12-1999

Analysis of mixtures containing isomeric hydrocarbons using process mass spectrometry

Martin L. Haddix

To the Graduate Council:

I am submitting herewith a dissertation written by Martin L. Haddix entitled "Analysis of mixtures containing isomeric hydrocarbons using process mass spectrometry." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Chemistry.

Kelsey D. Cook, Major Professor

We have read this dissertation and recommend its acceptance:

John E. Bartmess, Earl L. Wehry, Mary G. Leitnaker

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Martin L. Haddix entitled "ANALYSIS OF MIXTURES CONTAINING ISOMERIC HYDROCARBONS USING PROCESS MASS SPECTROMETRY". I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements of the degree of Doctor of Philosophy, with a major in Chemistry.

Professor Kelsey D. Cook, Major Professor

We also have read this dissertation
and recommend its acceptance:

Professor John E. Bartmess

Professor Earl L. Wehry

Professor Mary G. Leitnaker

Accepted for the Council:

Associate Vice Chancellor and Dean
of The Graduate School

Analysis of Mixtures Containing Isomeric Hydrocarbons using Process Mass Spectrometry

A Dissertation

Presented for the

Doctor of Philosophy Degree in Chemistry

at

The University of Tennessee-Knoxville

Martin L. Haddix

December, 1999.

This dissertation is dedicated to my parents,
Leo L. and Coralee A. Haddix
for all of their support, understanding and encouragement
over the past several years.

ACKNOWLEDGMENTS

There are many individuals who have contributed to the success of my research. I will acknowledge a few of them in the following paragraphs. I thank my research advisor, Dr. Kelsey D. Cook, for the opportunity to pursue my interest in mass spectrometry at The University of Tennessee-Knoxville (UTK). His guidance and financial support is appreciated over the past several years. I feel that he has helped to provide me with a well balanced education regarding the aspects of becoming a good research scientist. I also appreciate the hard work and dedication from staff and faculty in the department of chemistry at UTK. Special thanks go to Dr.'s John Bartmess, Al Tuinman and Gerry Seebach for their many helpful discussions. As director of the chemistry department's Center for Mass Spectrometry, Dr. Tuinman performs an excellent job of teaching the graduate student instrument operators the nuances of instrument maintenance, troubleshooting and general mass spectrometry. Professor Seebach has been a summer faculty visitor from Transylvania University in Kentucky since the beginning of this work in June of 1994. His help in starting the research discussed in this dissertation is gratefully acknowledged.

Also, thanks go to the friends I made at UTK. Their friendship seemed to make the time go much more quickly. Special thanks also to past and present members of the Cook research group, including; Mike Troutman, Bill Boggess, Mona Shahgholi, Scott Kriger, Steven Salvato, Shaolin Zhou and Kevin Bennett. Kevin's work on the Empirical Algorithm listed in this dissertation is especially noted.

Financial support was provided in part by grants from the University of Tennessee

Measurement and Control Engineering Center (an NSF Industry/University Cooperative Research Center) and the UTK chemistry department mass spectrometry center which is funded by the Science Alliance, a State of Tennessee Center of Excellence. I extend a special acknowledgment to Greg Cessna and Dwain Waguespack at ABB-Extrel for their generosity in making available to our research group the process mass spectrometer with which all of the mentioned research took place. Brooks / Rosemount and Union Carbide also contributed instrumentation which was much needed. Financial support from all of these agencies is gratefully acknowledged.

ABSTRACT

The feasibility of process mass spectrometry (PrMS) as a method for the rapid analysis of a piperylene stream is assessed beginning with a focus on resolving binary mixtures of isomeric hydrocarbons. This application was suggested by an industrial member of the University of Tennessee's Measurement and Control Engineering Center (MCEC). The piperylene stream consists mostly of cis- and trans-1,3-pentadiene (totaling over 55% of the stream) along with cyclopentene (~20%) and 2-methyl-2-butene (~10%), plus others at lower levels: cyclopentane, cis- and trans-2-pentene, *n*-pentane, isopentane, and 2,2-dimethylbutane. This particularly challenging PrMS application involved the need for resolution of hydrocarbon isomers which were constituents in the stream. This required the design and construction of a vaporization inlet since these components are liquids at standard temperature and pressure. It was found that parameterization (selection of which m/z signals to monitor in the deconvolution of the mixture mass spectra) was critical to the accuracy and precision of the mixture component concentration estimates. Use of the entire mass spectrum (an intuitive approach to parameterization) was found to be non-optimum since inclusion of some m/z signals deteriorated the accuracy and precision. A brute force approach to parameterization (whereby the accuracy is assessed using all possible combinations of ion intensities) was found to afford the most accurate concentration estimates but was useful for small datasets due to the large number of calculations required. The brute force method of parameterization was suspected of over-fitting the data due to the relatively small datasets utilized. In response to the large amount of calculation time required for the brute force method, other faster parameterization methods were evaluated. These

faster parameterization methods included a genetic algorithm, an empirical algorithm and an algorithm (process stream evaluator) provided with the instrument data acquisition system. The genetic and empirical algorithm provided parameterizations faster than the brute force method. However, the accuracy and precision of concentration estimates were not as good. The success or performance of mixture deconvolution was measured using validation plots and correlation coefficients (r-squared). Validation plots are graphs of estimated versus true mole fraction of gravimetrically prepared standard samples and the correlation coefficient is a number from 0 to 1 describing the accuracy and precision of estimated concentrations where a correlation coefficient of 1 is a perfect fit to the Y = X line. An r^2 of 0.9963 was obtained from analysis of binary mixtures of *n*-pentane / isopentane using the brute force method of parameterization. Similarly, correlation coefficients of 0.9054 and 0.9986 were obtained for binary mixtures of cis- / trans-piperylene and cis- / trans-2-pentene. The concentrations of these samples ranged from approximately 0.10 to 0.90 mole fraction. The lower correlation coefficient for the piperylene system was attributed to the greater similarity between the pure component mass spectra (reference spectra) of these isomers and their relatively greater chemical reactivity. The similarity of the reference spectra was measured using the Drahos-Vekey similarity index equation and the reactivity was compared by considering the double-bonds in the molecule. Finally, a nine-component standard mixture similar in composition to a piperylene stream grab sample was analyzed by GC-MS and PrMS. The additional complexity incurred by increasing the number of components to a total of nine required that the stereoisomer pairs each be treated as one quasi-component. In other words, to achieve a reasonable accuracy and precision each pair of cis- / trans- isomer

reference spectra were averaged and considered as one component leaving a total of seven components. Upon reducing the mixture complexity to a total of seven components it was found that cis- / trans-piperylene, 2-methyl-2-butene and cyclopentene were among the major components in the piperylene stream that could be quickly analyzed with reasonable accuracy and precision by PrMS. The remaining components were at levels that appeared to be present below the limit of detection.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION TO PROCESS MASS SPECTROMETRY

1.1: ECONOMIC BENEFITS OF ON-LINE ANALYSIS

The chemical industry is a multibillion dollar-per-year business where billions of pounds of many different kinds of products are manufactured annually [1]. The handling, transport and measurement of very large amounts of starting materials and products must be performed safely and efficiently if the manufacturing process is to be profitable. In the chemical processing plant an improvement in the cost-efficiency of the production method or "shaving-off" as little as one-tenth of a cent-per-pound to conserve starting materials translates into millions of dollars saved. An ultimate objective of a chemical process plant is to produce products of a specified quality. Chemical processing plants are expensive to construct and maintain. If the final products are intended to afford a significant profit they must meet market purity and quality standards, otherwise they will not be saleable. If the products are not saleable at a reasonable profit the cost required to construct and maintain the chemical processing plant will not be recovered. Conversely, manufacturing an excessively pure or high quality product will incur unnecessary cost and thus reduce profit.

One can conceive of a continuous chemical process plant as a collection of tanks, in which materials are heated/cooled, and reacted/stored with several connecting pipes through which the materials flow. Typically such a processing system will not naturally

maintain itself in a state such that the optimum temperature is achieved, the pressure remains below a safe limit or a flowrate of materials is kept in which an economically optimum product composition is achieved. Thus accurate and precise chemical measurements of feedstock, intermediate streams and finished products are key to monitoring and controlling these processes economically [2]. In some applications, dynamic product quality control, through on-line, real-time analysis, can significantly reduce operating costs and improve profitability [2]. For example, operating costs can be reduced through the elimination of reprocessing or further downstream processing of off-specification products or by blending off-specification products, which generally requires intermediate storage facilities and the sale of products at reduced prices because of reduced quality. Profitability can be directly improved by maximizing the production of the highest-priced product within its quality specification limits. Overall, a chemical processing plant is a dynamic system which requires a dynamic monitoring and control system. The use of continuous on-line, real-time monitoring of the product and the use of a dynamic process control system to maintain the product within those guidelines can improve the cost-efficiency of the process as well as safety of the personnel who operate the instrumentation and the safe operation of that instrumentation.

In some chemical process plants, the basic information obtained from pressure, temperature and flow measurements is sufficient to afford economical, cost-efficient control of the process. One example of simple flow monitoring is in the operation of a 400 megawatt coal-fired power generating facility [3]. Here, monitoring the fuel/air mixture ratio (a basic flow measurement) to the steam boiler that participates in the operation of the electrical generators afforded information regarding the heat rate of the

steam boilers. This monitoring information allowed adjustments in the heating process such that a 1% fuel savings of nearly $500,000 per year could be recognized [3]. However, for other applications, in order to insure that some product falls within the specified criteria stated by the consumer, an on-line, real-time analyzer may be the most reasonable solution [4]. A specific example in the coal mining industry involves utilization of an on-line analyzer to select differing grades of coal for washing [5]. The strategy of by-passing the washing step for high-grade coal affords an increase in profit since washing involves loss of up to 5% of the material. This step of "scalping away" part of the feed affords the following advantage. The feed rate to the washing mechanism can be increased thus allowing processing of the low-grade material at a faster rate, thus reducing the marginal operating costs. The reported savings from this application of on-line analysis approached 1.6 million dollars per year [5].

Another example of improvement in processing efficiency using on-line analysis is in the production of amines [6]. In this case, the process of purifying the amine via distillation (re-generation of the amine) to a higher quality than necessary was avoided by using an on-line analyzer. In this chemical processing example the amine samples were approaching 0.005 to 0.008 moles of an acid-gas (a contaminant) per mole of amine. The acceptable specification limit of this impurity was an order of magnitude larger than that obtained. By cutting back on the re-generation process and maintaining a consistent and higher (but tolerable) level of impurity in the product, a cost-improvement of 1.1 million dollars per year was obtained [6]. Basically, disturbances and uncertainties in process parameters during processing operations offer opportunities of producing unsatisfactory product or unsafe process conditions and warrant appropriate monitoring and control [7].

1.2: WHY REAL-TIME ANALYSIS?

The primary need for real-time feedback in a chemical processing plant arises from the attempt to make the chemical processing more cost efficient. For example, one method that is used to improve the control in a chemical processing plant is to provide plenty of storage capacity especially of intermediates between processing steps. This requires that key items of equipment be designed to take on a much larger capacity than is required (*i.e.*, these items are "over-engineered"). An over-sized intermediate storage container between a reactor and, for example, a separation process downstream, makes the chemical processing plant easier to control because the large storage tank acts as a "time buffer" and affords a smoothing or averaging effect on fluctuations in the chemical processing procedure. This provides the process control system "thinking time" and allows adjustments to be invoked in the chemical processing procedure [8].

However, large intermediate storage tanks can be a significant capital investment and the resources with which to fill them can be another non-trivial monetary burden especially when processing high-value materials. In a gasoline-blending application of on-line analysis a reduction of inventory (storage tank material) was realized due to the ability to blend in-line as opposed to blending by batch mode. Batch-mode blending (optimization of the octane-rating of gasolines by blending higher octane material with lower octane material in a non-continuous fashion) required time to recirculate the materials in the large tanks, acquire the analytical data and re-blend when necessary [9]. This is an additional cost in terms of time, storage capacity and energy to circulate the material in the tanks.

More importantly, safety issues arise when processing excessively large volumes of toxic or flammable materials. Therefore, the tendency to cut back on the amount of storage capacity is an important issue regarding the processing cost and the safety of the chemical processing plant. This approach of cutting back on storage capacity forces the process control system to react more quickly and thus requires a monitoring system that can send real-time or near real-time data to the process control system allowing more time for adjustments to be made to the chemical processing procedure.

In some applications of on-line analysis, safety is of higher priority than profit. Such is the case for a process stream in which the product is an organic peroxide [10]. In this example the possibility of a "run-away" reaction is avoided by using an on-line mass spectrometer [10]. Here an exothermic process which has the potential to self-accelerate to a point where the cooling capacity of the system is exceeded can create a dangerous increase in the temperature of the reactants. This leads to a higher number of moles of reaction product and an increase in pressure (reaction run-away) such that the limit of the tensile strength of the reaction vessel walls is approached. Thus, the possibility for an explosion is increased and a danger to personnel as well as process equipment is a significant threat. It was found that the use of real-time on-line analysis afforded a greater state of safety (to the process personnel as well as the process equipment) compared to the conventional method of temperature monitoring due to a considerably earlier detection of the onset of reaction run-away. Overall, the main purpose of rapid analysis in this case is to provide increased safety to process personnel as well as processing equipment, to maximize profit and to afford reasonable environmental safety.

The latter may also involve monitoring cooling or waste water discharge into the

environment, although in some cases where an on-line analyzer is employed it is possible to have too much knowledge about the process stream being monitored [11]. For example, in one particular installation of an on-line analyzer the instrument indicated low-level (ppm) presence of hazardous materials that were previously undetected in the chemical processing stream waste [11]. Since these dangerous materials were now known to exist in the processing stream, precautions to remove them and follow proper disposal procedures were required. By installing the on-line analyzer the cost of production had increased due to the need to remove the low-level hazardous materials in the process stream waste. The outcome in this case was contrary to the intent of installing the on-line process stream analyzer in the sense that the cost of producing the product increased rather than decreased.

1.3: TYPES OF ON-LINE PROCESS ANALYSIS

Process analyzers have been in use in the chemical industry for a number of years. Some of the earliest applications of process analysis in industry date back to the 1940's and 1950's [12][13][14][15][16][17][18][19][20]. A process analyzer offers a great deal of additional information as compared to process monitoring utilizing basic temperature, pressure and flow sensors [21][22][23]. During the 1940's and 1950's different instrumentation was tested for use as process analyzers. Sampling the process stream was an issue of debate discussed extensively at that time and remains so today. Deciding which type of analyzer to utilize requires consideration of the chemical system to be monitored and controlled as well as the technology that is available and affordable. Some

process analyzers lend themselves more readily to a particular chemical process stream than others. For example, sample issues such as product reactivity, volatility, viscosity and toxicity may have to be considered in the sampling step or introduction of the sample to the analyzer [24]. Sampling is addressed separately in a later section of this chapter. Also, location of the analyzer in relation to the process stream is often of importance with respect to the smooth operation and maintenance of the analyzer. There are a number of different process analyzers and analytical methods in use today which include a wide range of chromatographic, electrochemical and spectroscopic methods [25].

Process analyzers can be located off-line, at-line, on-line and in-line with respect to the process stream [25]. Sample removal from the process stream and transport to the instrument is required for the first situation. Instruments located off-line do not require enclosures to protect them from possible harsh conditions (high humidity, caustic or flammable atmospheres, etc.) that can be present on-line. Drawbacks regarding off-line analysis include a delay between sample acquisition and reporting of results. This leads to the at-line situation where an instrument is installed at a location that is close to the process stream allowing faster sample transport without the complexity of an on-line instrument. The sampling process can be automated to increase speed if a large number of samples per time are taken or a large number of points on the chemical processing stream are tested. This can afford the advantage of allowing the time-sharing of expensive instrumentation as well as expert staff for consultation between the process stream and, for example, a pilot plant stream. However, analysis cycle times can still be relatively long as in the application of gas chromatography (GC) for benzene, which required a 25 minute cycle time [26]. Hence, the off-line and at-line approaches can be

relatively slow although sampling automation can improve the speed of both.

An on-line process analyzer can be a complex instrument. It typically includes an automated sampling system which removes the sample from the process stream, conditions it and presents it to the analytical instrument for measurement [27]. An inconvenience associated with the on-line approach is the need to construct distinct analytical lines that properly sample the main stream, filter and present it to the instrument at a suitable temperature and pressure. This has led to the in-line approach to process analysis where the chemical analysis is done in-situ or inside the chemical process stream using a probe. This might include utilization of ion-selective electrodes or microsensors using field-effect transistor technology.

One final area involves analysis of the process stream in a non-invasive manner. The analytical methods in this arena include, but are not limited to, infrared absorption, infrared emission, microwave and X-ray spectroscopy. This approach to process analysis is reported to be "the ultimate in desirability" [25] since there is no physical contact between the instrument and the sample. This sampling arrangement reduces instrument/sample interactions and thus diminishes problems related to, for example, fouling of the sampling system. Often when problems such as the latter are reduced or eliminated, an improvement in the sampling reproducibility may be realized. However, problems can still arise in the spectroscopic methods such as fouling of the optical windows which lead into the process stream. Also, particulates and suspended solids (non-homogeneous samples) can be a problem in these methods. Overall, each sampling classification mentioned above has its own caveats and must be considered before the particular chemical process stream is to be monitored.

1.4: SAMPLING

In on-line process analysis, as in most of analytical chemistry, sampling is very critical and often the least accurate step in the entire method. It is estimated that 80 - 90% of all the maintenance problems experienced by online analyzers occur at the sampling system interface to the process analyzer [28]. Some of the most challenging systems to analyze involve liquids at room temperature and pressure [24]. For example, difficulties can arise from irreproducibilities incurred from: sample condensation at "cold spots" in the sampling system, solids forming from precipitation reactions at sampling system surfaces, or sample dimerization at areas in the sampling system where the temperature is elevated with the intent of preventing sample condensation.

Currently, there are two fundamentally different sampling approaches for such liquid systems: selective and non-selective sampling. In the former, the majority of the process stream is excluded by the interface. A good example of this is the analysis of organic contaminants in water. In this case water is selectively excluded by the sampling interface and the organic compounds are enhanced. Silicone rubber membranes are a common interface material for these types of applications since organic compounds are soluble in the rubber membrane and water is not. An example of a sampling system customized for the analysis of samples containing water can be seen in a sampling inlet which was designed to incorporate a shutter/chopper [29]. In this case signal lock-in techniques could be used and background signal could be readily identified.

The non-selective approach can be exemplified by process streams where a small portion of the total stream is analyzed and, ideally, there is little, if any, selectivity

imposed by the sampling interface. Thus the prime requirement is to transfer the liquid stream as reproducibly as possible to the analyzer [30]. The sampling accommodations that are taken for these types of systems include analyzing them via a direct analysis of the liquid (*e.g.* near-IR [31]), via their headspace, or via a total vaporization technique where the sample is completely volatilized [32][33]. An example of the latter case can be seen where modifications were made to a "liquid-sampling" valve such that it could be interfaced to the analyzer when a pentachloropyridine processing stream was examined [34]. In this case no sample dilution was necessary since a very small (μL) amount of sample was injected. This attribute of the sampling system was considered to be a major advantage compared to a sampling system utilizing a membrane interface since the amount of sample to be vaporized could be directly controlled whereas in the latter it is much more difficult if not impossible to control, dynamically, the amount of sample that permeates the membrane interface. Also, in the analysis of a liquid stream at a synthetic rubber manufacturing facility, sampling the stream in a reproducible manner was achieved by employing a heated valve system [35]. In this case the advantage of using this particular system was that a controlled amount of sample, metered by an injection device, could be volatilized and presented to the analyzer in a reproducible manner.

It has been said by experts in the profession of analyzing chemical processing streams that "the cleaner a sample, the better the process analyzer will perform" [36]. This idea has led to the development and testing of sampling techniques that extract, filter, cool and condense a sample and control its flow and pressure [37]. However, one should be cautious about any analytical technique that requires too sophisticated a sampling technique to produce accurate and precise results.

There has been a growing demand for analytical methodologies aimed at monitoring non-volatile materials such as additives, surfactants and biochemicals including proteins, hormones and antibiotics produced in the biotechnology process industry. It is anticipated that these relatively large molecules (with kilodalton molecular weights) possessing low volatilities can be analyzed on-line in the near future using "sampling" techniques such as electrospray ionization with mass spectrometry [38]. Overall, it is apparent that sampling is a very important issue in different applications of process analysis with an emphasis on the appropriate sampling of liquid mixtures in process streams.

1.5: PROCESS GAS CHROMATOGRAPHY

In the manufacture of many different types of products there are volatiles which are directly or indirectly related to the individual component compositions in the chemical process stream. Hence, an analysis of the volatiles or headspace can be a good measure of the composition of the process stream. Real-time information regarding the individual component compositions in the chemical process stream can be readily utilized to inform the dynamic control system of what adjustments need to be invoked to afford the highest quality product. Thus, information vital for process, quality, and/or waste control can be derived from continuous, on-line, real-time monitoring of volatiles in areas such as the headspace of a pressurized reactor, or within a fluid process or waste stream [35].

Process gas chromatography (PrGC) has traditionally been the method of choice

for process analytical work involving volatile compounds [39][40][41][42][43]. Relatively simple and robust chromatographic equipment can be adapted for determination of a wide range of volatiles. A PrGC with a conventional detector can be applied to the analysis of a large variety of complex mixtures of volatiles after optimization of the GC method. In one application of PrGC a complex mixture that resulted from the catalytic cracking of *n*-heptane over a zeolite catalyst was analyzed. High-resolution gas chromatography was required to resolve some components and a flame ionization detector was used to quantify the separated constituents. Some components in the mixture, such as tri-substituted naphthalene, demonstrated retention times of greater than 60 minutes [40]. This relatively long analysis time can be an issue in the control of some process streams.

Recent developments have shown that the speed of GC can be dramatically increased by the use of special inlet systems (such as cryogenic focusing sample collection) with relatively short capillary columns operated at unusually high carrier gas flow rates. Using these techniques, mixtures containing volatile organic compounds (VOC's) can be separated and analyzed in a matter of several seconds [44][45][46]. This represents an increase in the sampling throughput of 1 to 2 orders of magnitude [44]. This technique has been termed "High-Speed GC" or "Fast-GC" [47][48].

Recently, in response to regulatory issues, an analytical system was tested at stack emission sites to monitor polycyclic hydrocarbon effluents by fast-GC [49]. Detection limits at the ppb level were demonstrated with analysis times of 5 minutes. In an additional application of fast-GC, 41 different VOC's were monitored in air at concentrations as low as 0.1 ppb. Analysis times ranged from 5 to 30 seconds for

differing mixtures of the VOC's [50]. Some caveats associated with this technique include the inability to analyze methane due to its low adsorption coefficient on the short chromatographic column [50]. Also, the requirement of injection of a very narrow (short time) sample plug can be a complex issue if the proper equipment is not available. Often, this is satisfied by using a cryofocusing sample injector which also imposes a requirement for liquid nitrogen or other cryogenic material [44]. Although this requirement may not add significantly to the monetary cost of the analysis method it does add to its complexity [50].

There are, however, at least two potentially serious limitations to GC. First, the qualitative information that is available in the retention time data recovered from a chromatogram can be challenging to use as an objective proof of component identification. A match between standard and analyte retention times may be considered evidence affording an identification, but the chromatogram provides a limited amount of information useful for identifying a component eluting at an unexpected retention time. Furthermore, while seldom a problem in a well-regulated process, the possibility of co-elution of an unexpected component generally cannot be ruled out, even in cases where there is good match with an expected retention time.

A second limitation intrinsic to the separation step in GC arises from the "batch" nature of sampling. Although elution conditions can be optimized for analysis of the components of interest, it is usually necessary to wait for elution of all peaks from one chromatographic injection before initiating a second injection. As a result, the PrGC is not really a *continuous*, on-line analyzer. This can be a serious limitation in cases where continuous, real-time process analysis is particularly critical [10].

1.6: PROCESS MASS SPECTROMETRY

Mass spectrometry was initially used for process analysis in the 1940's [19][20]. By the 1950's its use was more fully documented [12][13][14][15][16][17][18]. The relative number of process stream monitoring applications using gas chromatography in the 1950's increased in industry on the grounds that it was of greater reliability and lower cost than mass spectrometry [51]. Hence, the relative importance of GC increased. The preference for GC over MS at that time was due to the lack of low cost, high-speed electronics (*e.g.*, cheap, fast computers) and reliable vacuum systems for MS. Thus, there was a lack of appropriately equipped mass spectrometry based process analyzers available. Improvements in these areas caused the number of PrMS applications to increase.

In the 1970's, the use of mass spectrometry as a process stream monitoring method became more popular for the control and automation of large-scale industrial processes such as in the iron and steel industry [52], in the manufacturing of polymer intermediates [53][54], and in the analysis and monitoring of off-gases produced from the processing of coal [55]. The increase in PrMS applications is also evident in the brief reviews and application notes in the area [56][57][58][59][60]. PrMS reached a stage where it is the preferred technique employed for a number of applications [61][62][63][64]. The use of PrMS and its caveats have been reviewed by our laboratory [65] as well as others [66][67]. As illustrated in the previous references, PrMS has been utilized to perform process analysis in the steel, coal, glass, semiconductor and chemical industries with ethylene oxide, ammonia, vinyl chloride, cracker gas and fermentation

processes. In some diverse applications, mass spectrometry has been used to probe the Venusian and Martian atmospheres [68][69][70] and has monitored the working environment of nuclear submarines [71]. Often mass spectrometry is used in environmental monitoring where sensitivity and selectivity are of prime consideration.

One important factor that has precipitated this diversity of chemical analysis using PrMS has been the availability of low cost, dependable and fast computers. With more powerful and lower cost computers, mass spectrometry can be automated and made to be more "user-friendly". In one application of process analysis, hand calculations were utilized to analyze the data from a heat exchanger to determine the required frequency of its mechanical cleaning [8]. In this case the calculations that were performed by hand were replaced by a much faster computer. Also, with the advent of faster computers, the use of non-linear analytical methods of data analysis, such as partial least squares and artificial neural networks, is more popular as these methods require the analysis of relatively large data sets and thus can be handled quickly and easily by a fast efficient computing system.

In addition to faster computers process mass spectrometers have been adapted for on-line applications by protecting the electronics from corrosive or flammable gases from the industrial processing streams by housing the instrumentation in appropriately sealed enclosures.

Like PrGC, mass spectrometry is inherently a gas sampling technique. In other words, the analyte under investigation must ultimately be in the form of a gas to be analyzed by the mass spectrometer. PrMS in the semiconductor industry, although now commonplace, has generally been restricted to measurements of background residual gas

levels and routine analysis of vacuum integrity (leak checking). Presently, the move in this particular industry is to reduce the feature sizes that are lithographed upon silicon wafer products [72]. This imposes stringent requirements upon the vacuum hardware and gas purity and requires that trace contaminant gases and their source be monitored closely [73]. Contaminants such as water vapor, chlorine and oxygen from unpurged gas lines, impure gas supplies or compromised integrity of vacuum fittings, valves and regulators can seriously upset process conditions. These problems may cause changes in process chemistry and ultimately lead to costly reductions in the wafer quality. The ability to identify these problems as early as possible (real-time analysis) will minimize the number of additional process steps needed to repair defective wafers and will reduce the cost of processing.

