data to Knowledge – Simplifying Modeling and Calibration of Chromatographic Separations

Owais Sulehria

Department of Chemical Engineering, Lund University, Sweden

31/05/12

Abstract

Today the mathematical modeling of chromatographic separations is a niche area limited to researchers who spend a great deal of time with understanding them. The basis for this article is a comparison of a simplified model against the commonly used advanced models. The simplified model is only valid at low solute concentrations or in the linear region of the isotherm, hence the name linear chromatography model. The comparison is made by doing a robustness analysis with both type of models and then charting the responses on the operating point in a chromatographic system. The conclusion from the robustness analysis is that trends can be detected with the simplified model.

Keywords: high-performance liquid chromatography, linear chromatography model, robustness analysis

Introduction

High-performance liquid chromatography has gained footing in the pharmaceutical industry because of its efficiency in purifying e.g. proteins, which also means that a need for optimization of the process has arisen. The optimization can be done by constructing a mathematical model and then finding the optimal settings for a given purity. Aside from saving time, there is another benefit of the computational optimization; leniency is given in choosing the operating points for the process, by the regulatory governments.

The mathematical models have been widely explored and the theory behind them has been proven valid. Although having benefits such as a high range and accuracy, there are a few downsides as well. They require a high step in theoretical knowledge to be fully understood and utilized, they also require much CPU time for a calibration.

There are simpler models that do not require as much knowledge and CPU power to handle. The validity of these linear chromatography models is however limited to a region of low solute concentrations.

The goal of this article is to see how these simpler models measure themselves against the more advanced ones in a robustness analysis.

Theory

Chromatography is a process designed to separate and detect different molecules in a solution. This is done by exploiting the interactions between a fixed substance in the chromatography column and the molecules in the solution. The amount of interactions determine how long a substance stays in the column, this means that the substances that are

to be separated must also have different amount of interactions with the column.

The way a solute interacts with the fixed substance, or the stationary phase, can be changed by the addition of a modifier, or eluent. Substances that bind irreversibly to the stationary phase are displaced by the modifier. The retention time of a substance is changed depending on the concentration of the eluent. An eluent is almost always present in High-performance liquid chromatography. The term, for when the eluent concentration is unchanged during the process, is isocratic elution, see Figure 1. The concentration can also increase, either in a step gradient, or in a linear gradient.

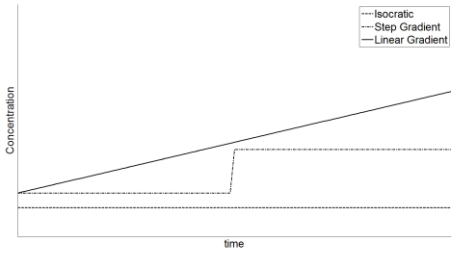

Figure 1 | Different elution methods

For reasons such as performance issues it is important to know the concentration of a substance that has been bound on the stationary phase. The relation between the concentration of a substance in the solution and in the stationary phase is given by the isotherm. The relationship is defined at equilibrium, constant temperature and pressure[1]. One widely used isotherm is the multi-component Langmuir isotherm, which has the following appearance

$$q_i = \frac{q_{max} \cdot b_i \cdot c_i}{1 + \sum_{j=1}^{n} b_j \cdot c_j} \quad (1)$$

The Henry's constant is used sometimes when simplifying the model:

$$H_i = q_{max} \cdot b_i \quad (2)$$

q_{max} is the maximum load capacity on the stationary phase and the Henry's constant can be seen as the slope of the isotherm at low concentrations of the solutes, see Figure 2.

