MODELING AS A PREDICTIVE TOOL FOR PERFORMANCE GAINS IN LIQUID CHROMATOGRAPHY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SIMGE ŞENCAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
CHEMICAL ENGINEERING

JULY 2023

Approval of the thesis:

MODELING AS A PREDICTIVE TOOL FOR PERFORMANCE GAINS IN LIQUID CHROMATOGRAPHY

submitted by **SIMGE ŞENCAN** in partial fulfillment of the requirements for the degree of **Master of Science in Chemical Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Pınar Çalık
Head of Department, **Chemical Engineering** _____

Assoc. Prof. Dr. Harun Koku
Supervisor, **Chemical Engineering, METU** _____

Examining Committee Members:

Prof. Dr. Naime Aslı Sezgi
Chemical Engineering, METU _____

Assoc. Prof. Dr. Harun Koku
Chemical Engineering, METU _____

Prof. Dr. Görkem Külah
Chemical Engineering, METU _____

Assoc. Prof. Dr. Selis Önel Kayran
Chemical Engineering, Hacettepe University _____

Prof. Dr. Göknur Bayram
Chemical Engineering, METU _____

Date:25.07.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Simge Şencan

Signature :

ABSTRACT

MODELING AS A PREDICTIVE TOOL FOR PERFORMANCE GAINS IN LIQUID CHROMATOGRAPHY

Şencan, Simge
M.S., Department of Chemical Engineering
Supervisor: Assoc. Prof. Dr. Harun Koku

July 2023, 88 pages

The essential purpose of this research is to investigate and compare peak characterization and parameter estimation methods applied to experimental data to establish their effectiveness for case studies in analytical chromatography.

Chromatography is a powerful separation method used to separate complex molecules. For separations, the chromatographic efficiency is a crucial parameter affecting resolution. The efficiency can be affected by many experimental variables, and is the consequence of the phenomenon of band broadening, also called dispersion. The theoretical plate number is one of the important measures of band broadening and calculating this parameter correctly is a crucial step in estimating chromatographic efficiency accurately. Robust and accurate results can be obtained for Gaussian peaks with the graphical theoretical plate number calculation methods. However, the accuracy of this calculation method remains poor for peaks that have tailing or fronting since it does not consider the non-Gaussian shape.

In this thesis study, experimental chromatograms and their associated peaks were obtained for a range of settings and variables, and the raw data was used to obtain

chromatographic parameters using MATLAB codes designed to implement moment analysis. These parameters, such as peak variance, theoretical plate number and plate height, provide illuminating information in terms of separation efficiency. The results from different theoretical plate number calculation methods were compared in van Deemter curve for different molecules, columns, and temperatures. Thus, accurate ways of efficiency calculation for both Gaussian and non-Gaussian peaks were investigated. Although accurate and consistent results can be obtained for Gaussian peaks when graphical methods are used, it has been observed that the results for non-Gaussian peaks, that is, peaks with tailing, deviate from reality. In the case of using graphical methods, we cannot see the effect of tailing or fronting in the results because the calculation is made by treating non-ideal peaks as if they are ideal. Also, a database for chromatographic parameters of different molecules in various conditions was created and by using this database an open- source prediction tool that can give results quickly with a few required parameters were developed. It is aimed to obtain a useful open-source tool that can predict the analysis result without performing chromatographic analysis, with further development of the database and MATLAB code as a future work.

Keywords: Chromatographic Efficiency, Peak analysis, van Deemter, non-Gaussian peaks

ÖZ

SIVI KROMATOGRAFİDE PERFORMANS KAZANIMLARI İÇİN ÖNGÖRÜ ARACI OLARAK MODELLEME

Şencan, Simge
Yüksek Lisans, Kimya Mühendisliği Bölümü
Tez Yöneticisi: Doç. Dr. Harun Koku

Temmuz 2023 , 88 sayfa

Bu araştırmanın temel amacı, deneysel verilere uygulanan pik karakterizasyonu ve parametre tahmin yöntemlerini araştırmak ve karşılaştırmak, analitik kromatografide vaka çalışmaları için etkinliklerini belirlemektir.

Kromatografi, karmaşık molekülleri ayırmak için kullanılan güçlü bir ayırma yöntemidir. Ayırmalar için kromatografik verim, çözünürlüğü etkileyen çok önemli bir parametredir. Verimlilik birçok deneysel değişkenden etkilenebilir ve dağılım olarak da adlandırılan bant genişletme olgusunun bir sonucudur. Teorik plaka sayısı, bant genişletmenin önemli ölçülerinden biridir ve bu parametrenin doğru bir şekilde hesaplanması, kromatografik verimin doğru bir şekilde tahmin edilmesinde önemli bir adımdır. Grafik teorik plaka sayısı hesaplama yöntemleri ile Gauss pikleri için sağlam ve doğru sonuçlar elde edilebilir. Bununla birlikte, Gauss olmayan şekli dikkate almadığından, bu hesaplama yönteminin doğruluğu kuyruklu veya cepheli pikler için zayıf kalır.

Bu tez çalışmasında, bir dizi ayar ve değişken için deneysel kromatogramlar ve bun-

larla ilişkili pikler elde edildi ve ham veriler, moment analizini uygulamak için tasarlanmış MATLAB kodları kullanılarak kromatografik parametreleri elde etmek için kullanıldı. Pik varyansı, teorik plaka sayısı ve plaka yüksekliği gibi bu parametreler, ayırma verimliliği açısından aydınlatıcı bilgiler sağlar. Farklı teorik plaka sayısı hesaplama yöntemlerinden elde edilen sonuçlar, farklı moleküller, sütunlar ve sıcaklıklar için van Deemter eğrisinde karşılaştırıldı. Böylece hem Gauss hem de Gauss olmayan pikler için doğru etkinlik hesaplama yöntemleri araştırılmıştır. Grafik yöntemler kullanıldığında Gauss pikleri için doğru ve tutarlı sonuçlar elde edilebilse de Gauss olmayan pikler yani kuyruklu pikler için sonuçların gerçeklikten saptığı gözlemlenmiştir. Grafik yöntemlerin kullanılması durumunda, ideal olmayan pikler idealmiş gibi ele alınarak hesaplama yapıldığından, sonuçlarda kuyruklama veya cephe alma etkisi gözlenmemiştir. Ayrıca, farklı moleküllerin çeşitli koşullardaki kromatografik parametreleri için bir veri tabanı oluşturulmuş ve bu veri tabanı kullanılarak gerekli birkaç parametre ile hızlı bir şekilde sonuç verebilen açık kaynaklı bir tahmin aracı geliştirilmiştir. Gelecekteki bir çalışma olarak veri tabanı ve MATLAB kodunun daha da geliştirilmesi ile kromatografik analiz yapmadan analiz sonucunu tahmin edebilen kullanışlı bir açık kaynak araç elde edilmesi amaçlanmaktadır.

Anahtar Kelimeler: Kromatografik Verim, Pik analizi, van Deemter, Gauss olmayan pikler

To family and friends,

ACKNOWLEDGMENTS

First of all, I would like to express my thanks to my thesis advisor Assoc.Prof. Harun Koku. The support and help he has given me in bringing my work to this stage is very important. His work ethic and knowledge have always impressed me and will continue to be.

I feel very lucky to have my family, who has always supported me in this process, as it has throughout my entire life. I would like to express my sincere thanks to all my friends who have encouraged me throughout this process and to 'Arwen' who has always been my biggest supporter.

This work would not have been possible without the support of my thesis advisor, Harun Koku, and my family and friends. Endless thanks.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

A	Eddy diffusion term
ASTM	American Society for Testing and Materials
F_c	Flow rate (ml/min)
IUPAC	International Union of Pure and Applied Chemistry
B	Longitudinal diffusion term
C	Mass transfer term
C_m	Mass transfer term in mobile phase
C_s	Mass transfer in stationary phase
c	Concentrations of the analyte in the mobile phase
C(t)	Concentration profile of an eluted peak
D	Diffusion Coefficient (m^2/s)
D_L	Axial dispersion coefficient
DAD	Diode array detector
EGH	Exponential- Gaussian Hybrid Function
F	Phase ratio
d_p	Particle diameter (μl)
H	Plate height (mm)
HETP	Height equivalent of a theoretical plate (mm)
h	Reduced plate height
k	Capacity factor
k_i	Rate constant
L	Column length (mm)
N	Number of theoretical plates
M	Molecular weight (g/mol)

$M_{1(x)}$	First statistical moments for peak x (min)
m_n	n^{th} statistical moments for peak
$m_n^{'}$	n^{th} central statistical moments for peak
Pe	Peclet number
q	Concentrations of the analyte in the stationary phase
R_s	Resolution factor
t_0	Hold up time of the mobile phase (min)
t_r	Retention time (min)
TF	Tailing factor
u	Flow velocity (m/s)
USP	United States Pharmacopeia
V_1	Molecular volume of solute (ml/mol)
V_2	Molecular volume of solvent (ml/mol)
V_0	Hold up volume of the mobile phase (ml)
V_r	Retention volume (ml)
$w_{1/2}$	Peak width at one-half peak height (min)
w_t	Peak width at one-tenth peak height (min)
w_v	Peak width in units of volume (ml)
$w_{50(x)}$	Peak width at half height of peak x (min)
$w_{4.4(x)}$	Peak width at 4.4% height of the peak x (min)
$w_{s(x)}$	Peak width derived from statistical moments for peak x (min)
σ	Peak standard deviation (min)
σ^2	Peak variance (min^2)
σ_g^2	Peak variance of the precursor Gaussian (min^2)
τ	Time constant of the precursor truncated exponential function (min)
$\tau_m^{'}$	The average time spent by a molecule in the mobile phase between the desorption event and the adsorption event

$\tau_s^{'}$	The a average time that it remains adsorbed in the stationary phase
γ	Obstructive factor
μ	Viscosity of solvent (cP)

CHAPTER 1

INTRODUCTION

Chromatography is the process of separating and purifying components in a mixture with the help of two phases, one of which is mobile and the other stationary. The basis of the separation lies in the differences in the interactions of the components in the mixture with the stationary and mobile phases. As the mixture passes through the column, how well the components of the mixture can be separated depends on two properties, which are the degree of attachment (affinity) of each compound to the stationary phase and how broad the peaks are so as to avoid overlap.

It can be said that one of the most important parameters measuring the chromatographic efficiency is the band broadening, i.e. dispersion when the chromatographic peak found in the chromatogram obtained as a result of the chromatographic analysis is examined. The dispersion can be directly correlated with the width of the chromatographic peak exiting the column and this width value can be quantified as the theoretical number of plates [10]. The column efficiency is proportional to the (theoretical) plate number [12].

The analyte to be analyzed is injected into the column as a pulse (i.e. a narrow band) and continues to expand with the effect of different parameters as it continues to move in the column. This phenomenon, which is called band broadening, arises mainly from three different effects: flow path inequalities, molecular diffusion, and mass transfer in and close to the vicinity of the stationary phase [25].

The performance of the chromatographic column is determined by examining band broadening parameters which are plate height (H) or plate number (N). The relation between plate height and plate number shown in Equation 1.1 below.

$$\mathbf{H} = \frac{\mathbf{L}}{\mathbf{N}} \tag{1.1}$$

In the literature, there are different calculation methods for theoretical plate number. All of the calculation methods have their own assumptions and limits. These methods will be examined in detail in Chapter 2 and Chapter 3.

The equation used in this thesis and the most famous one in the literature for modeling band broadening is the van Deemter equation. Van Deemter explained the band broadening with following Equation 1.2 and all of the terms have different contributions to band broadening of the chromatographic peak [26].

$$\mathbf{H} = \mathbf{A} + \frac{\mathbf{B}}{\mathbf{u}} + \mathbf{Cu} \tag{1.2}$$

Where the u term represents the flow velocity in m/s.

The value of A, which is independent of the flow rate, results from the substance molecules following different paths as they move through the column. The velocity of the substance molecules in the column that follow paths of different lengths will also be different. As all molecules reach the end of the column at different times, the band broadens. This phenomenon is called “Eddy Diffusion” [19]. The second term in the formula is the longitudinal diffusion term, which is particularly important at low flow rates. The contribution of this term increases in direct proportion to the values of the diffusion coefficients of the components in the mobile phases. The third term of the formula is the mass transfer term which means that the resistance between the stationary and mobile phase, becomes important at high flow rates. The sum of these three parameters was illustrated as van Deemter curve as shown in Figure 1.1 below [19].

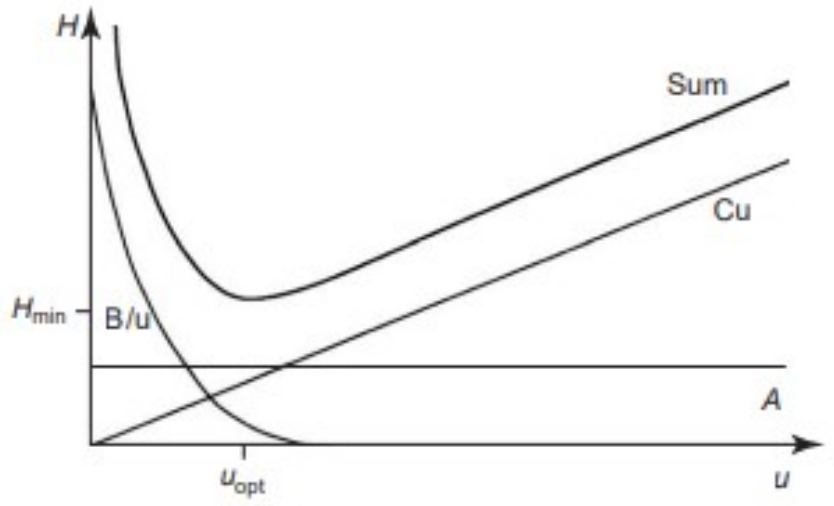

Figure 1.1: The van Deemter curve [19]

This thesis work is mainly focused on examining different calculation methods of theoretical plate numbers and their effect on the resulting van Deemter curve. In this way, it is aimed to find the most appropriate chromatographic efficiency calculation method in different scenarios. Commercial software generally uses graphical calculation methods that assume the chromatographic peak as a Gaussian peak. However, in real cases, mostly the chromatographic peak has tailing or fronting that should be considered. Obtaining the chromatographic parameters of the non-Gaussian peaks observed as a result of analyses in real situations is the main motivation of this study.

The remaining parts of the thesis are formed as follows;

In Chapter 2, a literature survey for providing more information in detail about the related topics about the thesis is presented. The methods section in Chapter 3 explains the experimental and computational approaches followed throughout the thesis work. This chapter includes the used theories and equations from the literature, the experimental methods and materials for chromatographic analyses, and the working logic of the created MATLAB codes.

Chapter 4 is Results and Discussion section, in which the comparison validation of the MATLAB codes, the comparison between van Deemter curves and plate numbers obtained for different case studies, and different estimation approaches are given and

discussed. Conclusion and future work ideas related to the thesis work are presented in Chapter 5.

CHAPTER 2

LITERATURE REVIEW

Chromatography is basically a process to separate, components in mixtures from each other. It is based on the distribution of components between a stationary phase and one or more bulk phases by means of countercurrent contact [3].

2.1 The Theory of Chromatography

2.1.1 The Chromatographic System

A chromatographic system comprises two different phases that are insoluble in each other, the stationary phase and the mobile phase flowing over a stationary phase. The mobile phase can be a liquid, or a gas and the stationary phase may consist of a liquid supported by a porous inert material, or molecules chemically attached to such a material or column wall. The stationary phase can be an inert or adsorptive solid which is usually porous, also it can be gel or an ion-exchange resin. This type of phase can form a planar layer or paper, as well as stuffed in a column, which is a tube with a flow of mobile phase [14].

Substances to be separated using chromatography must have different affinities because thus the substance with the higher affinity for the stationary phase moves at a much slower rate than the substance with the lower affinity. The difference between the migration rates of the components that make up the mixture provides the physical separation of the sample into its components [14].

There are many different combinations of phase and technical realizations for chromatographic methods. As a result of this situation, many techniques have emerged

that include individual naming and abbreviations. Some of these nomenclatures and abbreviations, approved by IUPAC and ASTM, are listed and briefly explained below in Table 2.1 [14].

Table 2.1: Names and Abbreviations for Chromatographic Methods Approved by IUPAC and ASTM (Jönsson & Poole, 1989)

Names describing phases	Abbreviation	Description
Gas chromatography	GC	Mobile phase is gas
Gas- liquid chromatography	GLC	GC with liquid stationary phase
Gas- solid chromatography	GSC	GC with solid stationary phase
Liquid chromatography	LC	Mobile phase is liquid
Reversed phase chromatography	RP	LC where the mobile phase is more polar than the stationary phase

2.2 Fundamental Parameters of Chromatography

2.2.1 Retention Parameters

In column elution chromatography, chromatograms are the outputs of chromatographic analyses. This output is a graph of the response from the detector device. These responses are directly proportional to the flow of material passing through the detector and are generally a function of time. The interval between the time the sample enters the column and the peak time formed by a component of the sample is termed the retention time and is denoted by t_R, also the corresponding gas or liquid volume is termed the retention volume and is denoted by V_R. If the flow rate, F_c, is constant, the relationship between the retention time and the retention volume can be represented by the following equation [14].

$$\mathbf{V}_R = \mathbf{t}_R + \mathbf{F}_c \tag{2.1}$$

In general, the maximum of the peak is used to define the retention time, but from a thermodynamic point of view, the mean of the chromatographic peak is considered a more accurate approximation to calculate it more precisely [14].

If the analyzed sample is not held by the stationary phase but moves entirely in the

solvent (unretained component), its retention time is called the holdup time of the mobile phase and is denoted by t_0, and its corresponding volume is the volume of the mobile phase in the column and is denoted as V_0. The adjusted retention time and volume can be derived by subtracting the holdup time and volume from the retention time and volume, respectively [14]. The holdup time can also be expressed in different terms such as dead time or void time in the literature. During this thesis, the term holdup time will be used for the exit of the molecules that cannot hold on the stationary phase side from the column together with the mobile phase.

$$\mathbf{t}^{'}_{R} = \mathbf{t}_R - \mathbf{t}_0 \tag{2.2}$$

$$\mathbf{V}^{'}_{R} = \mathbf{V}_R + \mathbf{V}_0 \tag{2.3}$$

The term capacity factor can be stated as follows,

$$\mathbf{k} = \frac{\mathbf{t}^{'}_{\mathbf{R}}}{\mathbf{t_0}} = \frac{\mathbf{V}^{'}_{\mathbf{R}}}{\mathbf{V_0}} \tag{2.4}$$

Thus, by using the capacity factor definition, the retention time and retention volume can be expressed as below in Equations 2.5 and 2.6 [14]:

$$\mathbf{t}_R = \mathbf{t}_0(\mathbf{1} + \mathbf{k}) \tag{2.5}$$

$$\mathbf{V}_R = \mathbf{V}_0(\mathbf{1} + \mathbf{k}) \tag{2.6}$$

2.2.2 Dispersion Parameters

Besides the retention parameters that depend on peak positions (i.e., retention time or retention volume), a chromatographic peak may also be expressed by its width. The second statistical central moment is usually used to express the width, the dispersion parameter used to characterize the peak. The "moment" term which characterizes statistical distributions will be examined in detail in the following parts of this chapter and Chapter 3. A dimensionless form of dispersion parameter is the number of theoretical plates, N, and is defined as [14]:

$$\mathbf{N} = \frac{\mathbf{t_R^2}}{\sigma^{\mathbf{2}}} \tag{2.7}$$

The term height equivalent of a theoretical plate, HETP, or simply plate height, H can be used to compare the efficiencies between columns that have different column lengths. This term can be obtained by dividing the column length, L to the theoretical plate number, N as shown in Equation 1.1 [2].

The number of theoretical plates has different representations in the literature and these representations can be summarized below in Table 2.2 [2].

Table 2.2: Different empirical representations of calculating number of theoretical plates from chromatographic data partially adapted from [2]

Relationship	Applicability
N = 16(t_R^2/w_t^2)	Commonly used plate equation with units of time
N = 16(V_R^2/w_v^2)	Used when retention volume is required
N =5.54($t_R^2/w_{1/2}^2$)	Used for asymmetric, skewed, or partially resolved peaks

In these relationships for the number of theoretical plates, t_R denotes retention time and V_R denotes corresponding retention volume. Also, while w_t represents peak width in units of time, w_v represents peak width in units of volume. $w_{1/2}$ means peak width at one-half peak height in units of time.

2.2.3 Resolution

Compounds of a mixture can be separated into individual peaks by using chromatographic systems. The efficiency of the separation can be measured using the resolution factor (R_s) between consecutive peaks. Generally, the resolution factor is used as a system suitability parameter for chromatographic systems. If the resolution factor value is over 1.5, the correct integration of the peaks and the quantities of the peaks obtained in the system is obtained. One of the most used calculation methods for this parameter is the United States Pharmacopeia (USP) accepted tangent method shown in Equation 2.8 below [23]:

$$\mathbf{R}_s = \frac{(\mathbf{t_2} - \mathbf{t_1})}{(\frac{\mathbf{W_2}}{\mathbf{2}} + \frac{\mathbf{W_1}}{\mathbf{2}})} \tag{2.8}$$

where t_2 and t_1 are the retention times of the two components, and W_2 and W_1 are the corresponding widths at the bases of the peaks obtained by extrapolating the relatively straight sides of the peaks to the baseline as shown in Figure 2.1 [23].

