To the Graduate Council:

I am submitting herewith a thesis written by Eloise Marie Joyce entitled “Process Mass Spectrometric Parameterization of Butene Isomer Mixtures.” I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science in Chemistry.

Kelsey D. Cook
Major Professor

We have read this thesis
and recommend its acceptance:

Frank Vogt

Robert Compton

Accepted for the Council:

Carolyn R. Hodges
Vice Provost and Dean of the Graduate School

(Original signatures on file with official student records)
Process Mass Spectrometric Parameterization of Butene Isomer Mixtures

A Thesis

Presented for the

Masters of Science Degree in Chemistry

at

The University of Tennessee, Knoxville

Eloise Marie Joyce

August 2007
ABSTRACT

Process mass spectrometry was used for the simultaneous quantitation of hydrocarbon mixtures with either six or four components. The four component hydrocarbon mixture contains isomers to increase the complexity of the mixture spectrum. The differences in relative intensities of pure component electron ionization mass spectra provided a basis for quantitation. Quantitation accuracy and precision were found to decrease as the spectral similarities among the components increased.

Selection of which ions to monitor (parameterization) was critical to optimize the analysis accuracy, precision, and speed for all mixtures tested. An empirical parameterization algorithm based on comparison of pure reference spectra of mixture components was developed for ion selection. A parameterization which monitors all the masses was used as a basis for comparison. The empirical algorithm parameterization gave analysis accuracy and precision statistically equal to the all-mass parameterization. Empirical algorithm square matrices (each component is assigned a single ion) were compared to the square matrices that were determined by the software provided with the mass spectrometer.

The six component hydrocarbon mixture which contained methane, ethane, propane, propene, isobutane, and isobutene had an $r^2$ value of 0.99920 ± 0.00007, an average correlation coefficient value ($<CV>$) of 0.43% ± 0.051, and a root mean squared error ($E_{RMS}$) of 0.43 when analyzing all masses in the spectra. Using the empirical algorithm and selecting 35% of the masses in the spectra, the accuracy and precision are statistically similar with $r^2 = 0.99913 ± 0.00001$, $<CV> = 0.49% ± 0.59$, and $E_{RMS} = 0.47$. 
A more complex mixture consisting of the four butene isomers also gave similar results. When analyzing all of the masses in the butene isomers spectra the figures of merit were $r^2 = 0.986 \pm 0.007$, $<CV> = 5.54\% \pm 6.66$, and $E_{RMS} = 1.86$. When using the empirical algorithm and selecting 26% of the masses in the spectra, that accuracy and precision were statistically similar with $r^2 = 0.983 \pm 0.008$, $<CV> = 6.20\% \pm 7.38$, and $E_{RMS} = 2.03$. As shown, the empirical algorithm successfully chooses a portion of the mass spectra with excellent precision and accuracy while decreasing analysis time.
## TABLE OF CONTENTS

CHAPTER 1  PROCESS MASS SPECTROMETRIC ANALYSIS................................. 1
  1.1  Introduction to Process Analytical Chemistry .................................................... 1
  1.2  Mass Spectrometry Theory ................................................................................. 3
    1.2.1  Sampling ..................................................................................................... 3
    1.2.2  Ionization .................................................................................................... 6
    1.2.3  Mass Analyzers........................................................................................... 7
    1.2.4  Detectors ..................................................................................................... 8
    1.2.5  Data Analysis and Chemometrics............................................................... 9
  1.3  Parameterization ............................................................................................... 10
  1.4  Using Process Mass Spectrometry.................................................................... 11
  1.5  Applications of Process Mass Spectrometry..................................................... 12
  1.6  Summary and Objectives .................................................................................. 14

CHAPTER 2  EXPERIMENTAL.................................................................................... 15
  2.1  Instrumentation and Procedures........................................................................ 15
  2.2  Gas mixture preparation.................................................................................... 17
  2.3  Concentration Calculations............................................................................... 22
  2.4  Parameterization Computations........................................................................ 24

CHAPTER 3  DEVELOPMENT OF A PARAMETERIZATION ALGORITHM ...... 26
  3.1  Background....................................................................................................... 26
  3.2  Algorithm Improvements.................................................................................. 30
LIST OF FIGURES

Figure 1.1 Diagram of molecule sampling, ionization, mass analysis, and detection [7]. . 4

Figure 2.1 Direct capillary or *Quick Inlet* to the process mass spectrometer. .................. 16

Figure 2.2 Mass flow controller setup. ...................................................................................... 21

Figure 3.1 A plot of the pure component reference spectra for the isobutene stream
(methane, ethane, propane, propene, isobutene, and isobutane).......................... 35

Figure 3.2 “Validation plot” of calculated versus actual concentration for 32 mixtures of
methane (■), ethane (X), propane (Δ), propene (◇), isobutene (●), and isobutane
(▲) using CLS with all-mass parameterization......................................................... 37

Figure 3.3 Plot of Stein similarity index using the ranked masses of the isobutene stream
shown in Table 3.3................................................................................................ 41

Figure 4.1 Chemical structures for 1-butene, isobutene, cis-2-butene, and trans-2-butene.
........................................................................................................................................ 48

Figure 4.2 Pure component reference mass spectra for all four butene isomers (1-butene,
isobutene, cis-2-butene, trans-2-butene)....................................................................... 49

Figure 4.3 Validation plot for mixtures of cis-2-butene (■), trans-2-butene (X), isobutene
(Δ), and 1-butene (♦) using all 19 peaks with >2% relative abundance. ............... 52

Figure 4.4 Plot of Stein similarity index using the ranked masses of the butene isomer
stream shown in Table 4.1 .......................................................................................... 55
LIST OF TABLES

Table 2.1 List of gases, purities, and suppliers ................................................................. 18

Table 2.2 Certified flow ranges for the mass flow controllers used along with the
calibration gas for the controller .................................................................................. 19

Table 2.3 List of sensitivities used for the mass flow calibration................................. 20

Table 3.1 Approximate concentration ranges for the six components in the isobutene
sample stream ............................................................................................................. 33

Table 3.2 Pure component reference spectra intensity values for isobutene stream
(methane, ethane, propane, propene, isobutene, and isobutane)............................... 36

Table 3.3 Empirical algorithm and similarity index scores for the pure component spectra
in Table 3.2 for the isobutene stream ......................................................................... 40

Table 3.4 Ions selected and figures of merit obtained using the empirical algorithm and
Extrel’s 'Stream Evaluator' parameterizations for triplicate measurements of 32
isobutene stream samples .......................................................................................... 44

Table 4.1 Pure component reference spectra intensity values for butene isomers (1-
butene, isobutene, cis-2-butene, and trans-2-butene) ............................................. 50

Table 4.2 Approximate concentration ranges for the four components in the butene
isomer mixtures ........................................................................................................... 53

Table 4.3 Empirical algorithm and similarity index scores for the pure component
spectra in Table 4.1 for the butene isomers ............................................................... 54

Table 4.4 Ions selected and figures of merit obtained using the empirical algorithm and
Extrel's 'Stream Evaluator' parameterizations for triplicate measurements of eight
butene isomer samples ................................................................................................. 57
Table 7.1 Percentage sample compositions for mixtures of the isobutene stream. ........ 65

Table 7.2 Percentage sample compositions for mixtures of the butene isomers. ............ 66

Table 7.3 Percentage sample compositions for small set of mixtures of the butene isomers. .............................................................................................................................................................. 67
CHAPTER 1

PROCESS MASS SPECTROMETRIC ANALYSIS

1.1 Introduction to Process Analytical Chemistry

Process analysis is a growing field targeted on improving efficiency of industrial processes and product quality by analyzing sample streams throughout the process in real-time [1]. With the continuing competition amongst chemical companies, it is essential for their production processes to become increasingly efficient and profitable. To operate efficiently and profitably, these companies must monitor and control their processes in one of two ways: off-line data acquisition or real-time, on-line data analysis. The latter is more time efficient and can give information about the process sample stream at any given time whereas the former allows for the data to be taken from the industrial plant floor to a controlled environment and studied by a trained chemist.

Better information about a process leads to better process control to increase the overall efficiency, which can be measured in terms of product yield, product quality, utilization of resources, composition of process discharge, or process safety [2]. If the processes are being monitored on-line, any potential problems may be remedied more quickly than if the data must be analyzed later. For example, if a product’s quality is drifting to an unacceptable range, an adjustment to the production process can be made to
correct the problem before more bad product is made or in time to compensate downstream (this is called ‘feed forward’). This prevents costly disposal of products that are below market standards and avoids the use of more chemicals to replace the bad product. One way of obtaining information about the chemical process is to use inexpensive, real-time sensors that can measure the temperature, pressure, and flow of a sample stream to infer the chemical composition of the mixture. While these measurements are useful, they lack more chemical-specific information that can be gained by using process analyzers.

There are several excellent reviews on the theory and applications of process analytical chemistry [1] [3] [4]. The analytical fields spanned include chromatography, spectroscopy, biosensors, and electrochemistry. The variety of techniques in process analytical chemistry is due to the varying needs in current industrial processes and the improvements of the current chemical instrumentation. Most recently, process analytical chemistry is expanding into the pharmaceutical industry for drug discovery and better understanding of metabolites. Process instrumentation is also heading toward miniaturization for speed and portability while still using a high-performance instrument [5].

Many challenges accompany process analysis as opposed to lab analysis. Process chemistry is usually performed in the non-sterile environment on the plant floor in industry. Therefore the process analyzers need to be robust and simple enough for any operator to use. Also, industries need to be able to do quick analysis that is precise and reliable to stay competitive with other companies. Process analysis is often more concerned with detecting and correcting for small changes in concentration in the
chemical process than knowing its true absolute concentration values. Therefore precision is critical in all steps of the analysis of the sample stream.

1.2 Mass Spectrometry Theory

The mass spectrometric monitoring of a chemical process involves the sampling of an analyte, the ionization of that analyte, the mass analysis of the ions produced, and the detection of those ions (an example is shown in Figure 1.1). The mass spectrum of a compound is simply a plot of the intensity versus the mass-to-charge ratio (m/z) of ions detected from the ionization of the compound [6]. Each chemical process and sample has its own unique conditions and chemistry so there are many options available for analyte sampling, ionization, mass analysis, and detection (see subsequent discussion below).

1.2.1 Sampling

The process MS sampling system must continuously deliver a sample that is representative of the process to the inlet of the instrument. Process MS is therefore utilized most often to analyze gaseous analytes from volatile or semi-volatile samples [8]. Typically, a sample of gas is extracted from a process stream via a transfer line to the sampling valve of the mass spectrometer. This can be a single sampling valve or more often a multi-position sampling valve that can select any of
Figure 1.1 Diagram of molecule sampling, ionization, mass analysis, and detection [7].
several transfer lines connected to the inlet. The advantage of using a multi-position valve is that it allows for the analysis of one or more calibration gases and multiple related test analytes thereby exploiting the analysis speed to enable multiplexing.

Liquid streams can also be analyzed using process MS. The liquid stream must either be vaporized completely or sampled from the headspace over a volatile liquid sample. This requires the heating of the MS sampling system to a temperature high enough for complete vaporization but not so high as to cause thermal decomposition of the sample. The high temperature will have to remain steady to insure that the liquid sample will not recondense before reaching the inlet and the pressure will not fluctuate excessively.