Mass spectrometry is a good choice for volatiles analysis and PrMS can offer a faster, more flexible alternative for continuous, real-time, on-line monitoring of volatiles in a process stream as compared to PrGC [74]. In PrMS, sample introduction and ionization (usually by "electron ionization" involving bombardment of sample vapor with high-energy electrons) can be continuous, and the rapid mass spectral scan times that are readily attainable (on the order of a few seconds per scan or less) enable virtually instantaneous or real-time analysis. Watson [75] illustrates, in **Figure 1.6.a**, how a gas phase analyte is analyzed by mass spectrometry. In part A the analyte is bombarded by high energy electrons (typically in the region of 70 - 100 electron Volts (eV)) which imparts enough energy into the molecule to invoke fragmentation which is illustrated in part B. The fragments produced are filtered or separated by the mass spectrometer based upon their mass-to-charge-ratio (m/z), detected and recorded (part C of **Figure 1.6.a**). A

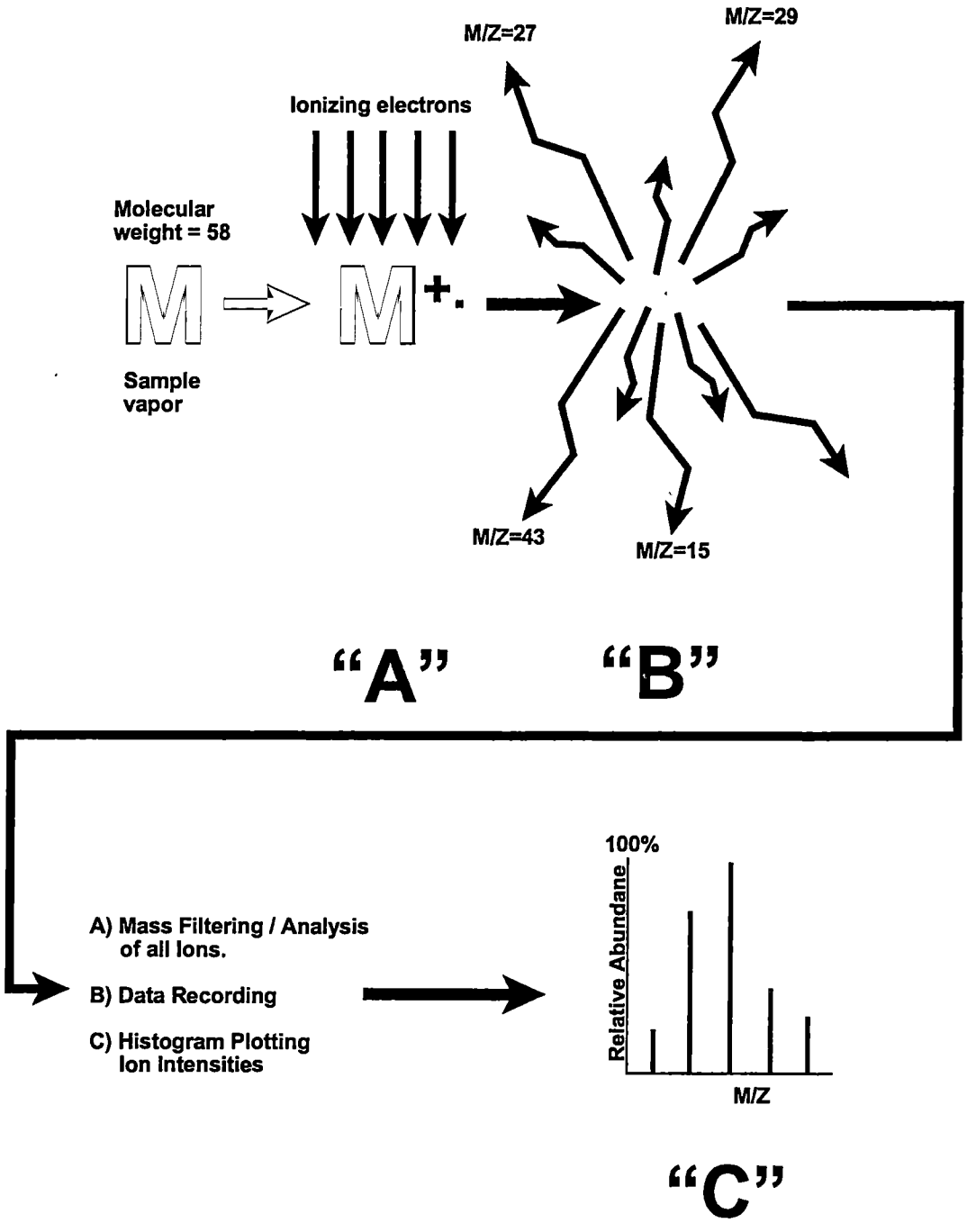


Figure 1.6.a: Illustration of analyte ionization, fragmentation and data recording to afford a mass spectrum. Taken from reference [75].

"fingerprint" of the pure component or mixture is produced as a mass spectrum which is usually presented as a histogram of intensity or abundance versus m/z. Typically, the m/z ratio with the highest intensity (the "base peak") is arbitrarily assigned an abundance of 100 and the remaining m/z signals are normalized to the resulting "base peak".

In many cases, the improvements in process efficiency attainable by real-time feedback more than offset the extra cost and sophistication of the PrMS analyzer (compared with the PrGC analyzer) [76][77]. Some consider a significant drawback of a process mass spectrometer to be the high cost necessary to make it capable of operating in hazardous chemical processing environments [78]. To compensate for this some applications take advantage of the "field portability" of the instrument by utilizing the analyzer to follow a chemical process stream from its conception in the laboratory to the pilot plant to the processing line in the factory. This avoids the need for individual dedicated analyzers at each location described above [79].

Finally, rapid scan times allow multiplexing of a single PrMS analyzer, making it even more attractive as a tool for volatiles analysis. This means that several process streams can be monitored by a single PrMS analyzer which can make capital repair and maintenance of the PrMS system more economical as compared, for example, to acquisition, repair and maintenance of several PrGC systems [79].

1.7: MIXTURE SPECTRUM DECONVOLUTION

Much of the speed of PrMS is derived from its avoidance of a slow separation step. When subjected to ionization, each component of a mixture gives a "signature" or "fingerprint" mass spectrum which is a fragmentation pattern characteristic of the functional groups present and their arrangement within the molecule. The mass spectrum of a *mixture* generally represents a convolution of the individual pure component mass spectra. In other words, the mixture mass spectrum can be assumed to be a linear combination of the pure component mass spectra. For this to occur, the pressures maintained in the source region of the mass spectrometer must be such that essentially no interactions occur between the different components. Typically, a pressure in the region of 1×10^{-5} torr or lower is sufficient to validate the assumption of no inter-component interactions [75]. At pressures significantly greater than this (above the linear dynamic range) or significantly lower (below the limit of quantitation) the linear assumption will not be true. This allows the deconvolution of mixture mass spectra using matrix algebra to estimate the relative concentrations of each component. In the mass spectra of pure materials (reference spectra), the intensity or abundance of a peak at a given m/z signal ($i_{A,\,m/z}$) is proportional to the pressure of the analyte (p_A) over a range of pressures as illustrated in **Equation (1.7.a)**:

$$i_{A_{m/z}} = k_A \times p_A \qquad \textbf{(1.7.a)}$$

where k_A represents a proportionality constant related to physical characteristics of the analyte such as ionization, fragmentation cross-sections, instrument transmission and

detection sensitivity. Thus, k_A is basically a measure of instrument sensitivity (the slope of a plot of signal versus analyte concentration, quantity or pressure) to different analytes. A first order treatment of mixture data assumes simply that the spectrum represents a linear combination of the component reference spectra. For example, for an *n*-pentane / isopentane mixture:

$$i_{Mixture_{m/z}} = k_N p_N + k_I p_I \qquad \textbf{(1.7.b)}$$

where *N* and *I* refer to *n*-pentane and isopentane, respectively. From Raoult's Law:

$$p_A = X_A \times p_A^{\circ} \qquad \textbf{(1.7.c)}$$

the analyte pressure can be related to the vapor pressure of pure *A* (p_A°) and its mole fraction in the liquid mixture (X_A). Substituting into **Equation (1.7.b)** yields **Equation (1.7.d)**:

$$i_{Mixture_{m/z}} = k_N X_N p_N^{\circ} + k_I X_I p_I^{\circ} \qquad \textbf{(1.7.d)}$$

Combining the constant terms according to **Equation (1.7.e)**:

$$j_A = k_A p_A^{\circ} \qquad \textbf{(1.7.e)}$$

yields the simplified form, **Equation (1.7.f)**:

$$i_{Mixture_{m/z}} = j_N X_N + j_I X_I \qquad \textbf{(1.7.f)}$$

When signals at more than one m/z are considered, **Equation (1.7.f)** represents a matrix equation, the inversion of which can yield mole fractions or sensitivity factors (j_N and j_I) from measured intensities. The constants j_N and j_I can be obtained for each m/z considered from spectra of pure components and a mixture of known composition assuming that there are no interactions (a reasonable assumption at low pressure). Since

the measurement of absolute intensity is of larger variance than normalized intensity and use of the vapor pressure of pure components (p_A°) is cumbersome, *relative* sensitivities for each component are calculated and used in **Equation (1.7.f)**. This is done to account for variability in absolute intensities since even an equimolar mixture of isomers (components with the same base peak) under ideal stability conditions will give base peaks in which the absolute intensities may differ significantly. So, normalized intensities (normalizing the most intense m/z signal in the mass spectrum to 100%) are utilized. However, the process of normalization removes important information from the base peak (the base peak in different reference spectra are both 100%) in all reference mass spectra. Thus, a calibration mixture (with known mole fractions X_N, and X_I) is used and j_N and j_I are calculated from **Equation (1.7.f)** using these known mole fractions. By incorporating the relative sensitivities into the calculations described in Chapter 2, the normalized reference spectra can be used without losing information contained in the base peak.

The system of equations represented by **Equation (1.7.f)** is "over-determined" if there are more equations (one for each m/z to be considered) than unknowns (*i.e.*, two concentrations for a binary mixture). Thus, any *pair* of equations like **Equation (1.7.f)** (*i.e.*, data at any two m/z values) can be used to derive concentration estimates for a binary system; matrix algebra provides a simple means of solving such simultaneous equations.

To illustrate in further detail the idea of individual component mass spectrum addition, imagine some component "A" which may afford a mass spectrum such as the one illustrated in the top portion of **Figure 1.7.a**, where all the ions have m/z ratios less

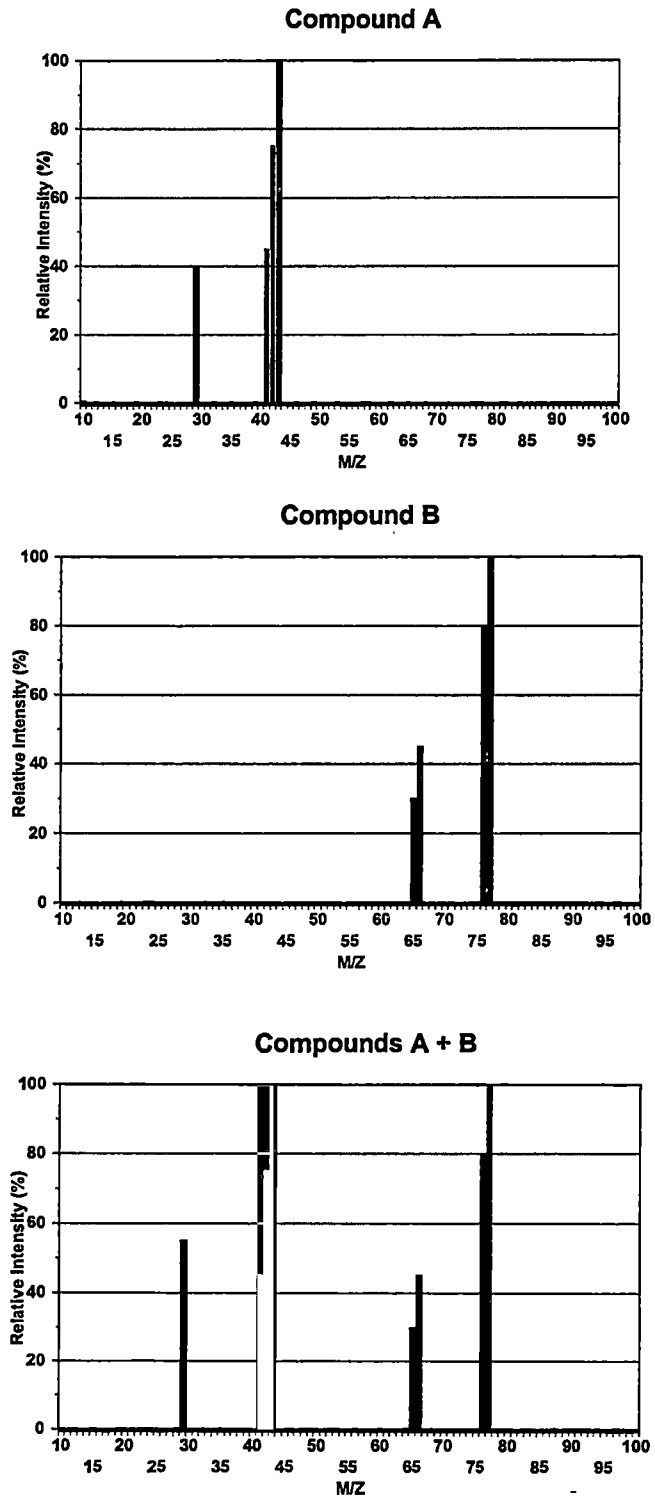


Figure 1.7.a: Pure component spectra of components "A" & "B" and an equal-molar mixture of "A" + "B".

than m/z = 50. Similarly, some component "B" may afford a mass spectrum such as the one illustrated in the center portion of **Figure 1.7.a** where all the ions appear at m/z ratios greater than m/z = 50. The resulting mass spectrum that is produced upon mixing an equal-molar amount of the two pure components "A" and "B" is illustrated in the lower portion of **Figure 1.7.a**. This spectrum is a linear combination of the two pure component reference spectra, assuming that the instrument has the same sensitivity to both components.

Commercial PrMS instruments rely on matrix algebra (using matrix inversion and least squares methods) to recover the concentrations or relative abundances of each component in the mixture by deconvoluting the mixture mass spectrum. Data acquisition and analysis are important aspects of process monitoring. Data feedback to the control system must be rapid to accommodate real-time control. In light of the high speed of modern computers, the rate-limiting step is generally the data acquisition rather than the data-processing step. The data-acquisition is generally transparent to the user of commercial instruments. It is typically buried in the software which both controls the instrument (including tuning of the mass spectrometer ion source and calibration functions) and deals with the data (including acquisition and presentation to a user and/or to a distributed process control system) [100]. Tuning the ion source in a mass spectrometer is not entirely controlled by the computer and is a critical, technical point regarding process analysis. It is more important in quantitative than in qualitative mass spectrometry. Unreasonable tuning parameters (noted by the operator) generally indicate when it is time to clean the ionization source. Still, the choices in tuning made by the instrument operator will affect the accuracy and precision of the analysis.

Least squares analysis is the most common quantitation approach in PrMS. This requires acquisition and storage of reference spectra of pure components, and analysis of calibration mixtures for determination of relative responses (sensitivities) to each component. Mixture spectra are then treated as linear combinations of the sensitivity adjusted (sensitivity weighted) reference spectra. The concentrations returned are those values which, when multiplied by the corresponding reference spectra and relative sensitivities, then summed for each m/z, provide the smallest sum of squares difference between the calculated and experimentally measured mixture spectra. Because of the difficulty of precisely controlling parameters such as electron emission from the hot filament in EI MS, nearly all analyses utilize *normalized* spectra to determine *relative* concentrations, rather than attempting to determine *absolute* analyte concentrations from *absolute* signal intensities; relative concentrations are generally sufficient for process analysis. Relative concentrations will demonstrate n-1 degrees of freedom where the concentrations add to one.

In cases where the assumption of linearity does not apply, other methods of mixture deconvolution can be applied to the data in the interest of improving accuracy and/or precision. These include Factor Analysis [80][81], Principal Components Analysis [82][83], Partial Least Squares [84][85][86] and Artificial Neural Network treatments [87][88]. These are multivariate methods which utilize a larger number of known variables than predicted variables as well as large "training sets" as a calibration step (*i.e.*, data from large numbers of mixtures of known composition) rather than single-component reference spectra with just one or a few mixture(s) for calibration. They have been proven to work well with several differing types of process stream data including

infrared and near-infrared data [89][90][91]. The initial training can be time-consuming, but reduces the reliance on a relatively small number of reference spectra. Continuous "retraining" can occur by addition of data to the training set.

The alternative methods have been used widely in optical spectroscopy [92], but their relative merits for interpreting mass spectra (with their intrinsic "histogram" character and potential for long-term drift in relative intensities and "tuning" due to the effects of source contamination) have apparently not been addressed. For artificial neural networks, extraction of many ($> \sim 2$) outputs (concentrations) from complex data was not prevalent in the literature a few years ago [93] but appeared to be feasible [94]. An increase in the number of applications of artificial neural networks in microbiology has taken place over the past couple of years. This is illustrated by the use of pyrolysis mass spectrometry and the application of an artificial neural network to rapidly monitor the growth of bacteria [95][96][97]. These deconvolution methods often demonstrate a high tolerance for data that contain outliers and other anomalies [98][99].

1.8: PARAMETERIZATION

In principle, for a mixture comprised of n components, a deconvolution can be accomplished by monitoring the signals ("peaks" in the mass spectrum) derived from as few as n distinct ions, provided that there is at least one ion derived from each component (intact or fragmented). In the case of relatively simple chemical process streams, commercial instruments will generally resort to this "square-matrix" solution where n m/z ratios are employed to deconvolve a mixture of n components [100]. Even in the event

that more than one component contributes to a given m/z signal, accurate quantitative deconvolution can still succeed, provided that relative intensities for each monitored signal are unique for each component. So, in principle, it is possible to accurately and precisely resolve isomeric mixtures provided that the relative intensities in the reference spectrum of each pure component differ by more than the precision of the analyzer. Thus, the work described in this thesis is aimed at finding out if this is possible in practice. But, which m/z signals should be selected for monitoring?

As noted above, the rate-limiting step for "real-time" analysis using any PrMS quantitation method is data acquisition. Three independent experimental parameters which affect acquisition time must be addressed: integration times for each signal measurement, the number of replicate measurements at each m/z signal, and the number of m/z ratios to monitor. The first two factors (integration time and replicates) have important influences on the balance between precision and analysis time. The *selection* of m/z ratios to be monitored in a selected ion monitoring (SIM) acquisition ("parameterization") has a more subtle effect. Accurate and precise analysis using PrMS requires monitoring at least one ion signal (*i.e.*, one m/z signal) from the spectrum of each component of interest (*e.g.*, at least six different m/z ratios must be monitored for simultaneous determination of six components). Accuracy and precision can often be enhanced by monitoring more than this minimum. Monitoring *too many* signals compromises the precision/speed trade-off by more than just increasing acquisition time. Attempting to fit data which contributes to the noise but does not constitute informative analyte signal can actually *degrade* accuracy and precision. There may even be m/z signals with relatively large intensities that may not help distinguish between

components.

Thus, if a target accuracy and/or precision is desirable, there will be an optimum subset of m/z ratios in the spectrum that can be monitored to achieve it. Factors to consider in selecting that parameterization (subset) include relative signal intensities in reference spectra, interference (overlap) among the spectra of various components, relative concentrations, and the required precision. Although some instrument manufacturers provide algorithms to select "the best peak" to monitor for each component [100], routines to select "the best" overall subset are not widely available. Discerning the optimum parameterization based on a "brute force" evaluation and comparison of all possibilities is impractical for complex mixtures, since the number of possibilities increases very rapidly with spectral and mixture complexity (*e.g.*, there are over 10^6 possible parameterizations when considering 20 different m/z ratios in a simple binary mixture). Furthermore, as noted above, the true "optimum" is likely to depend on concentration; the best parameterization for determining trace "A" in "B" may not be the best for determining trace "B" in "A". Since wide swings in concentration are not the norm in a process environment, this concentration dependence will not preclude there being an optimum parameterization for a given stream, but it may limit the *general* applicability of a given parameterization.

Empirical parameterization algorithms (EA), under development in our laboratory, are designed to utilize the information available in the pure component mass spectra to establish a parameterization that is statistically equivalent to the best parameterization. A parameterization that is statistically equivalent to the best parameterization is one that affords concentration estimates that lie within the error bars

of the best parameterization. Algorithms that perform functions which are similar to our EA compare the m/z ratios between reference spectra and some "unknown" spectrum with the intent of indicating the likelihood of a spectral match or identification [101]. Our EA compares reference spectra between the known components with the intent of identifying an optimum parameterization. This is done by comparing properties in m/z ratios between each component in the mixture such as intensity, relative standard deviation and uniqueness (the existence of a m/z signal in one component that may be absent in another).

A "genetic algorithm" (GA) [102][103] may offer an alternative approach to identifying those m/z ratios for which intensity is best correlated with component concentration in a partial least squares or multiple linear regression training set. GA's have been applied to data obtained from analysis methods including infrared spectroscopy [104], pyrolysis mass spectrometry [105] and liquid-chromatography [106]. In these applications the optimum selection of variables (wavelengths in spectroscopic applications) is identified for which the root-mean-square of the error between calibration mixtures and the predicted values for the same mixtures (from the model equation produced by the GA) is minimized. Datasets are usually divided in half. One half is used for calibration and the other is used for comparison with estimated spectra calculated from the model. The model or model equation is produced from the GA model equation.

In some cases the total number of variables monitored was reduced by about half resulting in a significant reduction in the sum of squares error of prediction (estimating known concentrations of the samples) [105]. This reduction in variables allows more rapid acquisition of analytical data and thus improved monitoring and control of a

chemical process stream. Often these applications utilize data from analytical methods that include chromatography as an extra dimension of analytical information. However, such a treatment has yet to be reported for PrMS (mass spectrometry without chromatography). The addition of chromatography to the mass spectrometry can compensate for the lack of precision of m/z signal intensities that PrMS utilizes for quantitation.

Once questions of quantitation methodology have been decided and the parameterization for a given application has been determined, routine operation of the analyzer may proceed with little operator intervention. Even fault detection and calibration checks may be automated, although there are aspects of calibration and preventative maintenance that require periodic operator intervention. The period of that intervention constitutes an important figure of merit for evaluation of a particular on-line analysis instrument.

1.9: ISOMER ANALYSIS BY MASS SPECTROMETRY

Recall that the mass spectra of isomeric components usually contain the same m/z ratios; therefore, the differentiation between isomeric components must be based upon the utilization of signal intensity differences, separation of the mixture or derivatization of the analyte. Derivatization is helpful when the analyte volatility is low and the derivatized products demonstrate a higher vapor pressure. This is the case when analyzing allenic fatty-acids [117]. Also, some isomers selectively react to form unique products that can be distinguished utilizing unique m/z ratios in the mass spectrometer whereas before derivatization no unique m/z ratios were available for monitoring.

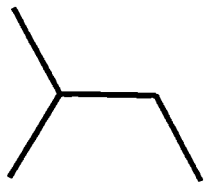
Deuterium labeling can be considered a type of derivatization. In one particular application deuterium was utilized to aid in the analysis of mixtures of positional isomers of glucose in the blood of rats [107]. The information obtained from deuterium labeling proved useful in the elucidation of gluconeogenesis pathways of synthesis and degradation. Isomers are compounds of identical molecular formula but the order in which the individual atoms are connected differs. An example of the structural isomers *n*-pentane and isopentane is illustrated in the top portion of **Figure 1.9.a**. Cis- / trans- isomers are stereoisomers. Stereoisomerism describes isomers whose atoms are connected in the same order, but their spatial arrangement differs such as that illustrated in the lower section of **Figure 1.9.a**. Stereoisomers are a topic of analysis described later in this dissertation.

Typically, the quantitative analysis of isomeric mixtures using mass spectrometry has been performed with chromatography and/or derivatization of the analyte as described later in this section. This is due to the similarity in the mass spectra of isomers. The research described in this dissertation illustrates how PrMS (MS without chromatography or derivatization) can serve as an excellent method for monitoring a process stream comprised of hydrocarbon isomers.

In some isomer analysis applications a rapid chromatographic step can be used to afford the partial separation of a complex mixture and the information obtained from the mass spectrometer is used to complete the analysis toward a positive identification of the analytes (GC-MS) [108]. The "fingerprinting" capability of the mass spectrometer is useful in cases where an incomplete separation of a complex mixture exists due to the lack of analysis time or the complexity (number or type of components) of the sample.

Pentanes (structural isomers)

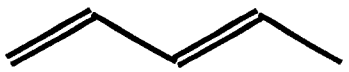
n-pentane



isopentane

Piperylenes (stereoisomers)

cis-



trans-

Figure 1.9.a: Top: Structural isomers of n-pentane / isopentane. bottom:Stereo-isomers of cis-, and trans-piperylene.

Drawbacks to this approach may include the increased time of analysis and the additional complexity of instrumentation for the chromatographic step as compared to performing the mass spectrometry analysis without chromatography. In these cases it is necessary to combine the separation power of GC and the "fingerprinting" ability of MS (GC-MS) to afford a hyphenated method that carries the advantages and handicaps of both. In other applications the hyphenated method is still insufficient; thus, derivatization *and* GC-MS is necessary.

Isomer distinction is important in many different fields of analytical chemistry. The bio-analytical field contains several different examples of isomer analysis. One reason that isomer identification is important in this field is because there are instances when subtle differences in molecular structure lead to major differences in chemical properties or reactivity. For example the degree of carcinogenicity can be dependent upon the position of substitution on some parent polycyclic aromatic hydrocarbons. Specifically, 5-methylchrysene is a potent carcinogen while the other monomethyl isomers are either noncarcinogenic or much less carcinogenic [109]. Hence, the quantitation of these isomers without providing isomer-specific identifications of the individual compounds may overestimate the carcinogenic hazard.

The same idea applies to other isomer-specific identifications such as nitrogen heterocycles or polychlorinated biphenyls [110][111]. Environmental samples containing PCB's were analyzed by GC-MS and an automated method of analysis was developed such that up to 52 PCB isomers could be quantitated [112]. The automated version of this analytical method involved the use of a computer to automate the sampling equipment. This allowed the opportunity of assigning the rigorous data analysis portion

of the analytical procedure to be handled by the same computer that operated the sampling system. Improvements in the accuracy and consistency of the new automated method, compared to the original method of manual analysis, was noted. It was concluded in a previous, related, publication that data interpretation via the manual method of data analysis was the practical limitation to the PCB analysis [113]. Thus, the automated version of the isomer analysis proved to be the more accurate, precise and preferred approach.

Diastereomeric isomers of a β-blocking drug have been quantitatively analyzed in human plasma and saliva by mass spectrometry but required the use of derivatization *and* GC/MS [114]. The authors noted that previous analytical methods utilizing a chiral chromatographic column and fluorescence detection were not successful due to a lack of sensitivity. Thus in this case the mass spectrometric technique proved to be a more sensitive probe for these drugs at the ~1 to 20 ng/mL level. In another biomedical application, quantitative isomeric analysis is important in the detection of 12 ubiquitous, highly toxic, polychlorinated dibenzo-*p*-dioxins known as "the dirty dozen." Derivatization, sample clean-up and chromatography were utilized to quantitatively analyze mixtures of these materials present in milk at levels in the parts-per-quadrillion range [115].

Also in the biomedical area, the peroxidation of lipids was studied because of its suspected role in the development of diseases including coronary artery disease [116]. This involved the quantitative GC/MS analysis of individual positional isomers of fatty acids. On-column derivatization of the hydroxy groups gave fatty acid methyl esters which were subjected to electron impact mass spectrometry (EI-MS) for quantitative

analysis. In a similar investigation of isomeric fatty acids it was found that picolinyl ester derivatives of the fatty acids gave much more useful fragmentations (in the sense of accuracy and precision of analysis) in the EI mass spectra than the methyl esters or the underivatized fatty acids [117].

Reactivity differences between different isomers are important thus making their differentiation and quantitation in mixtures a significant issue. Chloropyridine isomers are important in industry because they participate as pivotal intermediates in the production of pesticides, herbicides, dyes, pharmaceuticals and polymers. Their separation and quantitative identification were performed off-line by GC-MS with an accuracy of approximately 16% relative error [118].

Low detection limits are not frequently required for chemical processing analysis applications since the sample occurs at a relatively abundant level. However, in chemical processing streams where *impurities* can be a sensitive issue, low detection limits are important (*e.g.*, in the semiconductor industry reagent gases are monitored for purity using process mass spectrometry [119]). Typically, the accepted level of contaminants in high purity gases is below 1 ppm [119]. In this specific application previous quality control methods in the fabrication of electronic devices using PrGC were replaced by mass spectrometry [120].

In other applications analytical methods such as Fourier transform infrared spectroscopy are added to the chromatographic step of the analysis [121][122]. There have also been instances utilizing off-line *tandem* mass spectrometry (MS/MS) to quantitatively analyze isomeric mixtures of ketenes [123] and isomer products from the high-energy collisions of butan-1,3-diol [124] and in the collisionally activated

dissociation in the analysis of isomeric ion mixtures [125]. Scientists at The Environmental Protection Agency's Characterization Research Division recently analyzed a mixture of substituted naphthalene isomers in a groundwater sample obtained from a Superfund Site near the Tom's River in New Jersey. In this case high-resolution mass spectrometry was required to quantitatively distinguish between the differing isomers [126]. These latter methods have not been adapted for on-line application at this time.