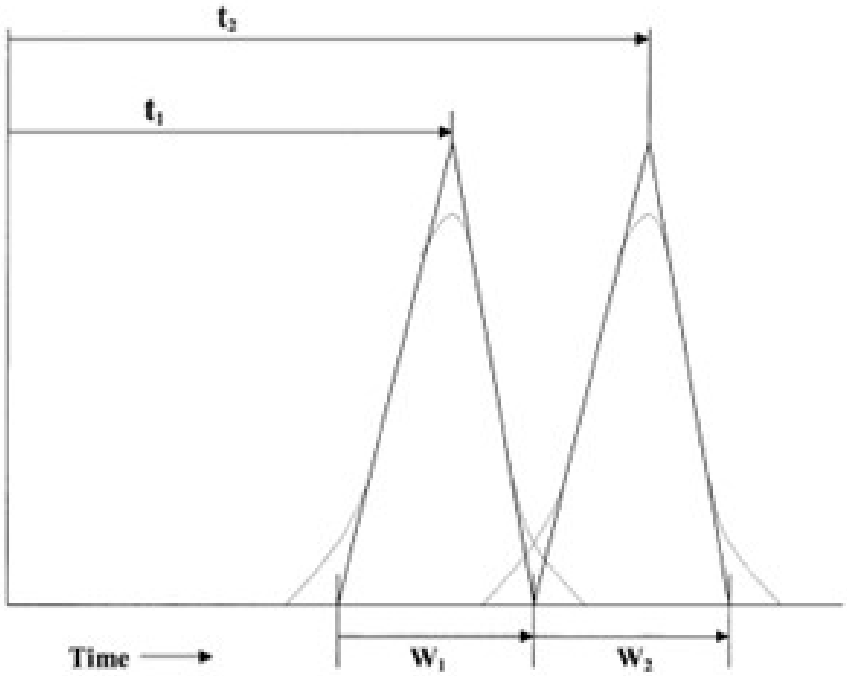

Figure 2.1: Illustration for USP resolution factor calculation method [23]

In the USP method, the most important assumption is that peak shapes are symmetrical, in other words, Gaussian peak shape. However, in practice, most of the chromatographic peaks are not perfectly symmetrical. In order to quantify the degree of peak asymmetry, the term tailing factor was proposed. The tailing factor is defined as the following by USP:

$$\mathbf{TF} = \frac{(\mathbf{A} + \mathbf{B})}{\mathbf{2A}} \tag{2.9}$$

where A and B measured at 5% peak height as shown in Figure 2.2 [23].

For a perfectly symmetrical peak, the retention time is expressed as the midpoint of the width of the peak on the baseline and the separation of two peaks in a chromatographic system is described as the difference between these midpoints. The peak midpoints on the baseline and the peak retention times are identical for symmetrical

peaks, so the difference in peak retention times can be used instead of the difference between midpoints on the baseline. However, as mentioned earlier, in practice peaks are not perfectly symmetrical. Accordingly, the representations of separation of two peaks cannot be the difference between the retention times and instead, the retention time of each peak should be adjusted as the peak's midpoint on the baseline [23].

The USP tailing factor can be revised as following the equation by accepting the peak shape as an asymmetrical triangle and neglecting the peak dispersion, as shown in Figure 2.2.

$$\mathbf{TF} = \frac{(\mathbf{A}' + \mathbf{B}')}{\mathbf{2A}'} \tag{2.10}$$

where A and B are measured at 5% of peak height, A' and B' are measured at baseline and W is the peak width at baseline [23].

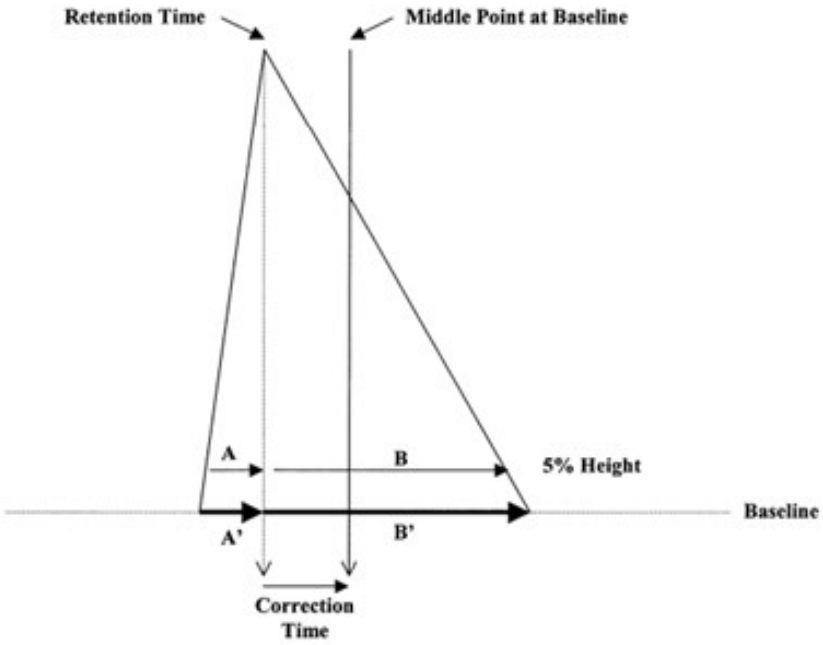

Figure 2.2: Illustration for USP tailing factor calculation method and retention time correction based on USP tailing factor [23]

The revised formula, by using the tailing factor modified retention time which is the peak's midpoint on the baseline, for resolution factor is shown in the formula below [23]:

$$\mathbf{R}_s = \frac{\mathbf{2}(\mathbf{t_2} + \frac{\mathbf{W_2}}{(\mathbf{2}(\mathbf{1}-(\mathbf{1}/\mathbf{TF_2})} - \mathbf{t_1} - \frac{\mathbf{W_1}}{(\mathbf{2}(\mathbf{1}-(\mathbf{1}/\mathbf{TF_1})})}{\mathbf{W_2} + \mathbf{W_1}} \tag{2.11}$$

In the literature, there are different available methods for calculating resolution. Some of these methods were summarized in Table 2.3 [23]:

Table 2.3: Different methods for calculation resolution from chromatographic parameters [23]

Name of the method	Equation
Half-Width	$\mathbf{R}_s = \frac{\frac{\mathbf{2.35}}{\mathbf{2}}(\mathbf{t_2}-\mathbf{t_1})}{\mathbf{W_{50(2)}}+\mathbf{W_{50(1)}}}$ (2.12)
Five Sigma	$\mathbf{R}_s = \frac{\mathbf{2.5}(\mathbf{t_2}-\mathbf{t_1})}{\mathbf{W_{4.4(2)}}+\mathbf{W_{4.4(1)}}}$ (2.13)
Statistical	$\mathbf{R}_s = \frac{(\mathbf{M_{1(2)}}-\mathbf{M_{1(1)}})}{\mathbf{W_{s(2)}}+\mathbf{W_{s(1)}}}$ (2.14)
Tangent (USP)	$\mathbf{R}_s = \frac{(\mathbf{t_2}-\mathbf{t_1})}{\frac{\mathbf{W_2}}{\mathbf{2}}+\frac{\mathbf{W_1}}{\mathbf{2}}}$ (2.15)
Tangent (Modified)	$\mathbf{R}_s = \frac{\mathbf{2}(\mathbf{t_2}+\frac{\mathbf{W_2}}{(\mathbf{2}(\mathbf{1}-(\mathbf{1}/\mathbf{TF_2})}-\mathbf{t_1}-\frac{\mathbf{W_1}}{(\mathbf{2}(\mathbf{1}-(\mathbf{1}/\mathbf{TF_1})})}{\mathbf{W_2}+\mathbf{W_1}}$ (2.16)

In these methods, $M_{1(x)}$ denotes the first statistical moments for peak x. Also, $W_{50(x)}$ denotes width at half height of peak x in units of time, $W_{4.4(x)}$ denotes width at 4.4% height of the peak x in units of time and $W_{s(x)}$ means width derived from statistical moments for peak x in units of time [23].

2.3 Methods for Examining Dispersion and Peak Shapes in Chromatography

The width of a sample zone expands as it passes through a column or layer due to several different mechanisms called dispersion processes. These dispersion processes can be listed as the diffusion of the solute along the column, the resistance to mass transfer between and/or within the phases, and the effect of various flow distributions. On the other hand, the efficiency of a chromatographic analysis is highly dependent on the shape as well as the width of the peak. There are many different methods in the literature to analyze peak shapes and dispersion. In this part of the literature review, some of the existing theories to examine the peak shape and dispersion in the chromatographic analysis will be summarized. These theories are based on the "rate theory" in their approach. Rate theory is a theory that uses partial differential equations to explain the phenomena that take place inside the chromatographic column [10]. The chromatographic process can also be formulated by using a stochastic method. This method uses discontinuous random walk models and statistical techniques based

on the treatise of Giddings [10]. The stochastic methods will be examined in detail in the following parts of this chapter.

2.3.1 The Statistical Moments for Analyzing Peak Shape

The most accurate way to analyze the shape of a chromatographic peak would be to derive a functional form that expresses the sample concentration detected by the detector as a function of time or volume, but unfortunately, many peak shape functions found in the literature cannot be obtained directly. As an example, for this situation, the equations for explaining the chromatographic peaks are obtained in the Laplace domain and its transformation to explicit form can be difficult. In this case, as an alternative, statistical moment analysis for the sample distribution can be used to examine the peak. Moments describe the shape of the distribution and allow chromatographic parameters such as peak retention time, width, and skewness to be calculated [14].

Although, usually, all the chromatographic peaks are assumed to have a Gaussian peak shape, this ideal assumption cannot be valid in most of the cases in the results of real cases. The chromatographic peaks obtained from the analysis can show asymmetric shapes such as fronting and tailing depending on the type of chromatographic columns used. For this reason, the assumption of the Gaussian peak can lead to deviation in the results of the chromatographic peak parameters. For instance, if the asymmetry of the peak is assumed as 1, which is the Gaussian peak assumption, for a tailed peak, the result of the chromatographic efficiency calculation will not give an accurate result [20].

A chromatographic peak can be characterized as a probability distribution function of the molecule over time. [14]. For a distribution function which is the concentration profile of an eluted peak in chromatography, C(t), the n^{th} statistical moment can be expressed as [12];

$$\mathbf{m}_n = \frac{\int \mathbf{t^n C(t)\,dt}}{\int \mathbf{C(t)\,dt}} \tag{2.17}$$

The normalized area of the chromatographic peak is simply the zeroth moment where

n is equal to zero in Equation 2.17. Also, the first moment of a chromatographic peak means its center of gravity, in other words, the mean of the concentration profile. The first moment represents the peak retention time in chromatography. If the considered peak is symmetrical, its maximum and retention time are the same.

Central moments are used to express higher moments with respect to the first moment as shown in Equation 2.18, in this way they are normalized for convenience in comparison and have greater physical significance [12].

$$\mathbf{m}'_n = \frac{\int (\mathbf{t} - \mathbf{m_1})^{\mathbf{n}} \mathbf{C(t)}\, \mathbf{dt}}{\int \mathbf{C(t)}\, \mathbf{dt}} \tag{2.18}$$

The second central moment express the variance of the peak and it is a parameter related to plate height or efficiency of the chromatographic separation. The third central moment is a parameter that measures the asymmetry of the peak and its direction. If third central moment, i.e., skewness is negative it means there is fronting in the peak and conversely, if the value is positive, it means that the peak has tailing [12].

2.3.2 Mathematical Models for Peak Shape Analysis

A peak profile is obtained as a result of the various interactions of the molecules in the column in the chromatographic analysis. Quantitatively accurate examination of the chromatographic peaks is very important because the correct processing of the signal received from the detector and obtaining the chromatographic parameters depends on it. In the literature, there is still no theoretical model to describe exactly the chromatographic peak's shape although there are several empirical mathematical functions to represent the shape of the chromatographic peak [18].

Romanenko *et al.* (2000) states that many idealized models are based on elementary functions, such as the logistic function, the Gaussian or Lorentzian (Cauchy) function, or the derivative of the logistic curve. In addition, some modifications have been made to these models to consider the effect of non-ideal and real processes on the peak shape when examining a chromatographic peak. Many different mathematical models have been defined, depending on the peak to be examined and the results

sought. Some models are simple to study an ideal Gaussian peak, while others are more complex to study non-ideal cases such as peak tailing [22].

Three main peak definitions were considered in the overall published data, Gaussian, logistic curve derivative, and Cauchy. Modifications or combinations of these three peak definitions are used to examine most of the other non-ideal conditions for chromatographic peaks [22]. These three functions are summarized in Table 2.4 below. Also, in this study, an Exponential-Gaussian Hybrid Model [17] was used to examine the peak shape and this mathematical model will be studied in detail in Chapter 3.

Table 2.4: Summation of main three function to describe the peak shape

Gaussian function	$\mathbf{y = y_m e^{(-(k(x-x_m)^2))}}$ (2.19)
The derivative of a logistic function	$\mathbf{y = \frac{4y_m}{(e^{(k(x-x_m)}+e^{(-(k(x-x_m)))^2}}}$ (2.20)
Cauchy function	$\mathbf{y = \frac{y_m}{1+(k(x-x_m))^2}}$ (2.21)

2.3.3 Measuring Dispersion

An analyte injected into a column as a finite band continues to expand as it moves along the column. This situation is called dispersion or, in other words, band broadening. Band broadening is generally explained by three independent influences: flow path inequalities, resistance to molecular diffusion, and mass transfer. Many different equations have been derived to model this process mathematically. Band broadening is directly related to column efficiency, plate height, and plate count. The two most popular equations used to model the chromatographic band broadening are the van Deemter and Knox equations [25]. Also, there are more advanced models in the literature, however, basically all of them use the same contributions which are in van Deemter equation.

In 1956, van Deemter defined band broadening in a chromatographic system as a sum of three terms, A, B, and C [26]. In the van Deemter equation, the A term represents the eddy diffusion, the B term represents the molecular diffusion, and the C term represents all sources of resistance to mass transfer [25]. The relation of these terms is shown in Equation 1.2.

In addition, the reduced plate height can be expressed as Equation 2.22 below. Plate height is divided by the diameter of the particles in the column to obtain reduced i.e., dimensionless plate height.

$$\mathbf{h} = \mathbf{A}^{'} + \frac{\mathbf{B}^{'}}{\mathbf{Pe}} + \mathbf{C}^{'}\mathbf{Pe} \tag{2.22}$$

The A$^{'}$, B$^{'}$ and C$^{'}$ terms represent the same contribution to the band broadening as in the plate height equation. Peclet (Pe) number is a dimensionless parameter that is a ratio of advective mass transfer to diffusive mass transfer in the system and it can be stated as Equation 2.23.

$$\mathbf{Pe} = \frac{\mathbf{d_p u}}{\mathbf{D}} \tag{2.23}$$

Where d_p represents the diameter of the column particle, u represents the mobile phase flow velocity and D represents the diffusion coefficient of the analyzed molecule. Reduced parameters facilitate the scale up of the obtained results to real conditions and generalizes the particular problem.

One of the most important ways to increase the efficiency of chromatographic separation is to reduce the plate height. According to Equations 1.2 and 2.22, there are three different contributions to plate height. The first one is the eddy diffusion term, A, which is the effect of the numerous paths that solute can follow in the chromatographic column. All the molecules of the mixture do not follow the same path in the column, and these paths are not straight, so the residence time of the molecules in the column is variable. Initially, the solute is injected as a rectangular pulse. The molecules in the solute can follow different paths in the column as shown in Figure 2.3 and as a result in the output of the column, the band profile is broader [19].

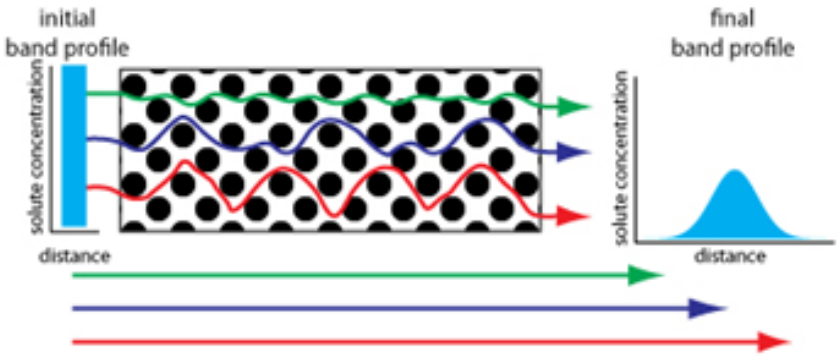

Figure 2.3: Illustration of A term in van Deemter equation [13]

The other contribution to plate height is the B term which represents molecular diffusion. While the molecules in the chromatographic column are passively transported by the effect of the mobile phase, they also move due to diffusion among themselves. Because of the concentration gradient on the outer edge of the rectangular band, it shows a tendency to disperse. Although the band broadening process happens for all piping systems, the main problem is longitudinal diffusion within the column [19].

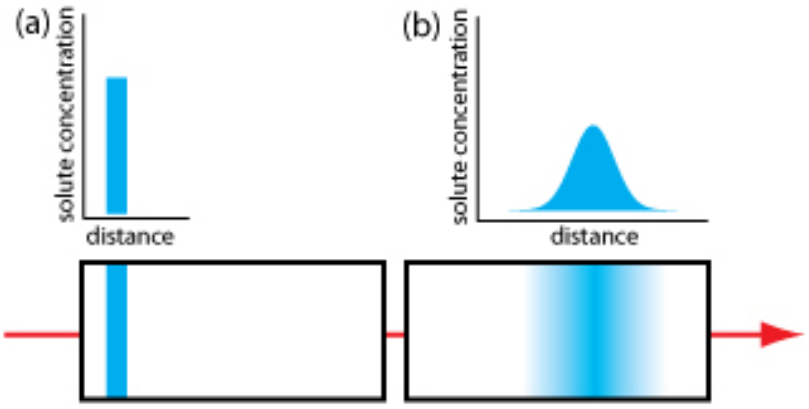

Figure 2.4: Illustration of B term in van Deemter equation [13]

The third effect is called as mass transfer effect. This term represents the resistance to mass transfer between the stationary and mobile phases. While some of the solute molecules enter and travel within the stationary phase, some of them spend little to no time in it. Molecules that cannot perform mass transfer in the same time period are located in different places along the chromatographic bed and cause the band

broadening [19].

After identifying the terms that contribute to the band broadening, van Deemter et al. (1956) obtained a curve that correlates the plate height with the velocity of the mobile phase flowing through the column. As shown in Figure 2.5, the curve can be used to interpret the terms that contribute to the band broadening, for example using this obtained curve it is possible to determine the optimum flow rate [19].

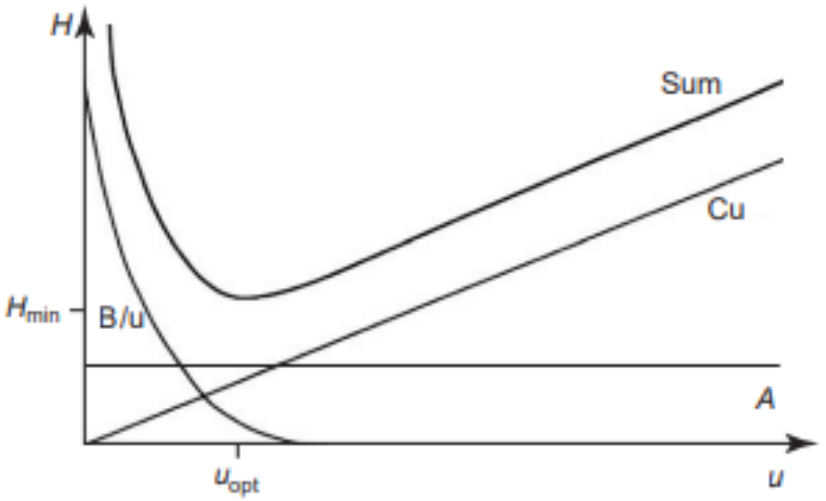

Figure 2.5: The van Deemter curve [19]

2.4 Macroscopic and Microscopic Models in Chromatography

The mass transfer between the stationary phase and the mobile phase is accepted as the bottom line for chromatographic processes. It is necessary to study the kinetics of the mass transfer that takes place within the chromatographic column, to accurately examine and describe the chromatographic separation process. It has very important effects on the shape, width, and asymmetry of the band profile. In the literature, there are various macroscopic and microscopic kinetic models to model the phenomena that occur during the chromatographic process [1].

Chromatographic separation processes are generally expressed using macroscopic (continuum) models. In these kinds of models, the separation process is expressed with differential mass balance equations which define the physico-chemical processes

in the chromatographic column. Generally, for mass balance models, an instantaneous equilibrium is assumed between the stationary phase and mobile phase, or a kinetic rate constant is used to indicate resistance to mass transfer. On the other hand, stochastic i.e., microscopic (particle-based) models describe the chromatographic process with random migration of the molecules along the column at a molecular level [6].

Stochastic approaches are a successful alternative to kinetic modeling. The application of molecular dynamics models in chromatography has been restricted because the mathematical analysis of this approach is quite complex. While the analytical or numerical solutions of the macroscopic models can be calculated easily, the solution and formulation of the microscopic –stochastic– models are complicated [5].