After the sample passes through the sampling valve it can be introduced into the instrument either through a direct or membrane inlet. Direct inlets typically have a micron-sized pinhole aperture or a 10-200 µm i.d. silica sampling capillary that is long enough to limit flow and to maintain low ion source (∼< 1 Torr) and analyzer (≤ 10⁻⁵ Torr) pressure [9]. This allows a small portion of the sample stream to enter directly into the low-pressure ion source while the rest of the sample stream is recycled back to the process or directed to a waste vent. A filter may be placed before the sample valve to keep particulates that may be present in the sample from entering and clogging the inlet. Membrane inlets can be used to enhance selectivity and sensitivity over the direct inlet system [10]. Typical membrane materials include silicone rubber, polypropylene, and Teflon™ for aqueous organics. A membrane is used to selectively omit portions of the sample stream, such as solvent molecules in a liquid matrix. This would enhance sensitivity to the materials of interest and reduce background contributions.
1.2.2 Ionization

Electron ionization (EI) is the simplest and most common method for ionizing samples in process MS. In EI, gaseous analyte molecules interact with electrons from a heated wire filament, such as tungsten or rhenium. These electrons are typically accelerated to 70-100 eV which causes the molecules to fragment extensively (most molecules only require about 5-20 eV to fragment). Operating above the threshold energy produces high precision intensity measurements and a spectrum that is relatively rich in fragment ions. The fragmentation pattern is distinct for each compound which enhances the informational content of the mass spectrum. However, if more than one component is being ionized at once, there may be some that give ions with the same mass-to-charge ratios. As the number of ion contributions at a given mass-to-charge ratio increases, interference and/or loss of sensitivity can result.

Chemical ionization (CI) is an alternative ionization technique in process MS though it is not as widely used due to lack of precision. The gaseous analyte interacts with ions present in the source causing ionization of the sample molecules but at a lower energy than EI causing less fragmentation. Other ionization methods that have been utilized include hyperthermal surface ionization, microwave plasma ionization, and nanoelectrospray [11]. Hyperthermal surface ionization has been paired up with laser desorption to allow for the ionization of nonvolatile samples [12]. Microwave plasma ionization has been used for the on-line detection of xenon and krypton isotopes while using both positive and negative ion modes [13]. Nanoelectrospray has been used for the
analysis of protein chemical reactions within the spray plume. The protein reactions are studied with the solvent changes that occur throughout the spray time [14].

1.2.3 Mass Analyzers

Quadrupole and magnetic sectors are the most common types of mass analyzers used in process mass spectrometry. The quadrupole mass analyzer consists of four cylindrical metal rods about 1 cm in diameter and 20 cm in length arranged in a parallel array at the four corners of a square (see Figure 1.1). The four metal rods act as electrodes with opposing rods connected electrically. One pair is attached to the positive side of a variable DC source while the other pair is connected to the negative side. Additionally, a variable RF potential is applied to each pair of rods 180° out of phase. Variations in the DC and RF voltages select the ions to be monitored while eliminating the rest [15]. The quadrupole selects the ions specifically due to their mass-to-charge ratios as opposed to their kinetic energy. Their time crossing through the analyzer is short compared to the time necessary for the mass being monitored to be changed. This makes the quadrupole an ideal mass analyzer for process MS.

The magnetic sector mass analyzer disperses the ions that have been accelerated to keV kinetic energies through a curved tube using a magnetic field perpendicular to the direction in which the ions are moving. The radius of the ion path curvature through the tube can then be related to the ion’s mass-to-charge ratio. Regardless of which mass analyzer is used, the selection of which ions to monitor is critical to the analytical
performance and will be discussed in detail in subsequent chapters. The ions of selected mass-to-charge ratio leaving the mass analyzer next enter the detector.

1.2.4 Detectors

The most common detectors for process MS instruments are a Faraday cup and a secondary electron multiplier. A Faraday cup is a conductive metal plate or cup that is connected through a high-impedance resistor to ground. As the ions come in contact with the plate, they become neutralized which creates a current. This current passes through the resistor creating a voltage which can be monitored as the signal. The Faraday cup has unit gain meaning there is a one-to-one relationship between the ionic charge and the neutralizing charge. This is an extremely simple detector that is low cost, accurate, low noise, and has no mass bias. However, the unit gain of the Faraday cup limits its dynamic range and sensitivity.

In a secondary electron multiplier, the incoming ions come in contact with an electro-emissive metal surface which causes secondary electrons to be emitted. The secondary electrons are accelerated and collide with additional electro-emissive surfaces causing an electron cascade. The electron cascade can cause a gain of up to $10^8$ electrons from one ion (the exact number of electrons gained depends on the velocity of the ion which may prove problematic with respect to mass bias), which lowers the limit of detection for the instrument [6]. The gain of secondary electron multipliers will deteriorate over time requiring recalibration or replacement. However, secondary
electron multipliers are often used because they extend the dynamic range as compared to the Faraday cup.

1.2.5 Data Analysis and Chemometrics

Data analysis is an extremely important part of process analysis because it must be quick enough to provide real-time control over the stream. Regardless of the detector, three independent experimental parameters influence the analysis time: the integration time for each signal measurement, the number of replicate measurements for each ion monitored, and the number of ions chosen to be monitored. The former two parameters obviously influence the balance between precision and analysis time. The latter will be discussed in Chapter 1.3.

Once the data have been collected, they need to be related to the chemical composition of the process stream. It should be noted that with the advances of modern computers, the data collection usually takes longer than the data workup. Chemometrics is defined as the application of mathematical and statistical techniques to the measurements to derive chemical information in scientific and industrial processes [16]. Since the relationship between measurements and their processes is so critical, chemometrics is an essential part in the success of process analytical chemistry. There are several chemometric tools that can be used to calculate and extract compound concentrations from the intensity data of mass spectra, two examples being PLS (partial least squares) regression and PCR (principle component regression) [17].
1.3 Parameterization

Simultaneous analysis requires monitoring at least one ion signal in a spectrum for each component of interest. For example, if the sample stream contains four components, then at least four different ion signals need to be monitored for simultaneous analysis. Accuracy and/or precision can be improved by monitoring more than the minimum number of ion signals. However, monitoring too many ion signals can decrease the precision by the addition of more noise without adding any more information. The data acquisition time will increase with the increase of ion signals being monitored which will also increase analysis time. Therefore, there must be an optimum number of mass-to-charge ratios in the spectrum to monitor that will give the best accuracy and precision for the analysis.

Most process MS parameterizations are determined by using algorithms developed by instrument manufacturers to be used with their instrument or else the individual operator will use experience and intuition to select the ions to be monitored. There are several factors the operator or algorithm will take into consideration when selecting masses such as signal intensity, concentration of the components, spectral interference (more than one component contributing to the signal at a given m/z) among the components, and the required precision. Additional approaches for the selection of ions in process MS other than instrument algorithms and operator intuition include various other algorithms such as brute force and genetic algorithms which will be discussed in Chapter 3.1 [18].
Parameterization methods can be classified into two types: those that use a sampling of data to determine the best parameterization based on quantitation performance and those that rank ions using pure component spectra with mathematical formulas to determine the best parameterization. The brute force and genetic algorithms require a data set of mixtures with known composition to test the effectiveness of different parameterizations to determine the optimum parameterization. Empirical ranking algorithms require only the pure component mass spectra to determine the optimum parameterization. An empirical ranking algorithm will be discussed in depth in chapters 3 and 4.

1.4 Using Process Mass Spectrometry

Once the method of quantitation and parameterization has been chosen, routine operation of the mass spectrometer can proceed. Tuning of the instrument is done by the operator (typically daily) to maximize the signal intensity and obtain the optimal peak shape. Rarely can both of these occur at the same time, so the operator will determine the best peak shape and relative intensity for best performance. Auto-tune features may exist for automatic operation, but these features may have limited capabilities such that either the intensity can be maximized or the peak shape can be optimized. Calibration of the instrument can be automated, but periodic maintenance (such as oil changes and source cleanings) does require hands-on operation. The interval between calibrations is dependent on the process and the instrument itself. Intensity calibration may be needed
to ensure that the intensities do not deviate much from day to day. The need for intensity calibration can be reduced by using relative intensity values (normalized so that $I_{\text{max}} = 100$) as opposed to absolute intensity values which are prone to drift. When calibration problems occur, the remedy could be as simple as recollecting reference spectra or could be more difficult with thorough cleaning of the instrument source.

### 1.5 Applications of Process Mass Spectrometry

Through recent reviews [1] [3] [4], the various uses for process mass spectrometric analysis becomes evident. The range of applications include drug discovery in pharmaceutical preparation, studies involving isotopes, on-line trace contamination detection, and combining process MS with lab-on-a-chip.

In the growing pharmaceutical industry, process mass spectrometry is being used to study the chemical structures of metabolites that are being synthesized in a variety of \textit{in vitro} systems. Such studies compare the structures of metabolites produced to database information and the resultant structures of other production methods. The metabolites are synthesized \textit{in vitro} and then separated by liquid chromatography. The mass spectrometer used [19] has an electrospray ionizer and various types of mass analyzers. The mass spectrometric data collected is used to determine the structure of the metabolites that were synthesized and confirmed by nuclear magnetic resonance (NMR). The understanding of the chemical structures of metabolites is critical to the understanding of the metabolic pathways in the human body [19].
Mass spectrometry is well known for its studies involving isotopes due to the characteristic isotopic pattern of ions in the mass spectrum. A recent isotopic study was used to evaluate methods of extracting mercury from soils and sediments and to compare these extraction methods for mercury with EPA Method 6800. The extracted mercury was usually present in two forms: inorganic ions ($\text{Hg}^{2+}$) and methylmercury ($\text{CH}_3\text{Hg}^+$). The extraction method determined in which form the mercury was found and whether it had been converted from one form to the other [20].

Another application for process MS is the on-line analysis of trace contamination and chemical species. In the study by Kingston et al. [21], they specifically study trace contamination in the semiconductor industry for automated analysis. The automated process is used to decrease manufacturing costs and process times, decrease dependence on non-product testing, and increase equipment utilization and effectiveness. To do this, the automated process must quantitatively analyze ppt samples of elemental species, ligands, and organic molecules [21].

Like many other areas of chemistry, process MS instrumentation is miniaturizing as a promising way to decrease time and cost. Lab-on-a-chip is an expanding form of miniaturization in the analytical field which allows processes to be portable. In a recent study, a lab-on-a-chip device was connected with a matrix assisted laser desorption ionization (MALDI) time-of-flight (TOF) mass spectrometer as an analytical screening system. The chip is fabricated on a standard MALDI sample plate and inserted into the sample chamber. On the chip, a chemical/biochemical reaction occurs and the products can then enter directly into the MALDI-TOF MS. These chips are small which keeps the sample size and cost down [22].
1.6 Summary and Objectives

The objective of this research is to design and test the performance of an empirical ranking algorithm for application to a mixture of isomers that have fragments at the same mass-to-charge ratios. Mixtures will be quantitated and results evaluated in terms of analysis accuracy and precision when compared with known mixture compositions. The influence of the number of mass-to-charge ratios will be addressed and the results of the empirical algorithm will be compared to other parameterizations discussed in subsequent chapters.
CHAPTER 2

EXPERIMENTAL

2.1 Instrumentation and Procedures

An ABB-Extrel (Pittsburgh, PA) Questor IV quadrupole process mass spectrometer (MS), with a scan range from mass-to-charge (m/z) 1-250, was used for sample analysis in all experiments. The samples were introduced through an Extrel Quick Inlet system (a T-splitter, Figure 2.1) which directs ~0.04% of the gas mixture through a 25µm i.d., 17 cm long fused silica capillary (Polymicro, Phoenix, AZ) directly into the electron ionization (100 eV) ion source. The excess gaseous mixture was directed from the Quick Inlet into a waste vent.