Overall, the quantitative analytical methods for isomeric mixtures found in the literature deal with the separation and/or derivatization of the isomers. However, the relative intensities in the mass spectra between some types of isomers should be distinct enough to allow their deconvolution in a complex mixture mass spectrum without chromatography or derivatization. Also, the precision of analysis should be such that the small differences between relative intensities of certain isomers can be used to deconvolve a mixture which contains isomers. Most importantly, the accuracy and precision of quantitation is anticipated to depend significantly on the *selection* of m/z ratios (parameterization) used in the deconvolution of the mixture mass spectrum. It is this aspect of quantitative analysis of isomeric mixtures based entirely upon the differences that exist between the relative intensities at selected m/z ratios in the pure component mass spectra that make the work in this dissertation novel in the process industrial as well as the analytical forum.

Although the hydrocarbon-isomer identification for the research presented in this dissertation is not as extensively documented in relation to the applications mentioned above, quantitative isomer analysis remains an important factor in the economic

monitoring and control of the chemical process streams investigated.

1.10: SUMMARY AND OBJECTIVES

The objectives of this research are to discern the feasibility of process mass spectrometry as an on-line analytical method for process streams containing mixtures of hydrocarbon isomers. Information regarding the process streams modeled were brought to our laboratory by industrial members of The University of Tennessee's Measurement and Control Engineering Center (MCEC). Since industrial process streams can often contain several components at differing composition levels, issues such as the total number of isomeric components that can be simultaneously quantitated and their limits of detection are addressed. The research begins with analysis of simple binary mixtures of isomers and progresses to more complex mixtures. Finally, mixtures of 9 components are used to mimic a process stream. Estimated relative component concentrations from several mixtures calculated from the data obtained from the process mass spectrometer are compared with gas-chromatographic and gravimetric results. The latter methods serve as validation methods of analysis to compare against the PrMS results. This data is presented as validation plots (plots of estimated concentrations vs. true concentrations). The work in this thesis is unique because of the absence of chromatography or derivatization to quantitatively analyze isomeric hydrocarbon mixtures by mass spectrometry. With the knowledge obtained from these studies it is intended that an on-line mass spectrometric analyzer be installed at the corresponding factory or pilot plant to rapidly monitor the process stream.

CHAPTER 2

EXPERIMENTAL

2.1: INSTRUMENTATION

A quadrupole-based process mass spectrometer (ABB-Extrel Mass Spectrometry, Pittsburgh, PA), operated in the selected ion monitoring mode was used to perform the process mass spectrometric analysis measurements in this work. The lens voltages in the mass spectrometer electron ionization source was tuned for "flat-topped" peaks with at least 10% valley between adjacent (isotope) peaks throughout the spectrum. The electron energy was 100 eV with 3 A filament current (instrument default). The trap current or emission current was set to 2.0 mA (±0.1 mA). For the *n*-pentane / isopentane and the 9-component systems *n*-pentane was utilized for tuning the ionization source voltages. Specifically, this involved adjusting the ion lens voltages to a point where the base peak of the tuning component was symmetrical and demonstrated a "flat-top" appearance. Cis-piperylene was used to tune the piperylene system and cis-2-pentene was used to tune the cis- / trans-2-pentene system. Typically, for an instrument with a clean ion source the tuning parameters (lens voltages) are as follows. Ionizer: 18 V, Extractor: 10 V, Lens #1: -115 V, Lens #2: 0 V, Pole Bias: 0 V.

Peak heights were used to represent ion abundances. The dwell time for each selected ion was approximately 300 milliseconds (instrument default). The m/z signals for all ions greater in intensity than ~0.1% relative to the base peak of an approximately equimolar mixture mass spectrum were used for the *n*-pentane / isopentane system giving a total of 21. The 2-pentenes and piperylene systems were limited to a total of 23 m/z signals in the brute force and intuitive parameterization methods by utilizing relative

intensities exceeding approximately 1 and 3% respectively. In the genetic algorithm (GA), empirical algorithm (EA) and stream evaluator (SE) tests a total of 30 m/z signals were utilized in the 2-pentenes and piperylenes system which included relative intensities greater than approximately 0.3 and 0.1%, respectively. The difference in the number of m/z signals utilized in these parameterization methods (23 vs. 30) was due to the efficiency with which the GA, EA and SE handle a larger number of m/z signals. Samples were analyzed in triplicate over a period of three consecutive days using the same source tuning parameters for each replicate. In order to correct for the possibility of fluctuations in day-to-day precision, pure component reference spectra *and* mixture mass spectra were acquired each day.

A Hewlett Packard "6890 plus" series gas chromatograph with a Hewlett Packard 5973 mass selective detector was used to acquire the GC-MS data presented in this dissertation. The column head pressure was 15 p.s.i.g,

Parameters for the GC-MS analysis method for 9-component mixtures were as follows. Injection volume: 0.1 μL, injection inlet temperature: 60° C, injection mode: split (500:1), column: DB1-MS (30 meters) (Alltech, Deerfield, IL) and DB1-MS (50 meters) (Quadrex, New Haven, CT) oven temperature program: isothermal at 25° C for 20 minutes, ramp 25° C/minute to 150° C, hold for 5 minutes, ramp 30° C/minute to 250° C, hold for 5 minutes [127]. The GC-column used in these experiments consisted of two separate columns connected in series using a quartz capillary connector (Alltech, Deerfield, IL). The resulting 80 meter column was required to afford an acceptable resolution between the different components since the lowest oven temperature that could be obtained was ambient (25° C), whereas separation temperatures recommended by the

MCEC industrial members' laboratory were sub-ambient using a 50 meter column. MS detector parameters were as follows. Solvent delay: off, acquisition mode: Scan, Mass range: 10 - 250 amu, Intensity threshold: 500 (default), Sampling: 2 (default), transfer line: 65° C.

2.2: REAGENTS AND MIXTURES

Standard liquid mixtures were prepared gravimetrically from the reagents listed in **Table 2.2.a** and are listed in **Table 2.2.b**. Mixtures were prepared by dispensing each component through a Teflon-lined silicone septum (Wheaton, Millville, NJ) into a 4 mL sample vial (Wheaton, Millville, NJ) with a 1 mL syringe (Hamilton, Reno, NV) connected to a 3 inch side-hole SS needle (Hamilton, Reno, NV). Each component in the sample vial was weighed after it was administered into the vial using a digital analytical balance (± 0.0001 g) (Sartorius, Model # BP2110, Gottingen, Germany). The septum-sealed vials were stored at -4° C for not longer than 24 hours prior to analysis. The weights of components in each mixture and compositions of mixtures tested are listed in **Table 2.2.b.**

Table 2.2.a: List of reagents and suppliers.

Reagent	Purity	Reagent	Purity
n-pentane	99+%	cyclopentane	99%
isopentane	99%	cyclopentene	99%
cis-piperylene	98%	2,2-dimethylbutane	99%
trans-piperylene	98%	2-methyl-2-butene	99+%
cis-2-pentene	99%	helium*	Ultra-pure 99+%
trans-2-pentene	99%		

*: Helium supplied by National Speciality Gases (Durham, NC), remaining reagents supplied by Aldrich Chemical Company (Milwaukee, Wisconsin).

Table 2.2.b: Sample compositions

Sample #:	Components*	Weight of sample (grams of component 1, 2, ...)	Mole fraction (Component 1,2, ...)
N1	1, 2	1.1521, 0.1260	0.9014, 0.0986
N2	1, 2	1.0442, 0.2936	0.7805, 0.2195
N3	1, 2	0.9044, 0.4204	0.6827, 0.3173
N4	1, 2	0.7703, 0.5700	0.5747, 0.4253
N5	1, 2	0.6252, 0.6273	0.4992, 0.5008
N6	1, 2	0.5150, 0.8411	0.3798, 0.6202
N7	1, 2	0.4008, 0.8687	0.3157, 0.6843
N8	1, 2	0.2669, 1.0183	0.2077, 0.7923
N9	1, 2	0.1405, 1.2316	0.1024, 0.8976
P1	3, 4	0.0539, 0.3126	0.1471, 0.8529
P2	3, 4	0.0845, 0.1574	0.3493, 0.6507
P3	3, 4	0.0897, 0.0863	0.5097, 0.4903
P4	3, 4	0.4550, 0.0763	0.8564, 0.1436
P5	3, 4	0.3216, 0.0522	0.8604, 0.1396

*Key to components: 1: *n*-pentane, 2: isopentane, 3: cis-piperylene, 4: trans-piperylene

Table 2.2.b: Sample compositions, continued.

Sample #:	Components*	Weight of sample (grams of component 1, 2, ...)	Mole fraction (Component 1,2, ...)
PE1	5, 6	0.0583, 0.2957	0.1647, 0.8353
PE2	5, 6	0.0739, 0.1533	0.3253, 0.6747
PE3	5, 6	0.0830, 0.0795	0.5108, 0.4892
PE4	5, 6	0.1529, 0.0716	0.6811, 0.3189
PE5	5, 6	0.3025, 0.0464	0.8670, 0.1330
G1	1,3,4, 5,6,7, 8,9,10	0.0064, 0.2833, 0.4529, 0.0058, 0.0054, 0.2879, 0.0070, 0.1388, 0.0118	0.0051, 0.2375, 0.3797, 0.0047, 0.0044, 0.2414, 0.0046, 0.1130, 0096
G2	1,3,4, 5,6,7, 8,9,10	0.0815, 0.0840, 0.0809, 0.0769, 0.0759, 0.0783, 0.0963, 0.0788, 0.0805	0.1100, 0.1201, 0.1157, 0.1068, 0.1054, 0.1120, 0.1088, 0.1094, 0.1118

*Key to components: 1: *n*-pentane, 2: isopentane, 3: cis-piperylene, 4: trans-piperylene 5: cis-2-pentene, 6: trans-2-pentene, 7: cyclopentene, 8: 2,2-dimethylbutane, 9: 2-methyl-2-butene, 10: cyclopentane

2.3: CALCULATIONS

Matlab version 4.2 (The Mathworks Inc., Natick, MA) was used to calculate least-squares concentration estimates from normalized mixture and single-component reference spectra using component relative sensitivities determined from calibration mixtures. Spectral normalization (which is standard for PrMS) requires normalization of concentrations (i.e., concentrations add to 100%), reducing by one the number of independent concentrations that can be determined, but greatly improving precision. Calibration samples were chosen to be near the mean composition of the test samples.

The "best" parameterization (giving the lowest sum of squares error (SSE) for all components in all test samples in a given set) was determined by "brute-force" analysis of all possible parameterizations using the 23 most intense ion signals. 21 m/z signals were utilized in the analysis of the *n*-pentane / isopentane system. Using 23 signals (instead of 30 in the systems other than *n*-pentane / isopentane) reduced the total number of possible combinations to slightly more than eight million, requiring approximately 20 hours of computation time using a SUN (Palo Alto, CA) computer system running under UNIX OS 5.6 with dual 300 MHZ processors and 1 gigabyte of RAM. This computing system is a time-shared server; if it were dedicated solely to the brute force computations the total calculation time would be less. It is this relatively long calculation time that prompts investigation of faster, more efficient methods of calculation to perform parameterization which are described in the following sections of this work.

2.4: SAMPLING

The *n*-pentane / isopentane mixtures were sampled by submerging one end of a deactivated (silanized) fused-silica "liquid sampling" capillary (100 μm i.d. x 60 cm long; Polymicro, Phoenix, AZ) into a liquid hydrocarbon sample contained in a glass reservoir pressurized to ~3 p.s.i.g. with ultra-pure helium (**Figure 2.4.a**), resulting in a liquid flow rate through the capillary of ~30 μL/min. The other end of the liquid transfer capillary was connected to an 1/8th inch stainless steel (SS) tee (Swagelok Inc., Solon, OH) in a flash vaporization (FV) box (**Figure 2.4.b**) where the sample was vaporized by mixing with hot ultra-pure helium (National Speciality Gases, Durham, NC) flowing at a rate of 20 mL/min. The helium was heated by flowing through a hot 4 meter coil of 1/8th inch i.d. SS tubing. The SS "tee" section of the FV box is illustrated in more detail in **Figure 2.4.c**.

Helium flow was metered using a mass flow controller (MFC; 5850 E-series, Brooks / Rosemount, Hatfield, PA) and controller boxes 0154E and 0152E. The vaporized hydrocarbon / helium mixture was supplied to the ABB-Extrel "Quick Inlet" (**Figure 2.4.d**) via an 1/8th inch i.d. SS heated transfer line. The transfer line was heated to a temperature of 100° C utilizing a flexible electric heating tape (Thermolyne-Barnstead, Briskheat): 120 V, 620 W maximum, powered by a Variac alternating current variable autotransformer (Staco Energy Products Corporation, Dayton, OH). Finally, the "Quick Inlet" allowed a portion of the vaporized hydrocarbon mixture to be directed to the process mass spectrometer's source, which was heated to 120° C, through a 25 μm x 25 cm fused silica quartz capillary (Polymicro, Phoenix, AZ) while the remainder of the

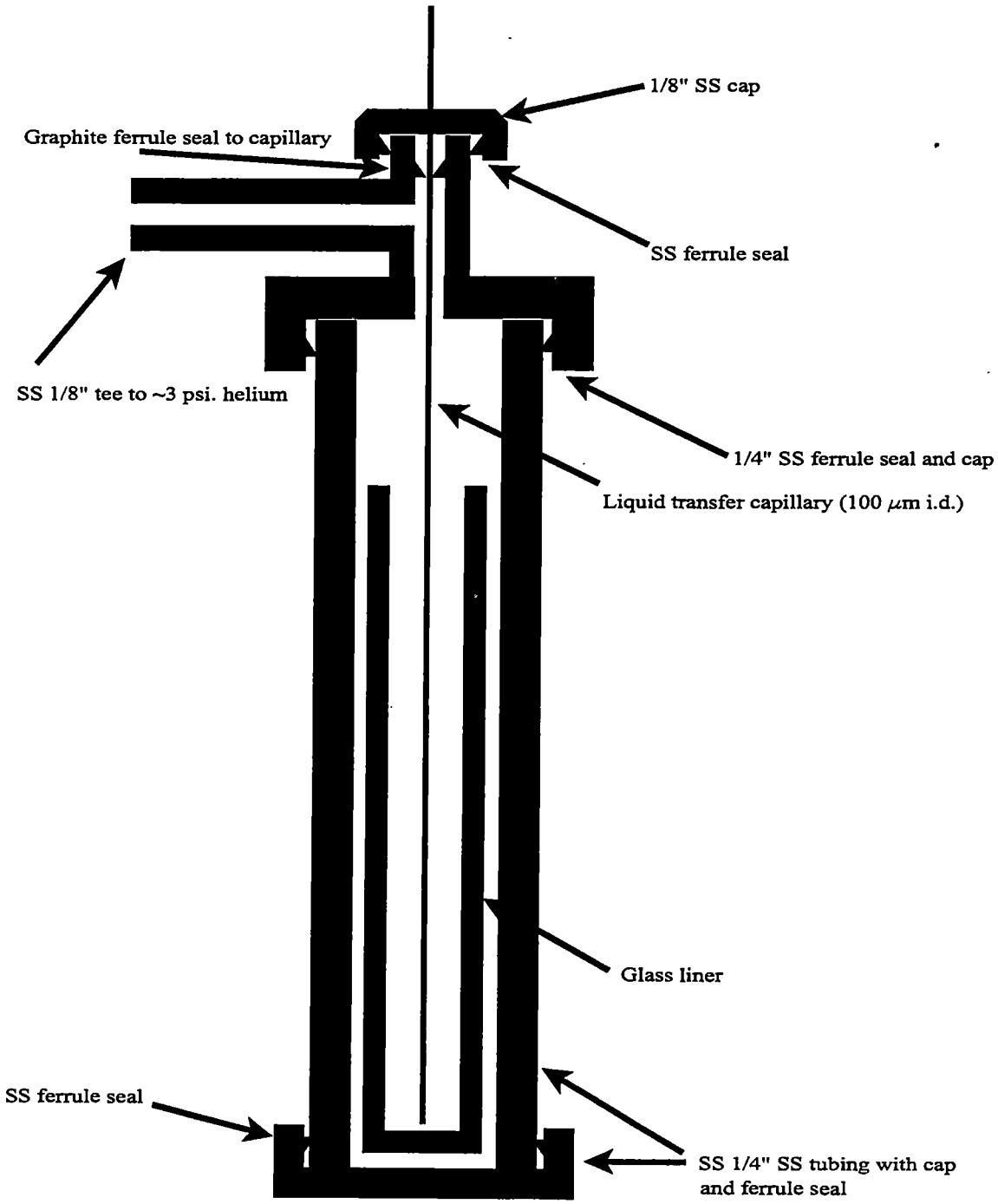


Figure 2.4.a: Pneumatic delivery system used to pressurize liquid hydrocarbon mixtures for delivery into the flash vaporization inlet.

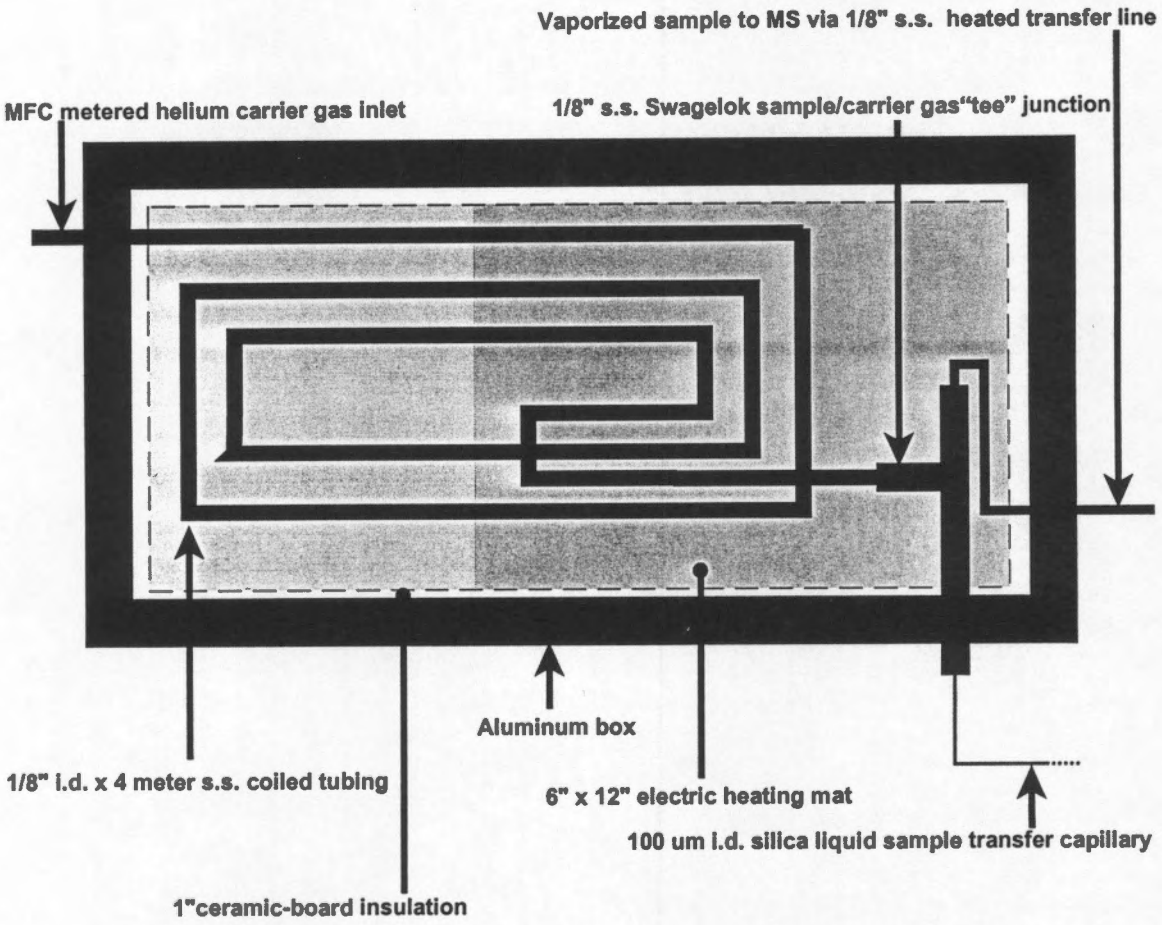


Figure 2.4.b: Continuous flow flash vaporization inlet illustrating carrier gas heating coil and SS vaporization tee.

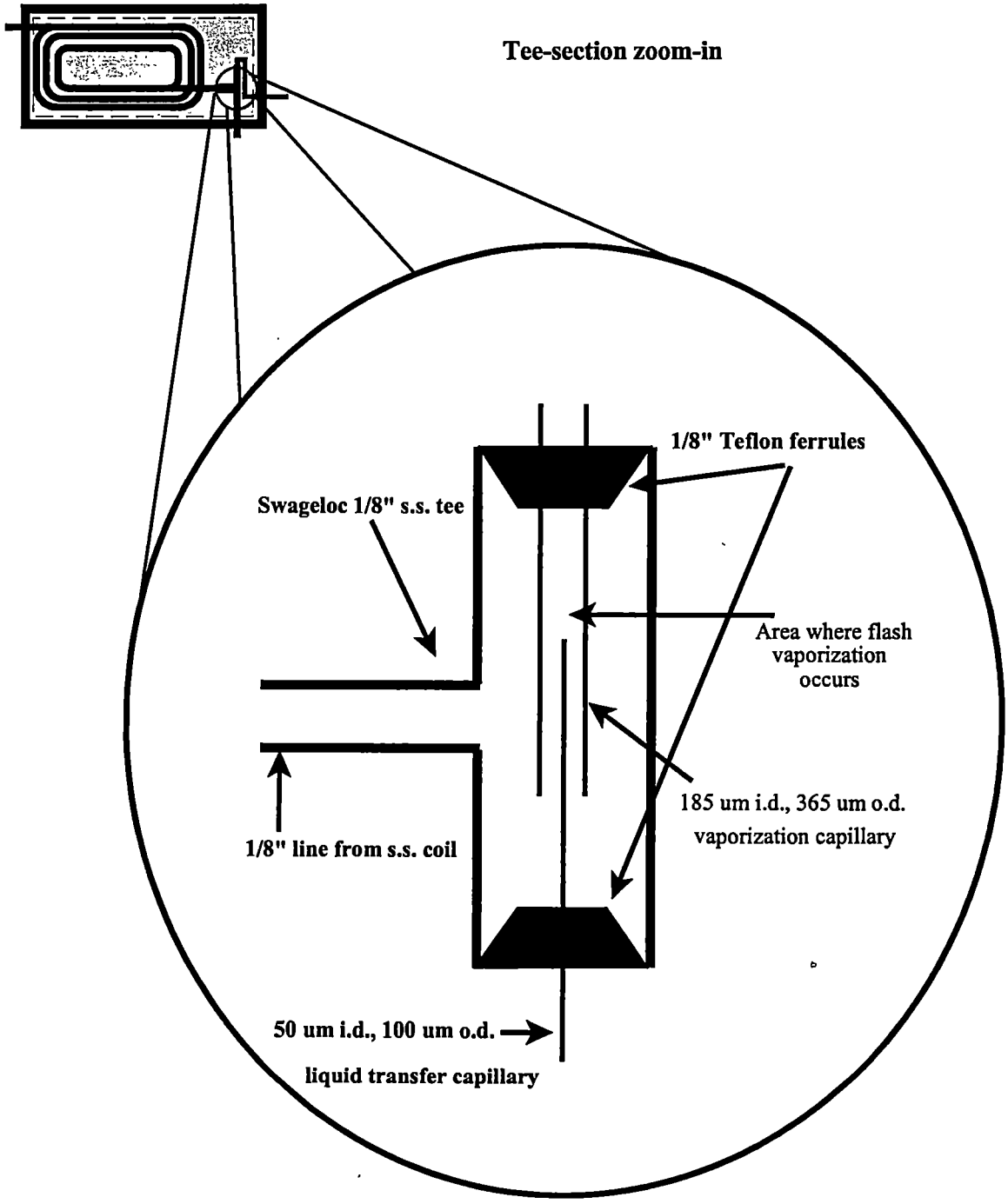


Figure 2.4.c: Continuous flow flash vaporization inlet highlighting SS vaporization tee, location of capillaries and area of vaporization.

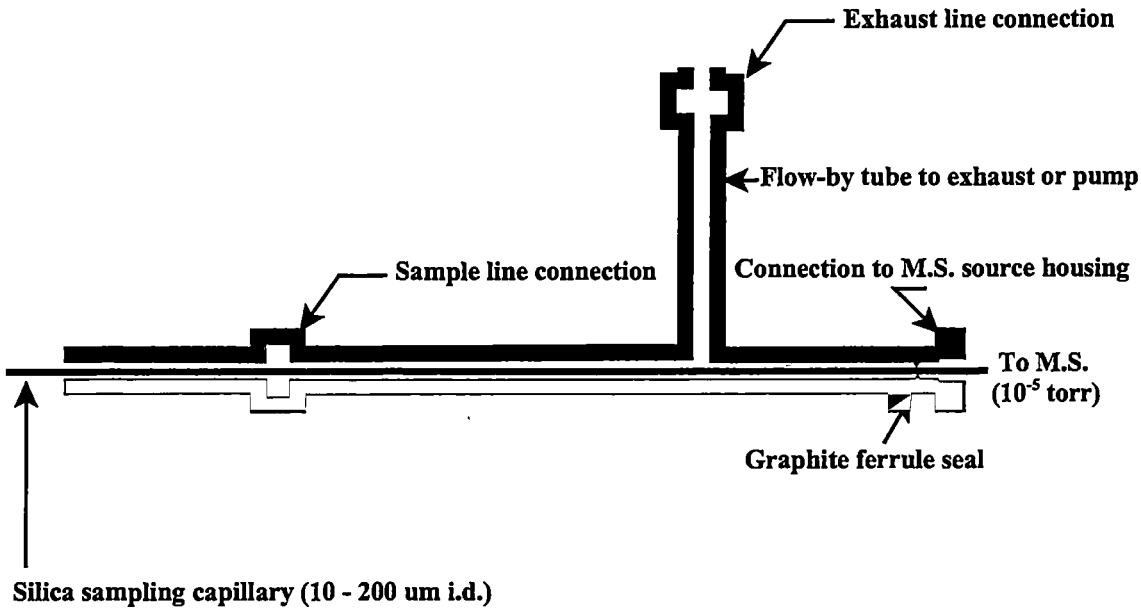


Figure 2.4.d: "Quick inlet" illustrating the process mass spectrometer sampling capillary and exhaust tee.

sample mixture was vented. This resulted in an ion source pressure below 5 x 10^{-6} torr, low enough to avoid molecular interactions that could introduce non-linearities in the mass spectrometer response.

The interior of the FV box was insulated with 1 inch thick ceramic board (Thermal Ceramics, Augusta, GA) and heated using a 12 inch x 6 inch heating blanket (Thermolyne-Barnstead, Briskheat model, 180 W maximum, 1.5 A) and a separate Variac. The temperature of the vaporization inlet was adjusted for different chemical systems based upon the highest boiling point of the components in the mixture. Approximately 15 to 20° C was added to the highest boiling point. For the *n*-pentane / isopentane (boiling point 33° C) system a temperature of 50° C was employed. The stereoisomer systems, including the 9-component mixtures were run at 80° C. The temperature of the flash vaporization box was monitored using a Chromel-alumel Type-K thermocouple connected to a digital voltmeter (Wavetek DM23XT, Research Triangle Park, NC).