2.4.1 Macroscopic Kinetic Theories

Macroscopic theories have been developed based on the phenomenon of migration of concentration zones along the chromatographic column. A mass balance formula was obtained by considering the convection and dispersion processes

$$\frac{\partial \mathbf{c}}{\partial \mathbf{t}} + \mathbf{F}\frac{\partial \mathbf{q}}{\partial \mathbf{t}} + \mathbf{u}\frac{\partial \mathbf{c}}{\partial \mathbf{z}} = \mathbf{D_L}\frac{\partial \mathbf{c^2}}{\partial \mathbf{z^2}} \tag{2.24}$$

In Equation 2.24, c and q represent the concentrations of the analyte in the mobile and stationary phase, F represents the phase ratio, D_L represents the axial dispersion coefficient (it considers both eddy dispersion and longitudinal diffusion). t and z show the temporal and spatial coordinates, and u means mobile phase velocity [5].

Some assumptions have been made to use the above mass balance equations. These assumptions can be listed as the column being radially homogeneous, neglecting the compressibility of the mobile phase, accepting the temperature as constant, and finally assuming that there is no diffusion in the stationary phase. The radial diffusion effect within the column is neglected, assuming the column is radially homogeneous [5].

Mass transfer and dispersive mechanisms in a packed column are illustrated in Figure 2.6. The numbered parts shown in the figure can be summarized as follows [5]:

1. Pore diffusion
2. Solid diffusion
3. Reaction kinetics at phase boundary
4. External mass transfer
5. Fluid mixing

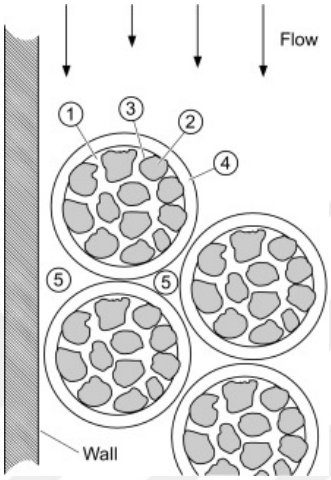

Figure 2.6: Illustration of mass transfer and dispersive mechanisms in a packed bed column [5]

2.4.1.1 Lumped Kinetic Model

In the lumped kinetic model, simple first order kinetic equation was used to explain the nonequilibrium parameters generating band broadening. The generic form of this kinetic model is shown in Equation 2.25 below. In this equation, k_I and k_{II} terms are rate constants [5].

$$\frac{\partial \mathbf{q}}{\partial \mathbf{t}} = \mathbf{k_I c} - \mathbf{k_{II} q} \tag{2.25}$$

Depending on what the rate-limiting step of the chromatographic process is, this kinetic model can be used in many distinct forms. For instance, reaction- dispersive model associates the reason for the non-equilibrium is due to slow adsorption-

desorption, on the other hand, the transport-dispersive model accepts that the mass transfer process is slow while the adsorption-desorption process is fast enough [5].

2.4.1.2 General Rate Model

The chromatographic process is a combination of complex phenomena of kinetic and thermodynamic origin, and these phenomena are in constant interaction with each other. Although significant simplifications have been made in the models examined so far by focusing on the phenomenon believed to be the most important and making some omissions, when band broadening is considered, many different processes occur simultaneously as the molecule moves through the column and it becomes almost impossible to treat a single phenomenon as a rate-controlling factor. The general rate model is the physically most detailed model of chromatography since it tries to simultaneously consider all the phenomena that contribute to the mass transfer kinetics [8]).

According to Guiochon G. et al. (2006), the general rate model consists of certain transport and adsorption-desorption mechanisms on different scales of the chromatographic column as shown in Figure 2.7.

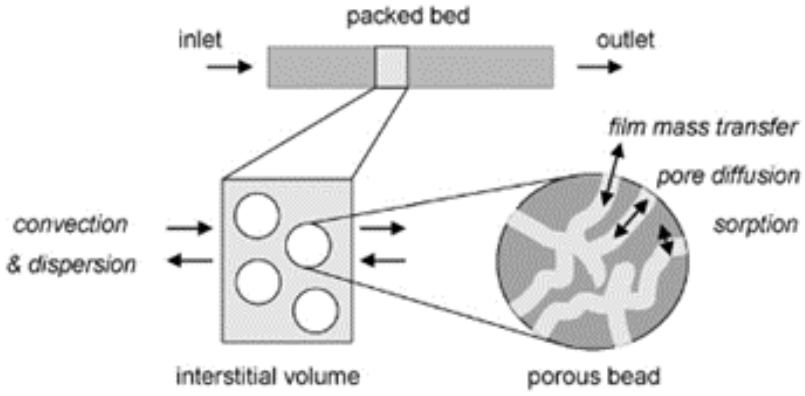

Figure 2.7: Illustration of the general rate model of column chromatography [27]

In accordance with this model, the mobile phase percolates through the interstitial volume between the stationary phase particles. The sorption processes appear within the pores on the stationary phase. This model studies the behavior of solute molecules

in the stagnant mobile phase in the pores and in the mobile phase flowing through the particles by using several mass equations [1].

2.4.2 Microscopic Kinetic Theories

Giddings and Eyring first revealed the microscopic i.e., stochastic model in the 1950s [11]. According to this theory, the random migration of a molecule along the chromatographic column is based. Two processes are formed in the stochastic model: the sorption processes of the molecules in the stationary phase and the random migration of the molecules in the mobile phase. While the diffusion of the molecules in the mobile phase can be modeled with a simple 1D random walk, the adsorption-desorption process was modeled by assuming the exponential distribution of the stopover time of the molecules in both phases, and this process was called the Poisson process [1].

2.4.2.1 Stochastic Theory

In this model, the random walk of a molecule through the chromatographic column is examined. The chromatogram is obtained as the probability density function of the molecules' residence time in the column [7].

The average time spent by a molecule in the mobile phase between the desorption event and the adsorption event that will take place during its random walk in the column is called $\tau_m{}'$, and the average time that it remains adsorbed in the stationary phase is called $\tau_s{}'$. The random migration of the molecules through the chromatographic column is illustrated in Figure 2.8. [5].

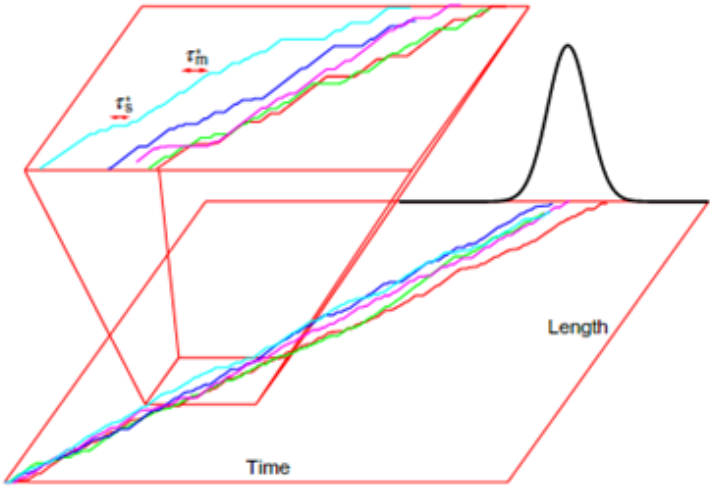

Figure 2.8: Illustration of random migration of the molecule through the chromatographic column [5]

As time passes, an adsorbed molecule remains fixed in the same position in the column, but a desorbed molecule moves with the mobile phase until it is adsorbed again. Since the residence times of each molecule individually in the stationary phase and the mobile phase are random values, each molecule follows a different path during the chromatographic separation process and as a result, they are separated at different times at the column exit. The chromatographic elution profile is a probability density function of the retention times of each molecule in the column [5].

According to Giddings (1965), the advantage of the random walk model lies in the simplicity of its laws. The standard deviation of the resulting concentration profile for many molecules that start the random walk of many steps together in a column is stated below formula:

$$\sigma = \mathrm{l}\sqrt{\mathrm{n}} \tag{2.26}$$

where n shows the number of steps taken and l shows the fixed step length [10].

In statistic, the final band broadening can be determined by using σ, if there are several simultaneous and independent random processes and these processes have their own standard deviation value [9], σ^2 as in Equation 2.27.

$$\sigma^2 = \sum \sigma_{\mathbf{i}}^{\mathbf{2}} \tag{2.27}$$

The above equation can be used to calculate the total effect of the kinetic, flow, and diffusional influences on the band broadening [9].

Diffusion can be explained as a result of the random movement of molecules first in one direction and then in other directions, and this phenomenon can be associated with the random walk model. This relation can be stated as below well-known formula:

$$\sigma = \sqrt{\mathbf{2Dt}} \tag{2.28}$$

where D represents the diffusion coefficient and t represents the time that the process takes place [9].

According to these equations, chromatography can be defined as a random process using the standard deviation value or as a diffusion process using the diffusion coefficient value. Recent simulations employ microscopic particle-based tracking methods based on the application of this principle [4][9] [15] .

2.4.2.2 Giddings Plate Height Equation

A plate height equation was formed, by using the theory of random walk of the molecules along the chromatographic column, by Giddings (1965). There are three relative processes that affect band broadening: longitudinal diffusion, adsorption-desorption, and flow and diffusion in the mobile phase [5].

The contribution of molecular diffusion to plate height can be represented in the following equation:

$$\mathbf{H_D} = \frac{\mathbf{2\gamma D}}{\mathbf{u}} = \frac{\mathbf{B}}{\mathbf{u}} \tag{2.29}$$

where D represents the diffusion coefficient, u represents the flow velocity and γ

represents the obstructive factor [5].

On the other hand, the effect of the adsorption-desorption process can be expressed by using the below formula in the mass transfer or non-equilibrium term:

$$\mathbf{H}_S = \mathbf{2}\frac{\mathbf{k}}{\mathbf{(1+k)^2}}\tau_\mathbf{s}\mathbf{u} = \mathbf{C_s u} \tag{2.30}$$

where k is the capacity factor [5].

Giddings (1965) realizes that there is a coupling between diffusion and flow. The coupled term is expressed below [5].

$$\mathbf{H}_C = \frac{\mathbf{1}}{\frac{\mathbf{1}}{\mathbf{H_f}} + \frac{\mathbf{1}}{\mathbf{H_D}}} \tag{2.31}$$

In above equation, H_f is the eddy dispersion term and H_D is formulated as:

$$\mathbf{H}_D = \mathbf{w}\frac{\mathbf{d_p^2}}{\mathbf{D}} = \mathbf{C_m u} \tag{2.32}$$

where w is a constant. The coupled term can be rewritten as [5]:

$$\mathbf{H}_C = \frac{\mathbf{1}}{\frac{\mathbf{1}}{\mathbf{2\lambda d_p}} + \frac{\mathbf{D}}{\mathbf{w d_p^2 u}}} = \frac{\mathbf{1}}{\frac{\mathbf{1}}{\mathbf{A}} + \frac{\mathbf{1}}{\mathbf{C_m u}}} \tag{2.33}$$

Finally, by combining the contributions to the band broadening the Giddings plate height equation is obtained as follows [5]:

$$\mathbf{H} = \frac{\mathbf{B}}{\mathbf{u}} + \frac{\mathbf{AC_m u}}{\mathbf{A + C_m u}} + \mathbf{C_s u} \tag{2.34}$$

In this thesis, it is aimed to develop a free and open-source code to analyze peaks and obtain key parameters using several empirical calculation methods in the literature and compare the results of these different methods. It is hoped that this code will aid researchers in the area to carry out quick and reliable estimates of chromatographic parameters.

CHAPTER 3

METHODS

In this study, moment analysis was used in order to obtain chromatographic parameters such as peak variance, skewness, and theoretical plate number. A MATLAB code was developed at the end of the study to apply the Exponential-Gaussian Hybrid Model [17].

3.1 Exponential-Gaussian Hybrid Model

The Gaussian and the truncated exponential functions are shown as follows, respectively:

$$\mathbf{f}_{gau}(t) = \begin{cases} H \exp(\frac{-(t-t_r)^2}{2{\sigma_g}^2}) & 2{\sigma_g}^2 > 0 \\ 0 & 2{\sigma_g}^2 \leq 0 \end{cases} \tag{3.1}$$

$$\mathbf{f}_{exp}(t) = \begin{cases} H \exp(\frac{-(t-t_r)^2}{\tau(t-t_R)}) & \tau(t-t_R) > 0 \\ 0 & \tau(t-t_R) \leq 0 \end{cases} \tag{3.2}$$

The way the two different equations are expressed is quite similar. It can be said that their denominators are the only difference between them. If the value of the denominators is zero, both functions will be limitless i.e., undefined. The exponential-Gaussian hybrid function is obtained by combining the functions mentioned above, and the result of this combination is shown in the function below [17].

$$\mathbf{f}_{egh}(t) = \begin{cases} H \exp(\frac{-(t-t_r)^2}{2{\sigma_g}^2+\tau(t-t_R)}) & 2{\sigma_g}^2 + \tau(t-t_R) > 0 \\ 0 & 2{\sigma_g}^2 + \tau(t-t_R) \leq 0 \end{cases} \tag{3.3}$$

While if $\tau \rightarrow 0$ the exponential- Gaussian hybrid function approaches to the Gaussian function, if ${\sigma_g}^2 \rightarrow 0$, the function approaches to the truncated exponential function [17].

3.1.1 Parameters of the EGH

Graphical measurements can be used to obtain the parameters in the exponential-Gaussian hybrid function. The data as concentration or absorbance vs. time is required to make the measurements on the chromatographic peak. The retention time and height of the peak can be found by using the coordinate of the peak. According to Lan & Jorgenson (2021), the variance of the Gaussian function and the time constant of the truncated exponential function can be determined graphically by using the following equations.

$$\sigma_{\mathbf{g}}^{\mathbf{2}} = \frac{-\mathbf{1}}{\mathbf{2ln}\alpha}(\mathbf{B}_\alpha \mathbf{A}_\alpha) \tag{3.4}$$

$$\tau = \frac{-\mathbf{1}}{\mathbf{ln}\alpha}(\mathbf{B}_\alpha - \mathbf{A}_\alpha) \tag{3.5}$$

The α sign in these equations indicates the fraction of the peak height which is the distance from the retention time to the peak's right and left sides. This fraction is determined as 0.1 [17]. When 0.1 of the peak maxima is calculated, the differences between corresponding x-axis values and retention time give the A_α and B_α values as shown in Figure 3.1.

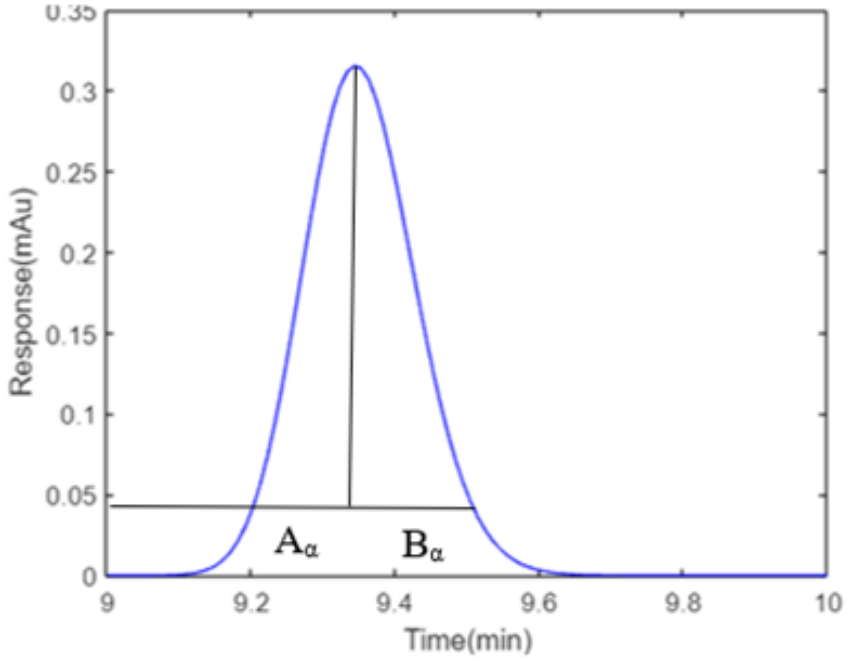

Figure 3.1: Illustration of Graphical Measurements from Chromatogram

3.2 Moment Analysis Theory

Moment term was used first by Karl Pearson almost more than a century ago to characterize the statistical distributions and asymmetric distribution curves in physical measurements. This concept originated from mechanics, which explains the zeroth moments as the total mass, the first moment as the moment of force, and the second moment as the moment of inertia. Most of the distributions have zeroth, first, second, third, and nth moments [20].

The chromatographic parameters are examined in the moment analysis method, by linking various moments to peak parameters such as retention time, peak area, peak variance (or standard deviation), peak skew, and peak excess [20]. In this study, moment analysis is used to examine the chromatographic behavior and obtain chromatographic parameters. This method is related to the general rate model of chromatography, which is described in detail in Chapter 2.

3.2.1 Moments of a Chromatographic Peak

The shape of the distribution of a function is measured quantitatively by using moments. Probability distribution and a chromatographic peak can be characterized as functions that exhibit the same behavior. Thus, according to the definitions of moments in mathematics, the weighted average is obtained from the first moment, the variance of the probability density function is obtained from the second moment and the skewness of the distribution is obtained from the third moment [20].

The moments for chromatographic peaks that are expressed by the distribution of concentration, C(t) over time, t, can be calculated with the moment definition mentioned in Chapter 2 [12].

zeroth order:

$$\mathbf{m}'_0 = \int \mathbf{C(t)}\,\mathbf{dt} \tag{3.6}$$

first order:

$$\mathbf{m}'_1 = \frac{\int \mathbf{tC(t)}\,\mathbf{dt}}{\mathbf{m_0}} \tag{3.7}$$

n^{th} order:

$$\mathbf{m}'_n = \frac{\int (\mathbf{t} - \mathbf{m_1})^{\mathbf{n}}\mathbf{C(t)}\,\mathbf{dt}}{\int \mathbf{C(t)}\,\mathbf{dt}} \tag{3.8}$$

These moments can be linked to peak parameters as peak area = m'_0, average retention time = m'_1, peak variance = m'_2, and peak skew = $m'_3/m'^{3/2}_2$ respectively [20].

The statistical moment approach provides quantitative measurement based on raw data without making any assumptions, but still, depending on the integral method used to calculate the moment, large errors may occur in the calculations. The integral in the moment calculation refers to the area under the distribution curve, so the numerically calculated value varies depending on the integration calculation method to be used. The number of data points selected for calculation affects the size of the error in the calculated area. In addition, the baseline shifts must be properly removed so that the area value of the probability distribution function can be calculated correctly. Moreover, the points where the start and end points of the chromatographic peak are

determined are also factors that significantly affect the area calculation. In literature, several methods about how to choose the start and end of the peak to calculate the area of the peak properly [20]. In the moment analysis calculation in this study, the MATLAB code was arranged so that the peak start and end points can be set manually by the user. In addition, the Trapezoidal rule was used for the integral calculations in the moment definitions.

3.3 Experimental Methods and Materials

3.3.1 Materials

Four different molecules, toluene, acetone, dimethylformamide, and ethyl acetate were used for chromatographic analysis. The reason for the use of these molecules in the study is that they are small and easily accessible molecules, they are frequently used in studies in the literature and they have different polarity values. In order to detect the presence of these molecules, a GL Sciences reverse phase (RPLC) Inertsil ODS3V C18 (150 mm, 4.6 mm, 5 µm) column and GL Sciences reverse phase (RPLC) Inertsil C8 (250 mm, 4.6 mm, 5 µm) column were used. Since the columns are nonpolar, a 60:40 volume ratio acetonitrile/ultrapure water mixture was used as the mobile phase. The same mixture is also used as a solvent for the solution preparation of molecules. The analyses were carried out with the Agilent 1260 Infinity II LC System. Openlab CDS software was used in order to examine the obtained chromatograms from the analysis. The chemicals for the chromatographic analysis were obtained from Merck. All the materials for the experiments in this thesis were obtained via Elixir Pharmaceuticals Research and Development Corporation.

3.3.2 Solution Preparation

For the analysis of the four molecules mentioned above, a mixture of acetonitrile and ultrapure water was prepared as a mobile phase in a 60:40 volume ratio. This ratio was determined based on sharper peak shape and speed of analysis. After mixing the reagents, the mobile phase was filtered through a 0.2 µm PTFE filter to prevent

corrosion and clogging of the lines of the HPLC device. In order to keep the chromatographic column clean and safe after use, an 80:20 acetonitrile/water solution and an 80:20 water/acetonitrile solution were prepared. After mixing and filtration, the mobile phase was left in the sonicator for 10 minutes to remove air bubbles. Sample solutions for acetone, toluene, dimethylformamide, and ethyl acetate to be analyzed by liquid chromatography were prepared by dissolving 0.5 mL of sample in 100 mL of mobile phase, thus solutions with a concentration value of 5 ppm were obtained for all molecules. Again, these solutions were kept in the sonicator for 10 minutes in order to eliminate the air bubbles that may be present and to ensure dissolution.