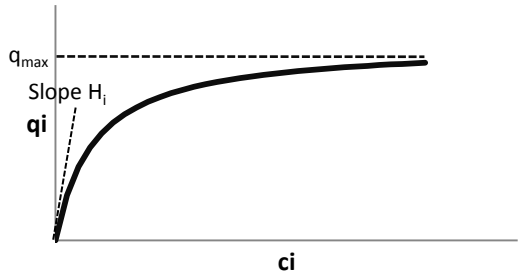

Figure 2 | The Langmuir isotherm with the initial slope H_i and the maximum loading capacity q_{max}

One of the advanced models is based on the following partial differential equation:

$$\frac{\partial c_i}{\partial t} = D_{ax} \cdot \frac{\partial^2 c_i}{\partial z^2} - u_{int} \cdot \frac{\partial c_i}{\partial z} - \frac{1-\varepsilon_t}{\varepsilon_t} \cdot F \quad (3)$$

F is dependent on various parameters, e.g. particle diameter and the isotherm.

The simpler model, also known as the linear chromatography model only calculates the retention volumes of a component. The equation used at isocratic elutions is[2]:

$$VR = V_{NA} + (1 - \varepsilon_c) \cdot \varepsilon_p \cdot K_d \cdot A \quad (4)$$

With A as the slope of the isotherm:

$$A = \frac{dq_i}{dc_i} \quad (5)$$

As previously stated, one of the downsides of this model is that it is only valid at low solute concentrations. Another downside is that the concentration profile of the component cannot be determined with the model, which the more advanced models are able to do.

If gradient elution is used then a modification of the equation has to be made. The retention volume, X, is calculated through:

$$\int_0^X \frac{1}{VR} dCV = 1 \quad (6)$$

Where

$$VR = VR(c_s) \quad (7)$$

$$c_s = c_s(CV) \quad (8)$$

(6) is found by splitting the gradient into infinitesimal parts in the time direction. The elution is then assumed to be performed isocratically with the concentration in each one of these parts. By adding up the fraction of how much each part moves the retention volume (6) is derived. The substance is eluted when it has been moved a whole column volume. The retention volume is thus calculated via (4) and (6).

Methods

A robustness analysis on the chromatographic system is done by subjecting it to different kind of disturbances in the critical process parameters, and then map the response of these disturbances.

A protein mixture consisting of insulin and impurities is run through the chromatographic step to purify the protein. Ethanol is the eluent used in the system. The response on the purity is mapped in this case, the normal operating point is at a purity of 99 % and the parameters changed can be seen in Table 1.The advanced model that is based on equation (3) has previously been optimized by Karin Westerberg et al.[3].

The range of the disturbances are chosen by following a multivariate normal distribution with the MATLAB function *lhsnorm*[4]. 500 different simulation points are generated and a correlation coefficient is determined between the purity and each single varied parameter. The purity is calculated by first generating concentration profiles and then calculating the areas of these from one cut point to another, see equation (9) and Figure 3.

$$Pu_{ins} = \frac{A_{ins}}{A_{ins}+A_{imp}} \quad (9)$$

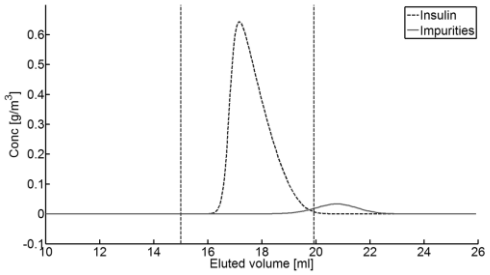

Figure 3 | The concentration profiles of the components at the end of the column. The vertical lines are the cut-off points.

A correlation coefficient is calculated to decide how much the single parameters influence the purity. MATLAB has a built in function *corr* [4] to do this. The correlation coefficient can vary between -1 to 1 where 0 indicates a weak or no correlation between the parameter and the purity.

The linear chromatography model cannot generate concentration profiles, instead the shapes are directly taken from data and then moved sideways considering the new retention volumes that have been calculated through (6).

One of the changed parameters is the protein load, in that case the shape of the concentration profile is changed by multiplying the concentration data with the fraction of the change. For instance, if the new concentration is 2% according to the disturbance, then the concentration vector that contains the data is multiplied by 1.02. This means that the area beneath the concentration profile is increased with a factor of 1.02.