All ion signals which comprised >1% relative abundance in the full scan pure component reference spectra were monitored in the selected ion monitoring mode (300 ms dwell time for each ion) using a Faraday cup detector. To obtain analysis precision information, triplicate measurements of each sample mixture were collected in a random sequence. The spectrometer source lenses were tuned to give a flat-top ion peak with unit mass resolution across the scan range. Adjustments to the source voltages were made daily to optimize the intensity first and then the peak shape. These adjustments were repeated until the optimum intensity and peak shape was obtained. An example of the
Figure 2.1 Direct capillary or *Quick Inlet* to the process mass spectrometer.
typical voltages for the five adjustable parameters in the instrument source were: Ionizer: 21 V, Extractor: 8 V, Lens 1: -103 V, Lens 2: -7 V, and Pole Bias: 15 V (*Note: all voltages are within standard operational ranges*).

2.2 Gas mixture preparation

The mixtures were prepared on-line from compressed gas cylinders (see Table 2.1) using mass flow controllers with 2µm filters (Model 5850E Brooks Instruments, Hatfield, PA). The gas flow through the mass flow controllers were manually adjusted by microprocessor controls and read out units (Models 0154E and 0152E Brooks Instruments). The mass flow controllers have limited flow ranges which can be found in Table 2.2 along with the gas for which each controller is calibrated and certified by the factory. The factory certified accuracy and precision for all flow controllers is 1% and 0.25% of the flow rate respectively. Though these mass flow controllers are calibrated for a particular gas, the factory provides certified sensitivities for other gases which can be found in Table 2.3. To correct for the differences between the calibrated gas and the gas being monitored, Equation 2.1 can be used.

\[
\text{Actual Flow Rate} = \text{Flow Rate Setting} \times \frac{\text{Sample Gas Sensitivity}}{\text{Calibration Gas Sensitivity}}
\]

Equation 2.1

This calibration correction for the flow controller was used in all experiments.

The gas mixture flows from the mass flow controllers through 1/8” stainless steel tubing into 1/8” stainless steel mixing T’s (Swagelok, Solon, OH) as shown in Figure 2.2.
<table>
<thead>
<tr>
<th>Compound</th>
<th>Grade</th>
<th>Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helium</td>
<td>99.999+</td>
<td>Matheson Gas Products (Marrow, GA)</td>
</tr>
<tr>
<td>Cis-2-Butene</td>
<td>95+</td>
<td></td>
</tr>
<tr>
<td>Trans-2-Butene</td>
<td>95+</td>
<td></td>
</tr>
<tr>
<td>Methane</td>
<td>99+</td>
<td></td>
</tr>
<tr>
<td>Ethane</td>
<td>99+</td>
<td></td>
</tr>
<tr>
<td>Isobutane</td>
<td>99+</td>
<td>National Specialty Gases (Durham, NC)</td>
</tr>
<tr>
<td>Isobutene</td>
<td>99+</td>
<td></td>
</tr>
<tr>
<td>1-Butene</td>
<td>99+</td>
<td></td>
</tr>
<tr>
<td>Propane</td>
<td>99+</td>
<td></td>
</tr>
<tr>
<td>Propene</td>
<td>99+</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.2 Certified flow ranges for the mass flow controllers used along with the calibration gas for the controller.

<table>
<thead>
<tr>
<th>Controller Number</th>
<th>High Flow Rate (SCCM)</th>
<th>Low Flow Rate (SCCM)</th>
<th>Calibration Gas</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>500</td>
<td>10</td>
<td>Helium</td>
</tr>
<tr>
<td>2</td>
<td>500</td>
<td>10</td>
<td>Helium</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0.08</td>
<td>Methane</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>0.06</td>
<td>Ethane</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>0.20</td>
<td>Propene</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>0.20</td>
<td>Propane</td>
</tr>
</tbody>
</table>

1 Standard Cubic Centimeters per Minute.

2 Controller can deliver less, but not with certified precision from the factory.
Table 2.3 List of sensitivities used for the mass flow calibration.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Sensitivity$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methane</td>
<td>0.81</td>
</tr>
<tr>
<td>Ethane</td>
<td>0.55</td>
</tr>
<tr>
<td>Propane</td>
<td>0.47</td>
</tr>
<tr>
<td>Propene</td>
<td>0.39</td>
</tr>
<tr>
<td>Isobutene</td>
<td>0.34</td>
</tr>
<tr>
<td>Isobutane</td>
<td>0.31</td>
</tr>
<tr>
<td>1-Butene</td>
<td>0.33</td>
</tr>
<tr>
<td>Cis-2-Butene</td>
<td>0.32</td>
</tr>
<tr>
<td>Trans-2-Butene</td>
<td>0.36</td>
</tr>
<tr>
<td>Helium</td>
<td>1.39</td>
</tr>
</tbody>
</table>

$^1$Factory provided, used in Equation 2.1.
Figure 2.2 Mass flow controller setup.
Mixture compositions for all experiments are listed in Appendix 1. The total sample flow going into the Quick Inlet was 15 standard cubic centimeters per minute or milliliters per minutes (SCCM or mL/min) for all mixtures. A constant flow of He (10 SCCM) was used as a sweep gas to move the mixtures to the mass spectrometer. The base peak of He, m/z 4, was not monitored and its fractional contribution was ignored. After the mixing T’s, the gas traveled through a 1/8” stainless steel tube to the Quick Inlet on the process MS for simultaneous measurements. The dead volume of the sample system is about 3.5 mL (based on the i.d. of the stainless steel tubing and the mixing T’s). Thus with a flow rate of 15 SCCM (mL/min) the clearing time of the system would be about 14 seconds. The manufacturer suggested a clear time for the instrument of 1 minute between two samples. To be more conservative, a clear time of 2 minutes was used in these experiments. This allowed for 1 minute of clear time plus 1 minute for the equilibration of the new sample.

2.3 Concentration Calculations

Classical least squares (CLS) was used to calculate the concentration estimates by treating the mixture spectra as linear combinations of sensitivity-weighted reference spectra. The relative sensitivity for each component was found simultaneously using the reference spectra of the components along with the average spectrum of a test mixture of known composition. As part of the data work up, each spectrum was normalized to its base peak to minimize the effects of fluctuations of the absolute intensities. The data
work-up was performed using MATLAB (The Mathworks Inc., Natick, MA). A sample matrix calculation for the CLS routine is shown in Equation 2.2.

\[
\begin{bmatrix}
p_{1m1} & p_{2m1} & p_{3m1} \\
p_{1m2} & p_{2m2} & p_{3m2} \\
p_{1m3} & p_{2m3} & p_{3m3} \\
p_{1m4} & p_{2m4} & p_{3m4} \\
\end{bmatrix} \begin{bmatrix}
s_1 & 0 & 0 \\
0 & s_2 & 0 \\
0 & 0 & s_3 \\
\end{bmatrix} \begin{bmatrix}
c_1 \\
c_2 \\
c_3 \\
\end{bmatrix} = \begin{bmatrix}
I_1 \\
I_2 \\
I_3 \\
\end{bmatrix}
\]

2.2

\[
R \cdot S \cdot c = i
\]

Here, the first matrix is the reference matrix where \(p_{1m1}\) indicates the normalized intensity at m/z 1 (m1) of pure component 1 (p1). In the reference matrix, each column is the normalized spectrum of a pure component and each row represents one of the mass-to-charge ratios being monitored. The reference matrix is then multiplied by (dot product) a diagonal sensitivity matrix (where \(s_1\) is the sensitivity of pure component 1) and by a column concentration matrix (where \(c_1\) is the concentration of pure component 1). The product of these three matrices is a column matrix containing the intensities I at each m/z being monitored in a mixture spectrum. Note that when any three of these four matrices are known, then it is possible to solve for the other values. For a square matrix (the number of components equals the number of masses in the reference matrix) an exact solution is calculated and the least squares routine is not needed. When there are more masses than components then the case is over-determined which leads to a least squares solution.

Normally the reference matrix is used with a ‘calibration’ spectrum with known concentration to determine the sensitivity values (\(S = c ./Ri\) in MATLAB code) for each component in the mixture. With the calculated sensitivity values and the reference matrix, the concentrations can be determined for test mixtures (\(c = (R \cdot S) \cdot i\) in MATLAB
code) to test performance of a given parameterization. The resultant concentrations are
generally plotted as calculated concentrations against the actual known concentrations of
a set of test mixtures (a ‘validation plot’). The validation plot allows comparison with a
line where the calculated concentration is equal to the actual concentrations. As a
measurement of accuracy, a correlations coefficient \( r^2 \) is determined for the data. This
reflects the covariance but not the slope and intercept of the plot. As a better measure of
accuracy, the root mean squared error between the calculated and actual concentrations is
also determined as shown in Equation 2.3 [15]:

\[
E_{\text{RMS}} = \sqrt{\frac{\sum (x - y)^2}{n}} \quad 2.3
\]

where \( x \) and \( y \) are the actual and calculated concentrations respectively and \( n \) is the
number of number concentrations being calculated.

### 2.4 Parameterization Computations

Parameterizations from the Extrel ‘Stream Evaluator’ (the parameterization
software supplied with the instrument) were obtained using pure component spectra from
the Extrel library. Some of the library spectra differed significantly from the
experimental spectra obtained but this did not affect the results of the experiment because
the pure component spectrum was collected daily. The pure component spectra were
used for all calculations to eliminate any bad parameterizations due to these deviations.
Estimated concentrations for each component were set equal in weighting to obtain the
parameterization with equal concentration dependence for each component. The other weighting options available were relative standard deviation, relative interference spectra, relative interference factor, and interference. The relative interference factor calculated using Equation 2.4 tells whether a chosen mass is mostly from the component of interest or interference.

\[ RIF_{m/z} = \frac{\sum \text{concentration - scaled intensity of all interferences}}{\text{concentration - scaled intensity for the component of interest}} \]  \hspace{1cm} 2.4

The relative interference spectra are just any spectra that may be interfering with the component of interest and the interference is the number of interferences at the m/z of an ion of interest [23]. All of these weighting options were set to 1 and the dual detector option was not selected.

A detailed explanation of the empirical ranking algorithm developed for this research follows in Chapter 3. All calculations for the empirical algorithm were performed using in-house programs written for MATLAB. The data was prepared for the MATLAB programs using Microsoft Excel. The data was imported into Excel and organized into matrices to be imported into MATLAB. A copy of the empirical algorithm program code with annotations and a sample data file can be found in Appendix 2.
CHAPTER 3

DEVELOPMENT OF A PARAMETERIZATION ALGORITHM

3.1 Background

Parameterization selects the specific mass-to-charge (m/z) ratios to be monitored for quantitation. It is a major part of the development of a process monitoring method. There are several parameterization methods that can be used. These methods include genetic algorithms (GA), brute force parameterizations, and software packages that come with commercial instruments, such as Extrel’s ‘Stream Evaluator’ program. Each method has its strengths and weaknesses in terms of calculation time and quality of the parameterization. The GA and brute force methods are based on large training sets to determine a parameterization, which can be time consuming. The ‘Stream Evaluator’ program is based on the distinctiveness of peaks in spectra of compounds in a mixture but its output is dependent on the order in which the compounds are entered into the program (see below).

Parameterization by brute force algorithms gives the ‘best’ set of m/z’s to be used for quantitation based on accuracy of analysis or some other figure of merit for a given data set by testing all possible parameterizations. As the number of components in a mixture increases, this method can become very time consuming due to its calculation-
intensive nature. For example, there are 8,344,056 possible parameterizations

\[
\left( \sum_{l=q}^{p} \frac{p!}{l!(p-l)!} \right)
\]

using \( q = 6 \) (the number of components) and \( p = 23 \) (the total number of ion signals monitored) ion signals. Also, this method requires a large training set of known mixtures for the optimum parameterization to be obtained. If the training set is too small, then over-fitting of the data can occur.