The pressurized flow system became fouled and eventually plugged when sampling pure piperylenes (reactive components which evidently polymerized based upon observations of tacky material recovered from the surfaces of the capillaries in the vaporization inlet). To avoid this problem, the sample size of these pure liquids was dramatically decreased. This was done by fabricating a vaporization inlet that utilizes a total sample volume of 15 μL rather than a flow of ~30 μL per minute (*n*-pentane / isopentane system). Reducing the amount of volatilized material avoided the plugging problem encountered upon using the piperylene components in the flash vaporization inlet. **Figure 2.4.e** illustrates the revisions to the flash vaporization inlet components that

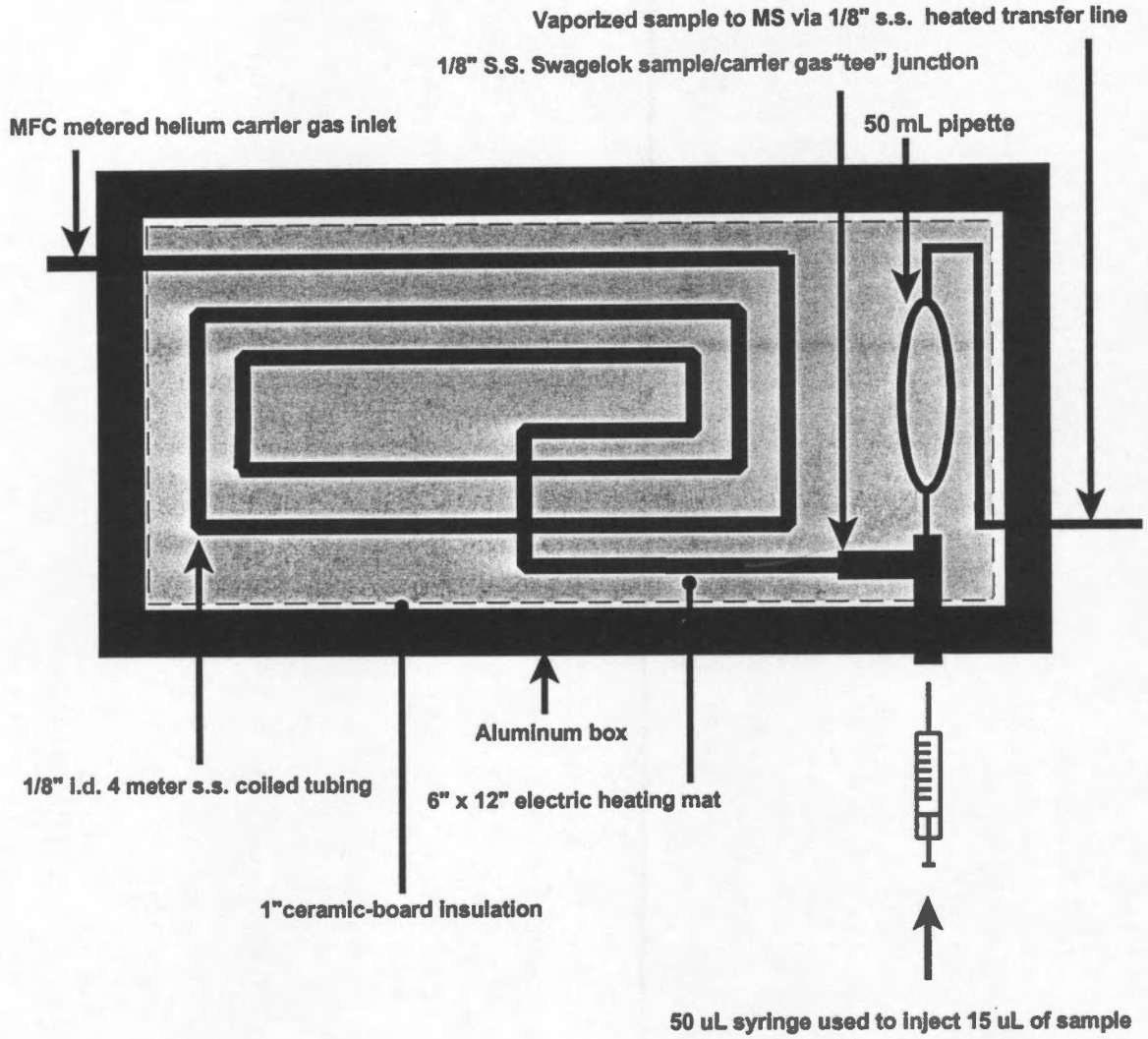


Figure 2.4.e: Injection flash vaporization inlet illustrating carrier gas, heating coil, SS vaporization tee and glass 50 mL pipette where vaporization occurs.

make up the injection inlet. These revisions basically involved replacing a 6 cm section of SS transfer tubing in the flash vaporization inlet with a segment of a 50 mL volumetric pipette. This allowed the 15 μL of injected sample to vaporize / expand in a glass lined environment and then swept into the quick inlet tee by the hot helium carrier gas. A Hamilton (#1705), Gas-Tight, 50 μL syringe (Reno, NV) was utilized to inject samples into the injection vaporization inlet. The two binary stereoisomers and the 9-component systems were sampled using the injection inlet. A helium carrier gas flow rate of approximately 0.9 mL/min was used for these mixtures.

2.5: PARAMETERIZATION

2.5.1: INTUITION; UTILIZING THE ENTIRE MASS SPECTRUM

This method of parameterization is the simplest that is discussed in this work. The intuitive parameterization method involves using all m/z signals that are greater than approximately 0.1% relative to the base peak in a spectrum of an equal-molar mixture of components. This includes 21 m/z signals for the *n*-pentane / isopentane system and 23 m/z signals for the remaining systems investigated. The intuitive nature of this parameterization method is based upon the idea that "the use of more information is better". In other words, the use of the entire mass spectrum (all m/z signals) in the parameterization is anticipated to afford concentration estimates that will be the most accurate. The intuitive parameterization was limited to the 23 total m/z signals to allow comparison to the brute force method of parameterization.

2.5.2: THE BRUTE FORCE METHOD

The brute force method of parameterization finds the most accurate parameterization that exists for a given data set at a targeted concentration. It does this by exhaustively considering all combinations of n or more m/z signals (n = number of components) out of a total number of m/z signals. Using each combination of m/z signals separately, it calculates concentration estimates. The combination of m/z signals that affords concentration estimates that are closest to the true or expected concentrations is considered to be the most accurate combination or the "brute force parameterization". This is done by calculating the sum of squares error (SSE) that exists between the estimated concentrations (from a given combination of m/z signals) and the true concentration which is measured gravimetrically. For several samples of a particular chemical system, the SSE is summed among the samples. The number of combinations can be calculated using **Equation 2.5.2.a**:

$$\textit{Number of combinations} = \sum_{l=m}^{n} \frac{n!}{(l!) * (n-l)!} \qquad \textbf{(2.5.2.a)}$$

In **Equation 2.5.2.a**, l is the loop-number (*i.e.*, 2 ion signals out of 23 total), m is the number of components in the mixture and n is the total number (21 or 23) of ion signals to be considered. 23 ion signals are chosen to afford a reasonable *total* number of combinations (8,388,534 for a 2-component case) with respect to calculation time, the required file space taken up by the resulting data and the total calculation time. In a 23

m/z signal dataset with 2 components, there are 253 possible combinations in the 2-loop case (*i.e.*, combinations of two different m/z's). Note that in the two component case at least 2 ion signals must be used to estimate the concentrations of each of the two components. In other words, there needs to be *at least* as many knowns as there are unknowns in the deconvolution calculation. Analogously, there exist loop-3, loop-4, loop-5, etc., where the number of combinations for each loop mentioned is (2,024), (10,879) and (44,528) respectively. A quick check of this equation for loop-23 and 23 ion signals affords *one* combination. Appendix 2 describes in detail the Matlab code that is used to perform the brute force method of parameterization.

2.5.3: THE EMPIRICAL ALGORITHM

The EA method of parameterization utilizes the differences between m/z signal intensities from the sensitized reference spectra. A mixture spectrum containing approximately an equimolar amount of components used and the entire mass spectrum (maximum of 23 m/z signals) is used to calculate the sensitivity factor. It is not constrained to a square-matrix parameterization (same number of m/z signals in the parameterization as the number of components). Typically, the EA will outperform the stream evaluator method of parameterization due to the lack of this constraint. This is because a square-parameterization has never been identified as an optimum selection upon application of a brute force analysis of the data presented in this work. The EA relies upon the assumption that relatively large differences in m/z signal intensities is the most important criterion upon which to base parameterization. The empirical

parameterization algorithm developed in our laboratory has met with some success but has not yet proven widely applicable and continues to evolve [128][129]. Other algorithms in the literature are intended to provide an estimate of the quality of the *spectral match* between a known analyte and an "unknown" [101]. Identifying the quality of a good spectral match is different from the approach of the EA. The EA approach is to identify those m/z signals with will provide the best parameterization with which to estimate accurate concentration estimates. The EA developed in our laboratory is intended to analyze the signal intensities and rapidly afford a parameterization that is accurate, precise and statistically equivalent to the brute force parameterization.

The empirical algorithm (EA) was developed to provide a more efficient (faster) approach to parameterization by ranking m/z signals in the mass spectra of pure components according to their distinctiveness as estimated by **Equation 2.5.3.a**:

$$Score_{m/z} = | i_{m/z,A} - i_{m/z,B} | \qquad \textbf{(2.5.3.a)}$$

where i is a sensitized (multiplied by the sensitivity factor) reference ion signal intensity at a particular m/z; the subscripts A and B refer to distinct compounds. Thus, the algorithm "scores" each m/z signal in a reference spectrum of a given compound (A), via a pair-wise comparison with intensities in the reference spectrum of a second compound (B). The algorithm is based entirely on comparison of reference spectra, and will generate a relatively large score for the best ions (ion signals with the greatest intensity differences). The largest score is normalized to 100. The average of three reference spectra replicates was used in the EA for this work. The final step in determining the

parameterization involves selection of the number of ranked (sorted m/z signals according to EA score) m/z signals to include in the parameterization. This was done by simply using all m/z signals with scores greater than three. This value was chosen based upon the trends observed upon plotting the scores in rank-order (highest to lowest). There is, typically, a rapid decrease to three in the plotted scores. Thus, at algorithm scores below three the *change* in algorithm score is typically less than above three, implying that as m/z signals are added to the parameterization (those with scores less than three) the addition of information to the matrix equations (solving for the concentration estimates) is probably also decreased. In cases where there are fewer m/z signals in the final parameterization than the number of components (not encountered in this work), additional m/z signals will need to be added to afford, at least, a square-matrix parameterization. Examples utilizing the EA method of parameterization are found in Chapter 3. Appendix 3 describes in detail the Matlab code that is used to perform the EA method of parameterization.

2.5.4: THE GENETIC ALGORITHM

Genetic algorithms are modeled after the principal of natural evolution. Whereas living organisms encode their characteristics into DNA, GA's encode required information into bitstrings (*i.e.*, a computer oriented representation of potential solutions). The bitstrings are comparable with biological chromosomes and each of them is a guess in terms of the optimal solution of the problem under investigation. A collection of these bitstrings is called a population. GA's reproduce and create new generations of

populations. During this process, the bitstrings in the population undergo changes according to the "struggle for life" and "survival of the fittest" principles. In the "struggle for life" a bitstring will suffer from positive and negative influences within its environment. A solution close to the optimum can be changed to a position further away from it and the reverse can also happen. These changes in a GA are caused by mutation and / or crossover which are both controllable operators set at the start of the GA computer program.

Crossover is the mutual exchange of information between the bitstrings while mutation is the random change of bitstring elements. Mutation enhances diversity in the population and thus is part of the optimization strategy that helps to quickly search a large portion of solution space (total number of combinations in the brute force parameterization method). A key to the successful operation of the GA is to guide these changes in such a way as to assure the convergence of the bitstrings upon an optimal solution (a parameterization statistically equivalent in this work). In "survival of the fittest", the bitstrings nearest to the best or the "fittest" solution (bitstrings with the highest fitness) have the highest probability of reproducing whereas the bitstrings with lower fitness have the highest probability of dying. Due to both of these principals, the next generation of bitstrings, on the average, will become more fit and indicate that the optimization process is approaching the optimum solution (a good parameterization). The life cycle of reproduction and dying bitstrings will be terminated when acceptable bitstrings in the population are obtained or when there is no significant improvement after a predefined number of generations. In this work more than 23 m/z signals are utilized in the operation of the GA since it is capable of handling this larger dataset with greater

efficiency than the brute force method.

The genetic algorithm (GA) used in this work is from the Eigenvector PLS toolbox version 2.0 (Eigenvector Research, Inc., Manson, WA). The following GA parameters are utilized. Population: 56, Window width: 1, Initial terms: 50, Maximum generations: 500, Convergence: 100%, Mutation rate: 0.01, Crossover: Double, Regression choice: Partial Least Squares, Number of latent variables: 10, Cross-validation parameters: Random with 5 subsets and 5 iterations. These parameters were recommended by Gemperline [130].

The GA is executed by using the following Matlab command: genalg(ms_all,Truedat(:,:),'outfit','selvar'). The "genalg" portion of this command invokes the GA Matlab program contained in a text-file named genalg.m. A graphical user-interface screen becomes available to adjust the GA parameters that are listed above. The variable ms_all contains the mixture mass spectral intensities and Truedat contains the mole fraction values of each mixture. The variable ms_all contains 27 mass spectra (9 samples with three replicates) for the *n*-pentane / isopentane data and 15 spectra (5 samples with three replicates) for the cis- / trans-piperylene, cis- / trans-2-pentene isomers. Separate files with data from each chemical system in each program were used. The variables outfit and selvar are used to store the results generated by the GA. The GA affords one or more parameterizations ranked by a fitness scalar.

Upon completion of the GA calculations the following commands are utilized to extract the parameterization from the outfit and selvar variables: ix = find(selvar(F,:)==1), this command sets the new variable ix equal to the selected variable (selvar, the parameterization) denoted by the F variable. F is entered into the previous

command by the user (*i.e.*, the user identifies the best fitness (F) indicated in the Matlab output). This value was typically 1 for the data tested in this work. The GA fitness scalar is analogous to the SSE used in the brute force parameterization method. The lowest fitness scalar implies that the associated parameterization affords concentration estimates with the least amount of error as compared to the true mole fractions in the variable Truedat. The following command is utilized to identify the m/z signal ratios associated with the parameterization selected by the GA: mzs = masses(ix). The variable masses is a list of m/z ratios in the same order as was entered in the variable ms_all. The variable mzs is the final parameterization determined by the GA method of parameterization.

2.5.5: DRAHOS - VEKEY SIMILARITY INDEX

The Drahos - Vekey similarity index (SI) [131] is used to measure the differences that exist between intensities in a pair of reference spectra. Since information in the base peak is lost (signals normalized to 100) upon normalizing the reference spectra the process of sensitizing the reference spectra with an equimolar mixture is used to regain that lost information. This sensitization step is the same as that taken for the empirical algorithm. This is the same step that is taken in the brute force parameterization method listed in Appendix 2, where lines 6 - 8 calculate a sensitivity factor and line 15 utilizes that factor in the multiplication of the normalized reference spectra.

$$Similarity\ Index = 100 - 100\left[\frac{\sum_{m/z=1}^{n} \left| i_{m/z,A} - i_{m/z,B} \right|}{\sum_{m/z=1}^{n} \left(i_{m/z,A} + i_{m/z,B} \right)}\right] \quad \textbf{(2.5.5.a)}$$

Equation 2.5.5.a illustrates how the SI is calculated from the reference spectra. The numerator and denominator in the SI are summed from $m = 1$ (the first m/z signal) to the total number of m/z signals (n) where i is the m/z signal intensity and A & B are the two different spectral components under comparison. Examples utilizing the SI are illustrated in Chapter 4. Appendix 4 describes in detail the Matlab code that performs the SI calculations.

2.5.6: THE STREAM EVALUATOR

The Stream Evaluator is a parameterization method which is included in the instrument operating software sold and distributed with the Questor IV process mass spectrometer by ABB-Extrel [100]. The stream evaluator utilizes information from normalized reference spectra. However, the stream evaluator requires additional information regarding the composition of the stream to be parameterized. For the two component systems tested in this work, each component was entered into the stream evaluator at a 50/50 % (wt/wt) composition. Other parameters used to afford a parameterization were default. Specifically, the weights used were; RSD = 100.0, RIF = 100.0, RIS = 5.0, Interference = 0.0, Dual detector = not selected.

CHAPTER 3

3.1: EFFECTS OF PARAMETERIZATION & CALIBRATION ON ACCURACY & PRECISION OF CONCENTRATION ESTIMATES

Whereas PrMS has been used successfully for analysis of multiple, distinct, volatile components in chemical processing streams, many important streams like those found in refineries or polymer processing include multiple isomers. For example, such is the case for a chemical processing stream containing a relatively large portion of piperylene (1,3-pentadiene), the major constituents of which are listed in **Table 3.1.a**. The approximate contents of the piperylene stream were made available to our laboratory by an MCEC collaborator who was interested in applying PrMS to this chemical

Table 3.1.a: Approximate compositions of piperylene stream.

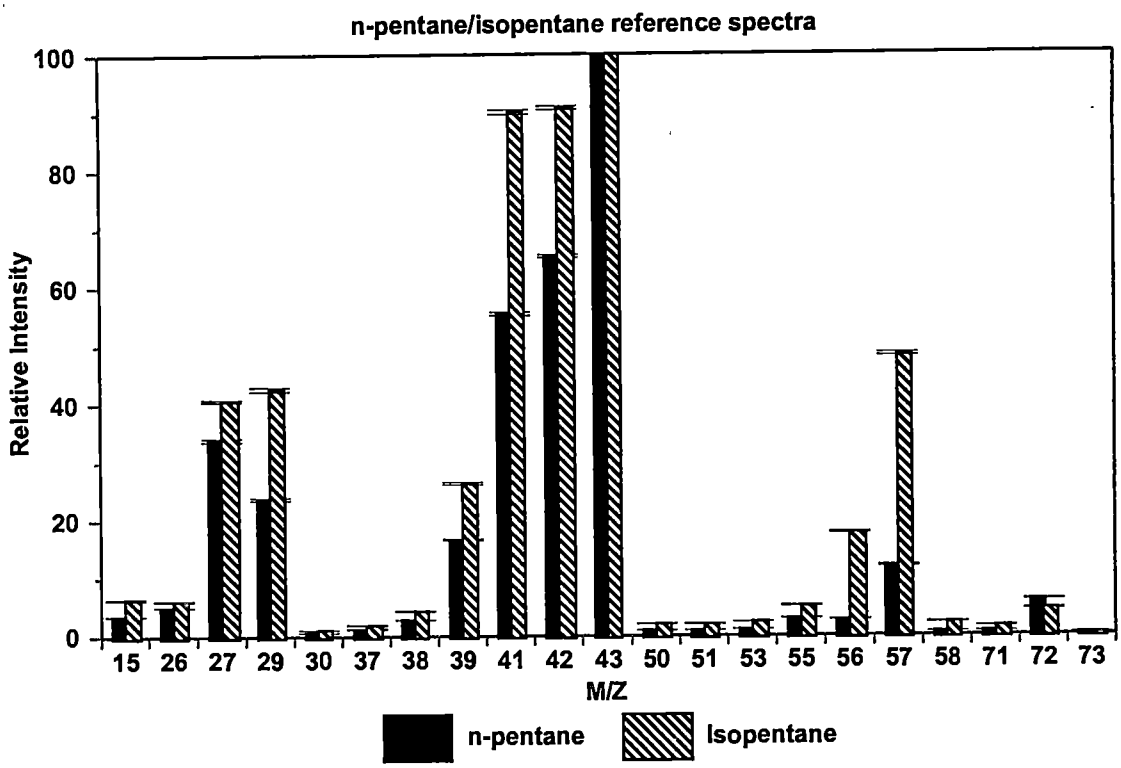
Component	Approximate Composition (mole fraction)	Component	Approximate Composition (mole fraction)
cis-piperylene *	0.21	2-methyl-2-butene **	0.11
trans-piperylene *	0.36	*n*-pentane ***	0.028
cyclopentene *	0.16	isopentane ***	0.0045
cis-2-pentene **	0.018	2,2-dimethylbutane	0.027
trans-2-pentene **	0.015	Unknowns	0.036
cyclopentane **	0.031	-	-

*:C_5H_8 isomers. **:C_5H_{10} isomers. ***:C_5H_{12} isomers.

processing stream. The feasibility of utilizing PrMS to analyze isomeric hydrocarbon systems as complex as those that are found in mixtures containing the components in **Table 3.1.a** without the use of chromatography, MS/MS or derivitization has not been found in the literature associated with the topics discussed in this thesis. Some of the most closely related work involves the use of principal components analysis with MS/MS data of isomeric monoalkylated naphthalenes [132] and the analysis of peptides containing D- or L- stereoisomers by electrospray ionization mass spectrometry [133]. This work involves use of tandem mass spectrometry and electrospray ionization mass spectrometry which remain to be proven as robust, reliable process analysis techniques.

First efforts at isomeric hydrocarbon mixture analysis using process mass spectrometry in our laboratory involved analysis of binary mixtures of *n*-pentane and isopentane [134][135]. These mixtures served as prototype mixtures of the piperylene stream. The initial approach to parameterizing (selecting m/z signals with which to perform the least-squares-calculations) mixtures of these components was simply to utilize all m/z signals that were approximately 0.1% or larger in intensity relative to the base peak in the reference spectra. This is considered an intuitive approach to parameterization since it is intuitive that the use of as much information as possible (avoiding the m/z signals that are comprised mainly of noise) is a reasonable approach. Selecting all m/z signals greater than 0.1% of the base peak (intuitive parameterization) in *n*-pentane / isopentane includes those illustrated in **Figure 3.1.a.**

When considering these two reference mass spectra it can be seen that, except for the m/z signal at 72, the relative intensities for isopentane are larger than for *n*-pentane. This is consistent with literature mass spectra for these components and is also reasonable

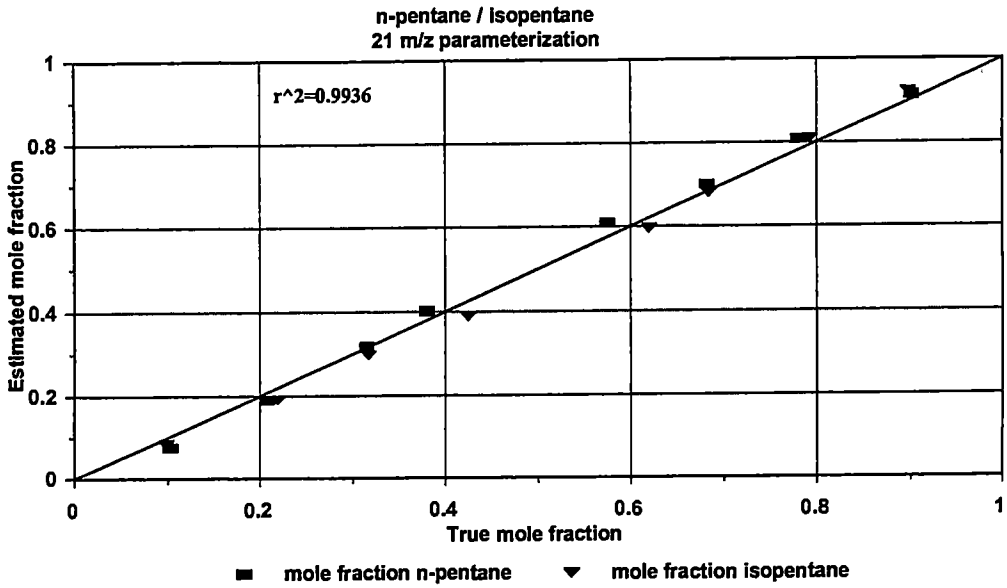


M/Z:	Average intensity, n-pentane	3 standard deviations	Average intensity, isopentane	3 standard deviations
15	3.579	0.054	6.455	0.123
26	5.080	0.008	6.153	0.081
27	33.910	0.277	40.687	0.201
29	23.747	0.227	42.619	0.351
30	0.629	0.006	1.104	0.015
37	1.285	0.010	1.941	0.011
38	2.835	0.019	4.326	0.067
39	16.640	0.096	26.370	0.168
41	55.391	0.288	90.177	0.377
42	65.318	0.281	90.853	0.289
43	100	0	100	0
50	0.933	0.003	2.055	0.012
51	0.935	0.007	2.005	0.018
53	1.107	0.013	2.370	0.019
55	2.974	0.035	5.155	0.021
56	2.719	0.020	17.763	0.093
57	12.016	0.110	48.271	0.228
58	0.569	0.006	2.283	0.012
71	0.631	0.006	1.553	0.009
72	6.071	0.072	4.485	0.073
73	0.350	0.004	0.256	0.004

Figure 3.1.a: E.I. mass spectra of *n*-pentane and isopentane. Error bars are 3 standard deviations calculated from three replicates and are tabulated above.

when considering the relatively larger ionization cross-section of isopentane as compared to that of *n*-pentane (a larger molecular cross-section will provide a larger target area for ionization and thus more fragmentation due to additional energy absorption). This effect is also evident in the sensitivity factors that are calculated from the normalized reference spectra and the 50 / 50 mixture (sample N5, Table 2.2.b). In this case the sensitivity factors for *n*-pentane / isopentane are 1.17 and 0.871 respectively. The factor for isopentane is *smaller* than one for *n*-pentane and the factor for *n*-pentane is *larger* than one. This is because, upon multiplication with the normalized reference spectrum of isopentane, the relative intensities (the sensitized mass spectrum) must be diminished while the intensities for *n*-pentane require an increase to be equally represented in the mass spectra on the basis of their differing sensitivities.

This parameterization method (21 m/z signals), when used to deconvolve the eight binary mixture spectra of different compositions listed in **Table 2.2.b** (N1 - N4 and N6 - N9), is found to afford both accurate and precise concentration results. Sample N5 is used for calibration of the component sensitivities. The validation plot in **Figure 3.1.b** illustrates the success in deconvolving these mixtures; the correlation coefficient (r^2) is 0.9936. This r^2 is calculated with respect to the Y = X line and is not the conventional r^2 which is calculated to the best fit line (Y = mX + b). This is because the compositions of the samples are known since they were prepared in the laboratory. The method of calculating r^2 is illustrated in Appendix 5. A perfect r^2 is equal to one. Since good accuracy and precision do not necessarily go together, the precision of concentration estimates was investigated for this parameterization. This was performed by calculating the standard deviation of concentration estimates from the three replicates and creating



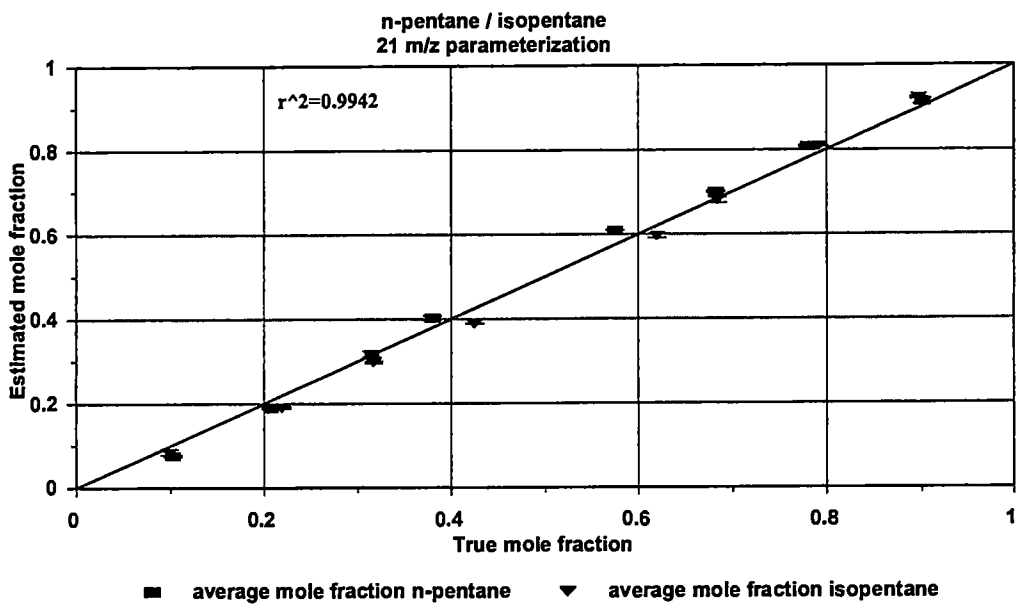


Figure 3.1.b: Top figure: Validation plot for *n*-pentane / isopentane. Eight samples, of two components with three replicates (48 points) are plotted. r^2 = 0.9936 indicates good correlation to the Y = X line. Parameterization utilized: 21 m/z signals from the intuitive parameterization method. Some points are not discernable on the plot due to overlap. Bottom figure is average of three replicates with error bars (16 points). Error bars are three standard deviations, some are not discernable due to overlap with the average estimated concentration symbol.

error bars. The error bars are three standard deviations with respect to the average of the estimated concentrations. The bottom validation plot of **Figure 3.1.b** illustrates the average concentration estimates with the error bars. The r^2 (0.9942) improved slightly upon averaging the concentration estimates. Some of the error bars are not discernable due to overlap with the average estimated concentration symbols. Upon investigating the percent relative standard deviations (%RSD) of concentration estimates for this system a range from 2.6% to 0.019% was observed. The highest %RSD was obtained from the extreme range of concentration estimates (sample N1). Whereas, the lowest %RSD was obtained from a sample (N4) that was close to the concentrations of the calibration sample (N5).