3.3.3 Chromatographic Analyses

Chromatographic analyses were carried out in a high-performance liquid chromatography device, the Agilent 1260 Infinity II LC System. A reverse phase column named GL Science Inertsil ODS3V C18 and GL Sciences Inertsil C8 column were used in order to provide separation between molecules. The lengths of these columns are 150 mm and 250 mm, respectively, and the particle size for both columns is 5 microns. Before the analysis, the column was washed with the 80:20 acetonitrile/water solution. Injections were made at different flow rates in order to obtain experimental data. As the detection means, the diode array detector (DAD) was used in the UV region for a specific wavelength of 254 nm for toluene and acetone as suggested in the article [17]. In addition, for dimethylformamide, the wavelength of 240 nm and for ethyl acetate the wavelength of 210 nm were used. The other chromatographic conditions such as flow rate, column temperature, and running time used can be seen in below Table 3.1.

3.4 Developing MATLAB Code to Calculate Chromatographic Parameters

A hybrid of exponential and Gaussian functions is developed as a mathematical model of asymmetric peak profiles. This exponential-Gaussian hybrid function (EGH) is mathematically simple and numerically stable, and its parameters are readily determined by making graphical measurements and applying simple equations. [17]. First,

Table 3.1: Chromatographic Conditions

Apparatus:	Agilent 1260 Infinity II HPLC
Column:	C18, 150 x 4.6 mm, 5 µm and C8, 250 x 4.6 mm, 5 µm
Detector:	UV-DAD at 210 nm, 240 nm, 254 nm
Flow Rate:	0.4-2.6 mL/min
Injection Volume:	10 µL
Column Temperature:	25°C, 35°C ,40°C
Running Time:	15-30 min

a MATLAB code was developed to test the exponential-Gaussian hybrid function (EGH). In this MATLAB code (EGH_function), by giving its parameters such as the height of the chromatographic peak (H), standard deviation (σ) of the Gaussian profile and, the time constant (τ) of the exponential profile, concentration versus time data and chromatograms for can be obtained. Data of toluene from the article was used to test the EGH function and the chromatogram obtained by using this MATLAB code was quite similar to the chromatogram in the article for toluene [17].

After validation of the EGH function, it was concluded that it is appropriate to make graphical measurements by using this function. A second MATLAB code which is 'chrom_loop_multiple' was created to make graphical measurements and calculate chromatographic parameters in chromatograms that include more than one peak. In this code, the time vs. absorbance data were processed and smoothed, and reconstructed using splines. The data was obtained from the OpenLab CDS software as the results of various experimental trials carried out. The chromatogram was reproduced by plotting the time vs. absorbance data.

In chrom_loop_multiple, the user can enter the program how many peaks will be identified and analyzed. Also, the resolution value can be calculated for pairs of peaks in the chromatogram. The user can specify the start and end point of the peak to calculate chromatographic parameters. After the range for the peak is stated, the code will recreate the chromatogram with indicated limits and carry out the calculations for the chromatographic parameters between these limits. These calculations include

width values in different positions of the peaks to calculate the number of theoretical plates, and the constants to be able to do the graphical measurements of a peak.

The graphical measurements, which are variance, σ^2, exponential constant, τ, tailing factor (TF), peak asymmetry factor, half width ($w_{1/2}$) and width (w_t) are done by using the constants at the different fractions of the height of the chromatographic peak. The determination of these constants was shown in Section 3.1.1 for the fraction of 1/10. The same method is also used to determine the constants at the other fractions of the height of the chromatographic peak. The equations to calculate the graphical measurements are as follows.

$$\mathbf{w}_t = \mathbf{A}_\alpha + \mathbf{B}_\alpha \tag{3.9}$$

where A_α and B_α are the constants at the 1/10 of the height of the peak.

$$\mathbf{w}_{1/2} = \mathbf{A}_{\alpha,1/2} + \mathbf{B}_{\alpha,1/2} \tag{3.10}$$

where $A_{\alpha,1/2}$ and $B_{\alpha,1/2}$ are the constants at the 1/2 of the height of the peak.

$$\mathbf{TF} = \frac{\mathbf{A}_{\alpha,1/20} + \mathbf{B}_{\alpha,1/20}}{\mathbf{2A}_{\alpha,1/20}} \tag{3.11}$$

where $A_{\alpha,1/20}$ and $B_{\alpha,1/20}$ are the constants at the 1/20 of the height of the peak.

$$\mathbf{PeakAssymetryFactor} = \frac{\mathbf{B}_\alpha}{\mathbf{A}_\alpha} \tag{3.12}$$

The illustration of the calculation of the constants can be seen in Figure 3.2.

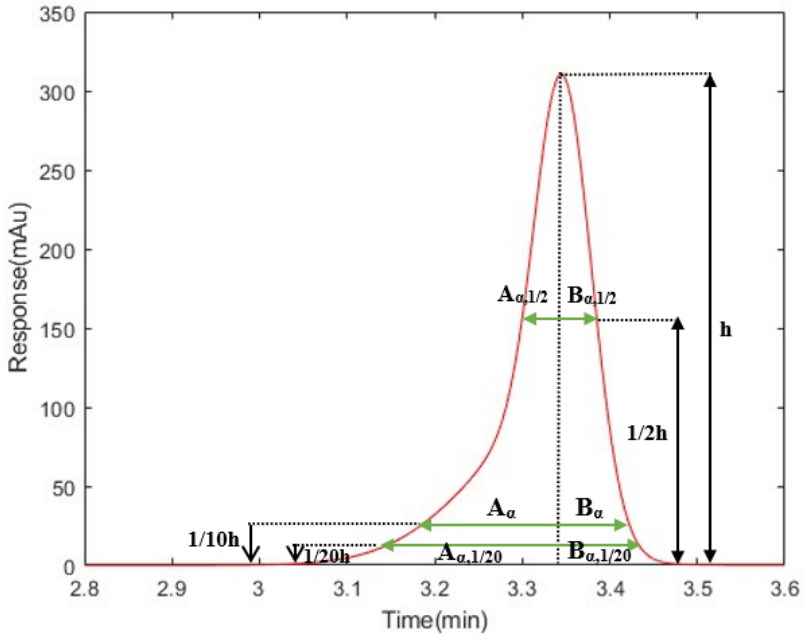

Figure 3.2: Illustration of the graphical measurements

The moment analysis is done by using the trapezoidal rule for integration. This rule was used since it is easy to use and has higher accuracy than the other numerical integration methods. In addition, the code is arranged so that other numerical integration methods can be applied. The integration is done on the chromatogram that is obtained via time vs. absorbance data from experiments. The "trapz" function is used to carry out the integration in the MATLAB code. The zeroth, first, second, and third moments are calculated by using spline smoothing to obtain peak area, peak retention time, peak variance, and peak skewness parameters, respectively. Also, by using the result of the calculations of graphical measurements and moments, three different calculations for the number of theoretical plates are done and the comparison for them will be shown in Chapter 4. The used calculation methods for theoretical plate number were shown in Table 3.2.

Table 3.2: Equations used for the calculation of theoretical plate number

Name of the Method	Relationship
Base-width	N = $16(t_R^2/w_t^2)$
Half-width	N =$5.54(t_R^2/w_{1/2}^2)$
Moment Analysis	N =$(m_1'^2/m_2'^2)$

It is calculated in different time limits as shown in Figure 3.3 below to check the reliability of the theoretical plate number calculation by moment analysis.

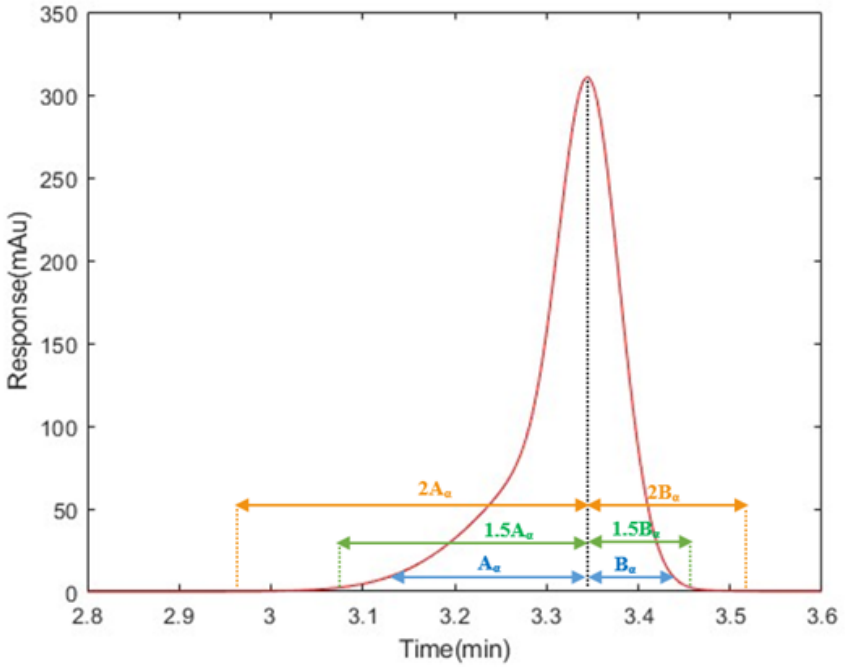

Figure 3.3: Illustration of the width of statistical moment calculations

These time limits are named as shown in Table 3.3 throughout the thesis.

Table 3.3: Time limits for calculating theoretical plate number with statistical moment

Time Range	Name
[t_R-A_α,t_R+B_α]	w
[t_R-1.5A_α,t_R+1.5B_α]	1.5w
[t_R-2A_α,t_R+2B_α]	2w

Finally, the resolution parameters for chromatograms that include more than one chromatographic peak can be obtained by using this MATLAB code. The code includes two types of calculation methods for the resolution value between two peaks. One of them uses the empirical relations from USP and the other method uses the statistical methods to obtain resolution parameter.

Also, another MATLAB code which is 'chrom_loop_single' was created for chromatograms that include a single chromatographic peak. This code works the same as the working principle of other code. The only difference between them is, all the chromatographic parameters are calculated for one peak and thus there is no resolution calculation. Users can choose one of these codes created according to their needs.

The flow chart for the algorithm of the MATLAB codes is given in Figure 3.4.

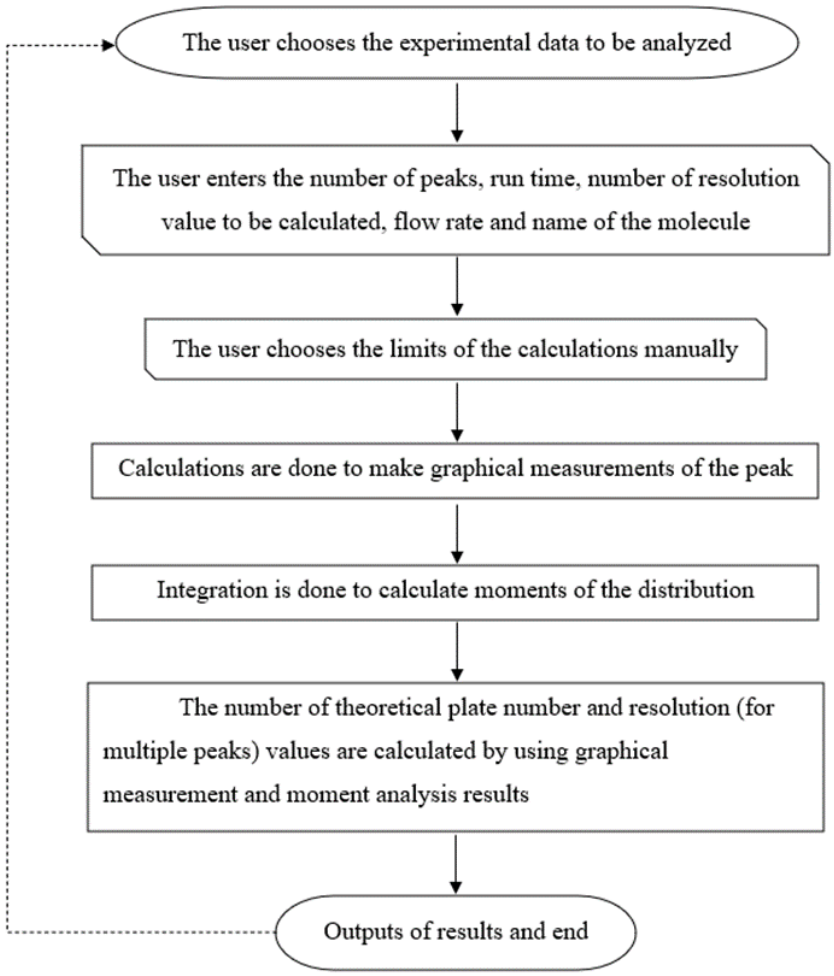

Figure 3.4: The simulation of the working of MATLAB code "chrom_loop_multiple" and "chrom_loop_single". Dashed lines indicate the loop to process multiple peaks for the chrom_loop_multiple code.

A tool to predict a chromatogram by using previously obtained chromatographic parameter data was developed for different molecules in various chromatographic conditions. This MATLAB code was named "test_read" and works with the principle of estimating chromatograms that may occur in different conditions by using the database created with the codes mentioned above. The code algorithm is summarized in the flow chart shown in Figure 3.5.

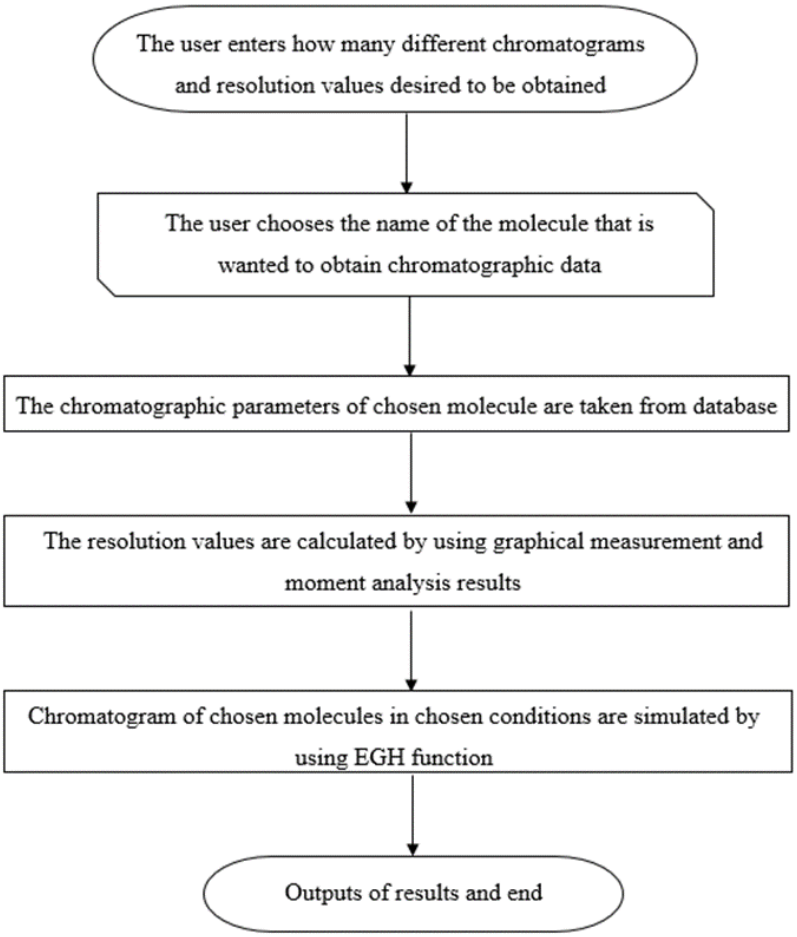

Figure 3.5: The simulation of the working of MATLAB code “test_read”

3.5 Obtaining van Deemter Curve for Molecules with Different Columns at Different Temperatures

The van Deemter curve is a curve that correlates the plate height with the velocity of the mobile phase flowing along the chromatographic column. This is an effective method to interpret the terms that contribute to the band broadening of the chromatographic column, as mentioned in Chapter 2. The curve obtained when the theoretical plate numbers calculated using the MATLAB code are plotted according to the flow rate can be called the van Deemter curve. The Excel program was used to plot the data and obtain the curve.

In this study, the van Deemter curve is obtained with dimensionless parameters which are reduced plate height and Peclet number. The particle size of the column particles and the length of the used column were used to calculate the reduced plate height as shown in the following equation,

$$\mathbf{h = \frac{L}{Nd_p} = \frac{HETP}{d_p}} \tag{3.13}$$

where HETP is the height equivalent to a theoretical plate, L is the column length and d_p is the size of the column particle.

Peclet number which is a dimensionless parameter of the mobile phase velocity is calculated by using the Equation 2.23.

The diffusion coefficient was calculated by using an empirical relation that is called Wilke and Chang equation.

3.6 Obtaining Diffusion Coefficient for Molecules with Wilke-Chang Equation

The diffusion coefficient of the analyzed molecule in liquid chromatography is one of the most important parameters of band broadening. There is a constant flow from high concentration region to low concentration region according to diffusion process principle and this leads to band broadening whether the band of the molecule in the chromatographic column is flowing or not.

In this study, a modified Wilke Chang equation [21] was used to calculate the diffusion coefficient of molecules in an acetonitrile-water mixture. According to the modified equation, there are two different empirical formulas by the fraction of molar volumes of solvent and solute as shown below [21],

$$\mathbf{D} = \begin{cases} \mathbf{10x10^{-8}} \frac{\mathbf{M^{1/2}T}}{\mu \mathbf{V_1^{1/3}V_2^{1/3}}} & \frac{V_2}{V_1} \leq 1.5 \\ \mathbf{8.5x10^{-8}} \frac{\mathbf{M^{1/2}T}}{\mu \mathbf{V_1^{1/3}V_2^{1/3}}} & \frac{V_2}{V_1} > 1.5 \end{cases} \tag{3.14}$$

Where M represents the molecular weight of the solvent, V_1 represents the molecular

volume of the solute, V_2 represents the molecular volume of the solvent and μ is the viscosity of the solvent.

The viscosity of the solvent is taken as 0.75 from Figure 3.6 below [24]. The experimental data which was shown with a dot symbol was used.

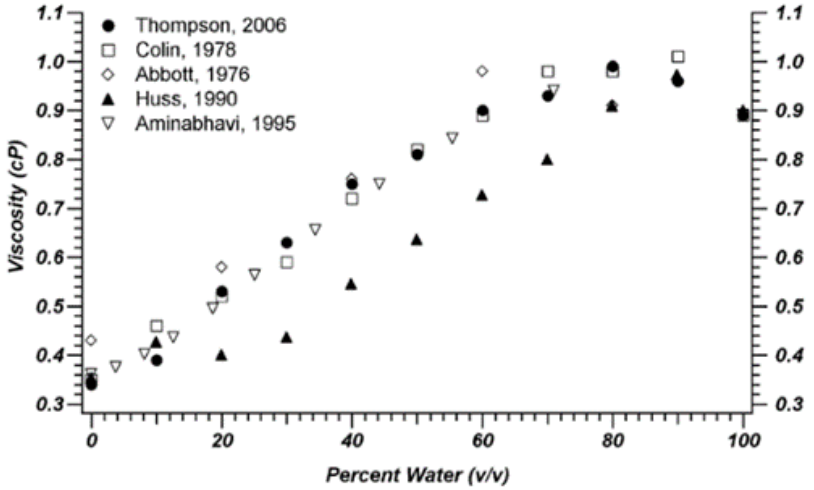

Figure 3.6: Viscosity of acetonitrile-water mixture according to composition of water [24]

The calculation of molecular weight and molecular volume of the solvent that is acetonitrile-water mixture was shown in Appendix A.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Comparing the Data and the Chromatogram of Toluene from the Article and from the MATLAB

The Exponential-Gaussian Hybrid (EGH) function [17] was used in a MATLAB code to obtain chromatographic parameters of experimental peaks. The EGH code was validated before use, and the results are as follows.

The chromatogram, which is obtained with graphical parameters from the article [17] and by using MATLAB code "EGH_function" that can be found in Appendix B.1, can be seen in Figure 4.1 below, respectively. This was carried out to confirm that the programmed EGH function produced the correct output, identical to that of the article.

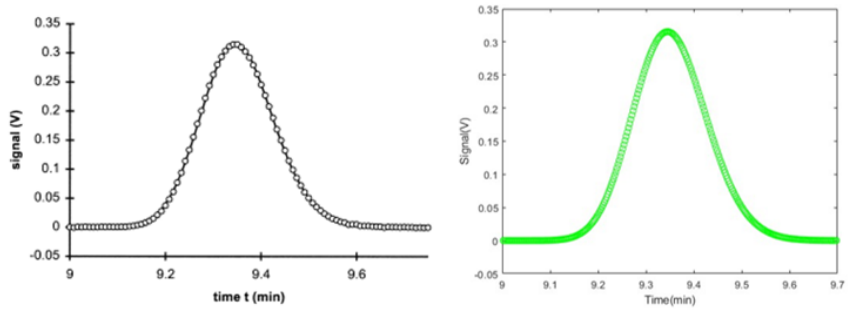

Figure 4.1: Comparison of chromatogram from article (Left) and simulated from MATLAB code (Right) for testing the EGH functionm

After the coded EGH was found to produce the correct output, the data reproduced in Figure 4.1 was used in the "chrom_loop_single" code to check whether the chromato-

graphic parameters could be recovered correctly. A comparison of original parameters from the article [17] and those recovered using the "chrom_loop_single" code is shown in Table 4.1 below. The results show that the equations are coded correctly.