Genetic algorithms (GA) sample a specified portion of all parameterizations and then continuously vary the sample population to converge on the ‘best’ parameterization for a given data set. By using only a specified portion of all the parameterizations, the number of calculations is less than the brute force algorithms but the GA still needs a large training set of known mixtures. Since the GA converges on the ‘best’ parameterization by using random variations, it can give a different ‘best’ parameterization each time it is used, even using the same data set. These may or may not be statistically equivalent, given the danger of over-fitting noted above.

The Extrel ‘Stream Evaluator’ software is used on the pure component spectra (either from the spectral library or user input if a component is not available in the library) which eliminates the large training sets that are needed for the brute force and genetic algorithms. This method usually selects the base peak to probe a given component. If the base peak is common to two or more components, the assignment for the second and subsequent components will depend on the order in which the components were entered into the software. This may not be the optimal parameterization for quantitative accuracy or precision but since no training sets are needed, the time for parameterization is minimal.
While these parameterization methods have positive points, their weaknesses leave room for improvements. Many mass spectrometrists use their intuition and experience when selecting ions because of the weaknesses of these methods. The research described in this thesis seeks to incorporate mass spectrometrists’ intuition into a simple but accurate empirical algorithm. It is desirable to have an empirical algorithm that does not need a large training set and selects a parameterization using only the pure component spectra. Unlike the ‘Stream Evaluator,’ the empirical algorithm should not be limited to a square matrix but should find an optimum parameterization.

An early version [24] of the empirical parameterization algorithm (EA) is given in Equation 3.1:

\[ \text{Score } A = \frac{1}{\sum_{\gamma} I_{A,m}} \left( \sum_{B \neq A} \text{SVD} \left[ \begin{bmatrix} I_{A,n} & I_{A,m} \\ I_{B,n} & I_{B,m} \end{bmatrix} \right] \right)_{\text{min}} \]

where I is the ion signal intensity, the subscripts A and B refer to the m/z ratios, and the subscripts n and m refer to the distinct compounds. Hence, this algorithm scores each ion (A, B …) in a reference spectrum of a given compound (m) compared with other components (n, et al.). The summations in the first two parts are taken over components \( \gamma \) and the SVD in the third part stands for singular value decomposition and will be discussed below.

This algorithm can be thought of in three successive parts: an intensity part, a distinctiveness part, and an orthogonality part. The first part, or intensity part, is found by summing the relative intensities for m/z A over all components \( \gamma \). The intuition behind this is that a large sum of intensities potentially means that there is a large amount
of information in that m/z. The second part, or the distinctiveness part, sums the ratio of relative intensities of m/z A over all components γ pairwise with the smaller relative intensity in the numerator and the larger in the denominator and then inverts the sum. The logic is that the larger the difference between the two relative intensities, the smaller the ratios. The reciprocal is taken because the smaller the sum of ratios, the larger the reciprocal meaning the larger the score and the larger the distinctiveness of this m/z.

The third part of the EA, or the orthogonality part, compares the relative intensities of m/z A to all the other m/z’s pairwise through singular value decomposition (SVD). SVD is a matrix decomposition method that separates a matrix (X) into three parts, two orthogonal matrices (U and V) and a diagonal matrix that contains the singular values (S) such that \( X = U S V^T \) (\( T \) is a matrix transpose) [25] [26]. For a 2 x γ matrix, the diagonal matrix S will contain two singular values due to the pairwise comparisons of m/z’s. The smaller of these two values is chosen for the singular value in the score. The SVD is used to test how similar m/z A is as a whole to each of the other m/z’s being monitored. The less similar the intensities are for the two m/z’s, the larger the singular value and potentially the more information. If two m/z’s were exactly the same, the singular value is close to zero whereas if they are completely different, the singular value is 100.

Once each m/z has been assigned a score, they are sorted from highest score to lowest. However, the empirical algorithm alone only assigns a number or score to each m/z; it does not provide a parameterization. Each m/z has been assigned a score such that its importance is known but the optimum number of mass-to-charge ratios to be included
in the parameterization needs to be determined. Stein’s spectral similarity index [27] was adapted to do just that and is shown in Equation 3.2:

$$SI = \left( \frac{\sum \sqrt{I_n I_m}}{\sum I_n \sum I_m} \right)^2$$

3.2

It is based on relative intensities (I, normalized so that $I_{\text{max}} = 100$ in a given spectrum) in the reference (n) and test (m) spectra. Contributions are usually included for all m/z at which there is significant intensity. A perfect match gives a similarity index (SI) of exactly 1; otherwise $0 \leq SI \leq 1$. For the purpose of parameterization, the SI scores are determined for a data set by adding m/z’s in rank order from the algorithm until the SI goes up (meaning an increase in similarity). Multi-component systems are handled by adding the scores from pairwise comparisons. The number of components determines the smallest number of m/z’s that should be entered into the SI giving a square matrix.

3.2 Algorithm Improvements

With the original empirical algorithm it was found that there was over-weighting of the intensity term which dominated the algorithm. There was also an ambiguity in the distinctiveness term which occurred when a pure component spectrum contained an intensity of zero for a m/z. The following corrections were made to alleviate this problem.
The intensity term takes the sum of all the intensities for a given m/z with the logic that a larger intensity should provide a more important ion. Logically, the base peak \((I = 100)\) of a component is extremely important in parameterization. If one or more of the components in the mixture share the same base peak, the score for that m/z will be dominated by the first term in Equation 3.1. To keep the importance of the base peak without dominating the score, the summation of intensities was replaced by the maximum intensity shown in Equation 3.3:

\[
\sum_{r} I_{A} \Rightarrow I_{A,\text{max}} \quad 3.3
\]

An intense peak, such as a base peak, will still have a large value but peaks prominent in multiple spectra will not be overweighted.

The distinctiveness term of the empirical algorithm addresses the interference between the contributions of multiple compounds at a given m/z. The original difference term did not fully address the different types of mixtures that could be used. If all the components in a mixture do not overlap in m/z, then some zeros will be introduced into the empirical algorithm. For example, if there was a mixture of methane and ethane, methane will only have four m/z’s monitored (m/z: 13, 14, 15, and 16) whereas ethane will have nine m/z’s monitored (m/z: 13, 14, 15, 16, 26, 27, 28, 29, and 30). For those m/z’s that belong only to ethane, their values would be zero for methane. The empirical algorithm would then entail division by zero which is not allowed mathematically. The original algorithm dealt with this by assigning a value of zero to be added to the summation. This can be replaced by the revised summation in Equation 3.4 where the
maximum intensity is now in the denominator compared to all other intensities for that m/z, including itself:

\[
\frac{1}{\sum_{\gamma} \frac{I_{A,n}}{I_{A,m}}} \Rightarrow \frac{1}{\sum_{\gamma} \frac{I_{A,n}}{I_{A,max}}} \quad 3.4
\]

This avoids the problem of having a varying number (larger value) in the denominator of the ratio. If there are zeros present, then a zero will just be added to the sum.

The orthogonality part was retained so that the new empirical ranking algorithm is given in Equation 3.5:

\[
\text{Score} \quad A = I_{A,\max} \times \frac{1}{\sum_{\gamma} I_{A,n}} \times \sum_{B \neq A} \left( \text{SVD} \begin{bmatrix} I_{A,n} & I_{A,m} & \cdots \\ I_{B,n} & I_{B,m} & \cdots \end{bmatrix} \right)_{\min}
\quad 3.5
\]

3.3 **Quantitative Analysis of Six-Component Hydrocarbon Mixtures**

A realistic and significant test of direct mass spectrometric analysis of complex mixtures was provided by the University of Tennessee Measurement and Control Engineering Center, a research center that partnered industrial companies with academic researchers. One member of the center had a need for rapid and continuous monitoring of a process stream comprised of six volatile C\textsubscript{1}-C\textsubscript{4} hydrocarbons. The specific stream, called the isobutene stream, contained methane, ethane, propane, propene, isobutane, and isobutene with concentrations in the ranges included in Table 3.1. The company was interested in improving their current processing methods of the isobutene stream by
Table 3.1 Approximate concentration ranges for the six components in the isobutene sample stream.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Concentration Range (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methane</td>
<td>3 – 16</td>
</tr>
<tr>
<td>Ethane</td>
<td>2 – 12</td>
</tr>
<tr>
<td>Propane</td>
<td>6 – 23</td>
</tr>
<tr>
<td>Propene</td>
<td>6 – 24</td>
</tr>
<tr>
<td>Isobutene</td>
<td>9 – 12</td>
</tr>
<tr>
<td>Isobutane</td>
<td>15 – 72</td>
</tr>
</tbody>
</table>
decreasing feedback time. Simultaneous analysis by process mass spectrometry to
monitor this mixture of volatile compounds seemed ideal.

The isobutene stream was originally studied in our lab by Dr. Kevin H. Bennett
using the original empirical ranking algorithm (Equation 3.1) as well as brute force and
genetic algorithms. Details of his results can be found in his dissertation [28]. As noted
above, his empirical ranking algorithm needed improvements to remove some ambiguity
in the difference term and the dominating power of the intensity term. To validate the
improvements, the isobutene sample stream was again used. Figure 3.1 shows the pure
component reference spectra and Table 3.2 shows the intensity values for all the
components of the isobutene stream. 32 different mixtures comprised a high/low
factorial design to span the wide range of concentrations. There were 23 m/z’s with
relative intensity above 2% available for monitoring. For each component, there was an
ion that was primarily due to the fragmentation of that compound (i.e., m/z 56 derives
primarily from isobutene). These ions correlate directly with the concentration of the
respective component. Some ions (i.e., m/z 27) have contributions from most or all of
the components which creates spectral interference which can lower quantitation
accuracy and precision.

The results for quantitation of all 32 mixtures of the isobutene stream applying
classical least squares (calculations described in Chapter 2.3) to the full spectrum are
shown in Figure 3.2. The points on the validation plot are the calculated concentrations
of each component plotted against the actual concentrations. All concentrations are
normalized so that the sums equal 100%. The theoretical line (a 45° line) is also shown
on the plot. Quantitation accuracy for this plot is excellent with a correlation coefficient,
Figure 3.1 A plot of the pure component reference spectra for the isobutene stream (methane, ethane, propane, propene, isobutene, and isobutane).
Table 3.2 Pure component reference spectra intensity values for isobutene stream (methane, ethane, propane, propene, isobutene, and isobutane).