Regarding an analysis of the accuracy, a table of percent relative errors calculated from the concentration estimates obtained from the intuitive parameterization was considered. These percent relative errors are listed in **Table 3.1.b**.
Similar to observations in the precision made previously, the largest percent relative error was obtained from samples that had concentrations that were farthest from the calibration sample (N1 and N9). The smallest percent relative error was obtained from samples that were close to the calibration sample (N3 and N4). Note that the largest mole fraction error was 2.6% for *n*-pentane and 1.6% for isopentane.

Thus, the accurate and precise deconvolution of this prototype binary mixture of positional hydrocarbon isomers is possible at the composition levels indicated in **Figure 3.1.b**.

Table 3.1.b: Percent relative errors for *n*-pentane / isopentane (intuitive parameterization).

	% relative error *n*-pentane	% relative error isopentane
Maximum error	26 (-0.0263)*	16 (-0.0158)*
Minimum error	0.076 (-0.000239)*	0.035 (-0.000240)*

*: Mole fraction error (non-relative) in parentheses

In PrMS knowing how often to re-calibrate is an important issue. It is for this reason that all chemical systems in this work were tested in triplicate. For the deconvolution of the *n*-pentane / isopentane mixtures, the reference spectra were re-acquired and used in each new replicate (*i.e.*, re-calibration was performed on a once-per-day frequency) to achieve optimum accuracy and precision. When re-calibration is not performed (*i.e.*, the reference spectra from replicate one are utilized for the data from replicates two and three) the results, illustrated in the validation plot of **Figure 3.1.c**, for the *n*-pentane / isopentane were obtained. As can be seen in the figure, the accuracy does not change appreciably without daily recalibration. In this case the correlation coefficient decreases from 0.9936 to 0.9935 which is negligable. However, in other chemical systems, larger decreases in correlation coefficient were observed. Thus, re-calibration on a day-to-day basis is not a necessity for the *n*-pentane / isopentane system. Recalibration at a frequency less than once-per-day was not pursued in this work since it was predicted that the day-to-day recalibration frequency may be appropriate.

For *non*-isomeric hydrocarbon chemical systems, daily recalibration is typically not necessary [136] and a more common recalibration frequency of once-per-month has been reported by MCEC industrial members. Greater calibration frequency was

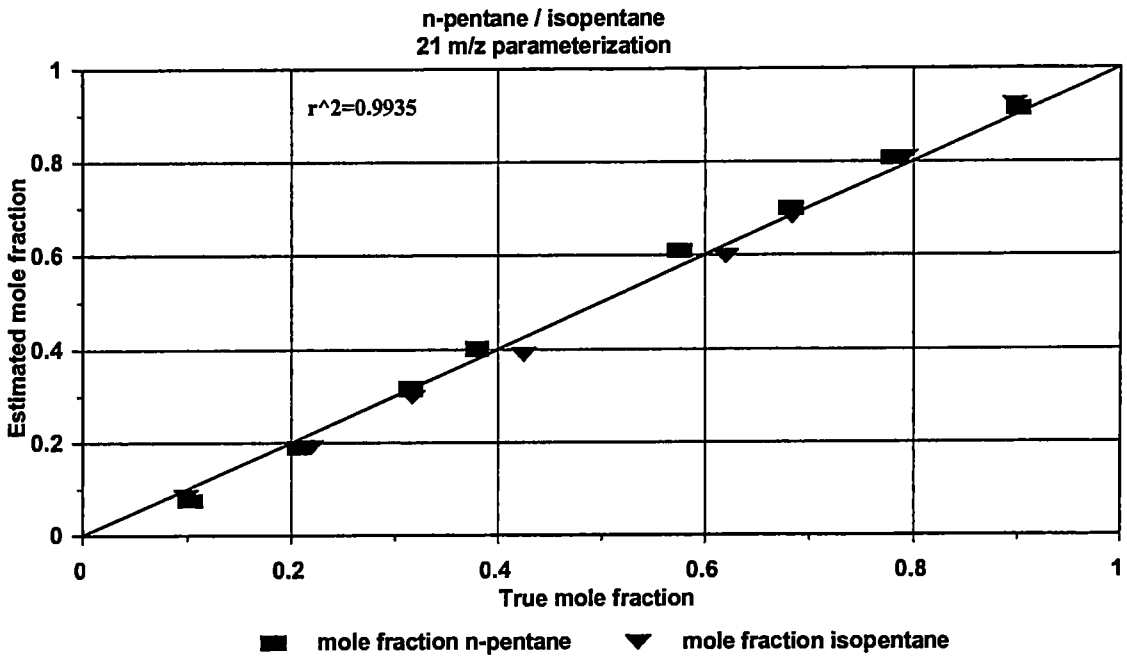
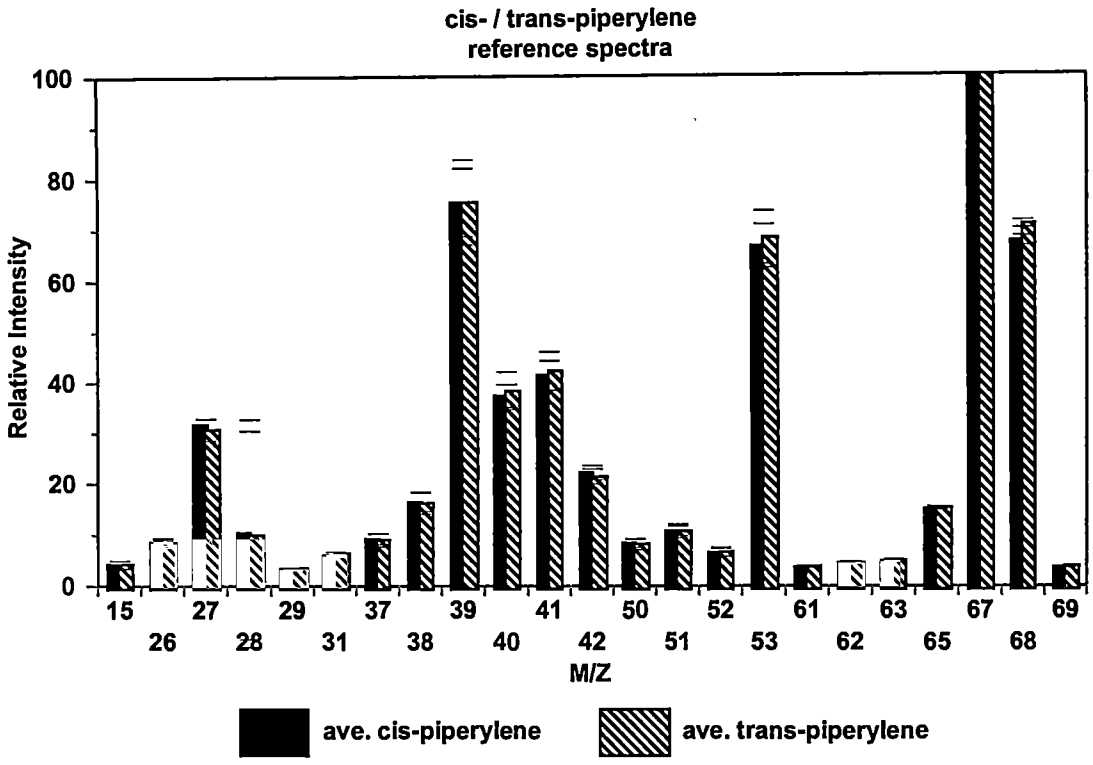


Figure 3.1.c: Validation plot for *n*-pentane / isopentane using reference spectra from replicate one. Eight samples of two components with three replicates (48 points) are plotted. Parameterization utilized: 21 m/z's from intuitive parameterization method. Some points are not discernable on the plot due to overlap.

anticipated for isomers due to the fact that the deconvolution (quantitation) of these mixtures is based upon the small differences that exist between the intensities at the same m/z signal. Thus a relatively small change in calibration was anticipated to be detrimental to the accurate quantitation of isomeric mixtures. The quantitation of *non*-isomeric systems is based upon *unique* signals typically without the interferences (intensity contributions from different components at the same m/z signal) that occur in isomeric systems. Thus, for the *n*-pentane / isopentane system the calibration is accurate for at least three days. However, a daily recalibration frequency is not unreasonable, in a practical sense, since the ABB-Extrel instrument can easily be programmed to automatically recalibrate.

In the interest of testing prototype mixtures of *stereoisomers* listed in **Table 3.1.a**, the five samples of cis- / trans-piperylene listed in **Table 2.2.b** (P1 - P5) were prepared and analyzed using the process mass spectrometer. **Figure 3.1.d** illustrates m/z signals in the reference spectra of the cis- / trans-piperylene pure components that exist at relative intensities greater than approximately 3% (the intuitive parameterization). Upon applying this parameterization to binary mixtures of these components and utilizing sample P2 as the calibration sample, the results in **Figure 3.1.e** were obtained. As can be seen in the figure the concentration estimates were not of high accuracy or precision; the r^2 was 0.4832 which indicates poor performance. Also apparent in the figure, this parameterization afforded some concentration estimates that were negative. While it is *mathematically* possible to obtain negative concentration estimates there is no reasonable *physical* definition of a negative concentration estimate and thus these results are non-sense solutions meaning that the intuitive parameterization is inappropriate in this case.



M/Z:	Average intensity cis-piperylene	3 standard deviations	Average intensity trans-piperylene	3 standard deviations
15	4.244	0.780	4.248	0.685
26	8.574	0.646	8.459	0.892
27	31.675	1.151	30.688	2.206
28	10.392	22.296	10.008	20.423
29	3.373	0.313	3.501	0.276
31	5.995	0.199	6.402	0.339
37	9.193	0.979	9.037	1.276
38	16.431	1.836	16.219	2.151
39	75.343	6.645	75.359	8.269
40	37.349	2.229	38.369	3.701
41	41.459	2.883	42.307	3.851
42	22.172	1.387	21.409	1.483
50	8.366	0.760	8.187	1.026
51	10.723	0.953	10.700	1.345
52	6.517	0.767	6.706	0.870
53	66.783	4.164	68.428	5.137
61	3.776	0.150	3.796	0.215
62	4.539	0.164	4.599	0.265
63	4.899	0.216	5.046	0.263
65	15.098	0.094	15.267	0.322
67	100	0	100	0
68	67.702	0.920	70.896	0.742
69	3.646	0.097	3.873	0.033

Figure 3.1.d: E.I. mass spectra of cis- and trans-piperylene reference spectra (23 m/z's). Error bars are 3 standard deviations from 3 replicates.

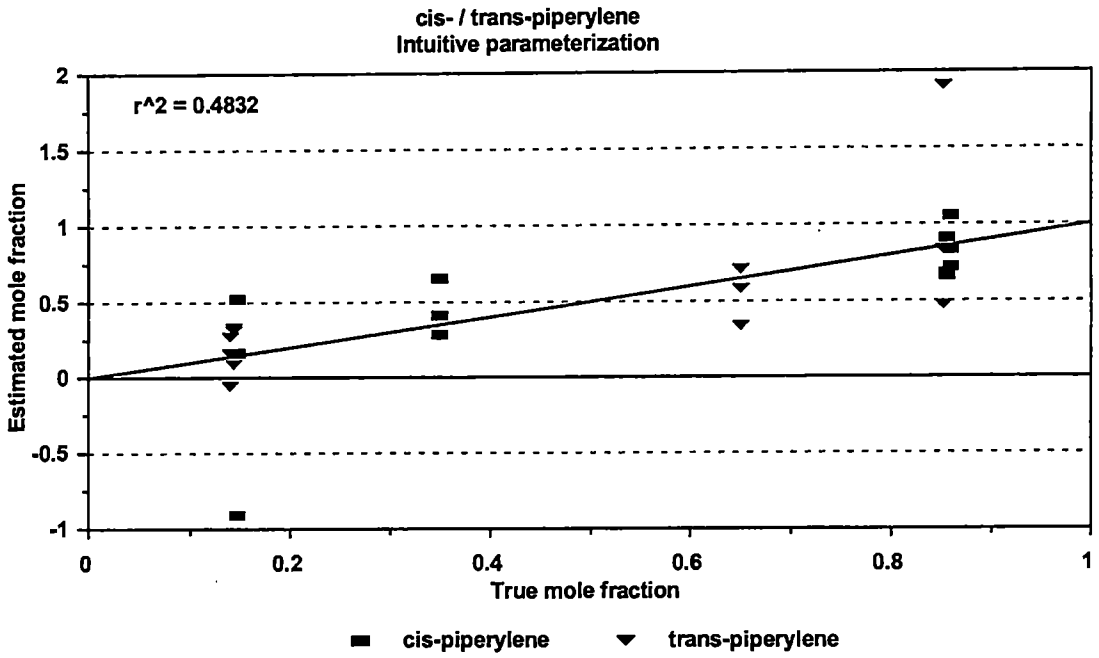


Figure 3.1.e: Validation plot utilizing intuitive parameterization method and 50/50 calibration sample. Four samples of two components with three replicates are plotted (24 points).

A similar argument applies to concentration estimates that are greater in value than a mole fraction of one.

Upon consideration of the reference spectra in **Figure 3.1.d** it can be seen that at m/z = 28 the standard deviation is relatively large. This is due to the air plug that is introduced into the syringe when dispensing the 15 μL of sample into the heated inlet. Upon loading the liquid sample into the syringe a 15 μL plug of air is drawn into the syringe to reduce the loss of sample from the end of the syringe needle upon transferring it into the vaporization inlet. It is from this step that the signal at m/z = 28 (nitrogen in air) is derived. The large standard deviation at m/z = 28 is due to the fact that the sample liquid was more precisely metered than the air plug. In other words the liquid sample was measured to 15 ± 0.5 μL whereas the air plug was 15 ± 5 μL.

Upon considering the possibility that the presence of m/z = 28 in the intuitive parameterization may afford a large contribution to the estimated concentration outliers in **Figure 3.1.e**, m/z = 28 was removed from the parameterization and the estimated concentrations were recalculated. The correlation coefficient value improved from 0.4832 to 0.4840; however, the negative concentration estimates (and estimates > one) persisted. Also, upon comparing the standard deviations between the reference spectra of *n*-pentane / isopentane system and the cis- / trans-piperylene systems, it is evident that the standard deviation is generally larger in the piperylene system even without considering the air peak at m/z 28. This was attributed to the difference in reactivity between the piperylenes and the pentanes (dienes vs. alkanes). This issue is discussed in Chapter 4. The ion signal at m/z = 28 was absent from the pentanes system due to the difference in sampling between the stereoisomers and the pentanes (injection inlet and continuous

inlet). Thus, the signal at m/z = 28 was among those cut out below the threshold of 0.1%.

Upon testing different parameterizations in the *n*-pentane / isopentane [135] and cis- / trans-piperylene system, it was found that the accuracy and precision of concentration estimates depended strongly upon the parameterization. For example, if an appropriate parameterization is selected, accurate and precise concentration estimates can be calculated. The use of some parameterizations afforded concentration estimates that were negative (as in the cis- / trans-piperylene system) and thus should be avoided. Thus, it was suspected that accurate and precise deconvolution of these stereoisomeric mixtures of hydrocarbons might indeed be obtained, given appropriate parameterization. The issue at hand involved *how* to select an appropriate parameterization that would give an accurate and precise set of concentration estimates.

One way to be sure of selecting the most *accurate* parameterization involves use a brute force approach to parameterization. This method consists of testing every possible combination of m/z signals by calculating a concentration using each one and evaluating its accuracy. For example, the intuitive parameterization method involved calculation of concentration estimates using all 21 m/z signals (one combination). Since it is necessary to use a minimum of 2 m/z signals for a binary mixture of components, the range of combinations stretches from all possible combinations of 2 to 21 with the total number of possibilities described by **Equation 2.5.2.a**. Thus after application of the brute force method of parameterization, one can be assured that the resulting parameterization is the most accurate for the data set at the concentration range defined by the samples tested.

Upon application of the brute force method of parameterization to the *n*-pentane / isopentane system, results illustrated by the validation plot in **Figure 3.1.f** were obtained.

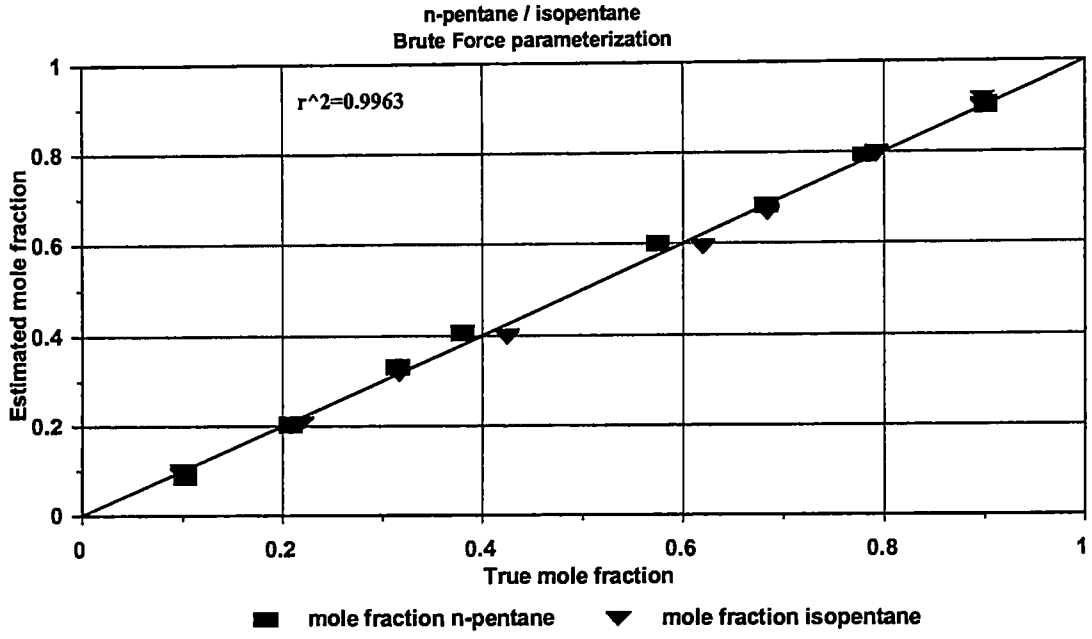


Figure 3.1.f: Validation plot for *n*-pentane / isopentane. Eight samples of two components with three replicates (48 points) are plotted. $r^2 = 0.9963$ indicates good correlation to the Y = X line. The brute force parameterization (m/z's: 26, 30, 72) was utilized to calculate concentration estmates. Some points are not visible on the plot due to overlap.

In the figure it can be seen that the correlation coefficient from the brute force method of parameterization (r^2=0.9963) is slightly improved relative to that obtained using the intuitive method (r^2=0.9936). This parameterization is the most accurate for all three replicates of the data. Also upon comparison of the number of m/z signals contained in the brute force parameterization vs. the intuitive method (3 vs. 21), it can be seen that the former parameterization makes better use of the process mass spectrometer's scanning time. This is an important issue since, in this case, fewer m/z signals in the parameterization afford better accuracy of concentration estimates. Also, the analysis time saved from scanning fewer m/z signals will provide the instrument more time for automated calibration or analysis of other chemical processing streams (multiplexing of the instrument).

The success of the brute force method of parameterization prompted its application toward the cis- / trans-piperylene system, the results of which are illustrated in the validation plot of **Figure 3.1.g**. In this figure it can be seen that the accuracy and precision of concentration estimates is drastically improved compared to the intuitive parameterization method (r^2 = 0.9054 vs. 0.4840). Similar to what was observed in the *n*-pentane / isopentane system, the number of m/z signals in the parameterization has largely decreased in comparison to the intuitive method of parameterization (4 vs. 21), affording the further advantage of reduced scanning time. Thus, the stereoisomers present an additional challenge as compared to the *n*-pentane / isopentane system.

In the interest of checking other stereoisomeric systems present in the piperylene stream, mixtures of a second pair of cis- / trans isomers (the 2-pentenes listed in **Table 3.1.a**) were prepared and tested. This system was intended as a check to discern the

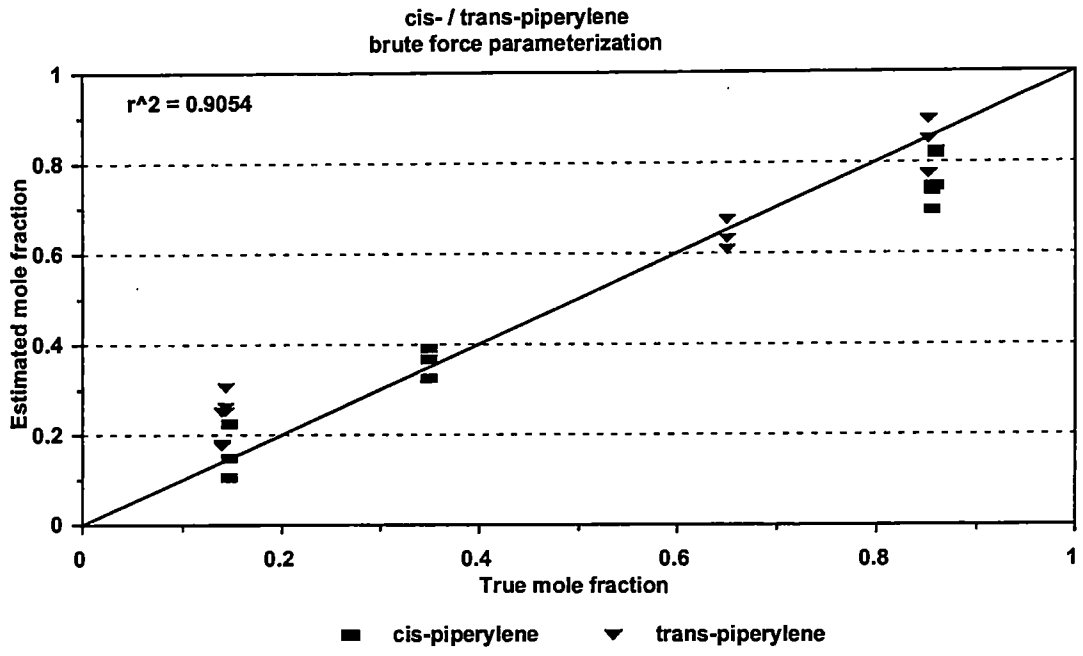
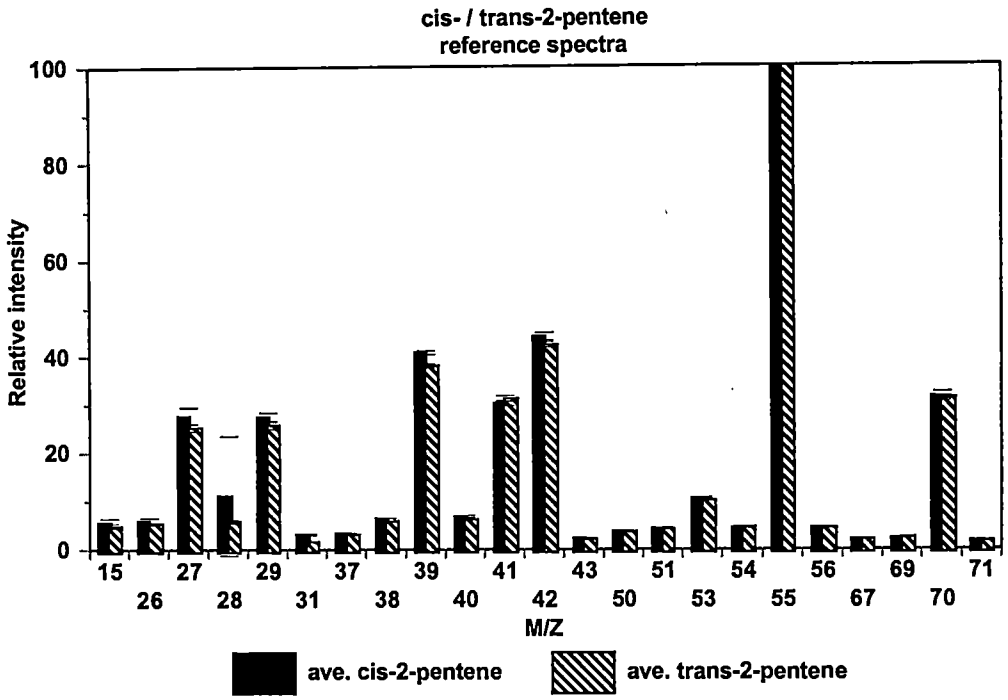


Figure 3.1.g: Validation plot utilizing brute force parameterization method with 50/50 calibration sample. Four samples of two components with three replicates are plotted (24 points). Brute force parameterization m/z's: 38, 61, 62, 63.

effects of spectral similarity and / or reactivity. The reference spectra of the cis- / trans-2-pentenes are illustrated in **Figure 3.1.h** and the relatively small differences between the isomer signal intensities are much like the differences observed in the reference spectra for the cis- / trans-piperylenes illustrated in **Figure 3.1.d.** However, the general precision for this system is comparable to the pentanes system. The signal at m/z = 28 remains the least precise and similar to the situation pointed out in the piperylene system, thus it was left out of the m/z signals considered for parameterization. The injection inlet was used for this system in anticipation that reactivity would cause fouling problems analogous to the piperylene system. Due to the relatively small differences between the signal intensities, it was anticipated that a brute force parameterization method would be required for the accurate and precise deconvolution of these mixtures. As expected, application of an intuitive parameterization (all 23 m/z signals) to the cis- / trans-2-pentene system afforded a relatively low correlation coefficient value ($r^2 = 0.4441$) as illustrated in **Figure 3.1.i.** Again, the application of the brute force method of parameterization indeed increased the correlation coefficient value in the validation plot to $r^2 = 0.9986$ as can be seen in **Figure 3.1.j.**

It was anticipated that, due to the small differences between the signal intensities in the piperylene isomers, with respect to what was determined for the pentanes system, a higher calibration frequency may be required. Thus, to test for this, the stereoisomer systems were analyzed using only the reference spectra from replicate one analogous to the manner in which the pentanes system was tested. The intuitive parameterization method was used for these comparisons because it represents the precision of a larger portion of the mass spectrum as compared to a few m/z signals which are derived from



M/Z:	Average intensity cis-2-pentene	3 standard deviations	Average intensity trans-2-pentene	3 standard deviations
15	5.592	0.762	4.896	0.449
26	5.994	0.529	5.388	0.195
27	27.702	1.796	25.293	0.761
28	11.139	12.374	5.814	0.372
29	27.538	0.856	25.736	0.777
31	3.133	0.125	1.670	0.092
37	3.384	0.090	3.073	0.054
38	6.428	0.115	5.884	0.080
39	40.827	0.378	38.110	0.449
40	6.798	0.222	6.229	0.103
41	30.328	0.517	31.263	0.565
42	44.197	0.809	42.468	0.527
43	2.335	0.061	2.059	0.041
50	3.787	0.153	3.633	0.025
51	4.311	0.142	4.145	0.032
53	10.568	0.131	10.113	0.073
54	4.467	0.140	4.565	0.195
55	1.00	0	1.00	0
56	4.524	0.007	4.464	0.031
67	2.044	0.041	2.012	0.014
69	2.271	0.066	2.446	0.080
70	31.596	0.962	31.333	0.372
71	1.758	0.042	1.734	0.061

Figure 3.1.h: E.I. mass spectra of cis- / trans-2-pentene reference spectra (23 m/z's). Error bars are 3 standard deviations and are tabulated above.