Table 4.1: Comparison of chromatographic parameters for testing the EGH function

Chromatographic Parameters	Results from Article [17]	Results from MATLAB Code
A_α (min)	0.0890	0.1513
B_α (min)	0.0966	0.1759
t_R (min)	9.3450	9.3440
H (V)	0.3155	0.3155
σ^2 (min^2)	0.0058	0.0058
τ (min)	0.0102	0.0107
$m_o^{'}$ (V.min)	0.0602	0.0602
$m_1^{'}$ (min)	9.3534	9.3530
$m_2^{'}$ (min^2)	$5.96x10^{-3}$	$5.90x10^{-3}$
$m_3^{'}$ (min^3)	$1.18x10^{-4}$	$0.93x10^{-4}$

4.2 Comparing van Deemter Curves that are Obtained from Software, Graphical Calculation and Statistical Moment Calculation

4.2.1 Toluene Analysis

After verifying that the chrom_loop_single code correctly extracted chromatographic parameters from a given chromatogram, it was applied to the experimental chromatograms. The first set was carried out using toluene.

The chromatographic analysis for toluene was carried out for two different columns (C18 and C8) and two different temperatures (25°C and 40°C). Toluene shows different characteristics in the two columns as can be seen in Figure 4.2. These two chromatograms were obtained at the same flow rate and temperature, and the only difference between them is the column type. While there is a fronting in the toluene

peak in the C18 column, there is a tailing in the toluene peak in the C8 column. Also, the width values that are used in different theoretical plate number calculations are presented in Figure 4.2. $w_{1/2}$ was used in the half width calculation method, w_t was used in the base width calculation method and w_m was used in the statistical calculation method for the number of theoretical plates. More differences due to the use of different columns will be pointed out in the following sections.

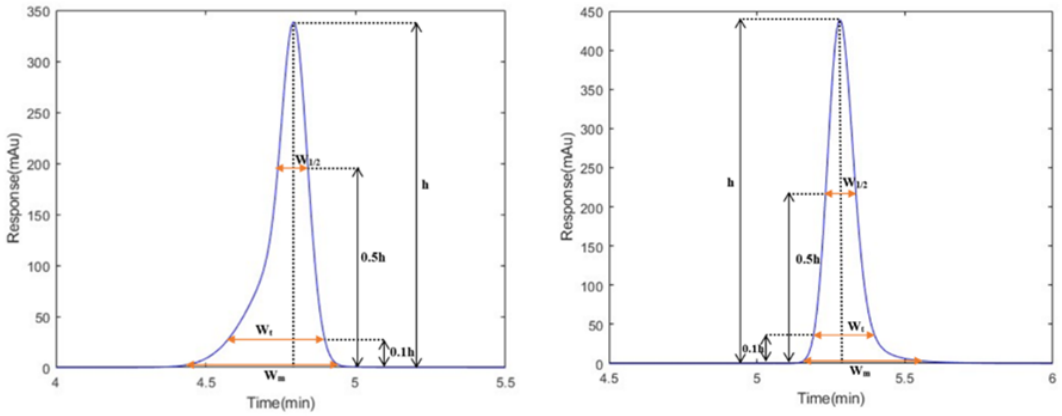

Figure 4.2: The chromatographic peak of Toluene in the C18 column (Left), Toluene peak in the C8 column (Right). Both chromatograms were at 40°C and 1.4 ml/min flow rate.

The van Deemter curves for toluene obtained with three different types of theoretical plate number calculations, as mentioned in Chapter 3, can be seen in the following figures.

4.2.1.1 Toluene Analysis in C18 Column at 25°C

In Figure 4.3, the van Deemter curves created by using the commercial OpenLab software and by using the half width theoretical plate calculation were almost identical, thus it can be concluded that this theoretical plate number calculation is used in OpenLab CDS Software.

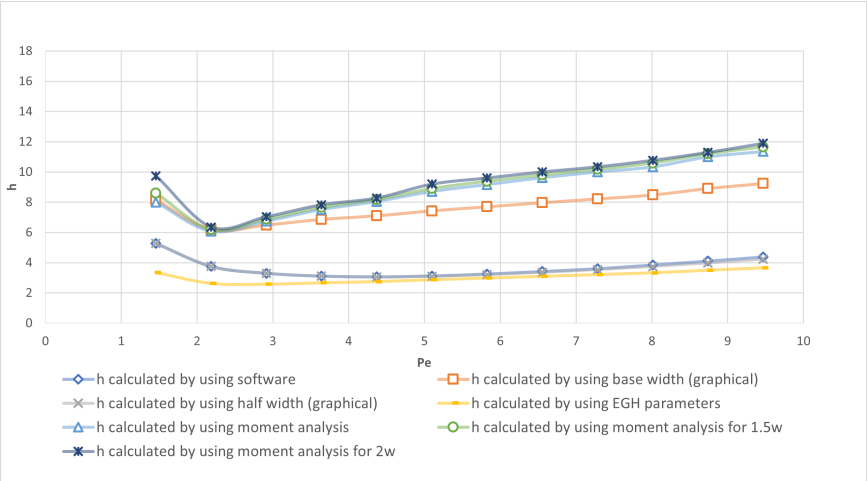

Figure 4.3: Comparison of van Deemter curves obtained with different calculation methods for Toluene in C18 column at 25°C

When van Deemter curves created by using the base width calculation method and moment analysis are examined, the results of reduced plate number from moment calculation are higher than results from base width calculation. This difference in reduced plate numbers can be caused by the assumptions for the width of the chromatographic peak. Although closer results to moment analysis were obtained with base width calculation than half width calculation, there is still a deviation from the real results because of the assumptions that were made. Since the shape at the base of the chromatographic peak is not taken into account when the half-width theoretical plate calculation method is used, and the shape of the peak is limited according to graphical constraints when the base-width theoretical plate calculation method is used, it might be obtained deviated column efficiency results with these methods. When the base width of the chromatographic peak is not considered, the possibility of tailing or fronting is ignored, and in this case, the column efficiency calculation for the peaks with tailing or fronting deviates from the correct result. From Figure 4.2, it

can be deduced that the chromatographic efficiency calculated using the base width or statistical moment will be lower than that based on half-width, since the fronting and tailing effects, which reduce efficiency and plate number are explicitly considered in the former case.

The calculations of theoretical plate numbers with statistical moment analysis were repeated in different time limits as shown in Figure 3.3 in the methods section to check the reliability of calculations. The results can be seen in Figure 4.3. The reduced plate height increases as the base width chosen for the peak integration increases. This increase is not pronounced, but it can be concluded that the results depend on the chosen time range.

The comparison of chromatograms that were obtained experimentally and by fitting the Gaussian function can be seen in Figure 4.4. This comparison was done in the flow rate of 2.6 ml/min (Pe = 9.46) at which the largest difference in reduced height was observed between the Gaussian assumption-based results and the moment-based results. In this comparison, the difference between the two chromatograms is seen. While no tailing or fronting was observed in the chromatogram obtained with the Gaussian assumption, a very clear fronting was observed in the chromatogram obtained with the experimental data. If the Gaussian assumption is used, as in graphical measurements, the fronting in the real chromatogram is neglected. In this situation, calculated theoretical plate numbers deviate from the real results because of the ideal chromatographic peak shape in the Gaussian assumption.

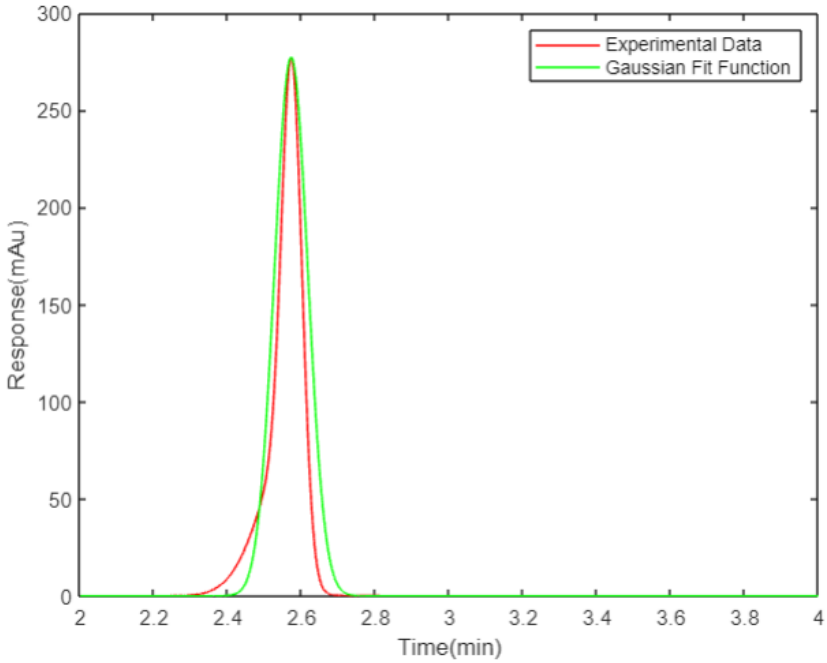

Figure 4.4: The comparison of chromatograms that is obtained experimentally and by fitting a Gaussian function

In contrast, in Figure 4.5, the comparison of chromatograms obtained experimentally and by fitting the EGH function can be seen. The comparison was done in the Peclet number 9.46 once again. Although the chromatographic peak obtained by fitting the EGH function is closer to the experimental one compared to the Gaussian fit, there are still differences. The reduced height obtained from the EGH fit of the peak is lower than the peak obtained from the moments obtained directly from the experimental peak since the assumed EGH function slightly underestimates the fronting value. It is concluded that, if any functional form is assumed for calculating chromatographic parameters, the calculated result deviates from real data.

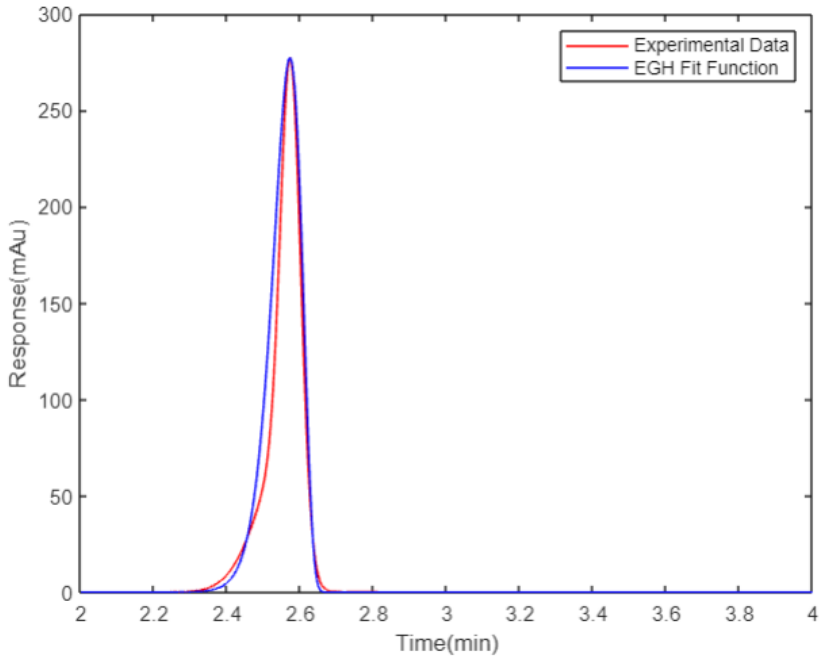

Figure 4.5: The comparison of chromatograms that is obtained experimentally and by fitting EGH function

4.2.1.2 Toluene Analysis in C18 Column at 40°C

It can be seen in Figure 4.6 that the van Deemter curve obtained by using the same chromatographic column at a different temperature supports the conclusion for the lower temperature chromatographic analysis, i.e. the major reason for the difference between curves is the method of calculation of the theoretical plate numbers. When tailing or fronting comes into play, assumptions for graphical measurements can cause deviations similar to the results mentioned above for 25 °C.

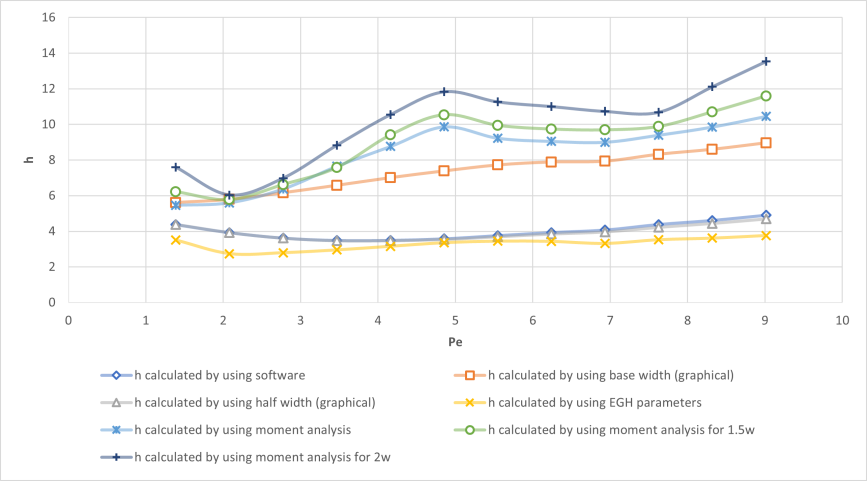

Figure 4.6: Comparison of van Deemter curves obtained with different calculation methods for Toluene in C18 column at 40°C

4.2.1.3 Toluene Analysis in C8 Column at 35°C

The van Deemter curve obtained for chromatographic analysis of toluene in C8 column by software and by using half width theoretical plate calculation were obtained exactly same as for chromatographic analysis of toluene in C18 as shown in Figure 4.7. The same trend between the reduced height calculated from graphical measurements and statistical moments was observed as in C18 column.

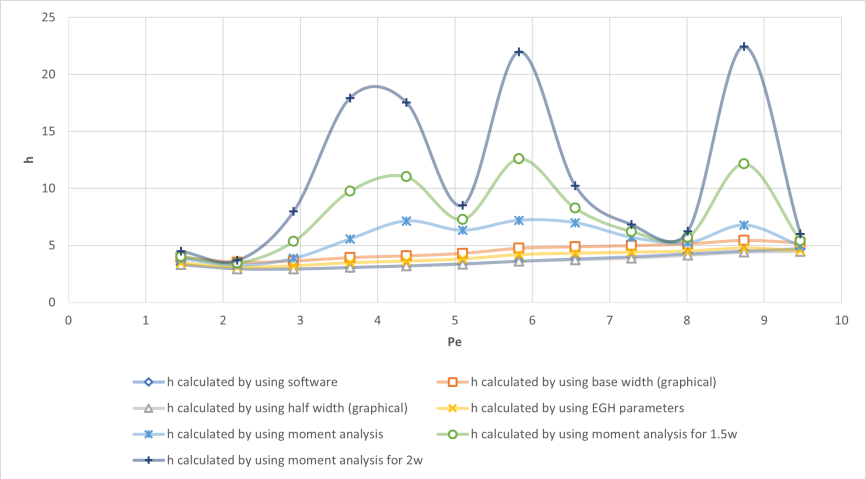

Figure 4.7: Comparison of van Deemter curves obtained with different calculation methods for Toluene in C8 column at 35°C

As shown in Figure 4.8, it was observed that the baseline of the chromatograms obtained at some flow rates was more distorted than the chromatograms at other flow rates. It can be concluded that the deviations and the fluctuations in the calculated theoretical plate numbers at flow rates may be a result of such distorted baselines because the integrated area deviates from the correct value when the baseline shifts. For future work, it is planned to work on obtaining more accurate results by correcting the baseline of the chromatogram.

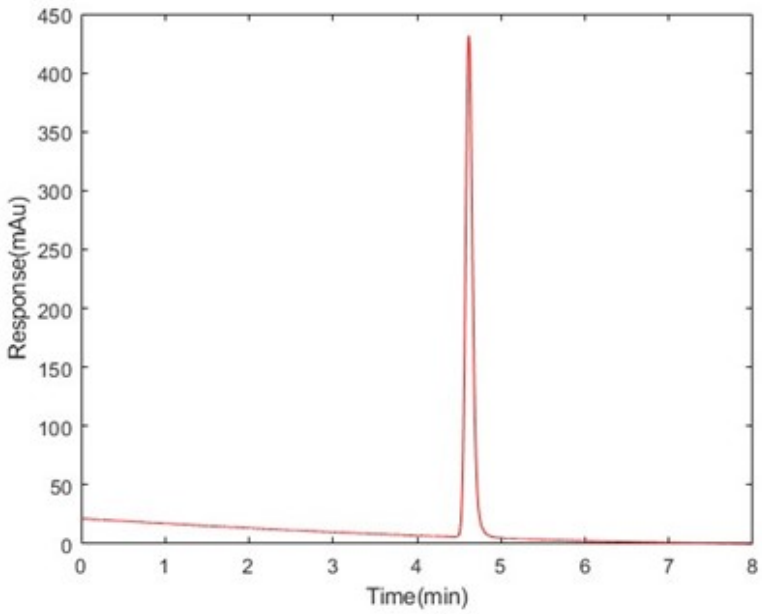

Figure 4.8: Distorted baseline obtained from Toluene analysis in C8 column at 35°C with the flow rate of 1.6 ml/min

4.2.1.4 Toluene Analysis in C8 Column at 40°C

The same trend between van Deemter curves obtained with different calculation methods was detected as in the other toluene analyses.

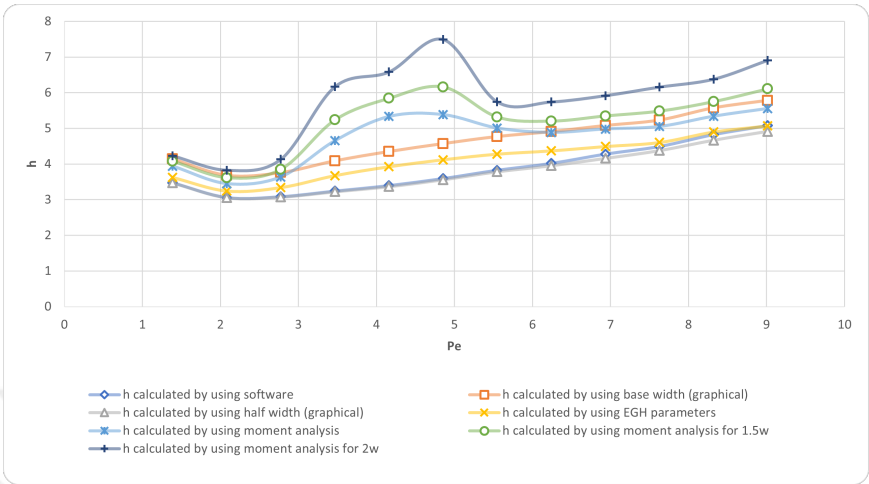

Figure 4.9: Comparison of van Deemter curves obtained with different calculation methods for Toluene in C8 column at 40°C

It was seen that the van Deemter curves obtained by the base width and moment analysis method in the toluene analyzes performed on the C8 column exhibit a closer trend than the curves obtained from the analyses performed on the C18 column. The reason for this similarity in the van Deemter curves, which are the results of the analyses made on the C8 column, is that the chromatographic peaks in these analyzes are narrower, that is, the variance value and the tailing value are lower. The Gaussian assumption made in the graphical calculation did not cause much deviation from the actual result because the experimental peak shape obtained is more suitable for the Gaussian assumption than the peak shape in the C18 column, as shown in Figure 4.10.

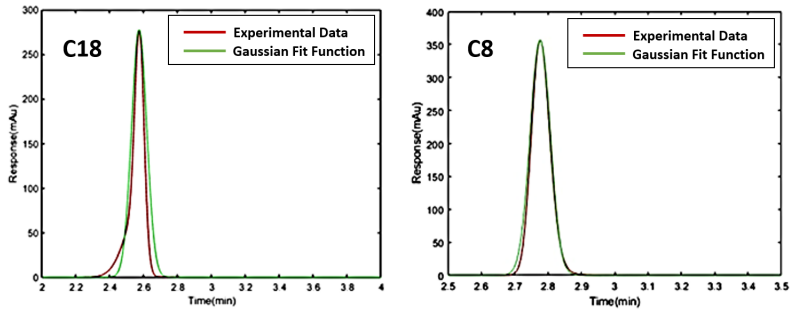

Figure 4.10: The comparison of chromatograms that is obtained experimentally and by fitting Gaussian function for C18 column (Left) and C8 column (Right)

4.2.2 Acetone Analysis

The same calculation methods were used as toluene to calculate chromatographic parameters of acetone molecule analysis. The polarity of acetone is higher than toluene and, in this study, reverse phase chromatography was used. Therefore, while the acetone molecule is a non-retained molecule for this analysis, toluene is retained molecule. As a result of this situation, the retention time and thus the peak variance is smaller for acetone than for the toluene molecule. van Deemter curve was affected by the retention as shown in the figures below.