<table>
<thead>
<tr>
<th>m/z</th>
<th>Methane</th>
<th>Ethane</th>
<th>Propane</th>
<th>Propene</th>
<th>Isobutene</th>
<th>Isobutane</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>6.191</td>
<td>0.736</td>
<td>0.527</td>
<td>1.012</td>
<td>0.489</td>
<td>0.245</td>
</tr>
<tr>
<td>14</td>
<td>13.326</td>
<td>2.311</td>
<td>1.436</td>
<td>2.030</td>
<td>1.129</td>
<td>0.783</td>
</tr>
<tr>
<td>16</td>
<td>100.000</td>
<td>0.154</td>
<td>0.194</td>
<td>0.226</td>
<td>0.320</td>
<td>0.186</td>
</tr>
<tr>
<td>26</td>
<td>21.147</td>
<td>8.726</td>
<td>8.699</td>
<td>4.523</td>
<td>2.303</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>100.000</td>
<td>60.793</td>
<td>3.094</td>
<td>19.384</td>
<td>3.294</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>21.036</td>
<td>100.000</td>
<td>0.433</td>
<td>9.327</td>
<td>5.652</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>25.946</td>
<td>2.254</td>
<td>0.038</td>
<td>0.212</td>
<td>0.124</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>4.260</td>
<td>12.124</td>
<td>4.315</td>
<td>1.813</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>6.613</td>
<td>17.314</td>
<td>7.097</td>
<td>3.477</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>20.323</td>
<td>70.382</td>
<td>49.525</td>
<td>18.423</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>3.263</td>
<td>26.225</td>
<td>12.564</td>
<td>2.956</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>16.620</td>
<td>100.000</td>
<td>100.000</td>
<td>47.721</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>5.747</td>
<td>62.796</td>
<td>3.522</td>
<td>38.429</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>30.606</td>
<td>2.078</td>
<td>0.036</td>
<td>100.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>37.303</td>
<td>0.839</td>
<td>1.031</td>
<td>3.846</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td>6.122</td>
<td>1.243</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td></td>
<td></td>
<td></td>
<td>4.531</td>
<td>0.871</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td></td>
<td></td>
<td></td>
<td>5.212</td>
<td>0.544</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td>18.048</td>
<td>0.512</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td>42.888</td>
<td>0.361</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td></td>
<td></td>
<td></td>
<td>0.021</td>
<td>4.415</td>
<td></td>
</tr>
</tbody>
</table>

Note: Those intensities in italics indicate the peaks that are distinctive in that they have contributions primarily from one component.
Figure 3.2 “Validation plot” of calculated versus actual concentration for 32 mixtures of methane (■), ethane (X), propane (Δ), propene (◇), isobutene (●), and isobutane (▲) using CLS with all-mass parameterization.
The correlation coefficient can be calculated using Equation 3.6 [29]:

\[
r^2 = \frac{\left( \sum_{i=1}^{n} \left( \frac{x_i - \bar{x})(y_i - \bar{y})}{n-1} \right)^2 \right)}{(S_x \times S_y)^2}
\]

where \(n\) is the number of measurements, \(x\) and \(y\) are the known and calculated concentrations, \(i\) is the individual concentration value, \(\bar{x}\) and \(\bar{y}\) connote the average of all the measurements for \(x\) and \(y\) respectively, and \(S\) is the standard deviation. The indicated deviation in \(r^2\) is based on the standard deviation in the correlation coefficient measured for triplicate runs of each of the 32 mixtures (i.e., the standard deviation of three \(r^2\) values from three separate validation plots). Besides the excellent accuracy, the precision is also excellent; one standard deviation error bars for the triplicate measurements of each sample are all smaller than the data points on the plot. The average coefficient of variation (\(<CV>\), the percent relative standard deviation values calculated for triplicate measurements and averaged across all components and samples) is \(0.43\% \pm 0.51\) and was found using Equation 3.7 [15]:

\[
CV = \frac{\sigma}{\bar{x}} \times 100\% \quad 3.7
\]

where \(\sigma\) is the standard deviation and \(\bar{x}\) is the average concentration. The root mean square error (\(E_{RMS}\)) value is 0.43 and can be calculated by using Equation 2.3.

Even with such excellent analytical performance using all 23 mass-to-charge ratios, there is still room for improvement. For example, not all of the data points in

38
Figure 3.2 are on the theoretical line. Even if this performance (represented by $E_{\text{RMS}}$) cannot be improved, it is possible that another figure of merit – analysis time – might be improved if the number of masses monitored can be reduced without affecting accuracy and precision. The empirical algorithm was utilized to search for an optimum parameterization to decrease the number of masses.

The empirical algorithm assigns a score to each ion based solely on the pure component spectra using Equation 3.5. The similarity index of Equation 3.2 is then used to determine the optimum number of masses to monitor. The algorithm scores, the individual parts that make up the score, and similarity index scores are found in Table 3.3. The selection of ions is based on local minima that occur in the similarity index scores as the number of masses increases. Plotted in Figure 3.3, the first minimum occurs after eight masses have been added to the Stein similarity index in Equation 3.2 determining that the top eight masses ranked in Table 3.3 is the optimum parameterization (m/z’s 15, 16, 28, 29, 41, 42, 43, and 56). Using this parameterization to quantitate all 32 sample mixtures gives a validation plot with an $r^2$ value of $0.99913 \pm 0.00001$, a $<\text{CV}>$ value of $0.49\% \pm 0.59$, and an $E_{\text{RMS}}$ value of 0.47. The empirical ranking algorithm quantitation performance was comparable to that of the all mass parameterization ($r^2 = 0.99920 \pm 0.00007$, $<\text{CV}> = 0.43\% \pm 0.51$, and $E_{\text{RMS}} = 0.43$). The empirical algorithm reduced the number of ions monitored for the isobutene stream by 65%. All figures of merit are statistically similar between the all-mass parameterization and the empirical algorithm parameterization.

Because the empirical algorithm is reliant on only spectral comparisons (one set of reference spectra) this method is potentially susceptible to error. For example, if there
Table 3.3 Empirical algorithm and similarity index scores for the pure component spectra in Table 3.2 for the isobutene stream.

<table>
<thead>
<tr>
<th>m/z</th>
<th>Intensity Part</th>
<th>Difference Part</th>
<th>Orthogonality Part</th>
<th>Algorithm Score (A. U.)</th>
<th>Similarity Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>100.00</td>
<td>0.99</td>
<td>874.36</td>
<td>86502.17</td>
<td>1.000</td>
</tr>
<tr>
<td>43</td>
<td>100.00</td>
<td>0.76</td>
<td>893.51</td>
<td>67322.65</td>
<td>0.435</td>
</tr>
<tr>
<td>29</td>
<td>100.00</td>
<td>0.74</td>
<td>835.27</td>
<td>61215.06</td>
<td>0.347</td>
</tr>
<tr>
<td>15</td>
<td>81.31</td>
<td>0.82</td>
<td>786.37</td>
<td>52538.10</td>
<td>0.373</td>
</tr>
<tr>
<td>28</td>
<td>100.00</td>
<td>0.53</td>
<td>866.02</td>
<td>46419.11</td>
<td>0.377</td>
</tr>
<tr>
<td>41</td>
<td>100.00</td>
<td>0.38</td>
<td>789.15</td>
<td>29853.64</td>
<td>0.261</td>
</tr>
<tr>
<td>42</td>
<td>62.80</td>
<td>0.57</td>
<td>689.31</td>
<td>24600.32</td>
<td>0.251</td>
</tr>
<tr>
<td>56</td>
<td>42.89</td>
<td>0.99</td>
<td>534.41</td>
<td>22729.78</td>
<td>0.224</td>
</tr>
<tr>
<td>39</td>
<td>70.38</td>
<td>0.44</td>
<td>695.98</td>
<td>21730.75</td>
<td>0.239</td>
</tr>
<tr>
<td>44</td>
<td>37.30</td>
<td>0.87</td>
<td>488.20</td>
<td>15791.72</td>
<td>0.231</td>
</tr>
<tr>
<td>30</td>
<td>25.95</td>
<td>0.91</td>
<td>393.53</td>
<td>9271.50</td>
<td>0.226</td>
</tr>
<tr>
<td>27</td>
<td>39.70</td>
<td>0.28</td>
<td>592.72</td>
<td>6575.98</td>
<td>0.257</td>
</tr>
<tr>
<td>40</td>
<td>26.23</td>
<td>0.58</td>
<td>361.26</td>
<td>5520.35</td>
<td>0.256</td>
</tr>
<tr>
<td>55</td>
<td>18.05</td>
<td>0.98</td>
<td>277.05</td>
<td>4862.28</td>
<td>0.250</td>
</tr>
<tr>
<td>26</td>
<td>21.15</td>
<td>0.47</td>
<td>343.56</td>
<td>3384.23</td>
<td>0.256</td>
</tr>
<tr>
<td>38</td>
<td>17.31</td>
<td>0.50</td>
<td>265.32</td>
<td>2305.32</td>
<td>0.257</td>
</tr>
<tr>
<td>14</td>
<td>13.33</td>
<td>0.63</td>
<td>228.99</td>
<td>1935.03</td>
<td>0.260</td>
</tr>
<tr>
<td>37</td>
<td>12.12</td>
<td>0.54</td>
<td>197.68</td>
<td>1290.73</td>
<td>0.261</td>
</tr>
<tr>
<td>50</td>
<td>6.12</td>
<td>0.83</td>
<td>107.39</td>
<td>546.470</td>
<td>0.259</td>
</tr>
<tr>
<td>13</td>
<td>6.19</td>
<td>0.67</td>
<td>117.12</td>
<td>487.95</td>
<td>0.260</td>
</tr>
<tr>
<td>53</td>
<td>5.21</td>
<td>0.91</td>
<td>90.77</td>
<td>428.37</td>
<td>0.258</td>
</tr>
<tr>
<td>58</td>
<td>4.42</td>
<td>1.00</td>
<td>91.91</td>
<td>403.85</td>
<td>0.256</td>
</tr>
<tr>
<td>51</td>
<td>4.53</td>
<td>0.84</td>
<td>80.40</td>
<td>305.56</td>
<td>0.255</td>
</tr>
</tbody>
</table>
Figure 3.3 Plot of Stein similarity index using the ranked masses of the isobutene stream shown in Table 3.3
is one bad pure component spectrum in the set of reference spectra, then the parameterization determined would be inaccurate. The ideal solution is to collect pure component spectra each day as part of the daily tuning to ensure that the reference spectra are have not deviated.

Some algorithm software such as Extrel’s ‘Stream Evaluator’ suggests a single ion per component to monitor. In this case, each of the six components will be assigned a single peak so that six ions would be monitored. When calculating the concentrations using the Extrel parameterization, the reference spectra matrix becomes a square matrix giving an exact answer through CLS. A square matrix would be desirable because it is the smallest number of masses to monitor which would minimize the analysis time. One drawback to using the ‘Stream Evaluator’ to select masses for the components is that the order in which the components are entered into the software affects the parameterization. However, examining all 720 possible orders in which the six components could be entered, only two different parameterizations were found. Each of the parameterizations came up 360 times. One parameterization (Extrel 1) chose m/z’s 16, 28, 29, 39, 41, and 43 giving an \( r^2 \) value of 0.99813 ± 0.00002, a \(<CV>\) value of 0.63% ± 0.86, and an \( E_{RMS} \) value of 0.66. The other parameterization (Extrel 2) chose m/z’s 16, 28, 29, 41, 43, and 56 giving an \( r^2 \) value of 0.99932 ± 0.00001, a \(<CV>\) value of 0.36% ± 0.41, and an \( E_{RMS} \) value of 0.44. There is only one mass difference between the two (m/z 39 or 56) and the difference is from propene and isobutene having the same base peak (m/z 41). If isobutene was entered first, m/z 41 was selected for isobutene and m/z 39 was selected for propene. If instead propene was entered first, m/z 41 was selected for propene and m/z 56 was selected for isobutene. Extrel 1 is slightly less accurate and precise than the
all mass and empirical algorithm ranking whereas Extrel 2 is slightly more accurate and precise.