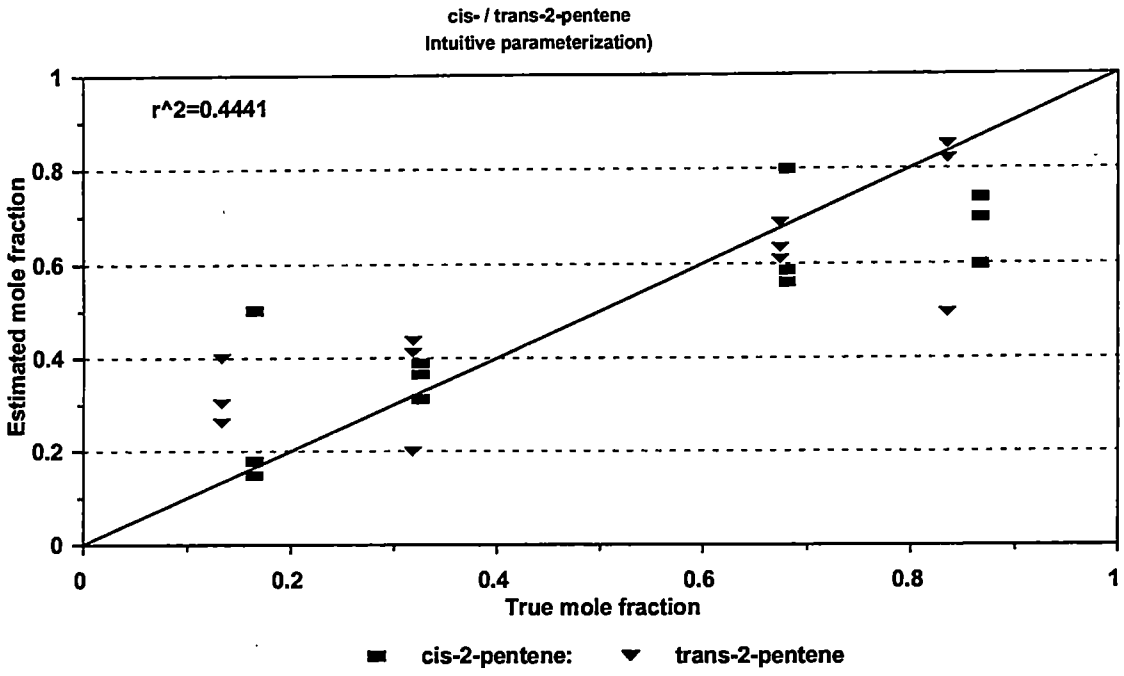


Figure 3.1.i: Validation plot utilizing intuitive full m/z spectrum parameterization method calibrated with 50/50 sample. Four samples of two components with three replicates are plotted (24 points).

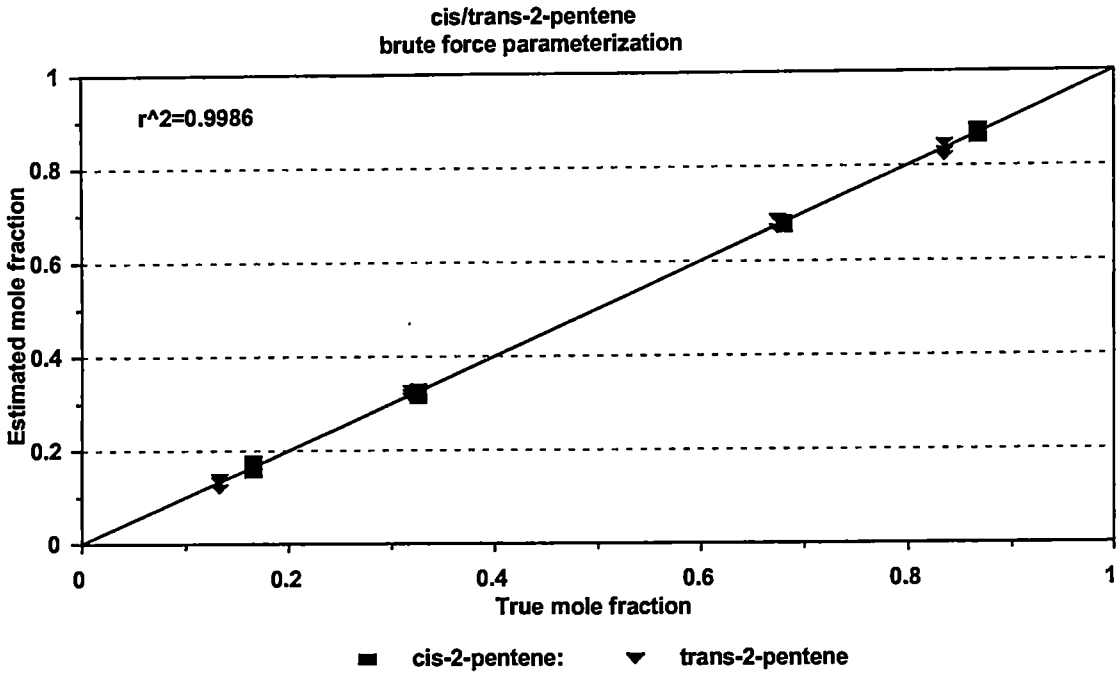


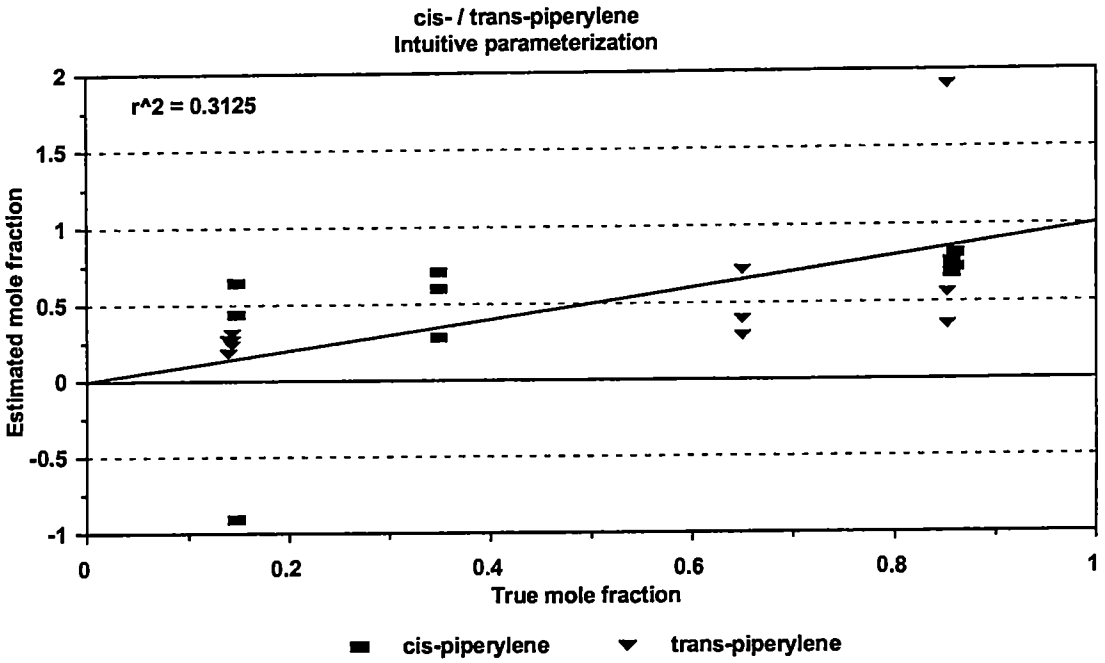
Figure 3.1.j: Validation plot utilizing brute force parameterization method calibrated with 50/50 sample. Four samples of two components with three replicates are plotted (24 points).

the brute force parameterization method. **Figure 3.1.k** illustrates the validation plots obtained for these cases. In the cis- / trans-piperylene validation plot, located at the top of the figure, there remains a problem with negative concentration estimates as well as concentration estimates exceeding a mole fraction of one. This reinforces the issue of appropriate parameterization.

In contrast to the results obtained from the pentanes system, higher accuracy and precision are obtained upon re-calibration in the stereoisomer systems. Thus, it is illustrated in **Figure 3.1.k** that correlation coefficient results of 0.3125 and 0.2119 are obtained for the piperylene and 2-pentene systems upon acquiring the reference spectra once for three consecutive days of data in the stereoisomer systems. It was shown earlier that correlation coefficient results of 0.4840 and 0.4441 could be obtained upon recalibrating (reacquiring the reference spectra) once per day. Hence, a calibration frequency of greater than once per three days is more accurate for the stereoisomer systems. A recalibration frequency greater than once-per-day was not attempted in this work but may afford more accurate results.

Thus, for optimum performance (highest accuracy of concentration estimates using the fewest m/z signals), brute force parameterization can be applied in addition to frequent (once per day) calibration, especially for stereoisomer systems.

A caveat to be noted in association with the brute force method of parameterization is the possibility that it may be over-fitting the data. For example, in the case of the pentanes system (21 total m/z signals and two components) there are over 8 million different parameterizations or treatments of 27 different spectra (9 samples, 3 replicates). In addition, the selection of m/z signals that the brute force method provides



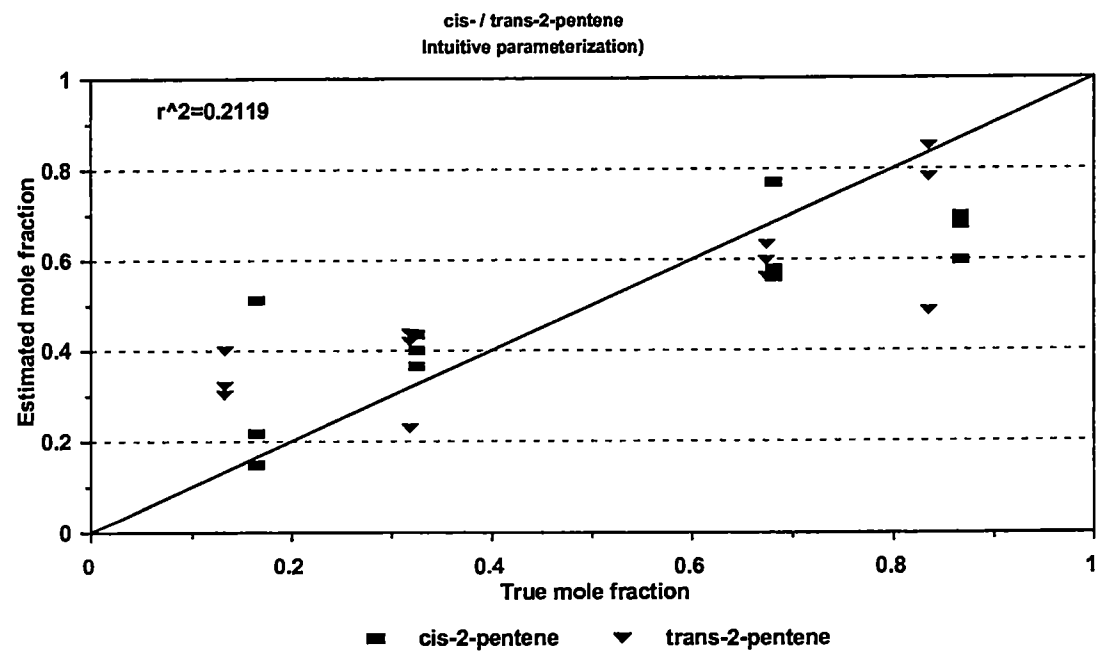


Figure 3.1.k: Validation plot for stereoisomer systems using reference spectra from replicate one. Four samples of two components with three replicates (24 points) are plotted. Intuitive parameterization was utilized. Some points are not visible on the plot due to overlap.

are not obvious with regard to spectral intensity or intensity differences between components. For example, **Figure 3.1.1** illustrates the brute force parameterization for the pentanes and 2-pentenes systems by highlighting the chosen m/z signals in the sensitized (reference spectra multiplied by the sensitivity factor) reference spectra. From the figure, it can be noted that, in general, the highlighted m/z signals do not have any obvious criteria for selection. In fact, out of the brute force m/z signals among both parameterizations, the intensities, except for m/z = 41, for pentene are less than 10%. It is counter-intuitive that the parameterizations which afford the most accurate concentration estimates would be derived from m/z signals with some of the lowest intensities and smallest intensity differences.

3.2: FASTER METHODS OF PARAMETERIZATION

An additional caveat associated with the method of brute force parameterization is its calculational exhaustiveness. For a two-component system with 23 m/z signals there are over 8 million different combinations for which concentration estimates and accuracy calculations (SSE) are performed as well as saved to file. Even with the rapid speed of modern day computers, this large number of calculations and operations requires a significant amount of time. In this case, approximately 20 hours was required on a time-shared computer. Often it is the case that there exist more than 23 m/z signals in the electron impact ionization mass spectrum of a complex mixture. **Table 3.2.a** illustrates the total number of combinations that is calculated when applying **Equation 2.5.2.a** to a two-component case with the indicated number of total m/z signals. From **Table 3.2.a**, it

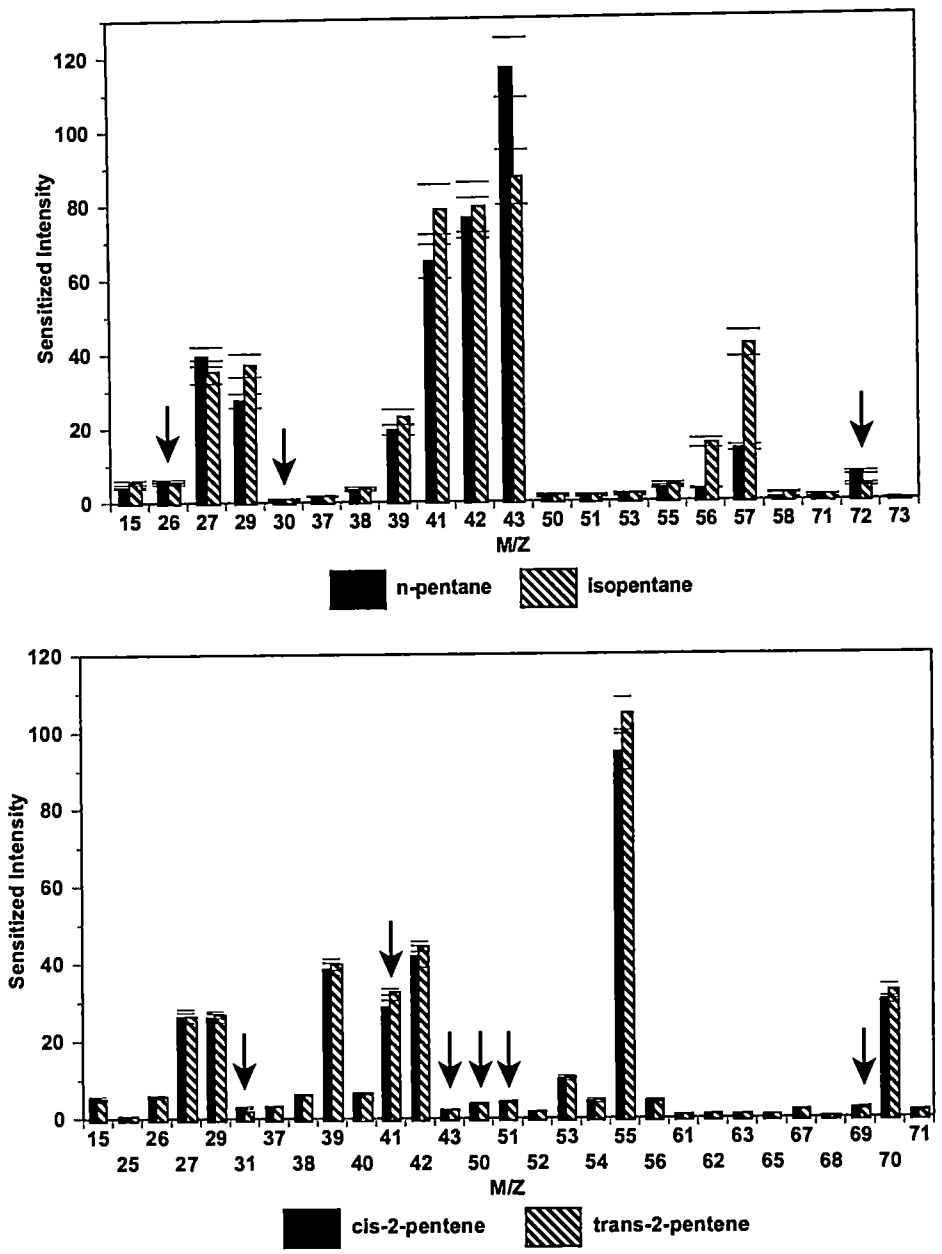


Figure 3.1.1: Pure components n-pentane / isopentane & cis- / trans-piperylene with sensitivity adjusted intensities. The brute force parameterization for each system is highlighted using arrows.

is evident that the total number of combinations becomes intractable at about 30 m/z signals. At a calculation / archive rate of approximately 20 hours for 8 million combinations, the one billion combination task for 30 m/z signals would require about 26,000 hours. This is why the total number of m/z signals is limited in the brute force parameterization method to 23 or fewer.

Table 3.2.a: Total number of combinations for a 2-component system with the indicated total number of m/z signals.

Number of m/z signals	Total number of combinations	Number of m/z signals	Total number of combinations
2	1	23	8,388,584
4	11	30	1.0737E09
8	247	50	1.1259E15
21	2,097,130	100	1.2677E30

Therefore, even though the brute force parameterization method is very useful for finding the most accurate parameterization for a particular data-set at a concentration range defined by the test samples, data-sets with greater than 23 m/z signals will require vast amounts of calculation time. This prompted investigation and testing of faster parameterization methods such as the genetic algorithm (GA), empirical algorithm (EA) and stream evaluator (SE). In **Table 3.2.b** the results from application of the EA to the *n*-pentane/isopentane data are illustrated.

Table 3.2.b: EA scores for *n*-pentane / isopentane system

M/Z	EA Score	M/Z	EA Score
43	100.00	27	3.90
57	40.69	41	3.01
56	8.26	29	2.46

In **Table 3.2.b** it can be seen that the EA score falls below three at m/z = 29; thus, the final parameterization includes only the top five scored m/z signals. Note that the highest score for m/z signal = 43 is normalized to 100.

The SE also uses sensitized reference spectra to predict a good parameterization; however, it is constrained to a square matrix solution. The square-matrix parameterization is useful for some applications but in this work it proves to be a limitation with regard to the accuracy of concentration estimates.

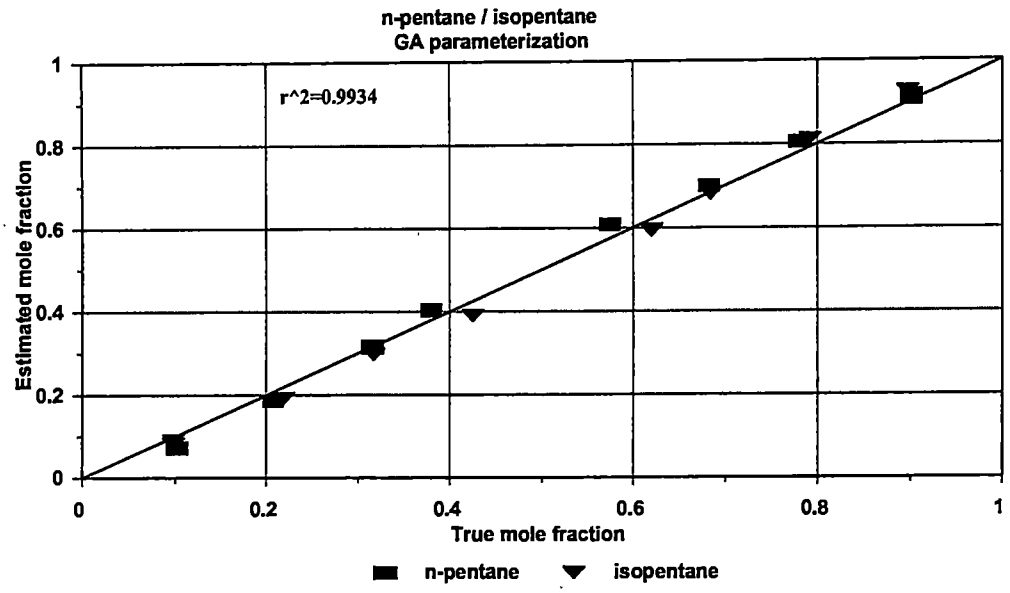
The GA, EA and SE all provide parameterizations very rapidly with respect to the calculation time required for the brute force method. The parameterizations that they afford are compared to those obtained by the brute force parameterization method (the most accurate) in the interest of finding a better (faster *and* equivalently accurate) parameterization method. This is done by constructing a confidence interval for the SSE's at the 99% level by using the SSE's from each replicate for the brute force parameterization and comparing it to the average SSE (from three replicates) obtained from the faster parameterization methods.

Table 3.2.c illustrates the results that were obtained upon applying the different parameterization methods to the pentane system. Since, by definition, no other

Table 3.2.c; Sum of square errors (SSE's), confidence intervals and correlation coefficients for *n*-pentane / isopentane system.

Parameterization / Replicate #:	**SSE (x 10-03)**	**Ave. SSE (x 10-03)**	**Std.devn. (SSE) (x 10-04)**	**Std.devn. *3 (x 10-03)**	**99% Confidence interval**
Brute Force / 1	4.751	4.284	5.438	1.631	**5.915E-03**
Brute Force / 2	3.687				
Brute Force / 3	4.414				
Brute force parameterization: 26, 30, 72, r^2=0.9963					
GA / 1	8.117	8.473	r^2=0.9934	Parameterization: 15, 26, 42, 43, 53, 55, 58, 72, 73	
GA / 2	8.306				
GA / 3	8.997				
EA / 1	8.938	8.861	r^2=0.9931	Parameterization: 27, 41,43, 56, 57	
EA / 2	8.911				
EA / 3	8.735				
SE / 1	8.511	8.517	r^2=0.9936	Parameterization: 41, 43	
SE / 2	7.911				
SE / 3	9.130				

parameterization will afford a smaller SSE than the brute force parameterization, the confidence interval limit contains only an upper limit (**5.915E-03**). This is calculated by multiplying the standard deviation of the SSE's, from each of the three replicates, by three and adding it to the average SSE to obtain the 99% confidence interval limit. If the average SSE determined from the parameterization of a faster parameterization method falls within the confidence limit defined by the brute force parameterization, the parameterization from that method is considered to be statistically equivalent to the brute force parameterization. As illustrated in **Table 3.2.c**, the average SSE's obtained from the alternative parameterization methods do not afford parameterizations that are statistically equivalent to the brute force parameterization. However, this does not indicate that the concentration estimates from each method are unusable. The validation plots in **Figures 3.2.a** and **3.2.b** illustrate that these parameterizations do afford reasonably accurate and precise concentration estimates ($r^2 = 0.9934$ for the GA, $r^2 = 0.9931$ for the EA and $r^2 = 0.9936$ for the SE). However, if the highest accuracy and precision (illustrated in the bottom validation plot in **Figure 3.2.b**) is desired utilizing the fewest number of m/z signals, then the relatively large amount of calculation time must be expended for execution of the brute force method of parameterization. However, similar to the development of an analysis method for GC, the brute force method of parameterization need only be applied once. The resulting parameterization can be utilized many times for the quick calculation of concentration estimates of a chemical system. An important point to be noted upon the comparison of the m/z signals among the different parameterization methods is the low amount of overlap or recurrence of the ame m/z signals in different good parameterizations. Note that there is no overlap



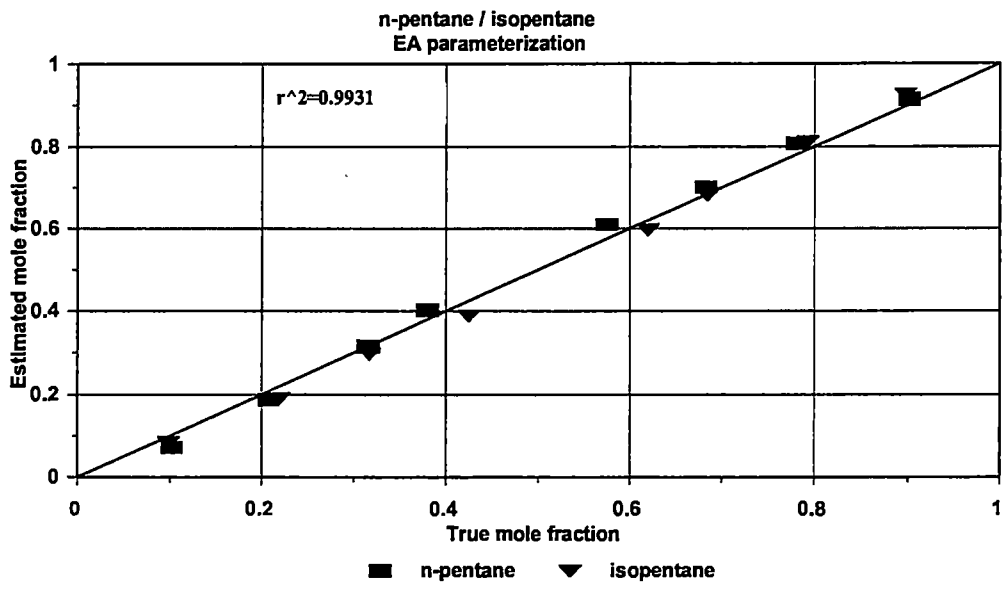
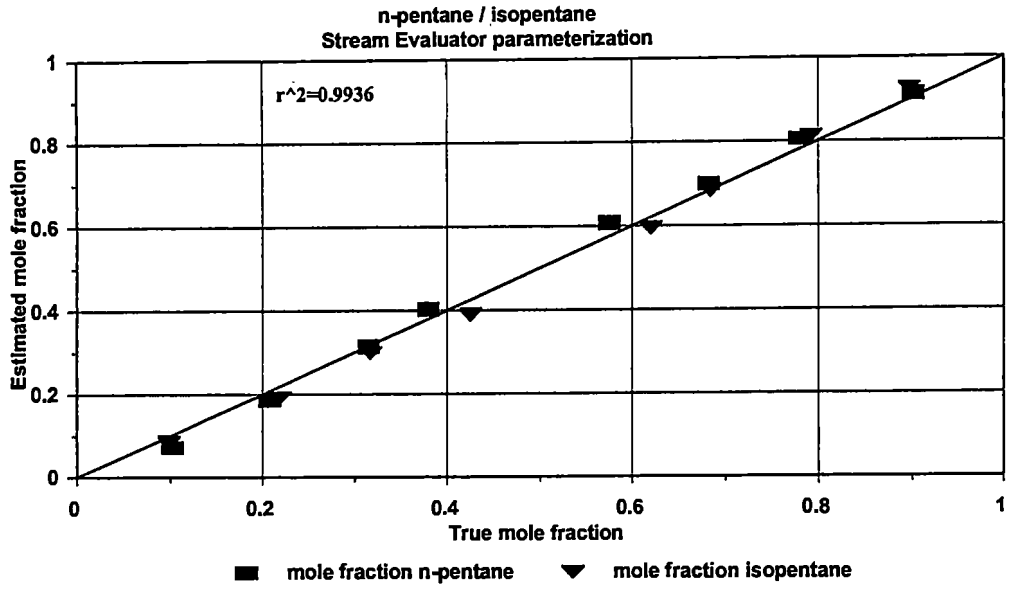


Figure 3.2.a: Validation plot for the GA and EA parameterization methods; m/z's: 15, 26, 42, 43, 53, 55, 58, 72, 73 (GA) m/z's: 27, 41, 43, 56, 57 (EA). Eight samples of two components with three replicates are plotted (48 points).