4.2.2.1 Acetone Analysis in C18 Column at 25°C

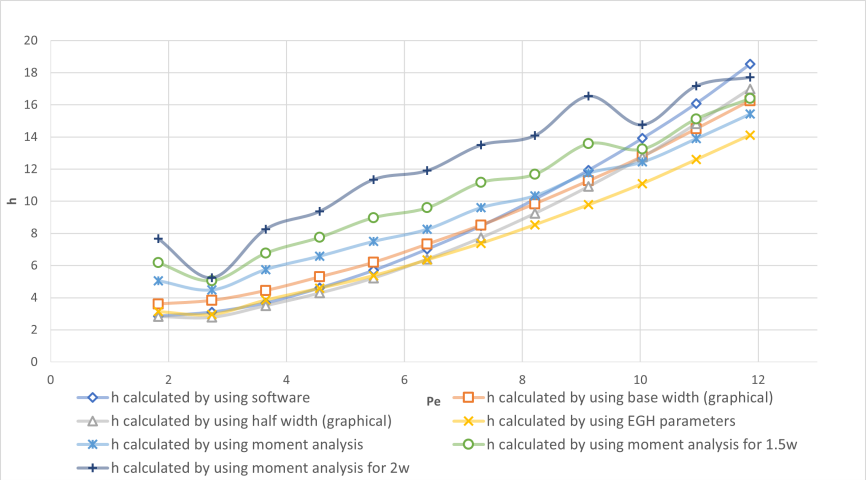

Figure 4.11: Comparison of van Deemter curves obtained with different calculation methods for Acetone in C18 column at 25°C

4.2.2.2 Acetone Analysis in C18 Column at 40°C

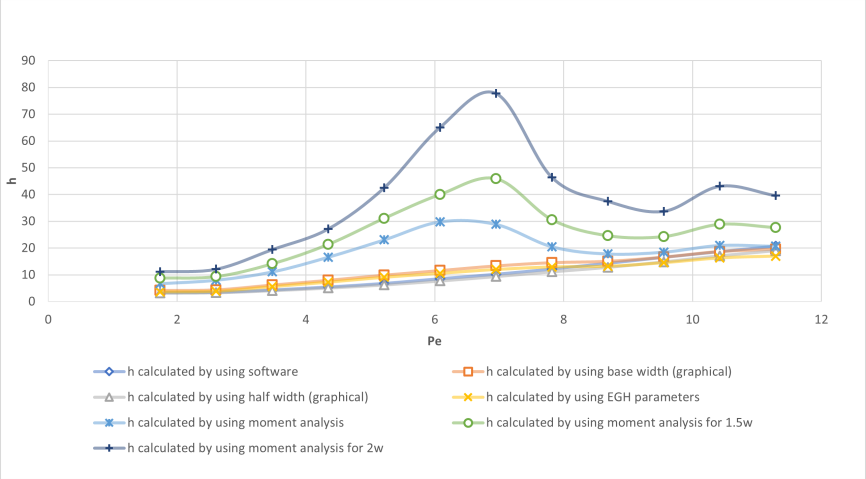

Figure 4.12: Comparison of van Deemter curves obtained with different calculation methods for Acetone in C18 column at 40°C

4.2.2.3 Acetone Analysis in C8 Column at 35°C

The reasons for the deviations in the van Deemter curve for acetone in different columns and temperatures are the distorted baseline and noise amount of the baseline.

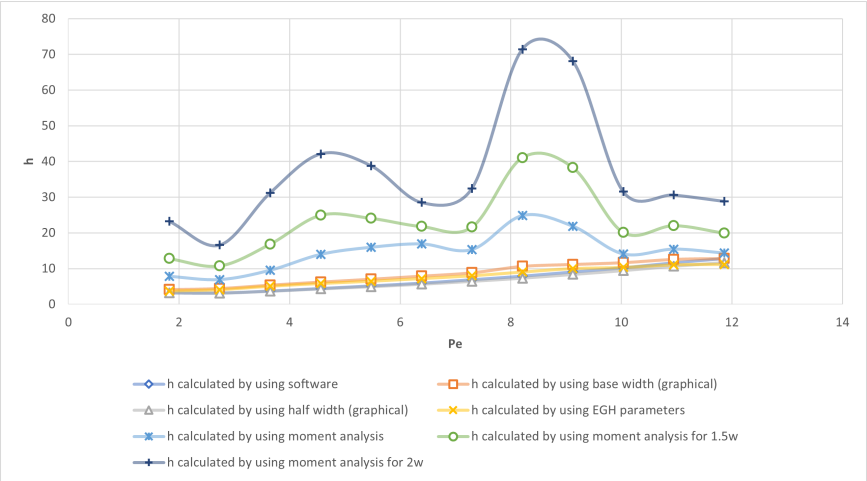

Figure 4.13: Comparison of van Deemter curves obtained with different calculation methods for Acetone in C8 column at 35°C

It can be seen in Figure 4.14 that the noise amount for acetone analysis in the C8 column at 35°C with the flow rate of 1.8 ml/min is higher than the noise amount for acetone analysis in the C8 column at 35°C with the flow rate of 0.4 ml/min. This difference affects the variance and thus theoretical plate number on related flow rate. Also, as the time range that the variance value is calculated broadens the fluctuation in van Deemter curve increases since the amount of noise that will affect the calculation increases.

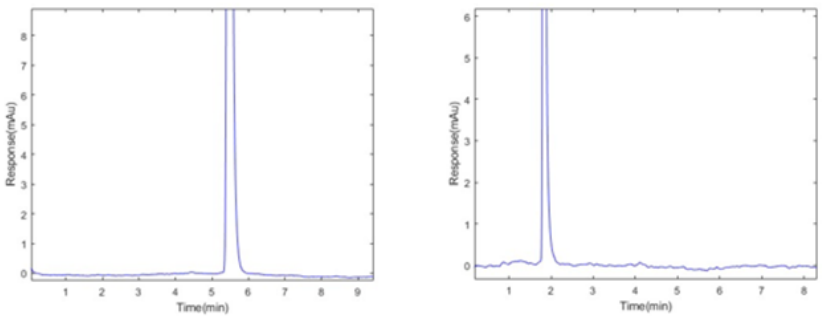

Figure 4.14: The comparison of the amount of noise for chromatograms obtained in the flow rate of 0.4 mL/min (Left) and 1.8 mL/min (Right)

4.2.2.4 Acetone Analysis in C8 Column at 40°C

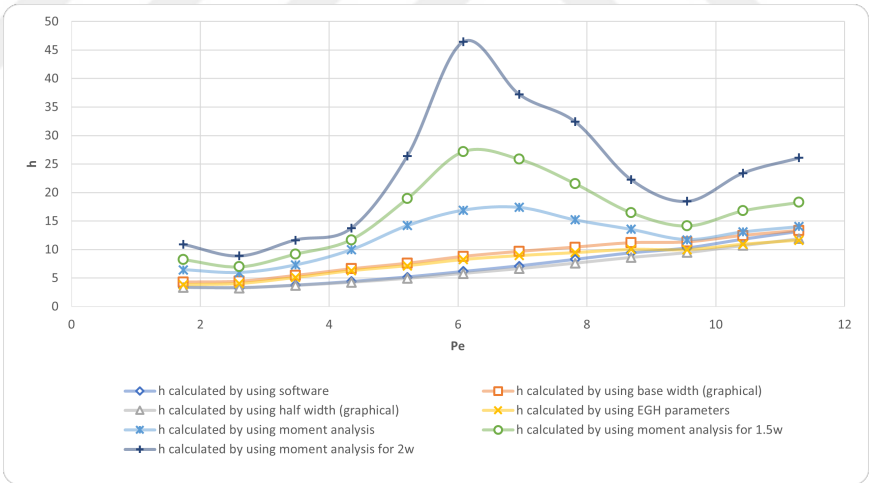

Figure 4.15: Comparison of van Deemter curves obtained with different calculation methods for Acetone in C8 column at 40°C

Analyzes carried out under the above-mentioned chromatographic conditions were also performed for dimethylformamide and ethyl acetate. van Deemter curves obtained for these molecules are shown in the Appendix C.

4.2.3 Examining the Effect of Different Chromatographic Conditions on van Deemter Curve

4.2.3.1 Effect of Temperature on van Deemter Curve

It is a known fact that peak times are reduced when liquid chromatography analysis is performed at higher temperatures. The main objectives of increasing the temperature of the column are to reduce the analysis time and to get higher efficiency. In this study, the used chromatographic columns do not permit to increase the temperature above 40°C because of their structure and sensitivity of the particle. Considering the results obtained in the temperature range allowed by the columns used, it was observed that the temperature had a reduced effect on the retention time, while there was no significant change in the obtained van Deemter curves, that is, in the chromatographic efficiency. It can be seen there is almost no difference between the van Deemter curves obtained in different temperatures for toluene.

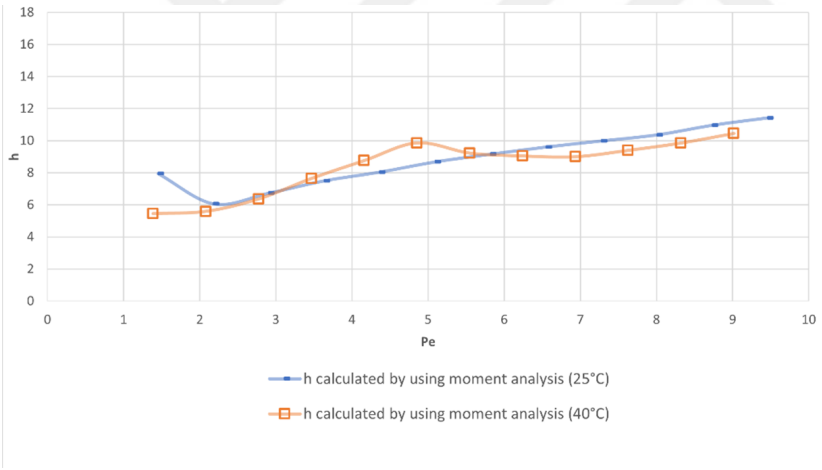

Figure 4.16: Comparison of van Deemter curves obtained in different temperatures for toluene

There is a shift in the x-axis since the temperature difference affects the diffusion coefficient value. This difference in diffusion coefficient shows itself in the retention time of the peak. Thus, although there are important effects on column efficiency for higher temperatures, in small temperature ranges, temperature has no significant impact on efficiency. The only advantage of higher temperature is shorter analysis time.

4.2.3.2 Effect of Column Type on van Deemter Curve

Two types of chromatographic columns which are C18 and C8 were used in this thesis study. All of the properties such as particle diameter and internal diameter are the same for these columns except their length. While the C18 column is 150 mm, the C8 column is 250 mm. These columns are used widely in reverse phase chromatography. Both of the columns are hydrophobic but there are certain differences between them and these affect the separation efficiency. The difference between the results from the two columns can be seen clearly in Figure 4.17. This figure includes van Deemter curves for toluene at 40°C in C18 and C8 columns.

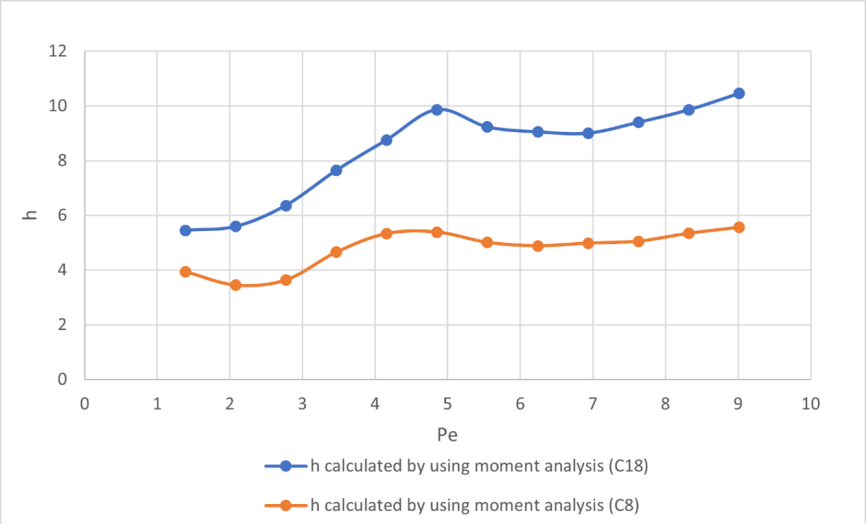

Figure 4.17: Comparison of van Deemter curves obtained in different columns for Toluene

According to the obtained van Deemter curve, the C8 column has better column efficiency for toluene than the C18 column. The major reason for the difference between column efficiencies is their hydrophobicity and their carbon chain lengths. It can be easily said that the carbon chain length of the C18 column is longer than the C8 column as shown in Figure 4.18. The retention time in the C8 column is lower than in the C18 column since C18 has a longer carbon chain bonded to silica and C8 has lower hydrophobicity than the other one. It can be concluded that if fast separation is required, the C8 column has more advantages. Also, the length of the column is an important parameter for the speed of the separation.

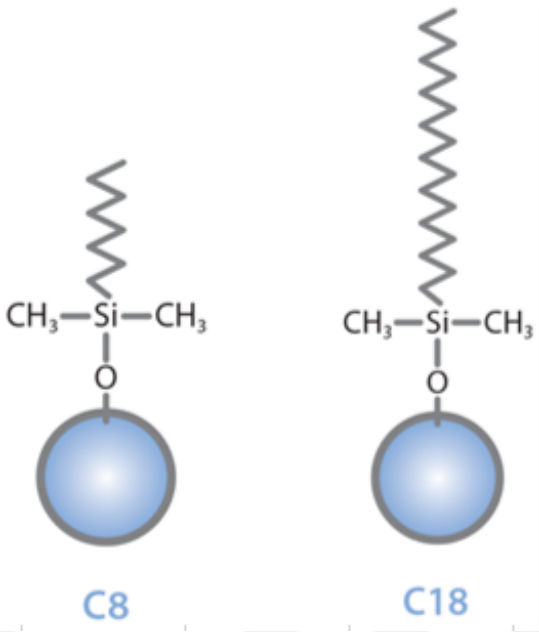

Figure 4.18: Representation of the structure of C18 and C8 columns

For this example, although the length of the C8 column is longer than C18, their retention time is almost the same at 0.8 ml/min flow rate, as can be seen in Figure 4.19.

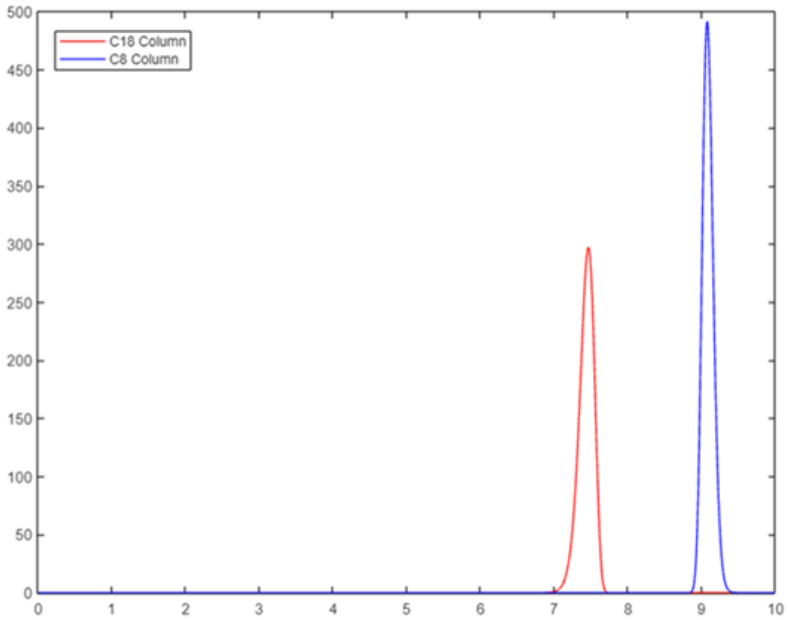

Figure 4.19: The comparison of chromatograms that is obtained for toluene at 40°C in flow rate of 0.8 mL/min (Pe = 2.77) for C8 and C18 columns

The variance value for the chromatographic peak from the C18 column is higher and also there is a tailing in this peak. As a result, for toluene analysis in these conditions, the C8 column is more suitable than the C18 column as can be seen from chromatograms in Figure 4.19. Also, the density of the C18 column is higher than the C8 column because of their carbon chain length. This causes better resolution in the case of closely eluting peaks. The column type to be preferred may vary according to the expected result from the analysis and with the open-source tool we have developed in this thesis, the necessary analysis parameters can be determined in advance according to the required result.

CHAPTER 5

CONCLUSION AND FUTURE WORK

In this thesis, peak characterization and chromatographic parameter estimation methods applied to experimental data were investigated, examined, and compared. For this purpose, experimental chromatograms were obtained for four different molecules in two different columns and at two different temperature values. MATLAB codes that estimate the chromatographic parameters by applying moment analysis were developed, and the chromatographic parameters and peak shape of the experimental chromatograms were examined by using these MATLAB codes. The theoretical plate number, which is an important criterion of separation efficiency, was obtained by different methods. By comparing these results on the van Deemter curve, the most accurate separation efficiency calculation was investigated. According to the results obtained, while graphical calculation methods give accurate and consistent results for Gaussian peaks, which are the ideal peak shape, it is observed that there is a deviation in the results in the case of fronting and tailing. The reason for this deviation is that non-ideality in chromatographic peaks is not considered while obtaining parameters. The result desired to be obtained from the analysis may vary depending on the purpose of the analysis. For example, in an experiment where a single molecule is analyzed, the aim is to increase the efficiency of the analysis by obtaining the most uniform peak shape in the shortest time, while in an experiment where more than one molecule is analyzed, the primary goal is to separate two molecules from each other in the most efficient way.

In this study, to determine the calculation method to be used for the target; an open-source estimation tool has been developed that can create a database from the analyses of different molecules made under different conditions, predict the results of the

chromatographic analyses planned using this database, and provide results quickly if a few necessary parameters are provided. In future work, it is planned to develop a predicted tool so that there will be no need for repeated chromatographic analyses. The database should be enhanced to get automatic results without experiments. As a result, it is aimed that this study will help researchers working in this field obtain fast and reliable results. Also, as a future work, it is aimed to consider the effect of the retention factor on separation efficiency. Although different from the terminology used throughout the thesis, the equation given in Appendix D can be used to observe the effect of the retention factor on the separation efficiency.

REFERENCES

[1] Ivett Bacskay and Attila Felinger. Macroscopic and microscopic analysis of mass transfer in reversed phase liquid chromatography. *Journal of Chromatography A*, 1216, 2009.

[2] Howard Barth. Chromatography fundamentals, part v: Theoretical plates: Significance, properties, and uses. *LCGC North America*, 36:830–835, 2018.

[3] Harold G. Cassidy. *Fundamentals of Chromatography*, volume 7. Oxford University Press (OUP), 11 1957.

[4] Anton Daneyko, Dzmitry Hlushkou, Vasili Baranau, Siarhei Khirevich, Andreas Seidel-Morgenstern, and Ulrich Tallarek. Computational investigation of longitudinal diffusion, eddy dispersion, and trans-particle mass transfer in bulk, random packings of core-shell particles with varied shell thickness and shell diffusion coefficient. *Journal of Chromatography A*, 1407, 2015.

[5] A. Felinger and A. Cavazzini. Kinetic theories of liquid chromatography, 2013.

[6] Attila Felinger, Alberto Cavazzini, and Francesco Dondi. Equivalence of the microscopic and macroscopic models of chromatography: Stochastic-dispersive versus lumped kinetic model. *Journal of Chromatography A*, 1043, 2004.

[7] Attila Felinger, Alberto Cavazzini, Maurizio Remelli, and Francesco Dondl. Stochastic-dispersive theory of chromatography. *Analytical Chemistry*, 71, 1999.

[8] Guiochon G., Felinger A., Shirazi D.G., and Katti A.M. Fundamentals of preparative and nonlinear chromatography" by g. guiochon, a. felinger, d.g. shirazi [elsevier, amsterdam, 2nd ed., 2006, ch. 1], 2006.

[9] J. Calvin Giddings. The random downstream migration of molecules in chromatography. *Journal of Chemical Education*, 35, 1958.

[10] J. Calvin Giddings. *Dynamics of chromatography: Principles and theory*. 1965.

[11] J. Calvin Giddings and Henry Eyring. A molecular dynamic theory of chromatography. *Journal of Physical Chemistry*, 59, 1955.

[12] Eli Grushka. Chromatographic peak shape analysis, 1975.

[13] David Harvey. Contributions to band broadening in chromatography, 8 2013.

[14] Jan Ake Jönsson and Colin F. Poole. *Chromatographic theory and basic principles*, volume 222. Elsevier BV, 1989.

[15] Harun Koku, Robert S. Maier, Mark R. Schure, and Abraham M. Lenhoff. Modeling of dispersion in a polymeric chromatographic monolith. *Journal of Chromatography A*, 1237, 2012.

[16] Michael R Ladisch. *Bioseparations engineering: principles, practice, and economics*. Wiley New York, 2001.

[17] Kevin Lan and James W. Jorgenson. A hybrid of exponential and gaussian functions as a simple model of asymmetric chromatographic peaks. *Journal of Chromatography A*, 915:1–13, 2001.

[18] Valerio B. Di Marco and G. Giorgio Bombi. Mathematical functions for the representation of chromatographic peaks, 2001.

[19] V. R. Meyer. Chromatography | principles. *Encyclopedia of Analytical Science: Second Edition*, pages 98–105, 1 2005.