A square matrix can also be derived from the empirical algorithm ranking by choosing just the six highest ranked peaks. The m/z’s in this parameterization are 15, 16, 28, 29, 41, and 43 with an $r^2$ value of 0.04 ± 0.02, a $<CV>$ value of -19.15% ± 61.04, and a $E_{RMS}$ value of 85.20. There is very little correlation between the calculated concentrations and the known concentrations and negative concentrations were calculated through CLS. Compared to the ‘Stream Evaluator’ square matrices, there is again only one mass difference in the parameterization. The empirical algorithm selected m/z 15 whereas the ‘Stream Evaluator’ selected either m/z 39 or 56. M/z 15 was chosen because it is mainly contributed to by one component (methane) and it has a large intensity (81.307). Though these are desirable when determining ions to monitor, this would suggest that this peak should be selected for methane. But m/z 16 is already being monitored as it is the base peak for methane meaning that the empirical algorithm is selecting two peaks for one component. The empirical algorithm seems to need a little more work when it comes to determining square matrices while its optimum parameterization performs excellently. A tabulation of all the isobutene parameterization performance can be found in Table 3.4.
Table 3.4 Ions selected and figures of merit obtained using the empirical algorithm and Extrel's 'Stream Evaluator' parameterizations for triplicate measurements of 32 isobutene stream samples.

<table>
<thead>
<tr>
<th>Parameterization Method</th>
<th>Ions Selected</th>
<th>( &lt;r^2&gt; )</th>
<th>( &lt;CV&gt; )</th>
<th>( E_{RMS} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>All mass</td>
<td>All</td>
<td>0.99920 ± 0.00007</td>
<td>0.43 ± 0.51</td>
<td>0.43</td>
</tr>
<tr>
<td>Empirical Algorithm</td>
<td>15 16 28 29 41 42 43 56</td>
<td>0.99913 ± 0.00001</td>
<td>0.49 ± 0.59</td>
<td>0.47</td>
</tr>
<tr>
<td>EA Square Matrix</td>
<td>15 16 28 29 41 43</td>
<td>0.04 ± 0.02</td>
<td>-19.15 ± 61.04</td>
<td>85.20</td>
</tr>
<tr>
<td>Extrel 1 Square Matrix</td>
<td>16 28 29 39 41 43</td>
<td>0.99813 ± 0.00002</td>
<td>0.63 ± 0.86</td>
<td>0.66</td>
</tr>
<tr>
<td>Extrel 2 Square Matrix</td>
<td>16 28 29 41 43 56</td>
<td>0.99932 ± 0.00001</td>
<td>0.36 ± 0.41</td>
<td>0.44</td>
</tr>
</tbody>
</table>
3.4 Conclusions

The use of the empirical algorithm can greatly reduce the number of masses to monitor and the analysis time without compromising the accuracy and precision. The empirical algorithm also saves time by relying on pure component spectra as opposed to large training sets of data. The lack of correlation in the square matrix of the empirical algorithm is a little unsettling but the algorithm set out to find an optimum number of m/z’s to monitor and not the smallest number of masses.
CHAPTER 4

SIMULTANEOUS ANALYSIS OF FOUR-COMPONENT BUTENE ISOMER MIXTURES USING PROCESS MASS SPECTROMETRY

4.1 Introduction to the Isomer Mixtures

Analysis of the six-component isobutene stream, described in Chapter 3, was accurate, precise, and fast. Its quantitation was influenced by which ions were chosen for the parameterization. One factor that contributed to the ease in this quantitation was that each component in the isobutene stream had a peak in the mixture spectrum with contributions from mainly one component. These peaks made the mixture mass spectrum deconvolution easy, thus the analysis accuracy was excellent.

While the performance of the empirical algorithm on the isobutene stream is encouraging, not all process sample streams have as relatively simple component composition as that of the isobutene stream. Process streams can contain complex mixtures of isobars (compounds that have different chemical compositions but the same nominal mass; for example, ethene and carbon monoxide) and isomers (compounds that have the same exact mass with identical elemental composition but different chemical structures; for example isobutene and 1-butene). Because of the similar chemical and/or
mass composition, simultaneous mass spectrometric analysis of isobars and isomers is more difficult.

Previous studies [24] [30] [31] [32] [33] in our lab have established that the high precision of ion intensity measurements of process mass spectrometry can enable distinction of relatively minor differences in relative ion intensities in the mass spectra of isomers. Hence, the high precision intensity measurements form a basis for simultaneous quantitation of mixtures containing isomers. The aim of this chapter is to further the study with the simultaneous quantitation of mixtures that include only isomers; in this case, all four butene isomers. The influence of the complexity of these mixtures will be studied.

The butene isomers were chosen for this study due to their use in industry. Butenes are commonly used in the production of high octane gasoline and synthetic rubbers. Figure 4.1 shows the chemical structure for each of the four butene isomers. Because these compounds have identical elemental composition but different structures, it can be expected that the mass spectra of the four butenes will be very similar in number of ions, mass-to-charge ratios, and their intensities. Figure 4.2 is a plot of the mass spectra and Table 4.1 lists the intensity values for the butene isomers. There are many spectral similarities between the compounds but they are not identical. Therefore simultaneous quantitation of such a complex mixture should be feasible but difficult.
Figure 4.1 Chemical structures for 1-butene, isobutene, cis-2-butene, and trans-2-butene.
Figure 4.2 Pure component reference mass spectra for all four butene isomers (1-butene, isobutene, cis-2-butene, trans-2-butene).
Table 4.1 Pure component reference spectra intensity values for butene isomers (1-butene, isobutene, cis-2-butene, and trans-2-butene).

<table>
<thead>
<tr>
<th>m/z</th>
<th>Isobutene</th>
<th>1-Butene</th>
<th>Cis-2-Butene</th>
<th>Trans-2-Butene</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>0.913</td>
<td>0.565</td>
<td>1.047</td>
<td>0.564</td>
</tr>
<tr>
<td>15</td>
<td>2.453</td>
<td>2.240</td>
<td>2.706</td>
<td>2.195</td>
</tr>
<tr>
<td>25</td>
<td>1.059</td>
<td>1.452</td>
<td>0.686</td>
<td>1.493</td>
</tr>
<tr>
<td>26</td>
<td>8.400</td>
<td>10.171</td>
<td>4.524</td>
<td>10.096</td>
</tr>
<tr>
<td>27</td>
<td>24.674</td>
<td>26.805</td>
<td>16.961</td>
<td>27.336</td>
</tr>
<tr>
<td>29</td>
<td>11.487</td>
<td>13.717</td>
<td>9.373</td>
<td>17.487</td>
</tr>
<tr>
<td>37</td>
<td>2.824</td>
<td>2.860</td>
<td>4.054</td>
<td>2.763</td>
</tr>
<tr>
<td>38</td>
<td>4.360</td>
<td>4.187</td>
<td>6.476</td>
<td>3.998</td>
</tr>
<tr>
<td>39</td>
<td>35.082</td>
<td>36.095</td>
<td>45.447</td>
<td>36.447</td>
</tr>
<tr>
<td>40</td>
<td>7.172</td>
<td>7.535</td>
<td>11.230</td>
<td>7.096</td>
</tr>
<tr>
<td>41</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
<td>100.000</td>
</tr>
<tr>
<td>42</td>
<td>3.399</td>
<td>3.348</td>
<td>3.648</td>
<td>3.352</td>
</tr>
<tr>
<td>50</td>
<td>5.429</td>
<td>7.612</td>
<td>6.092</td>
<td>7.801</td>
</tr>
<tr>
<td>51</td>
<td>4.069</td>
<td>5.850</td>
<td>4.560</td>
<td>5.943</td>
</tr>
<tr>
<td>53</td>
<td>5.393</td>
<td>7.998</td>
<td>5.123</td>
<td>8.576</td>
</tr>
<tr>
<td>54</td>
<td>1.920</td>
<td>3.007</td>
<td>1.755</td>
<td>2.953</td>
</tr>
<tr>
<td>56</td>
<td>35.108</td>
<td>44.927</td>
<td>40.352</td>
<td>44.684</td>
</tr>
<tr>
<td>57</td>
<td>1.550</td>
<td>1.969</td>
<td>1.800</td>
<td>1.946</td>
</tr>
</tbody>
</table>
4.2 Quantitative Analysis of Butene Isomers

The spectral differences between butene isomers may be used for the simultaneous quantitation of the mixtures by simple CLS deconvolution of the mixture mass spectra. Figure 4.3 shows the performance of the CLS deconvolution of mixtures of the four butene isomers using all 19 mass-to-charge ratios with intensities over 2% of the base peak in the spectra of the pure components. As before, the validation plot in Figure 4.3 is a plot of calculated concentration versus actual concentration with a theoretical line where the two concentrations are equal. As with the isobutene stream, high/low factorial design determined that eight mixtures should be studied. The concentration ranges studied are listed in Table 4.2 with specific concentrations for each mixture listed in Appendix 1, Table 7.2. Each sample was run in triplicate in a random order and concentrations were calculated using CLS with pure component reference mass spectra weighted with sensitivities determined using a calibration mixture (which consisted of equal concentrations of each isomer). Analysis accuracy and precision are not as good as in Chapter 3 (correlation coefficient $r^2 = 0.986 \pm 0.007$, average correlation of variation $<CV> = 5.54\% \pm 6.66$, and root mean square error $E_{RMS} = 1.86$) which is expected with the complexity of the isomer mixtures.

The algorithm scores with the individual parts of the scores and similarity index scores are found in Table 4.3 with a plot of the similarity index in Figure 4.4. The first minimum that occurs in the plot of the Stein index is after five masses have been added. The empirical algorithm thus selected a parameterization that included m/z’s 27, 29, 39,
Figure 4.3 Validation plot for mixtures of cis-2-butene (■), trans-2-butene (X), isobutene (∆), and 1-butene (♦) using all 19 peaks with >2% relative abundance.
Table 4.2 Approximate concentration ranges for the four components in the butene isomer mixtures.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Concentration Range (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cis-2-Butene</td>
<td>10 – 69</td>
</tr>
<tr>
<td>Isobutene</td>
<td>10 – 67</td>
</tr>
<tr>
<td>1-Butene</td>
<td>9 – 70</td>
</tr>
<tr>
<td>Trans-2-Butene</td>
<td>9 – 36</td>
</tr>
<tr>
<td>m/z</td>
<td>Intensity Part</td>
</tr>
<tr>
<td>-----</td>
<td>----------------</td>
</tr>
<tr>
<td>41</td>
<td>100.00</td>
</tr>
<tr>
<td>39</td>
<td>45.45</td>
</tr>
<tr>
<td>56</td>
<td>44.93</td>
</tr>
<tr>
<td>27</td>
<td>27.34</td>
</tr>
<tr>
<td>29</td>
<td>17.49</td>
</tr>
<tr>
<td>55</td>
<td>21.64</td>
</tr>
<tr>
<td>40</td>
<td>11.23</td>
</tr>
<tr>
<td>26</td>
<td>10.17</td>
</tr>
<tr>
<td>53</td>
<td>8.58</td>
</tr>
<tr>
<td>38</td>
<td>6.48</td>
</tr>
<tr>
<td>50</td>
<td>7.80</td>
</tr>
<tr>
<td>51</td>
<td>5.94</td>
</tr>
<tr>
<td>37</td>
<td>4.05</td>
</tr>
<tr>
<td>42</td>
<td>3.65</td>
</tr>
<tr>
<td>54</td>
<td>3.01</td>
</tr>
<tr>
<td>15</td>
<td>2.71</td>
</tr>
<tr>
<td>25</td>
<td>1.49</td>
</tr>
<tr>
<td>57</td>
<td>1.97</td>
</tr>
<tr>
<td>14</td>
<td>1.05</td>
</tr>
</tbody>
</table>
Figure 4.4 Plot of Stein similarity index using the ranked masses of the butene isomer stream shown in Table 4.1
41, and 56. This parameterization yielded an $r^2$ value of 0.983 ± 0.008, $<\text{CV}>$ value of 6.20% ± 7.38, and $E_{\text{RMS}}$ value of 2.03. The precision and accuracy are slightly worse than the all-mass parameterization while only using 26% of the spectra, with proportionate saving in analysis time.