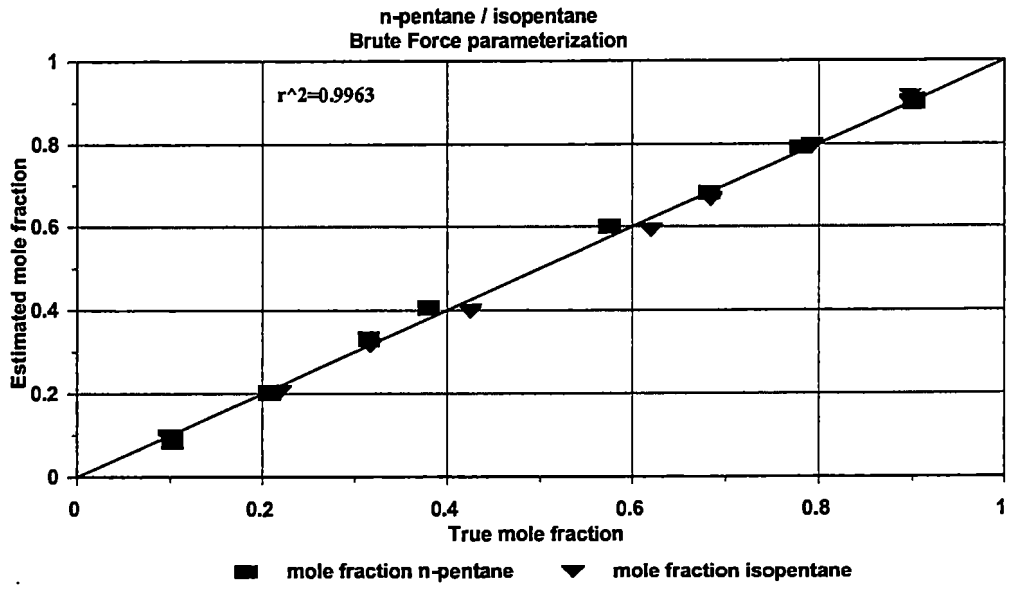


Figure 3.2.b: Validation plot for the SE parameterization method; m/z's: 41, 43 and brute force parameterization method m/z's: 26,30,72. Eight samples of two components with three replicates are plotted (48 points). The lower validation plot is copied from Figure 3.1.f.

between the three m/z signals of the brute force and the EA and SE parameterizations. Intuitively, one would expect that the brute force parameterization method would contain m/z signals that would be found to be common among good parameterizations. This apparent absence of overlap occurs for the other chemical systems tested in this work. For this reason, it is suspected that the brute force method of parameterization is overfitting the data as mentioned earlier. One way by which to address this problem of overfitting the data would be to increase the number of different samples. This problem has been addressed in our laboratory by preparing and testing a large number of samples of chemical systems that are gases at room temperature and pressure. These mixtures were prepared using a system of mass flow controllers (MFC's, gas metering devices). These MFC's allow the preparation of gas mixtures much more quickly than preparation of liquid samples. Extensive analysis of the data from these mixtures is still taking place in our laboratory.

The difference in performance (correlation coefficient) between the four methods of parameterization is not large for the *n*-pentane / isopentane system as illustrated above. However, larger differences in performance occur for the stereoisomer systems. This is illustrated in **Tables 3.2.d** and **3.2.e**. In **Table 3.2.d** the upper limit of the confidence interval for the cis- / trans-piperylene system is over 13 times larger than that for the *n*-pentane / isopentane system (**7.835E-02** vs. **5.915E-03**) even though the cis- / trans- system contains 4 fewer samples. If accuracy and precision are otherwise the same, the SSE should be smaller with fewer samples. Since this is the best performance that can be accomplished for this system (the brute force parameterization was used to calculate the SSE's) it is evident that there is a much larger amount of error in the system than that in

Table 3.2.d: Sum of square errors (SSE's), confidence intervals and correlation coefficients for cis- / trans-piperylene system.

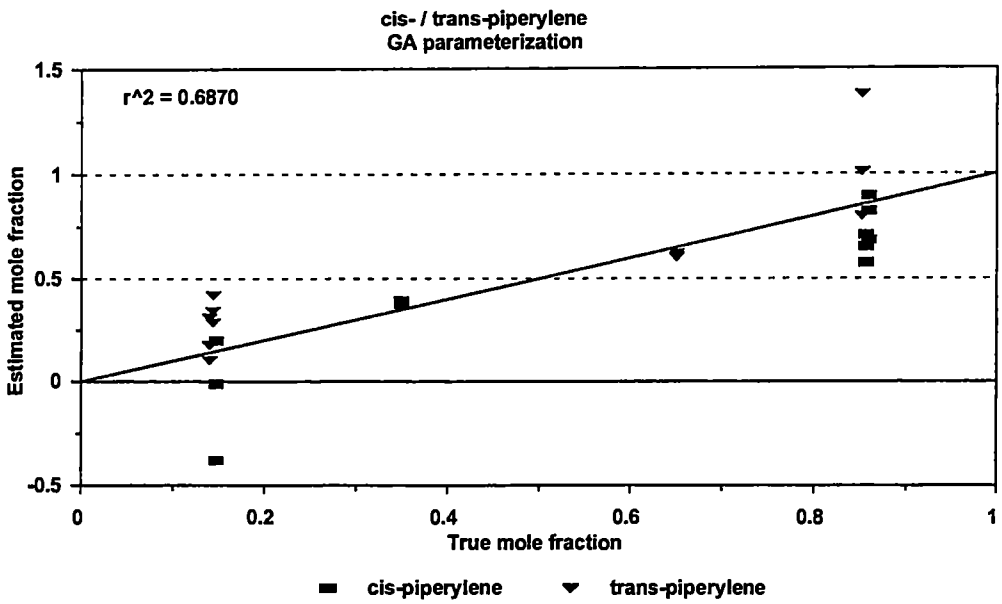
Parameterization / Replicate #:	**SSE (x 10-02)**	**Ave. SSE (x 10-02)**	**Std.devn. (SSE) (x 10-03)**	**Std.devn. *3 (x 10-02)**	**99% Confidence interval**
Brute Force / 1	6.210	5.259	8.586	2.576	**7.835E-02**
Brute Force / 2	5.028				
Brute Force / 3	4.540				
Brute force parameterization: 38,61,62,63; r^2=0.9054					
GA / 1	26.87	31.93	r^2= 0.6870	Parameterization: 31, 40, 43, 50, 54, 56, 62, 63, 65, 68, 71	
GA / 2	5.178				
GA / 3	63.73				
EA / 1	7.266	7.634	r^2= 0.8620	Parameterization: 27, 39, 40, 41, 53, 67, 68	
EA / 2	7.951				
EA / 3	7.686				
SE / 1	11680	4018	r^2=0.0148	Parameterization: 39, 67	
SE / 2	272.0				
SE / 3	97.69				

Table 3.2.e: Sum of square errors (SSE's), confidence intervals and correlation coefficients for cis- / trans-2-pentene system.

Parameterization / Replicate #:	**SSE (x 10-04)**	**Ave. SSE (x 10-04)**	**Std.devn. (SSE) (x 10-04)**	**Std.devn. *3 (x 10-04)**	**99% Confidence interval**
Brute Force / 1	5.915	8.671	2.904	8.713	1.738E-03
Brute Force / 2	11.70				
Brute Force / 3	8.395				
Brute force parameterization: 31,41,43,50,51,69; r^2=0.9986					
GA / 1	615.7	535.7	r^2=0.8634	Parameterization: 31, 39, 40, 51, 53, 56, 67, 68, 70, 71	
GA / 2	925.0				
GA / 3	66.22				
EA / 1	1031	1088	r^2=0.6350	Parameterization: 27, 29, 39, 41, 42, 55, 70	
EA / 2	1645				
EA / 3	587.6				
SE / 1	3756	2243	r^2=0.8296	Parameterization: 42, 55	
SE / 2	938.9				
SE / 3	2035				

the *n*-pentane / isopentane system. This is attributed to both the smaller differences between the m/z signal intensities (increased spectral similarity) and the increased reactivity (less energy required to lose an electron from the pi-bond) of the diene system (piperylenes) as compared to the alkane system (*n*-pentane / isopentane). Also evident from the table is the fact that the parameterization obtained from the EA method is statistically equivalent to the brute force parameterization (*i.e.*, 7.634E-02 < 7.835E-02). In this case, the EA parameterization provides concentration estimates that appear to be low in accuracy and precision ($r^2 = 0.8620$) but, since the average SSE is equivalent via the brute force confidence interval, the EA is considered to be successful in selecting a parameterization that is statistically equivalent to the brute force parameterization when applied to this dataset.

The GA and SE provide parameterizations which afford concentration estimates of poor performance as illustrated by their correlation coefficients. The performance of the GA and EA parameterizations are illustrated in the validation plots of **Figure 3.2.c**; the SE parameterization performance is illustrated in **Figure 3.2.d**. In this case, the EA is the only rapid method of parameterization that affords non-negative concentration estimates. The concentration estimates from the EA parameterization are also the most accurate and precise of the faster parameterization methods presented for the cis- / trans-piperylene system ($r^2 = 0.6870$ for the GA, $r^2 = 0.8620$ for the EA method and $r^2 = 0.0148$ for the SE method). However, the highest accuracy concentration estimates are obtained from the brute force parameterization illustrated in the validation plot at the bottom of **Figure 3.2.d**. Here, the $r^2 = 0.9054$ indicates that it is possible to measure binary mixtures of these components with reasonable accuracy.



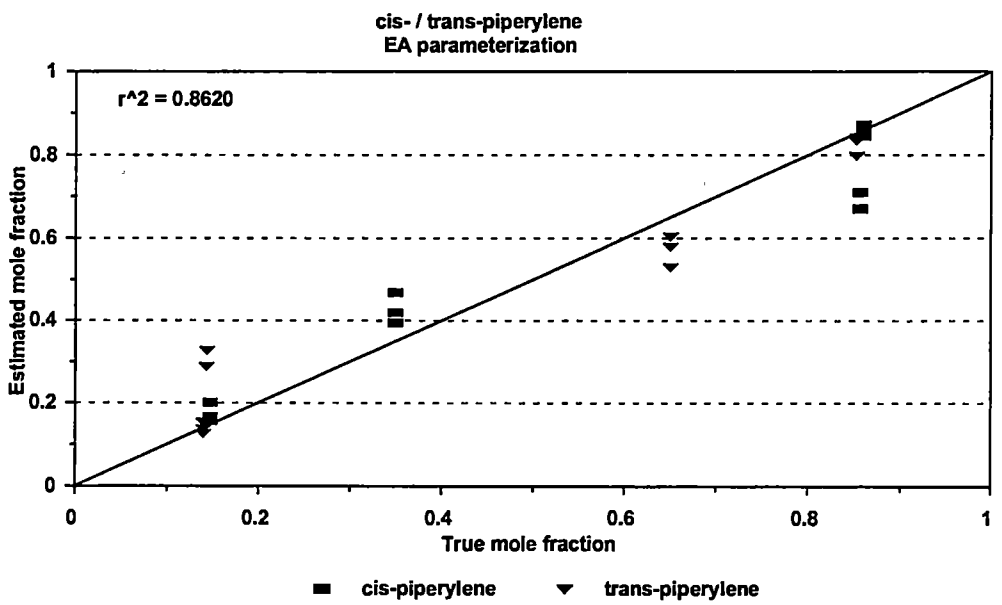
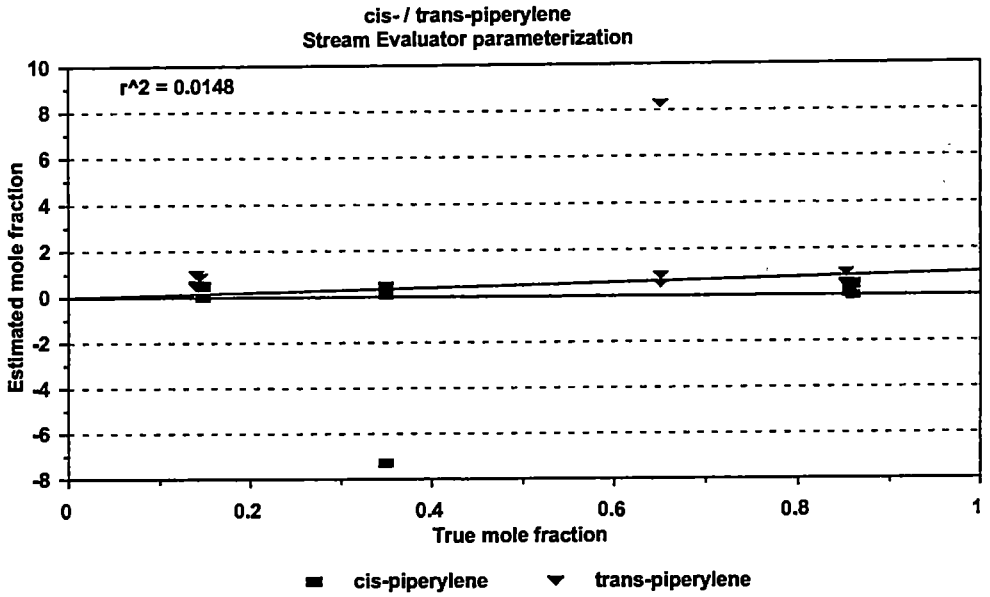


Figure 3.2.c: Validation plot for GA parameterization m/z's: 31, 40, 43, 50, 54, 56, 62, 63, 65, 68, 71 and EA parameterization m/z's: 27, 39, 40, 41, 53, 67, 68 for cis- / trans-piperylene system. Four samples of two components with three replicates are plotted (24 points).



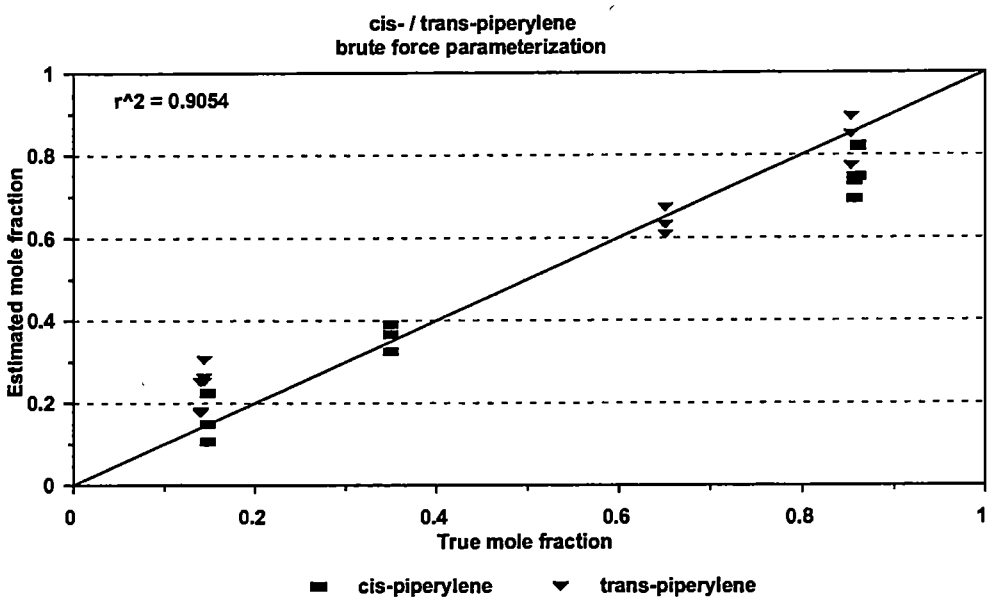
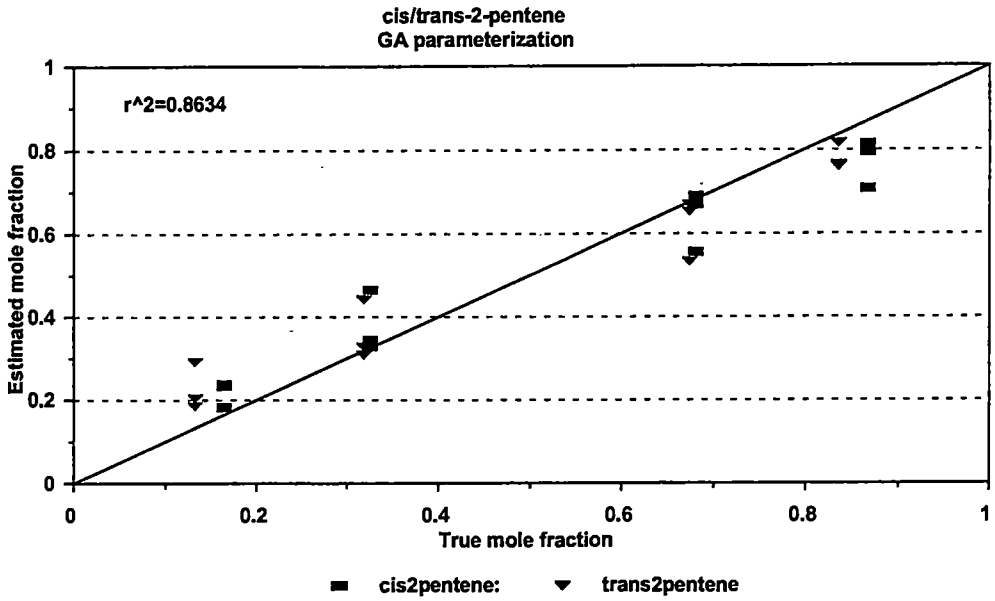


Figure 3.2.d: Validation plot for the SE parameterization method; m/z's: 39, 67 and brute force method; m/z's: 38,61,62,63. Four samples of two components with three replicates are plotted (24 points). The lower validation plot is copied from Figure 3.1.g.

Finally, in turning to the remaining stereoisomer system, the SSE results are reported in **Table 3.2.e** for the cis- / trans-2-pentene dataset. In **Table 3.2.e** it can be seen that the average SSE for the GA and SE parameterization methods are smaller than that observed for the piperylene system (**Table 3.2.d**). Also, the correlation coefficient for the brute force parameterization of the pentenes system is higher than for the pentanes system. This may be due to the smaller number of samples in this dataset as compared to that in the pentanes dataset (fewer samples implies lower SSE if accuracy and precision are comparable).

The EA parameterization method has not selected as accurate of a parameterization compared to the piperylene system. Also, none of the auxiliary methods has selected a parameterization that affords concentration estimates that are statistically equivalent to the brute force parameterization method. This is due to the smaller SSE's for the cis- / trans-2-pentene system as compared to the piperylene system. The performance of these auxiliary methods is evident in the validation plots of **Figures 3.2.e** and **3.2.f**. Among the auxiliary parameterization methods, the GA and EA afford concentration estimates that have low correlation coefficients ($r^2 = 0.8634$ for the GA, $r^2 = 0.6350$ for the EA and $r^2 = 0.8296$ for the SE). Due to its limited number of m/z signals in the parameterization the SE method, once again, affords negative concentration estimates. If the SE were able to select a larger number of m/z signals the resulting parameterization may be able to better predict the concentrations at the extreme ends of the validation plot in **Figure 3.2.f**.



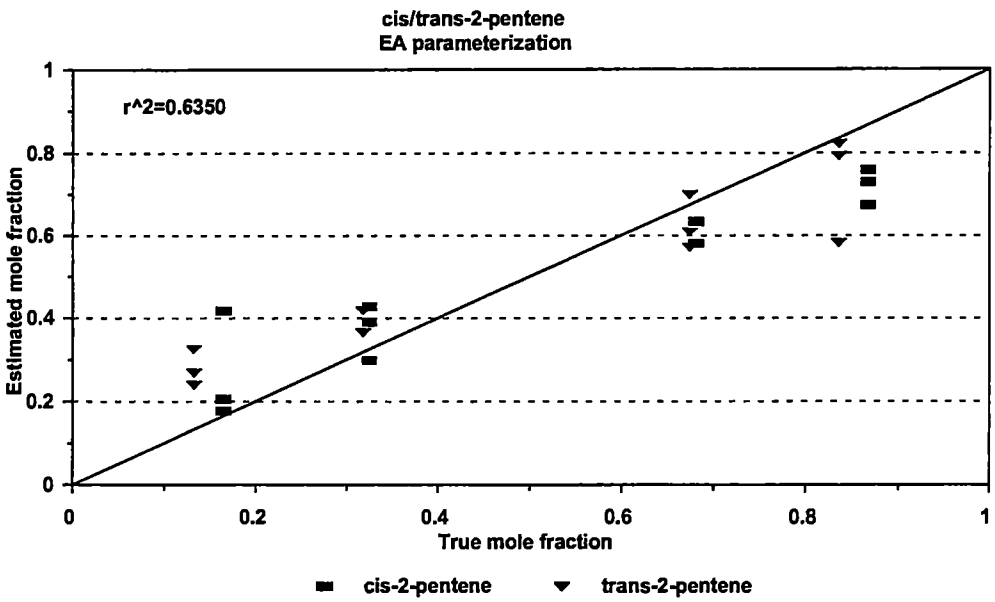


Figure 3.2.e: Validation plot for GA parameterization: 31, 39, 40, 51, 53, 56, 67, 68, 70, 71 and EA parameterization: 27, 29, 39, 41, 42, 55, 70 for cis- / trans-2-pentene system. Four samples of two components with three replicates are plotted (24 points).

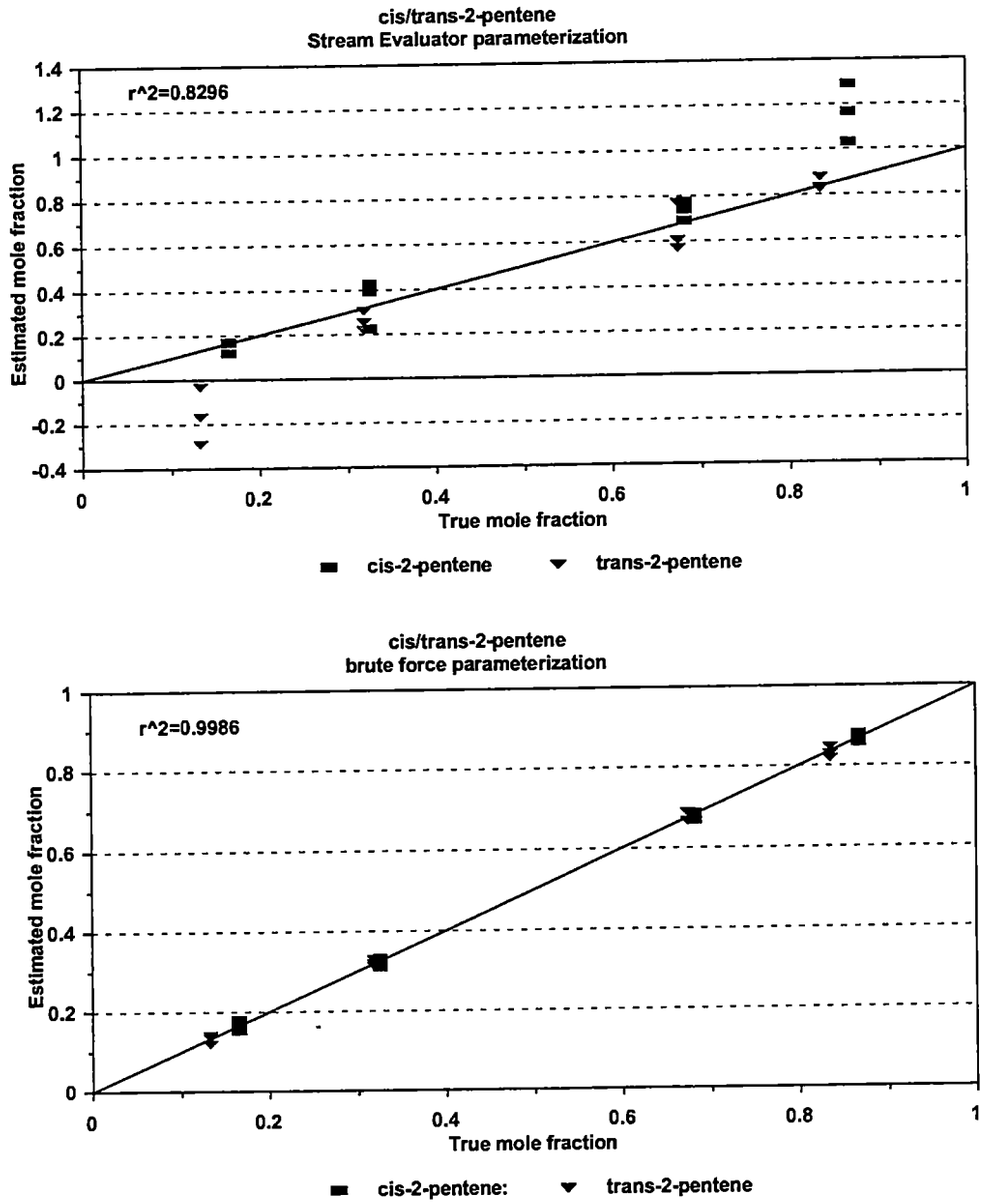


Figure 3.2.f: Validation plot for the SE parameterization; m/z's: 42, 55 and brute force parameterization; m/z's: 31,41,43,50,51,69. Four samples of two components with three replicates are plotted (24 points). The validation plot at the bottom of the figure is a copy of Figure 3.1.j.

3.3: SUMMARY AND CONCLUSIONS

Although it does not make efficient use of the instrument's scanning time, the parameterization that results from an intuitive approach (use of all m/z signals above a given intensity threshold) can afford reasonably accurate and precise concentration estimates for some isomeric hydrocarbon systems (*n*-pentane / isopentane in this case). A brute force analysis of the data requires a large amount of calculation time but is required for the highest accuracy of concentration estimates. This amount of time is comparable to the time that might be required to develop a gas chromatography method. So, although it may require a significant amount of time to calculate the parameterization, the *deconvolution* calculations are rapid (< 2 seconds on a 200 MHz computer). Thus, an exhaustive evaluation (brute force parameterization method based upon the accuracy) of the data will afford a parameterization that will calculate the most accurate concentration estimates for a particular dataset.

It is also possible to target the best *precision* in the brute force method as opposed to the accuracy with the interest of finding a parameterization that affords the most precise concentration results over a specified concentration range. The evaluation of a brute force method of parameterization based upon precision is recommended for future work.

For the *n*-pentane / isopentane system, the faster parameterization methods select parameterizations that afford reasonably accurate and precise concentration estimates but are not statistically equivalent to the brute force parameterization. Upon applying the EA to the piperylene system, it was illustrated that the resulting parameterization was indeed

statistically equivalent to the brute force parameterization. However, the success of this method with the piperylene system is likely due to the larger SSE thus affording a larger confidence interval. The SE method of parameterization is inherently limited in its ability to select an appropriate parameterization. In other words, the number of m/z signals in the parameterization is limited the number of components which affords a square parameterization. In observing the number of m/z signals that exist in the brute force parameterization of all the chemical systems tested, a square-parameterization is never selected. The SE tends to select good m/z signals (with relatively large intensities and / or intensity differences) but not enough of them. For the data sets tested in this work, the SE did not choose a parameterization that was equivalent to the brute force parameterization.

The GA is a powerful approach to variable selection that strategically tests many different parameterizations efficiently. In this work, it is a parameterization method that utilizes partial least squares as a quantitation (deconvolution) approach. Hence, the GA utilizes the mixture spectra and the known mole fractions of those spectra (the reference spectra are not required). Thus, the comparison of parameterizations between the GA and the other parameterization methods that utilize the reference spectra *and* mixture spectra may not be equivalent. One approach that may allow a more equal footing by which to compare the GA and the other parameterization methods is the development and application of a brute force parameterization method that is based on partial least squares deconvolution. Thus, testing every combination of m/z signals among the *mixture* spectra using partial least squares would be a useful topic for future study in this work.

Upon consideration of the r^2 values that are derived from the three different

chemical systems (alkane vs. alkene or stereoisomers) it is apparent that, independent of the parameterization, a higher r^2 is generally obtained from the pentane and pentene mixtures. This effect is attributed to both the higher reactivity of the piperylene system, compared to the alkane system, and the smaller differences in intensity that are apparent in the reference spectra. The pentene system served as a stereoisomer system with a lower reactivity as compared to the piperylene system. There were no indications of reactivity problems in the sampling inlet as compared to the piperylene system.

This effect of reactivity was addressed upon selection of a sampling method. A continuous infusion method, used for sampling the relatively non-reactive *n*-pentane / isopentane system, could not be used for the piperylene system due to the formation of a tacky material that was believed to be a product of piperylene thermal dimerization. Upon decreasing the amount of material analyzed (15 μL injected into a heated reservoir compared to ~30 μL / min continuous infusion), the problem was diminished to the point at which the analysis could be performed. This reactivity difference was suspected to affect the spectral similarity or differences in m/z signal intensity by degrading the precision of m/z signal intensities. With increased spectral similarity (decreased differences between the m/z signal intensities), it was suspected that a decrease in accuracy would be incurred since there is less differentiating information available in the mass spectrum for deconvolution. This issue of spectral similarity is addressed in the following chapter.