[20] Shobhit Misra, M. Farooq Wahab, Darshan C. Patel, and Daniel W. Armstrong. The utility of statistical moments in chromatography using trapezoidal and simpson's rules of peak integration. *Journal of Separation Science*, 42:1644–1657, 4 2019.

[21] K. A. Reddy and L. K. Doraiswamy. Estimating liquid diffusivity. *Industrial and Engineering Chemistry Fundamentals*, 6, 1967.

[22] S. V. Romanenko and A. G. Stromberg. Classification of mathematical models of peak-shaped analytical signals. *Journal of Analytical Chemistry*, 55, 2000.

[23] Di Song and Juan Wang. Modified resolution factor for asymmetrical peaks in chromatographic separation. *Journal of Pharmaceutical and Biomedical Analysis*, 32:1105–1112, 2003.

[24] J. Will Thompson, Theodore J. Kaiser, and James W. Jorgenson. Viscosity measurements of methanol-water and acetonitrile-water mixtures at pressures up to 3500 bar using a novel capillary time-of-flight viscometer. *Journal of Chromatography A*, 1134, 2006.

[25] Karyn M. Usher, Carolyn R. Simmons, and John G. Dorsey. Modeling chromatographic dispersion: A comparison of popular equations. *Journal of Chromatography A*, 1200, 2008.

[26] J.J. van Deemter, A. Klinkenberg, and F. J. Zuiderweg. Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. *Chemical Engineering Science*, 5, 1956.

[27] Eric von Lieres and Joel Andersson. A fast and accurate solver for the general rate model of column liquid chromatography. *Computers and Chemical Engineering*, 34, 2010.

Appendix A

DIFFUSION COEFFICIENT

A.1 Sample Calculation of Diffusion Coefficient

The molecular weights of the molecules that form the mixture taken from the literature are as follows.

MW_{ACN} = 41.05 g/mol

MW_{Water} = 18.02 g/mol

The average molecular weight of acetonitrile and water mixture can be calculated as below.

$$\overline{MW} = \sum_{i} x_i MW_i = 0.34x41.05 + 0.66x18.02 = 28.85 g/mol \quad \text{(A.1)}$$

Molecular volume of acetonitrile and water mixture can be calculated as below.

The total volume of the liquid mixture was chosen as 1000 mL as a basis.

V_{Total} = 1000 ml

For the mixture of 60:40 volume ratio acetonitrile- water,

V_{ACN} = 600 ml

V_{Water} = 400 ml

The densities of the molecules that form the mixture taken from the literature are as follows.

ρ_{ACN} = 0.786 g/ml

ρ_{Water} = 1 g/ml

According to the information given above,

ρ_{mix} = 0.872 g/ml

The moles of the molecules for 1000 ml mixture were calculated as follows,

$$\mathbf{m_i} = \rho_\mathbf{i}\mathbf{V_i} \tag{A.2}$$

m_{ACN} = 471.6 g

m_{Water} = 400 g

$$\mathbf{n_i} = \frac{\mathbf{m_i}}{\mathbf{MW_i}} \tag{A.3}$$

m_{ACN} = 11.5 mol

m_{Water} = 22.2 mol

Finally, the mole fraction for acetonitrile and water in a 1000 mL solution is 0.34 and 0.66, respectively.

Accordingly, molecular volume of the solution can be calculated as below,

$$\mathbf{V_m} = \frac{\sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{x_i MW_i}}{\rho_\mathbf{mix}} = \frac{\mathbf{0.34x41.05 + 0.66x18.02}}{\mathbf{0.872}} = \mathbf{29.64 ml/mol} \tag{A.4}$$

Sample calculation for toluene diffusion coefficient at 25°C in the mixture of 60:40 volume ratio acetonitrile- water:

$$\mathbf{D_{Toluene,25C}} = \mathbf{10x10^{-8}} \frac{\mathbf{25.85^{1/2}x298}}{\mathbf{0.75x106.3^{1/3}x29.64^{1/3}}} = \mathbf{1.38x10^{-5} cm^2/s} \tag{A.5}$$

Appendix B

MATLAB CODES

B.1 MATLAB Code for Testing EGH Function

```
function EGH_function
% syms t
% The chromatographic parameters are entered to obtain chromatogram
% Following chromatographic parameters are taken from article (Lan & Jorgenson, 2001) for toluene
H = 0.3155;
var = 0.0058;
tau = 0.01019;
tr = 9.345;
t = linspace(0,15,500);
%C = H*exp((-((t-tr).^2)./(2*var+tau.*(t-tr)))) --> EGH function equation
C = zeros(size(t));
for i=1:length(t)
    cond = (2*var+(tau*(t(i)-tr)));
    if  cond > 0
        C(i) = H*exp((-((t(i)-tr).^2)./(2*var+(tau.*(t(i)-tr)))));
    elseif cond <= 0
        C(i)=0;
    end
end
S = [t C]'
%whos A t S
figure(1)
plot(t,C,'b-')
xlabel('Time(min)');
ylabel('Response(mAu)');
```

Figure B.1: MATLAB Code 'EGH_function'

B.2 MATLAB Code for Calculating Chromatographic Parameters in Chromatograms that Include More than One Peak

```
function chrom_loop_multiple

% The user can enter the parts that are variable according to the analysis herself/himself at the beginning of the code.
Number_of_peak = input('Enter number of peak you want to simulate')
run_time = input('Enter run time for your system:')
Resolution_of_peak = input('Enter number of resolution value you want to calculate')

% The data to be used in the code to obtain the chromatographic parameters are given to the code and splined in this section.
Data = xlsread(['Acetone.5ppm.1.4.C18.xlsx']);
Flow_Rate = input('Enter the flow rate of the molecules in the column')
t = Data(:,1);
C = Data(:,2);
xx = linspace(0,run_time,10000); % (min)
splinedata = spline(t,C,xx);
minspline = min(splinedata);
splinedata_new = (splinedata - minspline);
figure(1)
plot(t,C,'b-');
xlabel('Time(min)');
ylabel('Response(mAu)');
figure(2)
plot(xx,splinedata_new,'r-');
xlabel('Time(min)');
ylabel('Response(mAu)');

% With the for loop, the user can enter how many different peaks are in the analysis and in this way, the resolution can be calculated.
for i = 1:Number_of_peak

    Molecule = input('Enter the name of the molecule you wish to simulate','s')

disp('Using left mouse button, click on the x-coordinate of the endpoints')
disp('of the data range you wish to focus on')
disp('Left border first and right border second...')

[xx1,splinedata_new1]=ginput(2);
figure(3)
if (xx1(1)<0)
    xx1(1)=min(xx);
end

if (xx1(2)>max(xx))
    xx1(2)=max(xx);
end
```

Figure B.2: MATLAB Code 'chrom_loop_multiple'(First Page)

```
dist_left = sqrt((xx-xx1(1)).^2);
[~,left_idx(i)] = min(dist_left,[],2);
dist_right= sqrt((xx-xx1(2)).^2);
[~,right_idx(i)] = min(dist_right,[],2);
start = xx(left_idx(i));
stop = xx(right_idx(i));
range=[start stop]

 % Plot data with updated range
 plot(xx(left_idx(i):right_idx(i)),splinedata_new(left_idx(i):right_idx(i)),'ro');
    xaxis_min=start; xaxis_max=stop; yaxis_min=min(splinedata_new); yaxis_max=max(splinedata_new);
    axis([xaxis_min,xaxis_max,yaxis_min,yaxis_max])
    xlabel('t');
    ylabel('UV280');
    grid on;
    hold on

C_truncated = splinedata_new(left_idx(i):right_idx(i));
xx_truncated_1 = xx(left_idx(i):right_idx(i));
[C_truncated ind] = unique(C_truncated, 'stable');
xx_truncated_1 = xx_truncated_1(ind);

[C_max, index] = max(C_truncated) % maximum ordinate value which means peak height (H)
tr(i) = xx_truncated_1(index) % retention time of the peak
cc = C_max*0.1; % It is calculated in order to make graphical measurements.
cc_HW = C_max*0.5; % It is calculated to find the half width of the peak and calculate the N.
RowMax(i) = find(C_truncated==C_max); % Calculate row where C_max falls.
C_End = C_truncated(end);
RowEnd(i) = find(C_truncated==C_End); % Calculate row where C_end falls.
C_Low = C_truncated(1:RowMax(i)); % Concentration values lower than C_max
C_High = C_truncated(RowMax(i):RowEnd(i)); % Concentration values higher than C_max
t_L = xx_truncated_1(1:RowMax(i)); % The time interval at which concentration values less than C_max coincide.
t_R = xx_truncated_1(RowMax(i):end); % The time interval at which concentration values higher than C_max coincide.
% Time range limits corresponding to a height of 0.10 of C_max.
t_Low = interp1(C_Low, t_L, cc, 'linear', 'extrap')
t_High = interp1(C_High, t_R, cc)

% Below calculations are for finding the range that mom2 is calculated.
Aalfa_mom2(i) = abs(tr(i) - start)
Aalfa_left(i) = tr(i) - Aalfa_mom2(i).*1.5
Aalfa_left_2(i) = tr(i) - Aalfa_mom2(i).*2
Balfa_mom2(i) = abs(stop - tr(i))
Balfa_right(i) = tr(i) + Balfa_mom2(i).*1.5
Balfa_right_2(i) = tr(i) + Balfa_mom2(i).*2

xx_mom2_15 = linspace(Aalfa_left(i),Balfa_right(i),10000);
```

Figure B.3: MATLAB Code ’chrom_loop_multiple’ (Second Page)

```
xx_mom2_15 = xx_mom2_15(ind);
C_truncated_15 = spline(xx,splinedata_new,xx_mom2_15);
xx_mom2_2 = linspace(Aalfa_left_2(i),Balfa_right_2(i),10000);
xx_mom2_2 = xx_mom2_2(ind);
C_truncated_2 = spline(xx,splinedata_new,xx_mom2_2);

% Below calculations are for calculating the widht at one-half peak height.
t_low_HW(i) = interp1(C_Low, t_L, cc_HW, 'linear', 'extrap');
t_High_HW(i) = interp1(C_High, t_R, cc_HW);
Aalfa_HW(i) = abs(tr(i) - t_low_HW(i));
Balfa_HW(i) = abs(t_High_HW(i) - tr(i));
Width_HW(i) = Aalfa_HW(i) + Balfa_HW(i) % Widht at one-half peak height


%Graphical measurements for chromatographic parameters.
Aalfa(i) = abs(tr(i) - t_Low)
Balfa(i) = abs(t_High - tr(i))
varianceGraphical(i) = ((-1/(2*log(0.10)))*Aalfa(i)*Balfa(i)) % variance calculated from the obtained peak from t, C graphical data.
tauGraphical(i) = ((-1/log(0.10))*(Balfa(i)-Aalfa(i))) % time constant calculated from the obtained peak from t, C graphical data.
AssymetryLevel(i) = (tauGraphical(i))/(varianceGraphical(i))^(1/2)
PeakAssymetryFactor(i) = Balfa(i)/Aalfa(i)
theta(i) = atan((abs(tauGraphical(i))/(varianceGraphical(i))^(1/2))); %theta the is absolute bounded assymetry
Width(i) = Aalfa(i) + Balfa(i)
Tailing_Factor(i) = (Aalfa_mom2(i)+Balfa_mom2(i))./(2.*(Aalfa_mom2(i)))

% Moment calculations for chromatographic parameters.
mom0(i) = trapz(xx_truncated_1,C_truncated) % moment0 = peak area (mAU.min)
func1 = C_truncated.*xx_truncated_1;
mom1(i) = trapz(xx_truncated_1,func1)./mom0(i) % moment1 = retention time (min)
func2 = (C_truncated.*((xx_truncated_1-mom1(i)).^2));
mom2(i) = trapz(xx_truncated_1,func2)./mom0(i) % moment2 = variance (min^2)
func2_15 = (C_truncated_15.*((xx_mom2_15-mom1(i)).^2));
mom2_15(i)  = trapz(xx_mom2_15,func2_15)./mom0(i)
func2_2 = (C_truncated_2.*((xx_mom2_2(i)-mom1(i)).^2));
mom2_2(i)  = trapz(xx_mom2_2(i),func2_2)./mom0(i)
N2(i)= ((mom1(i))^2/(mom2(i))) % Statistical theoretical plate number
N2_15(i) =((mom1(i))^2/(mom2_15(i)))
N2_2(i) =((mom1(i))^2/(mom2_2(i)))
Ng_16(i) = 16* (tr(i)/Width(i)).^2 % Graphically calculated theoretical plate number (1)
Ng_554(i) = 5.54* (tr(i)/Width_HW(i)).^2 % Graphically calculated theoretical plate number (2)
func3 = C_truncated.*((xx_truncated_1-mom1(i)).^3);
mom3(i) = trapz(xx_truncated_1,func3)./mom0(i) %moment 3 = skewness (min^3)
Skew(i) = mom3(i)./(mom2(i)).^(3/2)

% Moment calculations from article (Lan & Jorgenson, 2001)
theta(i) = atan(abs(tauGraphical(i))/(varianceGraphical(i))^(1/2))
```

Figure B.4: MATLAB Code ’chrom_loop_multiple’ (Third Page)

```
e0(i) = 4 + (-6.293724*theta(i)) + (9.232834*(theta(i))^2) + (-11.342910*(theta(i))^3) + (9.123978*(theta(i))^4) + (-4.173753*(theta(i))^5) + (0.827797*(theta(i))^6);
e1(i) = 0.75 + (0.033807*theta(i)) + (-0.301080*(theta(i))^2) + (1.200371*(theta(i))^3) + (-1.813317*(theta(i))^4) + (1.279318*(theta(i))^5) + (-0.326582*(theta(i))^6);
e2(i) = 1 + (-0.982254*theta(i)) + (0.568593*(theta(i))^2) + (0.512587*(theta(i))^3) + (-1.184361*(theta(i))^4) + (0.939222*(theta(i))^5) + (-0.240814*(theta(i))^6);
e3(i) = 0.5 + (-0.664611*theta(i)) + (0.706192*(theta(i))^2) + (-0.293500*(theta(i))^3) + (-0.083980*(theta(i))^4) + (0.200306*(theta(i))^5) + (-0.064264*(theta(i))^6);
e4(i) = 1 + (0.015695*theta(i)) + (-0.753849*(theta(i))^2) + (0.578422*(theta(i))^3) + (0.068108*(theta(i))^4) + (-0.019288*(theta(i))^5) + (-0.042403*(theta(i))^6);

mom1g(i) = tr(i) + tauGraphical(i)*e1(i)
mom2g(i) =e2(i)*((varianceGraphical(i)+((varianceGraphical(i))^(1/2)*tauGraphical(i)))+(tauGraphical(i))^2)
mom3g(i) = e3(i)*tauGraphical(i)*(3*varianceGraphical(i) + 4*(varianceGraphical(i))^(1/2)*tauGraphical(i)+ 4*(tauGraphical(i))^2)
N_EGH(i) = ((mom1(i))^2/(mom2g(i)))

% Tabulation of the results
fprintf(1,'Moving on to next peak\n\n');
close(3);

resulted_params = 'chrom_loop_multiple.txt';
results = fopen(resulted_params, 'a+');
fprintf(results,'Molecule\tHeight\tFlowRate\ttr\tvar\tta\tmom1\tmom2\tmom2_15\tmom2_2\tmom2g\tNg_16\tNg_554\tN2\tN2_15\tN2_2\tN_EGH\tWeight_HW\tWeight\n');
fprintf(results,'%s\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f' , ...
    Molecule, C_max, Flow_Rate, tr, varianceGraphical, tauGraphical, mom1, mom2, mom2_15, mom2_2,mom2g, Ng_16, Ng_554, N2, N2_15, N2_2, N_EGH, Width_HW, Width);
fprintf(results,'\r\r');
type chrom_loop_multiple.txt;
fclose(results)

if Number_of_peak <= 1
    fprintf(1,'Done\n\n');
end

% Calculation of resolution
for i = 1:Resolution_of_peak
if Number_of_peak > 1
Resolution(i) = (2*(tr(i+1)-tr(i)))/(Width(i+1)+Width(i))
%Resolution calculation with statistical moments.
Resolution_Statistical(i) = ((mom1(i+1)-mom1(i))/(Width(i+1)+Width(i)))
end
end
end
end
```

Figure B.5: MATLAB Code 'chrom_loop_multiple' (Fourth Page)

B.3 MATLAB Code for Calculating Chromatographic Parameters in Chromatograms that Include One Peak

```
function chrom_loop_single

% The user can enter the parts that are variable according to the analysis herself/himself at the beginning of the code.
run_time = input('Enter run time for your system:');
% The data to be used in the code to obtain the chromatographic parameters are given to the code and splined in this section.
Data = xlsread(['Acetone.5ppm.0.6.C18.xlsx']);
Flow_Rate = input('Enter the flow rate of the molecules in the column');
t = Data(:,1);
C = Data(:,2);
xx = linspace(0,run_time,10000); % (min)
splinedata = spline(t,C,xx);
minspline = min(splinedata);
splinedata_new = (splinedata - minspline);
figure(1)
plot(t,C,'b-');
xlabel('Time(min)');
ylabel('Response(mAu)');
figure(2)
plot(xx,splinedata_new,'r-');
xlabel('Time(min)');
ylabel('Response(mAu)');

Molecule = input('Enter the name of the molecule you wish to simulate','s');

disp('Using left mouse button, click on the x-coordinate of the endpoints')
disp('of the data range you wish to focus on')
disp('Left border first and right border second...')

[xx1,splinedata_new1]=ginput(2);
figure(3)
if (xx1(1)<0)
    xx1(1)=min(xx);
end

if (xx1(2)>max(xx))
    xx1(2)=max(xx);
end

dist_left = sqrt((xx-xx1(1)).^2);
[~,left_idx] = min(dist_left,[],2);
dist_right= sqrt((xx-xx1(2)).^2);
[~,right_idx] = min(dist_right,[],2);
start = xx(left_idx);
stop = xx(right_idx);
range=[start stop];
```

Figure B.6: MATLAB Code 'chrom_loop_single'(First Page)

```
% Plot data with updated range

plot(xx(left_idx:right_idx),splinedata_new(left_idx:right_idx),'ro');
    xaxis_min=start; xaxis_max=stop; yaxis_min=min(splinedata_new); yaxis_max=max(splinedata_new);
    axis([xaxis_min,xaxis_max,yaxis_min,yaxis_max])
    xlabel('Time(min)');
    ylabel('Response(mAu)');
    grid on;
hold on

C_truncated = splinedata_new(left_idx:right_idx);
xx_truncated_1 = xx(left_idx:right_idx);
[C_truncated, ind] = unique(C_truncated, 'stable');
xx_truncated_1 = xx_truncated_1(ind);

[C_max, index] = max(C_truncated); % maximum ordinate value which means peak height (H)
tr = xx_truncated_1(index); % retention time of the peak
cc = C_max*0.1; % It is calculated in order to make graphical measurements.
cc_HW = C_max*0.5; % It is calculated to find the half width of the peak and calculate the N.

RowMax = find(C_truncated==C_max); % Calculate row where C_max falls.
C_End = C_truncated(end);
RowEnd = find(C_truncated==C_End); % Calculate row where C_end falls.
C_Low = C_truncated(1:RowMax); % Concentration values lower than C_max
C_High = C_truncated(RowMax:RowEnd); % Concentration values higher than C_max
t_L = xx_truncated_1(1:RowMax); % The time interval at which concentration values less than C_max coincide.
t_R = xx_truncated_1(RowMax:end); % The time interval at which concentration values higher than C_max coincide.
% Time range limits corresponding to a height of 0.15 of C_max.
t_Low = interp1(C_Low, t_L, cc, 'linear', 'extrap');
t_High = interp1(C_High, t_R, cc);

% Below calculations are for finding the range that mom2 is calculated.
Aalfa_mom2 = abs(tr - start);
Aalfa_left = tr - Aalfa_mom2.*1.5;
Aalfa_left_2 = tr - Aalfa_mom2.*2;
Balfa_mom2 = abs(stop - tr);
Balfa_right = tr + Balfa_mom2.*1.5;
Balfa_right_2 = tr + Balfa_mom2.*2;
Width_mom = Aalfa_mom2 + Balfa_mom2;

xx_mom2_15 = linspace(Aalfa_left,Balfa_right,10000);
C_truncated_15 = spline(xx,splinedata_new,xx_mom2_15);
xx_mom2_2 = linspace(Aalfa_left_2,Balfa_right_2,10000);
C_truncated_2 = spline(xx,splinedata_new,xx_mom2_2);

% Below calculations are for calculating the widht at one-half peak height.
```