Using the empirical algorithm to obtain a square matrix, the top four m/z’s chosen are 27, 39, 41, and 56. With this parameterization the results are abysmal ($r^2 = 0.1 \pm 0.1$, $<\text{CV}> = 43.63\% \pm 320.19$, and $E_{\text{RMS}} = 517.42$). There is basically no correlation that can be found and there were negative concentrations that were being calculated. With further inspection into the chosen peaks, it seems that the four masses that have the largest intensities were chosen while none of them were really distinct from one component to the other. So the intuition that large intensities are good overshadowed the need for differences in intensities of the components at the m/z.

Extrel’s ‘Stream Evaluator’ once again chose two different parameterizations which were dependent on the order in which the components were entered into the software. Studying all 24 possible orders, 75% of the parameterizations matched that of the empirical algorithm and 25% of the parameterizations chose m/z’s 39, 41, 55, and 56. The results of this parameterization were equally as abysmal ($r^2 = 0.15 \pm 0.08$, $<\text{CV}> = -614.35\% \pm 2146.17$, and $E_{\text{RMS}} = 125.19$). Again, there is no correlation to be found and negative concentrations were being calculated. Once again, the peaks that are being chosen have large intensities but no significant differences in the intensities in the spectra of different components. A tabulation of this data is included in Table 4.4.
Table 4.4 Ions selected and figures of merit obtained using the empirical algorithm and Extrel's 'Stream Evaluator' parameterizations for triplicate measurements of eight butene isomer samples.

<table>
<thead>
<tr>
<th>Parameterization Method</th>
<th>Ions Selected</th>
<th>$&lt;r^2&gt;$</th>
<th>$&lt;CV&gt;$</th>
<th>$E_{RMS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All mass</td>
<td>All</td>
<td>0.986 ± 0.007</td>
<td>5.54 ± 6.66</td>
<td>1.86</td>
</tr>
<tr>
<td>Empirical Algorithm</td>
<td>27 29 39 41 56</td>
<td>0.983 ± 0.008</td>
<td>6.20 ± 7.38</td>
<td>2.03</td>
</tr>
<tr>
<td>EA Square Matrix</td>
<td>27 39 41 56</td>
<td>0.1 ± 0.1</td>
<td>43.63 ± 320.19</td>
<td>517.42</td>
</tr>
<tr>
<td>25% Extrel Square Matrix</td>
<td>39 41 55 56</td>
<td>0.15 ± 0.08</td>
<td>-614.35 ± 2146.17</td>
<td>125.19</td>
</tr>
<tr>
<td>75% Extrel Square Matrix</td>
<td>27 39 41 56</td>
<td>0.1 ± 0.1</td>
<td>43.63 ± 320.19</td>
<td>517.42</td>
</tr>
</tbody>
</table>
4.3 Future Work

The empirical algorithm does need some work with its square matrices. Their lack of correlation can come from the slight differences in the spectra from day-to-day tuning which is being amplified through the spectral overlap of this complex mixture. Another possibility could come from the third part of the empirical algorithm itself. The singular value decomposition is a pairwise comparison which is being summed. This part may need to be replaced with a mathematical function that will compare all the masses at once to alleviate the need for the sum of pairwise comparisons. Future work should study this lack of correlation to see if it is from the data being collected or whether part of the algorithm is to blame.

Future work of the empirical algorithm should also include a wider range of samples. These samples should include a larger number of components and a larger number of masses. The empirical algorithm has proven itself with isomers, so the study of some samples with isobars may prove helpful. There are many different options to study to determine various applications of the empirical algorithm.
CONCLUSIONS

For the isobutene stream in Chapter 3.3, the empirical algorithm parameterization gave results statistically similar to that of the all-mass parameterization. It decreased the number of masses by 65% without losing accuracy or precision. Unfortunately, the empirical algorithm’s square matrix’s performance was unsatisfactory when compared to the Extrel instrument’s software. When studying the more complex mixtures of butene isomers, the empirical algorithm parameterization gave similar results to the all-mass parameterization while decreasing the number of masses by 74%. When compared to the Extrel’s ‘Stream Evaluator’ parameterization, both sets of square matrices gave poor results.

In conclusion, the empirical algorithm, which was created to determine an optimum parameterization for a mixture, has been proven. The empirical algorithm paired with the Stein similarity index has ranked the peaks of components in a mixture and determined the optimum number of masses. This optimum number of peaks was used to determine the concentrations of the components with excellent accuracy and precision. These studies show how the parameterization is dependent on the difficulty of application and similarity of component spectra.


APPENDICES
## APPENDIX 1

### SAMPLE COMPOSITIONS

Table 7.1 Percentage sample compositions for mixtures of the isobutene stream.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Methane</th>
<th>Ethane</th>
<th>Propane</th>
<th>Propene</th>
<th>Isobutene</th>
<th>Isobutane</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calibration</td>
<td>9.598</td>
<td>6.998</td>
<td>14.637</td>
<td>15.197</td>
<td>10.760</td>
<td>42.810</td>
</tr>
<tr>
<td>1</td>
<td>15.852</td>
<td>1.806</td>
<td>6.140</td>
<td>6.421</td>
<td>9.816</td>
<td>59.965</td>
</tr>
<tr>
<td>2</td>
<td>3.364</td>
<td>11.775</td>
<td>6.128</td>
<td>6.408</td>
<td>9.797</td>
<td>62.527</td>
</tr>
<tr>
<td>3</td>
<td>3.366</td>
<td>1.803</td>
<td>23.203</td>
<td>6.412</td>
<td>9.802</td>
<td>55.413</td>
</tr>
<tr>
<td>5</td>
<td>3.364</td>
<td>1.802</td>
<td>6.126</td>
<td>6.407</td>
<td>11.753</td>
<td>70.548</td>
</tr>
<tr>
<td>9</td>
<td>3.361</td>
<td>11.765</td>
<td>23.169</td>
<td>6.403</td>
<td>9.788</td>
<td>45.514</td>
</tr>
<tr>
<td>11</td>
<td>3.358</td>
<td>1.799</td>
<td>6.117</td>
<td>24.387</td>
<td>11.735</td>
<td>52.605</td>
</tr>
<tr>
<td>12</td>
<td>15.824</td>
<td>1.803</td>
<td>6.129</td>
<td>6.410</td>
<td>11.759</td>
<td>58.075</td>
</tr>
<tr>
<td>13</td>
<td>15.827</td>
<td>1.803</td>
<td>6.130</td>
<td>24.441</td>
<td>9.801</td>
<td>41.999</td>
</tr>
<tr>
<td>14</td>
<td>15.837</td>
<td>1.804</td>
<td>23.215</td>
<td>6.415</td>
<td>9.807</td>
<td>42.921</td>
</tr>
<tr>
<td>15</td>
<td>3.359</td>
<td>11.755</td>
<td>6.117</td>
<td>6.397</td>
<td>11.736</td>
<td>60.636</td>
</tr>
<tr>
<td>16</td>
<td>3.359</td>
<td>11.757</td>
<td>6.118</td>
<td>24.393</td>
<td>9.781</td>
<td>44.592</td>
</tr>
<tr>
<td>17</td>
<td>3.360</td>
<td>1.800</td>
<td>23.163</td>
<td>6.401</td>
<td>11.743</td>
<td>53.533</td>
</tr>
<tr>
<td>18</td>
<td>3.355</td>
<td>1.797</td>
<td>23.127</td>
<td>24.365</td>
<td>11.724</td>
<td>35.632</td>
</tr>
<tr>
<td>19</td>
<td>15.799</td>
<td>1.800</td>
<td>6.120</td>
<td>24.399</td>
<td>11.740</td>
<td>40.142</td>
</tr>
<tr>
<td>20</td>
<td>15.801</td>
<td>11.761</td>
<td>6.120</td>
<td>6.400</td>
<td>11.742</td>
<td>48.176</td>
</tr>
<tr>
<td>25</td>
<td>15.812</td>
<td>1.801</td>
<td>23.178</td>
<td>24.419</td>
<td>9.792</td>
<td>24.998</td>
</tr>
<tr>
<td>26</td>
<td>15.810</td>
<td>1.801</td>
<td>23.175</td>
<td>6.404</td>
<td>11.748</td>
<td>41.062</td>
</tr>
<tr>
<td>27</td>
<td>15.803</td>
<td>11.762</td>
<td>6.121</td>
<td>24.405</td>
<td>9.786</td>
<td>32.122</td>
</tr>
<tr>
<td>30</td>
<td>15.785</td>
<td>1.798</td>
<td>23.138</td>
<td>24.377</td>
<td>11.730</td>
<td>23.172</td>
</tr>
</tbody>
</table>

*Note: Number of mixtures determined by high/low factorial design.*
Table 7.2 Percentage sample compositions for mixtures of the butene isomers.

<table>
<thead>
<tr>
<th>Sample</th>
<th>1-Butene</th>
<th>Cis-2-Butene</th>
<th>Isobutene</th>
<th>Trans-2-Butene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calibration</td>
<td>25.23</td>
<td>24.91</td>
<td>24.94</td>
<td>24.94</td>
</tr>
<tr>
<td>1</td>
<td>30.91</td>
<td>30.51</td>
<td>28.88</td>
<td>9.71</td>
</tr>
<tr>
<td>2</td>
<td>30.91</td>
<td>9.83</td>
<td>26.67</td>
<td>32.60</td>
</tr>
<tr>
<td>3</td>
<td>11.40</td>
<td>40.00</td>
<td>33.99</td>
<td>14.60</td>
</tr>
<tr>
<td>4</td>
<td>13.87</td>
<td>13.88</td>
<td>37.09</td>
<td>35.15</td>
</tr>
<tr>
<td>5</td>
<td>42.00</td>
<td>39.32</td>
<td>8.89</td>
<td>9.80</td>
</tr>
<tr>
<td>6</td>
<td>69.12</td>
<td>9.99</td>
<td>10.00</td>
<td>10.88</td>
</tr>
<tr>
<td>7</td>
<td>12.36</td>
<td>66.66</td>
<td>12.45</td>
<td>8.53</td>
</tr>
<tr>
<td>8</td>
<td>21.49</td>
<td>21.48</td>
<td>21.48</td>
<td>35.56</td>
</tr>
</tbody>
</table>

Note: Number of mixtures determined by high/low factorial design.
Table 7.3  Percentage sample compositions for small set of mixtures of the butene isomers.