Results & Discussion

The linear chromatography model had an inherent error, it gave a retention volume that was higher than it should be, and to adjust for this the new retention volume was calculated through

$$VR_{new} = VR_{linear} \cdot \frac{VR_{exper}}{VR_{linear,0}} \quad (10)$$

$VR_{linear,0}$ is the faulty retention volume calculated with the linear model, at the normal operating point. VR_{exper} is what it should be. So if this adjustment would not have been made then the retention volume would be faulty, even at the normal operating point. This means that the new retention volume is calculated with a fraction of how much it has been changed from the normal retention volume.

The correlation coefficients from the robustness analysis are written in the table below

The correlation coefficient of ethanol B and the purity is -0.90, this means that there is a strong negative linear correlation between them. It is natural that the concentration in buffer B has the strongest correlation because the components are released from the stationary phase when displaced by a higher ethanol concentration. The higher ethanol concentration depends on the concentration of buffer B.

The proximities of the other correlation coefficients to 0 imply that there is little to no correlation between the variations and the purity. A comparison between the values found by Westerberg [3] shows that there is a difference between the values. The linear model does however predict the same trends as the more advanced model, which is a negative correlation between parameters 2-5. Parameters 1 and 2 show no or little correlation, the reason for this is that the buffer concentration is very low during these stages and can thus not have a significant effect on the retention volumes. The eluent concentration at the step part of the gradient is obtained through a mixture of ethanol buffer A and ethanol buffer B why both of them still do have an effect on the purity.

A reason for why the linear model underestimates the effect of the protein concentration might be because of the overload that influences the shape of the elution profile, a shape change like this is not taken into account with the linear model.

The effect of elution buffer B concentration is a bit exaggerated with the linear model. The correlation is almost linear. The explanation for this is probably that there are additional effects that influence the elution and that the advanced model remediates.

Conclusions

The robustness analysis from case study 1 showed similar results as the advanced model. A compensation for an error found in the linear chromatographic model was made. It was found that the linear model along with this compensation could detect the same trends as the more advanced model. Although some of the effects were underestimated or overestimated the trends could still be seen.

Symbols Used

b	Equilibrium constant of the adsorption/desorption on the stationary phase
c_i	Solute concentration
CV	Column volumes
D_{ax}	Dispersion coefficient
ε_c	Void fraction of the column

Table 1 | Varied parameters with the correlation coefficients between them and the purity

Parameter	Set point	Variation [%]	Correlation coefficient, Westerberg	Correlation coefficient, Linear model
[1] Ethanol load	0.1 w/w	2	+0.01	+0.02
[2] Ethanol buffer A	0.1 w/w	2	-0.19	-0.13
[3] Ethanol buffer B	0.4 w/w	2	-0.51	-0.90
[4] Protein concentration load	6 g/l	10	-0.42	-0.18
[5] Amount impurity load	6.30%	20	-0.27	-0.29

ε_p	Particle porosity
ε_t	Total porosity in a column
F	Adsorption term
H_i	Henry's constant
K_d	Fraction of the pore volume that is available to a certain solute
q_i	Concentration on the stationary phase
$q_{max,i}$	Saturation concentration on the stationary phase
u_{int}	Interstitial velocity
V_{NA}	Retention volume of a non-adsorbed component
VR	Retention volume
z	length

References

[1] Guiochon, G., Felinger, A., Shirazi, D. G., Katti, A. M.: *Fundamentals of Preparative and Nonlinear Chromatography,* 2nd ed. 2006, San Diego, United States of America: Elsevier Inc.

[2] Mollerup, J. M., Budde Hansen, T., Kidal, S., Staby, A.: *Quality by Design – Thermodynamic Modelling of Chromatographic Separation of Proteins.* Journal of Chromatography A, 2007.

[3] Westerberg, K., Borg, N., Andersson, N., Nilsson, B.: *Supporting Design and Control of a Reversed-Phase chromatography Step by Mechanistic Modeling*.Chemical Engineering & Technology, 2012.

[4] The MathWorks Inc., Statistics Toolbox – For Use with MATLAB, MathWorks, Natick, MA 2006.