Figure B.7: MATLAB Code ’chrom_loop_single’(Second Page)

```
t_low_HW = interp1(C_Low, t_L, cc_HW, 'linear', 'extrap');
t_High_HW = interp1(C_High, t_R, cc_HW);
Aalfa_HW = abs(tr - t_low_HW);
Balfa_HW = abs(t_High_HW - tr);
Width_HW = Aalfa_HW + Balfa_HW; % Widht at one-half peak height


%Graphical measurements for chromatographic parameters.
Aalfa = abs(tr - t_Low);
Balfa = abs(t_High - tr);
varianceGraphical = ((-1/(2*log(0.10)))*Aalfa*Balfa); % variance calculated from the obtained peak from t, C graphical data.
tauGraphical = ((-1/log(0.10))*(Balfa-Aalfa)); % time constant calculated from the obtained peak from t, C graphical data.
AssymetryLevel = (tauGraphical)/(varianceGraphical)^(1/2);
PeakAssymetryFactor = Balfa/Aalfa;
theta = atan((abs(tauGraphical)/(varianceGraphical)^(1/2))); %theta the is absolute bounded assymetry
Width = Aalfa + Balfa;
Tailing_Factor = (Aalfa_mom2+Balfa_mom2)./(2.*(Aalfa_mom2));

% Moment calculations for chromatographic parameters.
mom0 = trapz(xx_truncated_1,C_truncated); % moment0 = peak area (mAU.min)
func1 = C_truncated.*xx_truncated_1;
mom1 = trapz(xx_truncated_1,func1)./mom0; % moment1 = retention time (min)
func2 = (C_truncated.*((xx_truncated_1-mom1).^2));
mom2 = trapz(xx_truncated_1,func2)./mom0; % moment2 = variance (min^2)
func2_15 = (C_truncated_15.*((xx_mom2_15-mom1).^2));
mom2_15  = trapz(xx_mom2_15,func2_15)./mom0;
func2_2 = (C_truncated_2.*((xx_mom2_2-mom1).^2));
mom2_2  = trapz(xx_mom2_2,func2_2)./mom0;
N2= ((mom1)^2/(mom2)); % Statistical theoretical plate number
N2_15 =((mom1)^2/(mom2_15));
N2_2 =((mom1)^2/(mom2_2));
Ng_16 = 16* (tr/Width).^2; % Graphically calculated theoretical plate number (1)
Ng_554 = 5.54* (tr/Width_HW).^2; % Graphically calculated theoretical plate number (2)
func3 = C_truncated.*((xx_truncated_1-mom1).^3);
mom3 = trapz(xx_truncated_1,func3)./mom0; %moment 3 = skewness (min^3)
Skew = mom3./(mom2).^(3/2);

% Moment calculations from article (Lan & Jorgenson, 2001)
theta = atan(abs(tauGraphical)/(varianceGraphical)^(1/2));
e0 = 4 + (-6.293724*theta) + (9.232834*(theta)^2) + (-11.342910*(theta)^3) + (9.123978*(theta)^4) + (-4.173753*(theta)^5) + (0.827797*(theta)^6);
e1 = 0.75 + (0.033807*theta) + (-0.301080*(theta)^2) + (1.200371*(theta)^3) + (-1.813317*(theta)^4) + (1.279318*(theta)^5) + (-0.326582*(theta)^6);
e2 = 1 + (-0.982254*theta) + (0.568593*(theta)^2) + (0.512587*(theta)^3) + (-1.184361*(theta)^4) + (0.939222*(theta)^5) + (-0.240814*(theta)^6);
e3 = 0.5 + (-0.664611*theta) + (0.706192*(theta)^2) + (-0.293500*(theta)^3) + (-0.083980*(theta)^4) + (0.200306*(theta)^5) + (-0.064264*(theta)^6);

mom1g = tr + tauGraphical*e1;
mom2g = e2*((varianceGraphical+((varianceGraphical)^(1/2)*tauGraphical))+(tauGraphical)^2);
mom3g = e3*tauGraphical*(3*varianceGraphical + 4*(varianceGraphical)^(1/2)*tauGraphical+ 4*(tauGraphical)^2);
```

Figure B.8: MATLAB Code 'chrom_loop_single'(Third Page)

```
N_EGH = ((mom1)^2/(mom2g));

%Tabulation of results.
fprintf(1,'Moving on to next peak\n\n');
close(3);


resulted_params = 'calc_params_single_mom2_den.txt';
results = fopen(resulted_params, 'a+');
fprintf(results,'Molecule\tHeight\tFlowRate\ttr\tvar\ttau\tmom1\tmom2\tmom2_15\tmom2_2\tmom2g\tNg_16\tNg_554\tN2\tN2_15\tN2_2\tN_EGH\tWidth_HW\tWidth\tWidth_mom\n');
fprintf(results,'%s\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f\t%7.4f' , ...
    Molecule, C_max, Flow_Rate, tr, varianceGraphical, tauGraphical, mom1, mom2, mom2_15, mom2_2,mom2g, Ng_16, Ng_554, N2, N2_15, N2_2, N_EGH, Width_HW, Width, Width_mom);
fprintf(results,'\r\r');
type calc_params_single_mom2_den.txt;
fclose(results);
```

Figure B.9: MATLAB Code 'chrom_loop_single'(Fourth Page)

B.4 MATLAB Code for Predicting Chromatograms from Chromatographic Parameters

```
function testread
clc;
clear all;
A=readtable('calc_params_single_mom2_den.txt')
TotalRow = height(A);
MoleculeName=A.Molecule
Chromatogram = input('Enter how many different chromatogram data you want to get')
Resolution_of_peak = input('Enter number of resolution value you want to calculate')
for i = 1:Chromatogram
Data = input('Enter the name of the molecule that is wanted to obtain chromatographic parameter data')
row(i) = find(strcmp(MoleculeName, Data) == 1)
Height(i) = A{row(i),2};
FlowRate(i) = A{row(i),3};
RetentionTime(i) = A{row(i),4}
VarianceGraphical(i) = A{row(i),5};
TauGraphical(i) = A{row(i),6};
Moment1(i) = A{row(i),7};
Moment2(i) = A{row(i),8};
NStatistical(i) = A{row(i),9};
Width(i) = A{row(i),19}
Width_mom(i) = A{row(i),20}

end
for i = 1:Resolution_of_peak
if Chromatogram > 1
Resolution(i) = (2*abs((RetentionTime(i+1)-RetentionTime(i))))/(Width(i+1)+Width(i))
Resolution_Statistical(i) = (abs((Moment1(i+1)-Moment1(i)))/(Width(i+1)+Width(i)))
end
end
Overlap_Chromatograms = input('Do you want to overlap the peaks? Type 1 for "yes" or 0 for "no"')
if Overlap_Chromatograms == 1
    t = linspace(0,10,1000);
    C = zeros(size(t));
    for i = 1:Chromatogram
        for a = 1:length(t)
            cond = ((2*VarianceGraphical(i))+TauGraphical(i).*(t(a)-RetentionTime(i)));
            if  cond > 0
                C(a,i) = Height(i).*exp((-((t(a)-RetentionTime(i)).^2)./((2.*VarianceGraphical(i))+TauGraphical(i).*(t(a)-RetentionTime(i)))));
            elseif cond <= 0
                C(a,i) = 0;

            end

        end
figure(1)
whos C
plot(t,C)
xlabel('Time(min)');
ylabel('Response(mAu)');
    end
end
```

Figure B.10: MATLAB Code 'test_read'

Appendix C

VAN DEEMTER CURVES

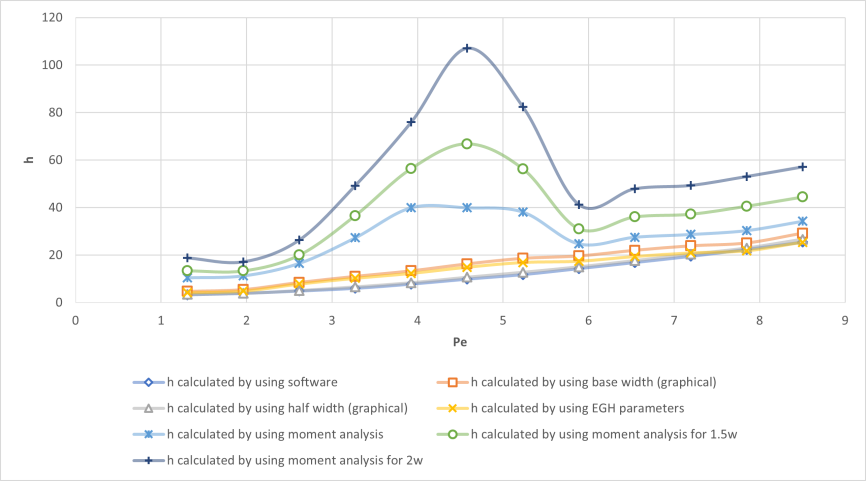

Figure C.1: Comparison of van Deemter curves obtained with different calculation methods for Dimethylformamide in C18 column at 25°C)

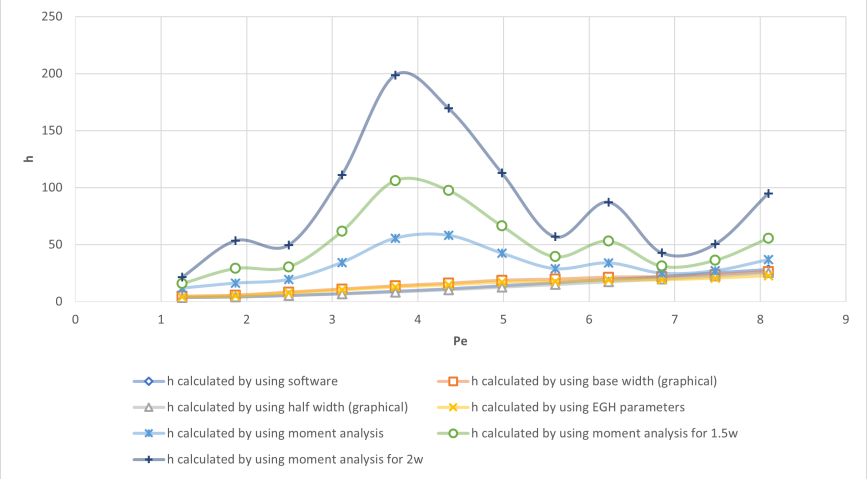

Figure C.2: Comparison of van Deemter curves obtained with different calculation methods for Dimethylformamide in C18 column at 40°C)

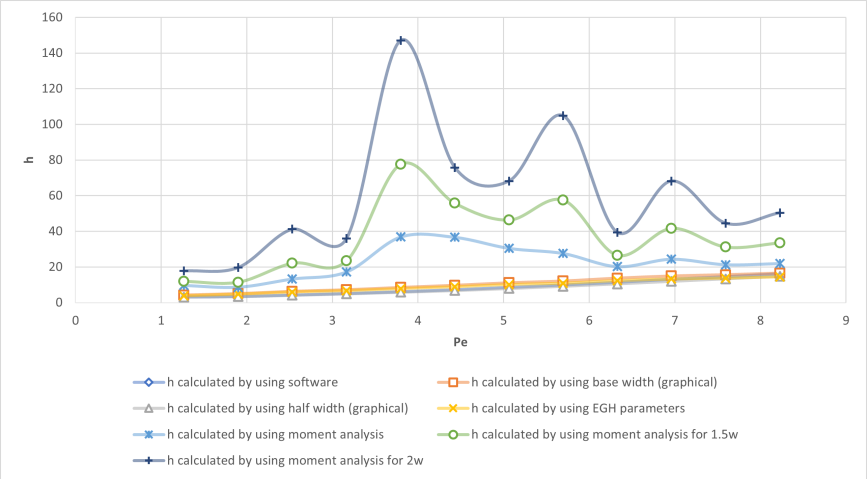

Figure C.3: Comparison of van Deemter curves obtained with different calculation methods for Dimethylformamide in C8 column at 35°C)

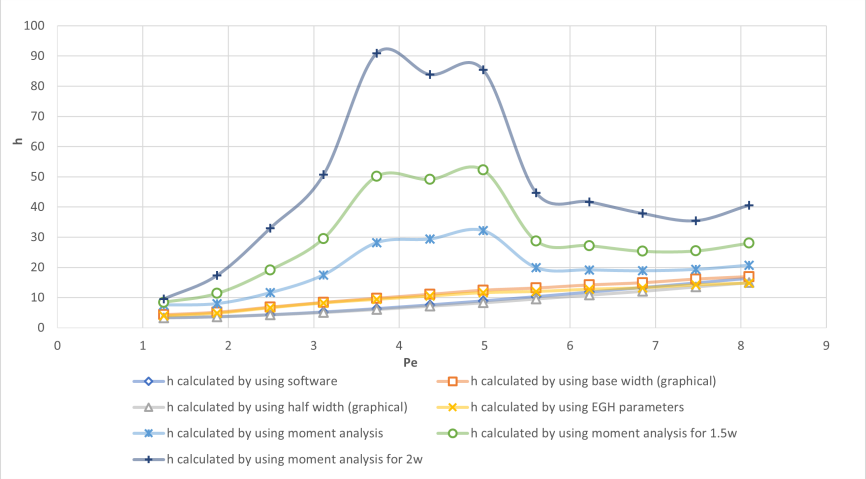

Figure C.4: Comparison of van Deemter curves obtained with different calculation methods for Dimethylformamide in C8 column at 40°C)

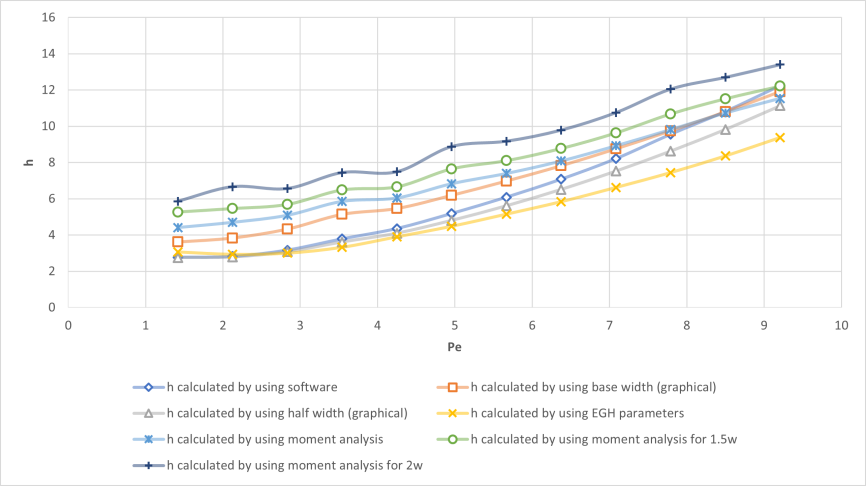

Figure C.5: Comparison of van Deemter curves obtained with different calculation methods for Ethyl Acetate in C18 column at 25°C)

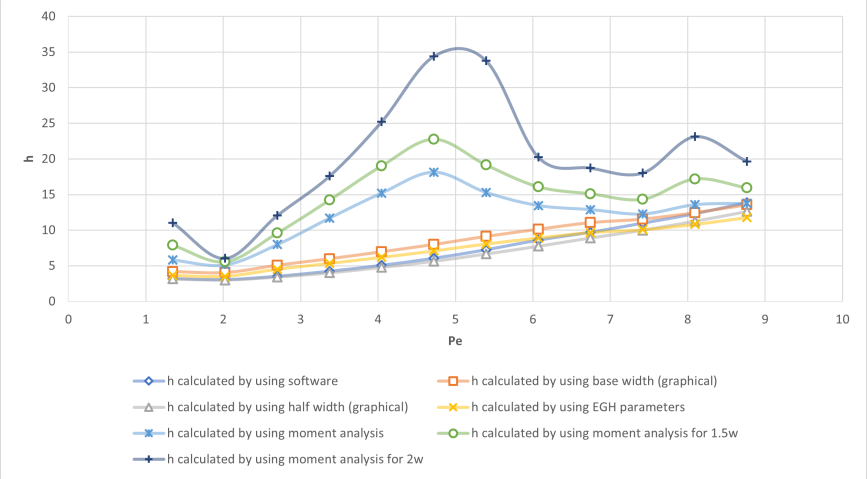

Figure C.6: Comparison of van Deemter curves obtained with different calculation methods for Ethyl Acetate in C18 column at 40°C)

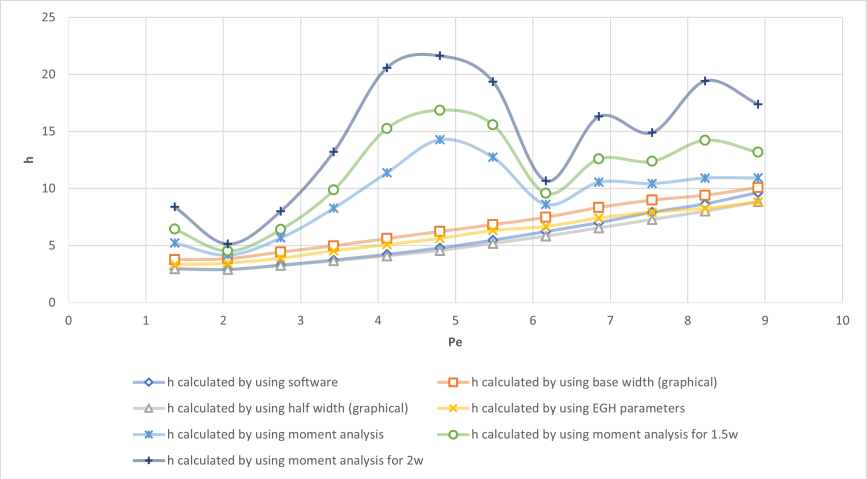

Figure C.7: Comparison of van Deemter curves obtained with different calculation methods for Ethyl Acetate in C8 column at 35°C)

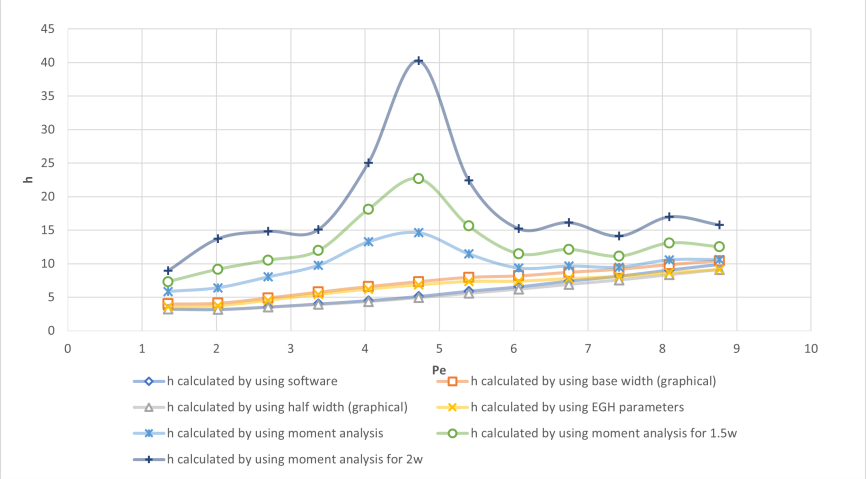

Figure C.8: Comparison of van Deemter curves obtained with different calculation methods for Ethyl Acetate in C8 column at 40°C)

Appendix D

EFFECT OF RETENTION FACTOR ON SEPARATION EFFICIENCY

$$\mathbf{h} = \frac{\mathbf{2}}{\mathbf{Pe}} + \frac{\mathbf{(1-x)^2}}{\mathbf{3(1-\epsilon_b)}}\mathbf{ReSc}(\frac{\mathbf{1}}{\mathbf{Nu}} + \frac{\mathbf{m}}{\mathbf{10}} + \frac{\mathbf{m^*}}{\mathbf{Da}}) \tag{D.1}$$

$$\mathbf{Pe} = \frac{\mathbf{2Rv}}{\mathbf{D}} \tag{D.2}$$

where R is the average radius of the stationary phase particles used to packed column [16]

$$\mathbf{Nu} = \frac{\mathbf{2Rk_c}}{\mathbf{D_m}} \tag{D.3}$$

$$\mathbf{Da} = \frac{\mathbf{R^2 k_a}}{\mathbf{D_p}} = \frac{\mathbf{adsorptive uptake}}{\mathbf{intraparticle pore diffusion}} \tag{D.4}$$

$$\mathbf{Re} = \frac{\mathbf{2Rv\epsilon_b\rho_b}}{\mu} = \frac{\mathbf{inertial force}}{\mathbf{viscous force}} \tag{D.5}$$

$$\mathbf{Sc} = \frac{\mu}{\mathbf{\rho_b D_m}} = \frac{\mathbf{diffusivity of momentum}}{\mathbf{mass diffusivity}} \tag{D.6}$$

$$\mathbf{ReSc} = \frac{\mathbf{2Rv\epsilon_b}}{\mathbf{D_m}} = \mathbf{reduced velocity} \tag{D.7}$$

$$\mathbf{m} = \frac{\mathbf{D_m}}{\mathbf{\epsilon_p^* D_p}} = \frac{\mathbf{D_m}}{\mathbf{\epsilon_p^* D_m/2.5}} = \frac{\mathbf{2.5}}{\mathbf{\epsilon_p^*}} \tag{D.8}$$

$$\mathbf{m^* = \frac{3m}{2}(\frac{K_D}{1+K_D})^2} \quad \text{(D.9)}$$

where

k_a = forward rate constant of adsorption, 1/s

K_D = equilibrium adsorption constant for ratio of mass of solute adsorbed to its concentration in the pore liquid

k_c = concentration-based fluid mass transfer coefficient , cm/s

v = interstitial fluid velocity, cm/s

x = fraction of solute in the mobile phase at long times

ρ_b = density of mobile phase, g/ml

μ = viscosity of mobile phase, g/(cm.s)

ϵ_p^* = intraparticulate void fraction

D_m = effective solute diffusivity in unbounded solution (mobile phase), cm^2/s

D_p = effective solute diffusivity in pore fluid, cm^2/s