<table>
<thead>
<tr>
<th>Sample</th>
<th>1-Butene</th>
<th>Cis-2-Butene</th>
<th>Isobutene</th>
<th>Trans-2-Butene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calibration</td>
<td>25.23</td>
<td>24.91</td>
<td>24.94</td>
<td>24.94</td>
</tr>
<tr>
<td>1</td>
<td>39.55</td>
<td>20.14</td>
<td>30.22</td>
<td>10.10</td>
</tr>
<tr>
<td>2</td>
<td>9.83</td>
<td>50.09</td>
<td>10.03</td>
<td>30.06</td>
</tr>
<tr>
<td>3</td>
<td>70.01</td>
<td>9.99</td>
<td>9.99</td>
<td>10.01</td>
</tr>
<tr>
<td>4</td>
<td>29.62</td>
<td>30.18</td>
<td>20.11</td>
<td>20.09</td>
</tr>
</tbody>
</table>
APPENDIX 2

EMPIRICAL ALGORITHM CODE

function [cookind6]=cookind6(m,d);
%This code ranks m/z for parameterization using an empirical algorithm.
%This algorithm scores m/z based on total intensity, difference in intensities, and
%orthogonality.
%Algorithm scores are then sorted by the Stein Similarity Index to obtain the
%optimum parameterization.
%The lower the Stein Similarity Index Score, the more dissimilar the spectra.
%To use the code, please have the data in the following format:
%m is a one column, multi-row matrix containing the m/z’s
%d is a multi-column, multi-row matrix containing intensities at each mass
%for pure component spectra, where the columns are the components and the rows correspond to m.

format long

mass=m; %Masses in the set
refdata=d; %Pure component reference data
comps=size(refdata, 2); %Number of components in the set
mas=size(refdata, 1); %Number of masses in the set

%ALGORITHM CODE BEGINS HERE!!!

for x=1:mas; %This cycles through each m/z
ty(x, :)=x; %Counter
score1=max(refdata(x, :)); %Chooses the maximum intensity per m/z
score2=0; %Resets the score
score3=0; %Resets the score
for tyu=1:comps; %This cycles through each component
    on3=refdata(x, tyu)/score1;
    score2=score2+on3; %Summation in the difference term
end
for zz=1:mas; %For SVD term
    if zz==x
        score3=0;
    else
        datas=[refdata(x, :) refdata(zz, :)]; %Matrix corresponding
38 %with SVD
39     sco=svd(datas);
40     score3=sco(2, 1);  %Chooses smaller of the two SVD values
41     end
42     score33=score33+score3;
43     end  %End of SVD
44     scorewe=[mass(x, :), score1*score33/score2];
45     score(x, :)=scorewe;  %Creates a matrix of the scores with the m/z values
46     end
47
48 %DATA WORK-UP
49
50 sumss=sortrows(score, 2);  %Sorts algorithm scores
51 sums=flipud(sumss);  %Ranks the scores with the largest at the top
52 normz=sums(:, 2)/sums(1, 2);  %Normalizes the algorithm scores to 1
53 alg=[sums(:, 1), normz];
54 format short
55 alg;  %This is the ranked, normalized scores
56
57 %END OF ALGORITHM SECTION OF CODE
58
59 %STEIN CODE BEGINS HERE
60
61 %DATA WORK-UP
62
63 format long
64 a=[score(:, :), refdata(:, :)];  %Creates a matrix of m/z, scores, and reference data
65 b=sortrows(a, 2);
66 c=flipud(b);  %Matrix is now sorted with largest on the top
67
68 yy=1;
69 for j=3:comps+2  %Cycles through the components
70     for k=j+1:comps+2  %Cycles through components comparing them to j
71         sum1=0;  %Resets the sum
72         sumx=0;  %Resets the sum
73         sumy=0;  %Resets the sum
74         for y=1:mas
75             level1=c(y,j)*c(y, k);
76             srlevel1=level1^0.5;
77             sum1=sum1+srlevel1;  %Numerator summation for Stein score
78             sumx=sumx+c(y, j);  %Denominator summation for Stein %score
79             sumy=sumy+c(y, k);  %Denominator summation for Stein %score
80             den=sumx*sumy;
81         end
82     end
83 end
84
if den==0  %This keeps from dividing by zero
    stein=0;
else
    stein=sum1^2/den*100;
end
steinc=[c(y, 1), stein];
stein1(yy, :)=steinc;
yy=yy+1;
end  %End of individual stein scores
end  end

%Summation of each Stein score
count=1;
for f=1:mas
    summ=0;
    r=0;
    for g=1:yy-1
        if stein1(f,1)==stein1(g,1)
            summ=summ+stein1(g, 2);  %Combines the individual Stein
            % scores
            r=r+1;
        end
    end
    summ=sum/r;  %Normalization of the Stein score
    steinf(count, :)=stein2;
    count=count+1;
end
%END OF STEIN SECTION OF CODE

format short
stein3=[steinf(:, :) , ty(:, :) ];  %Combines the algorithm and Stein scores in a matrix

figure(2)  %Plots the Stein score
plot(stein3(:, 3), stein3(:, 2))  %Plots the Stein score
xlabel(‘Number of masses added for algorithm ranked spectra’)
ylabel(‘Stein index score’)
title(‘Stein Score’)

%Prints the variance information
all=[alg(:, :) , steinf(:, 2)];  %Combines the algorithm and Stein scores in a matrix
128 disp('Algorithm Steinh ‘
129 disp('Mass Score Score ‘
130 disp('---------- ---------- ---------- ‘
131 format=' %3.0f %3.7f %7.3f ‘
132 for ti=1:max
133 tab=sprintf(format, all(ti, :)); disp(tab)
134 end

Line 1 defines the function which has been named cookind6a for Matlab. Lines 2-13 are comment lines explaining the use of the function. Line 15 formats the numbers to be outputted. Lines 17-18 convert the user input mass and reference spectra into variables for the calculation. Lines 19-20 determine the number of masses and the number of components in the data set. Line 20 is a comment line for the start of the algorithm code. Line 24 starts a loop that cycles through each mass. Line 25 starts a counter for data storage. Line 26 chooses the maximum intensity for each mass. Line 27-28 resets the previous scores to zero. Lines 29-32 determine the ratio of the ion intensities of each component's intensity over the maximum intensity for each mass. Line 33 starts a loop to score each ion pair with the singular value decomposition (SVD). Lines 34-35 give a score of zero when the loop compares the ion being scored in the loop to itself. Lines 36-41 calculate the SVD score for the ion pair. Lines 42-43 add all the SVD scores for each ion. Line 44 gives the total score for the ion. Line 45 stores the score in a matrix. Lines 46-49 end the calculation of the empirical algorithm and start the data workup. Lines 50-52 sorts the algorithm scores and normalizes them to the highest score. Lines 53-55 store the normalized algorithm in a matrix. Line 57 ends the scoring of the empirical algorithm. Lines 59-61 start the Stein similarity index scoring routine. Line 64 places into one matrix the empirical algorithm scores (unsorted) and the reference matrix of pure
components. Lines 65-66 sorts the new matrix of algorithm scores and pure component spectra by the empirical algorithm score, with the ion with the largest score listed first. Line 68 starts the counter for data storage. Lines 69-70 start the loop for scoring of the similarity index. Lines 71-73 set initial similarity index scores to zero. Line 74 starts the similarity index scoring loop for each mass compared pairwise with each component. Line 75 multiplies the two intensity values at the given mass. Line 76 takes the square root of the product of line 75. Line 77 then adds the value from line 75 to the previous values from line 75. Lines 78-81 sum the individual intensities for a given mass. Line 82 multiplies the two values from lines 78 and 80. Lines 83-84 give an ion a Stein score of zero if line 82 equals zero. Lines 85-87 calculate a similarity index score based on the values calculated in lines 77 and 82. Lines 88-89 place each mass with its calculated similarity index score in a matrix. Line 90 adds one to the counter variable yy. Line 91 ends the loop that cycles through each mass. Lines 92-93 end the loops that cycle through each component pairwise. Line 95 starts the summation of each pairwise comparison at each mass. Line 96 starts a counter variable. Line 97 starts a loop to cycle through each mass. Line 98 sets each mass initial similarity index score to zero. Line 99 starts a counter variable at zero. Line 100 starts a loop that cycles through all pairwise mass scores. Line 101 compares the mass of the current ion being summed to the mass from line 100. Lines 102-103, if the two masses match then the Stein score is added to the total similarity score for that mass. Line 104 increases the counter variable in line 99. Lines 105-106 end the comparison and addition of similarity index scores at each mass. Line 107 takes the average of the summed similarity index scores. Line 108 places the mass and the associated similarity index score into a matrix. Line 109 places the values
from line 108 into a larger data storage matrix. Line 110 increases the counter loop for
data storage. Line 111 ends the summing of the Stein scores for each pairwise
comparison. Line 113 ends the Stein similarity index portion of the code. Line 116 fills
a matrix with similarity index scores for plotting. Lines 117-121 plot the similarity index
scores to the screen. Line 124 brings into one matrix the empirical algorithm scores and
the Stein similarity index scores for output to the screen. Lines 123-134 format
command for the printout of the mass, algorithm score and Stein similarity index score to
the screen.

An example of the isobutene isomer data containing the masses m and the pure
component spectra d (components in order: isobutene, 1-butene, cis-2-butene, and trans-
2-butene) is shown below. Both should be in the form of text (.txt) files. For text file m,
each line will contain a m/z.

14
15
25
26
27
29
37
38
39
40
41
42
50
51
53
54
55
56
57
For text file d, each row contains the intensities for a m/z where the columns are the components. The rows in the text file are tab separated so that MATLAB will read the intensities in matrix form.

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.913</td>
<td>0.565</td>
<td>1.047</td>
<td>0.564</td>
<td></td>
</tr>
<tr>
<td>2.453</td>
<td>2.240</td>
<td>2.706</td>
<td>2.195</td>
<td></td>
</tr>
<tr>
<td>1.059</td>
<td>1.452</td>
<td>0.686</td>
<td>1.493</td>
<td></td>
</tr>
<tr>
<td>8.400</td>
<td>10.171</td>
<td>4.524</td>
<td>10.096</td>
<td></td>
</tr>
<tr>
<td>24.674</td>
<td>26.805</td>
<td>16.961</td>
<td>27.336</td>
<td></td>
</tr>
<tr>
<td>11.487</td>
<td>13.717</td>
<td>9.373</td>
<td>17.487</td>
<td></td>
</tr>
<tr>
<td>2.824</td>
<td>2.860</td>
<td>4.054</td>
<td>2.763</td>
<td></td>
</tr>
<tr>
<td>4.360</td>
<td>4.187</td>
<td>6.476</td>
<td>3.998</td>
<td></td>
</tr>
<tr>
<td>35.082</td>
<td>36.095</td>
<td>45.447</td>
<td>36.447</td>
<td></td>
</tr>
<tr>
<td>7.172</td>
<td>7.535</td>
<td>11.230</td>
<td>7.096</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>3.399</td>
<td>3.348</td>
<td>3.648</td>
<td>3.352</td>
<td></td>
</tr>
<tr>
<td>5.429</td>
<td>7.612</td>
<td>6.092</td>
<td>7.801</td>
<td></td>
</tr>
<tr>
<td>4.069</td>
<td>5.850</td>
<td>4.560</td>
<td>5.943</td>
<td></td>
</tr>
<tr>
<td>5.393</td>
<td>7.998</td>
<td>5.123</td>
<td>8.576</td>
<td></td>
</tr>
<tr>
<td>1.920</td>
<td>3.007</td>
<td>1.755</td>
<td>2.953</td>
<td></td>
</tr>
<tr>
<td>35.108</td>
<td>44.927</td>
<td>40.352</td>
<td>44.684</td>
<td></td>
</tr>
<tr>
<td>1.550</td>
<td>1.969</td>
<td>1.800</td>
<td>1.946</td>
<td></td>
</tr>
</tbody>
</table>
VITA

Eloise Marie (Raby) Joyce, daughter of Edward and Deborah Raby, was born in Bucyrus, OH on March 19, 1980. She graduated from Bucyrus High School in Bucyrus, OH in May of 1998. Upon graduation, Eloise attended Bowling Green State University in Bowling Green, OH and received a Bachelors of Science in chemistry and mathematics in May of 2002. She stayed at Bowling Green State University as a graduate teaching assistant for a year in mathematics. In the fall of 2003 she entered graduate school at the University of Tennessee, Knoxville. She held a teaching and research assistantship throughout her time there. While at the University of Tennessee, she was involved with the Tennessee State Science Olympiad and the Chemistry Camp for middle school students.