CHAPTER 4

4.1: EFFECTS OF SPECTRAL SIMILARITY ON ACCURACY & PRECISION OF CONCENTRATION ESTIMATES

In some process analytical applications, it would be useful to know beforehand the likelihood of success in deconvolving complex mixtures (such as those tested in this work) using process mass spectrometry. One method by which this prediction can be approached is utilization of the Drahos-Vekey similarity index (SI) detailed in **Equation 2.5.5.a**. Other SI equations have been presented in the literature [137][138] and may also be suitable. The Drahos-Vekey equation was chosen for this work based upon its success upon application toward matching tandem mass spectra of isomers and its relatively recent introduction in the literature (1996). This equation affords a general index that can be applied a pair of mass spectra to determine qualitatively the degree of similarity between the mass spectra. Thus, it is anticipated that the possibility of accurate deconvolution of an un-tested mixture of isomeric components can be estimated by comparing the SI of their pure component mass spectra to other systems that have been proven to be accurately deconvolved using PrMS.

Upon applying the SI to "sensitized" and "non-sensitized" (with and without the application of the sensitivity factor to the normalized reference spectra to account for differences in ionization cross section, ionization energy, etc.) reference spectra from the binary systems, the results in **Table 4.1.a** were obtained.

Upon consideration of the SI equation it can be predicted that, for spectra which

contain relatively *small* differences between their intensities, the calculated SI will be relatively *large*. Conversely, for spectral pairs that contain relatively *large* differences (*e.g.*, components "A" and "B") between their intensities, the calculated SI is relatively *small* listed in **Table 4.1.a**. This is demonstrated for the theoretical case of components "A" and "B" (two very *dis*similar components) described in Chapter 1.7. In this case these components give an SI of approximately 4 which is relatively small. Thus, the SI score for an identical pair of spectra would be 100 and for a pair of spectra with no overlap it would be zero. Thus, the SI should not exceed 100 or take on values less than zero. But, upon calculating error bars for the SI by evaluating the standard deviation of the SI and multiplying by three, the error bars may afford an SI greater than 100 or less than zero. Upon application of the sensitivity factor to the reference spectra, the SI becomes smaller or the spectra become more dissimilar. This is because of the additional information from the base peak. The intensity difference from the base peak is zero (since the base peak is normalized to 100) in the non-sensitized spectra and thus it does not contribute differentiating information to the calculated SI.

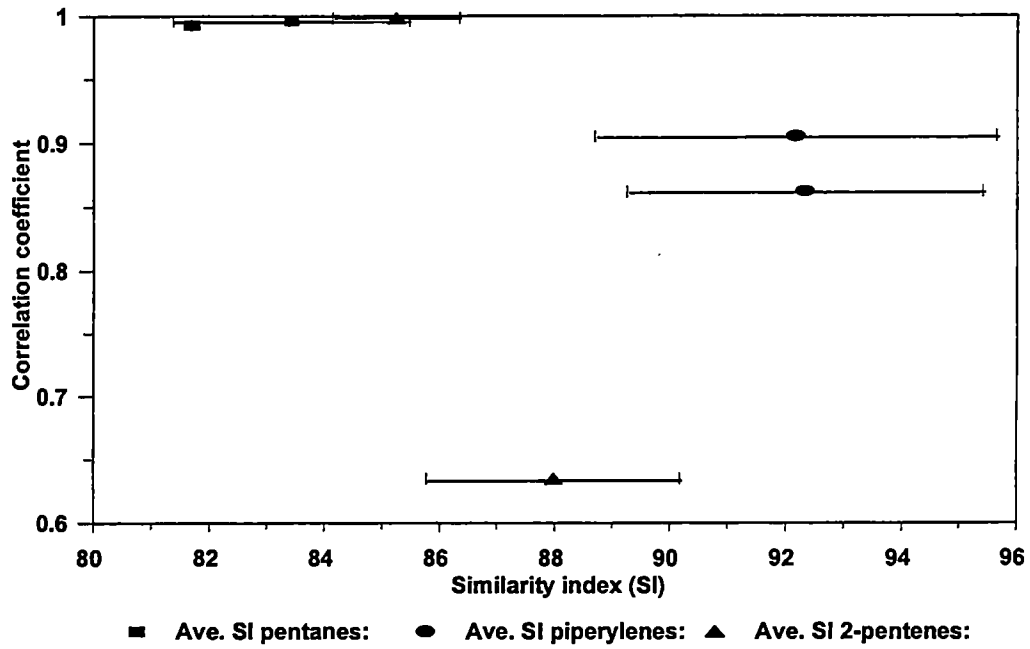
In using normalized reference spectra, the sensitivity factor is used in the deconvolution calculations to regain the information lost in the base peak; thus, there are no entries in **Table 4.1.a** for a non-sensitized correlation coefficient.

It was anticipated that, as spectral similarity (SI) increased for a particular system, the performance, as measured by the correlation coefficient, would decrease in a linear fashion. The anticipated linear case was not found in the plots of **Figure 4.1.a**. In fact, at r^2 values less than 0.9 it is apparent that the SI is independent of r^2. At r^2 values greater than approximately 0.9 there appears to be a threshold where the SI decreases quickly. In

Table 4.1.a SI for *n*-pentane / isopentane, cis- / trans-piperylene, cis- / trans-2-pentene systems.

Reference spectra / Parameterization:	**Average SI[1]**	**Standard deviation * 3**	**r^2**
pentanes (NS) / Brute force	86.68	0.4129	-
pentanes (S) / Brute force	83.43	2.042	0.9963
piperylenes (NS) / Brute force	98.21	1.994	-
piperylenes (S) / Brute force	92.17	3.476	0.9054
2-pentenes (NS) / Brute force	87.46	0.3056	-
2-pentenes (S) / Brute force	85.25	1.099	0.9986
pentanes (NS) / EA	81.46	0.1497	-
pentanes (S) / EA	81.69	0.007530	0.9931
piperylenes (NS) / EA	98.21	1.994	-
piperylenes (S) / EA	92.34	3.071	0.8620
2-pentenes (NS) / EA	98.35	0.4742	-
2-pentenes (S) / EA	87.99	2.205	0.6350
Components "A" and "B"[2]	4.1	-	-

(S): Sensitized spectra. (NS): Non-sensitized spectra. [1]: Calculated from three replicates.
[2]: Theoretical components described in Chapter 1.



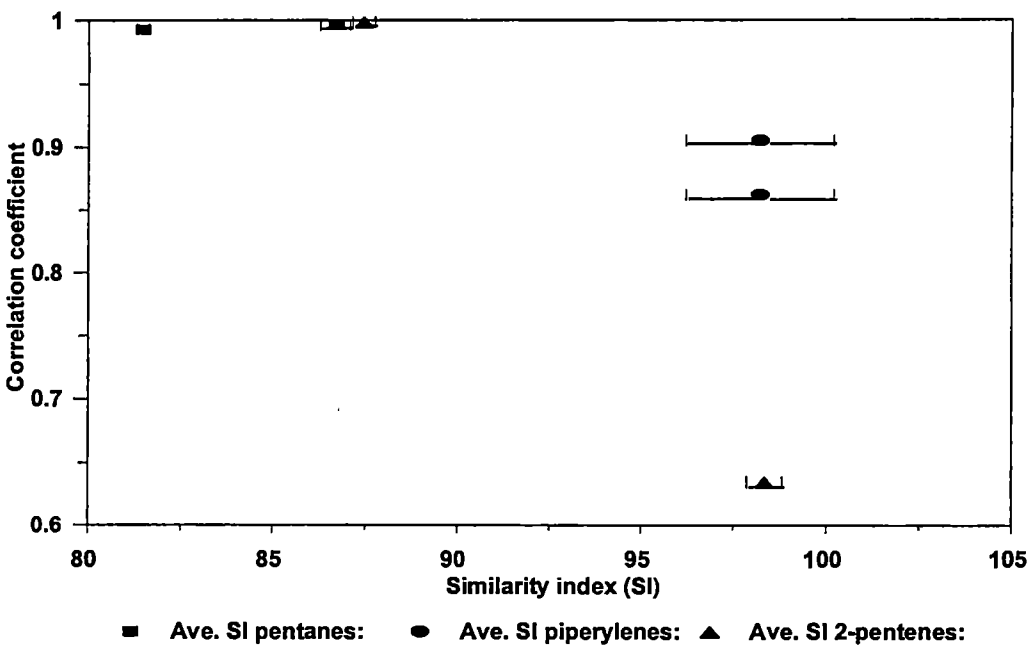


Figure 4.1.a: Correlation coefficient vs. similarity index calculated from brute force and empirical algorithm parameterizations (sensitized: upper, non-sensitized: lower).

the top portion of the figure, the SI from sensitized reference spectra (brute force and empirical algorithm) is plotted. Also, the performance of the three systems listed in **Table 4.1.a** via the correlation coefficient, as calculated by the brute force and empirical algorithm parameterization methods, are plotted. In the lower plot of **Figure 4.1.a** non-sensitized reference spectra are utilized to calculate the SI. The same correlation coefficient is used in both plots.

Parallel results are obtained for the SI from the sensitized and non-sensitized reference spectra. The error bars for the piperylene system (± three standard deviations added to the average SI) are smaller for the non-sensitized than for the sensitized SI. This is due to the convolution of error that occurs (between the reference *and* the calibration spectrum) upon sensitizing the reference spectra. This additional error is derived from treating the data in three separate replicates. Hence, there are three slightly different sensitivity factors that are used with the three slightly different reference spectra. The SI results obtained from these systems indicate that at an SI in the high nineties (sensitized or non-sensitized) mixtures of these components can be accurately resolved using process mass spectrometry. Use of the sensitivity factor is an advisable approach based upon the fact that the use of the base peak is made after sensitization. However, use of the sensitivity factor increases the standard deviation in the SI. The positive aspect of this is that use of either sensitized or non-sensitized reference spectra give about the same SI. Thus, since it is the case that mass spectral libraries do not usually contain sensitized reference spectra, the non-sensitized SI is useful.

Binary mixtures of cis-piperylene / cyclopentene (C_5H_8 isomers), cis-2-pentene / cyclopentane and cis-2-pentene / 2-methyl-2-butene (C_5H_{10} isomers) were not prepared;

thus, their sensitized reference spectra or correlation coefficients are not evaluated here, but their *non*-sensitized reference spectra can be compared using the SI. This is done in **Table 4.1.b**. These components were selected from **Table 3.1.a** because they are isomeric hydrocarbon components in the piperylene stream that remain to be tested as binary mixtures by PrMS and thus are done so here using the SI.

Table 4.1.b: SI for cis-piperylene / cyclopentene, cis-2-pentene / 2-methyl-2-butene, cis-2-pentene / cyclopentane systems.

System # / Reference spectra :	SI	Ave. SI	Std.devn.
1: cis-piperylene / cyclopentene (NS)	71.82	72.04	2.835
2: cis-2-pentene /2-methyl-2-butene (NS)	92.16	91.96	0.7251
3: cis-2-pentene / cyclopentane (NS)	65.39	66.77	1.200

(NS): Non-sensitized spectra

From **Table 4.1.b** it can be seen that systems 1 and 3 afford an SI that is relatively small, implying that mixtures of these components can likely be deconvolved with reasonable accuracy and precision. System 2 demonstrates an SI in the nineties and would likely prove to be among the most difficult systems listed. It is recommended that, for future work, binary mixtures of the components listed here be prepared and tested in an analogous manner by which the systems in Chapter 3 were examined.

4.2: SUMMARY AND CONCLUSIONS

The SI is a measure of the differences between the m/z signal intensities in mass spectra. Although the idea of utilizing *sensitized* mass spectra in the SI calculations is

inviting with regard to re-gaining the information lost in the base peak due to normalization, utilizing the sensitized reference spectra for the SI calculations increases it's standard deviation. It is suspected that the SI results from the cis- / trans- isomers (stereoisomers) are larger in comparison to those of the structural isomers for the *n*-pentane / isopentane system due to the contrast in energy required to fragment the different molecular structures. Although there is an excess of energy (100 eV) present during ionization it is presumed that the relatively smaller differences in m/z signal intensity observed in the stereoisomers is due to the relatively lower energy required to afford the same fragments (rotation of carbon-carbon bonds as compared to the breaking of bonds in the case of positional isomers). The relatively higher reactivity of the piperylene system is suspected to be the cause of compromised reproducibility due to reactions at sampling surfaces. This suspicion was derived from the observation of tacky material deposited, after an extended amount of sampling time, upon the surfaces of the vaporization inlet.

The EA, like the SI, is designed to relate spectral intensity differences to deconvolution performance (correlation coefficient). It is shown here that the spectral intensity differences, as measured by the SI, are correlated to the correlation coefficient. Thus, since the EA is also based upon the spectral intensity differences, it is a reasonable approach to quickly identifying an appropriate parameterization. The brute force parameterization was selected for the SI comparisons because it represents the highest in accuracy that can be demonstrated for a dataset. The EA parameterization was chosen in this comparison because it is a simplified measure of spectral intensity differences with respect to the SI. Since the EA was able to select a statistically equivalent

parameterization for only the piperylene system in the previous chapter, it may be necessary to increase the complexity of the EA parallel to that of the SI. Specifically, this may involve incorporation of a denominator much like that used in the SI. In addition, the number of components analyzed requires augmentation. In other words, the SI equation requires extension to more than a pair of components. In the case of three components, this may be approached by taking all possible pairs from three components and summing their EA score for each m/z signal.

Currently, an EA based upon a more sophisticated approach, as compared to the spectral intensity difference criterion, is being evaluated in our laboratory [139]. This new approach involves analyzing the m/z signals in the spectra as vectors. This method adds the criterion of m/z signal intensity as well as intensity differences. At this point, the new EA has demonstrated promising performance for several isomeric systems.

CHAPTER 5

5.1: ANALYSIS OF 9-COMPONENT HYDROCARBON ISOMER MIXTURES BY GC-MS

Upon demonstration of the success in deconvolving binary isomeric hydrocarbon mixtures in the previous chapters, mixture complexity was increased to nine components. The overall interest here was in analyzing a grab sample taken from a piperylenes stream. This grab sample was determined via GC (at the MCEC collaborator's instrumental analysis laboratory) to contain components at the levels previously indicated in **Table 3.1.a.**

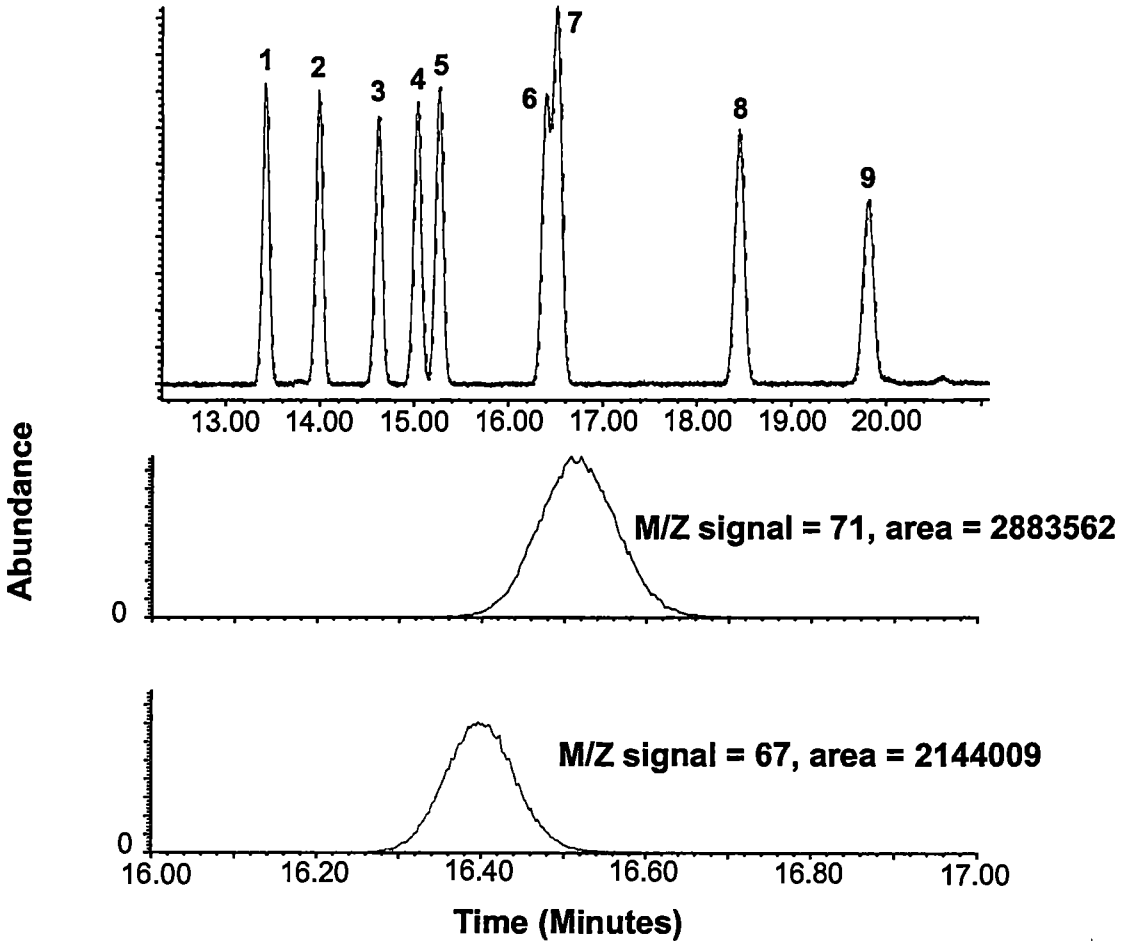
The first step in analyzing the piperylene grab sample involved preparation and analysis of two samples that served as standards in the PrMS as well as the GC-MS analysis (a GC-MS analysis was performed in our laboratory to confirm results obtained from the MCEC collaborator). Thus before testing the piperylene grab sample, the prepared standards were analyzed to confirm two points. The first point was the ability to repeat the GC-MS results at our laboratory. The second point was to confirm the ability to use PrMS to accurately deconvolve mixtures as complex as nine components containing hydrocarbon isomers.

Sample G1, the component level of which is listed in **Table 2.2.b**, was used as a sample that contained the nine major components in the piperylene grab sample but the unknown components (present in the grab sample) were not present. The isopentane component was left out of the standard samples since it was at a relatively low

concentration level as indicated in **Table 3.1.a**. This standard sample (G1) was utilized to determine if large discrepancies in the estimated component concentrations occurred between the piperylene grab sample and the standard sample (G1) due to the presence of the low-level unknowns in the piperylene grab sample. Note the unknowns listed at the bottom of **Table 3.1.a**. In other words, if the differences between estimated component concentrations of sample G1 and the piperylene grab sample were significant, they may be due to the presence of unknown components in the grab sample.

Standard sample G2, an equimolar mixture of the nine major components listed in **Table 3.1.a**, was used to calibrate the peak areas from the GC-MS data and as a calibration sample to calculate the component sensitivities in the PrMS. The MCEC collaborator's GC analysis procedure utilized a subambient temperature program to resolve the hydrocarbon mixture on a 50 meter column. Since our laboratory does not presently have cryogenic-GC capabilities, component resolution was augmented by increasing the column length to 80 meters and utilizing an ambient temperature in the GC oven.

Figure 5.1.a illustrates the gas chromatogram of the nine-component equimolar standard (G2). The identification of each component is tabulated in data below the figure. As illustrated in the figure, 2,2-dimethylbutane and cis-piperylene were not completely resolved by GC. The areas for these two components in the chromatogram were estimated by utilizing selected ions unique to each component as well as the gravimetric data obtained upon preparation of the sample. Specifically, this involved using m/z = 71 for the 2,2-dimethylbutane component and m/z = 67 for cis-piperylene. Sensitivities for each component were determined by calculating the ratio of the mole



Component #:	Component:	Peak area:	Sensitivity (moles / area) (x 10^{-3})
1	n-pentane	8394585	1.26
2	cis-2-pentene	8321594	1.24
3	trans-2-pentene	8231503	1.23
4	2-methyl-2-butene	8919773	1.18
5	trans-piperylene	8949858	1.24
6	cis-piperylene	2144009	3.63
7	2,2-dimethylbutane	2883562	5.39
8	cyclopentene	9629587	1.12
9	cyclopentane	7389203	1.46

Figure 5.1.a: Gas chromatogram (TIC) and associated peak areas of equimolar nine-component mixture (G2).

fractions of each component to the normalized peak areas. This gave the nine sensitivity factors listed in **Figure 5.1.a** that, when multiplied by the normalized peak areas for the G1 standard, afforded estimated moles for each component. Normalization of the moles in G1 then afforded the mole fraction values listed in **Table 5.1.a**. The same treatment (selected ions 67 and 71 for cis-piperylene and 2,2-dimethylbutane) was applied to the estimation of mole fractions for the piperylene grab sample. Results for the piperylene grab sample are listed in **Table 5.1.a**.

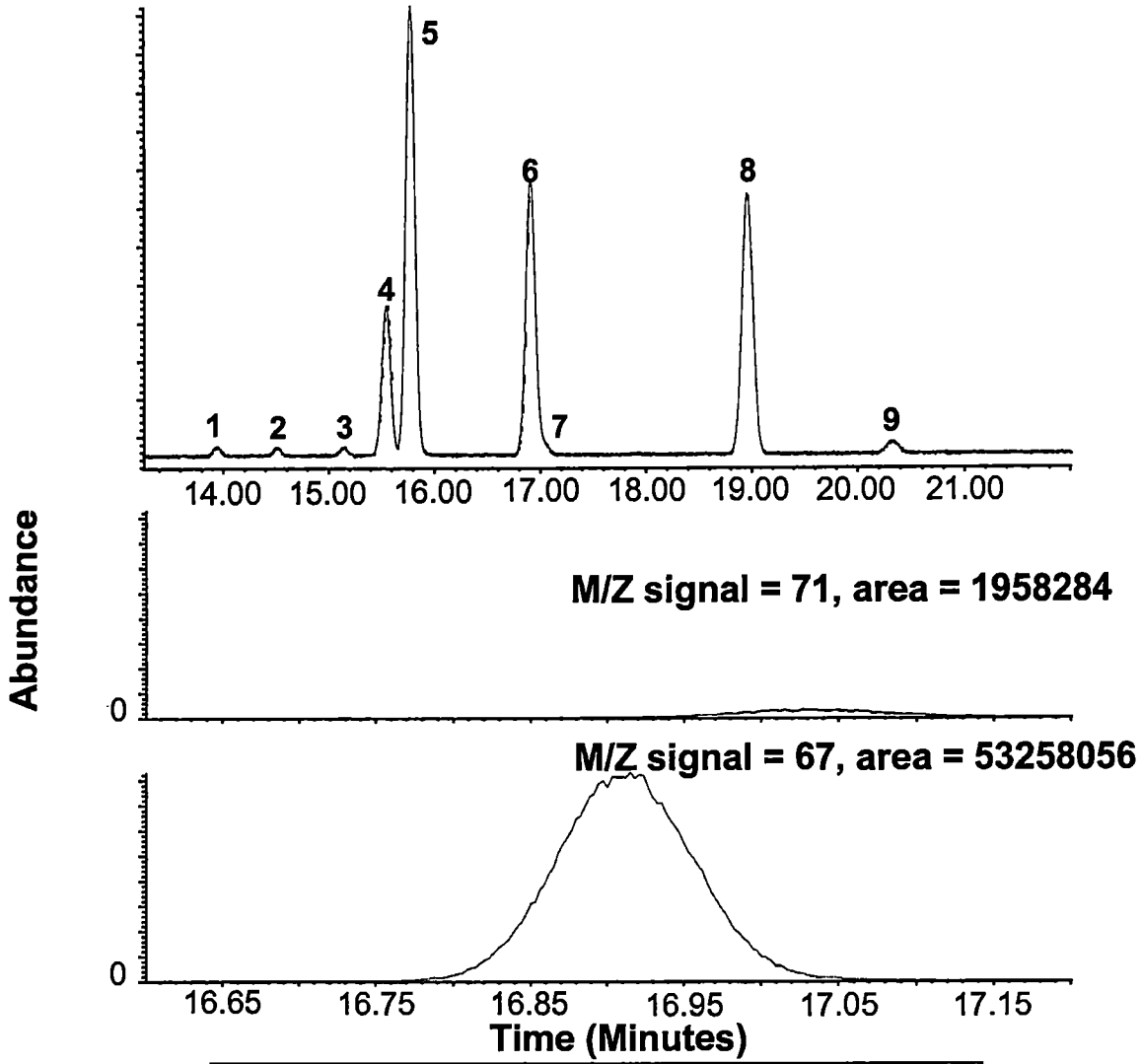
Figure 5.1.b illustrates the gas chromatogram for the standard sample G1 and the related peak areas. As was the case for the G2 standard, the chromatogram is not completely resolved for cis-piperylene and 2,2-dimethylbutane. The ability to sufficiently achieve the GC analysis of nine component isomeric hydrocarbon mixtures without the use of a cryogenic GC oven temperature program is illustrated in **Figure 5.1.c**. In this figure the GC-MS measured component mole fractions from the piperylene standard (G1) are plotted against the values that were determined gravimetrically upon preparation of the sample. The lower half of the figure provides a zoom-in on the mole fraction values near 0.005. The correlation coefficient value for this plot is 0.9990 which indicates good correlation to the Y = X line. In viewing the lower validation plot of **Figure 5.1.c** it is apparent that all five of the components are overestimated (greater than the 45 degree line) which may indicate a systematic error. This analysis requires further study (replicate measurements) to allow the analysis of these data points using error bars. It is possible that these components may be present at a level that is below the limit of detection for the GC-MS system.

Upon reasonable success in the GC-MS analysis of the standards (G1 and G2)

Table 5.1.a: Component compositions from GC-MS data for sample G1, G2 and piperylene grab sample.

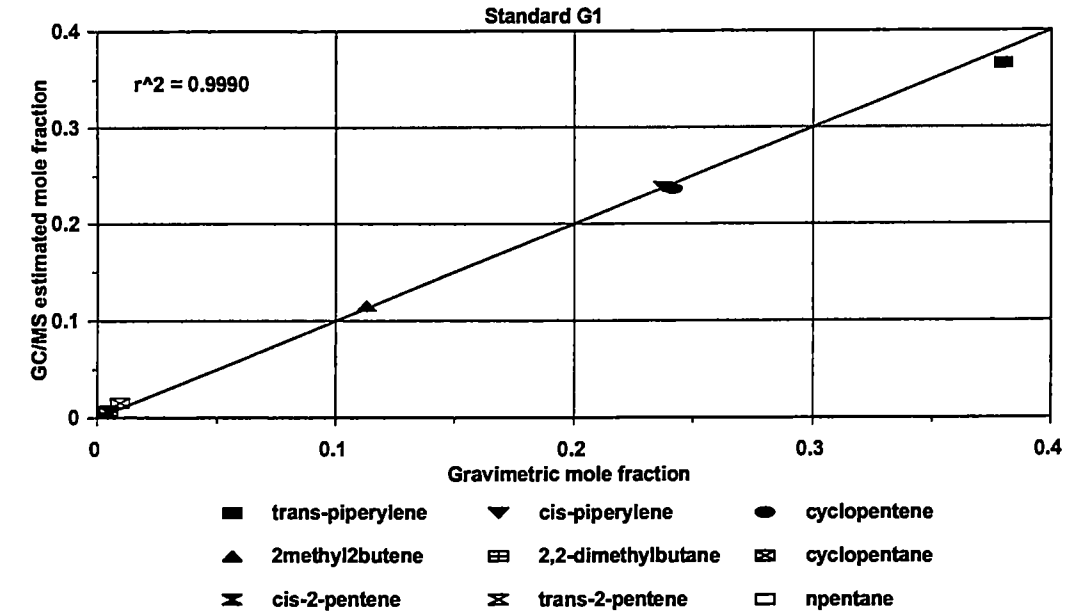
Component / #*:	**Mole fraction Sample G1:**	**Mole fraction Sample G2:**	**Mole fraction grab sample:**
n-pentane / 1	0.007752	0.1100	0.01613
cis-2-pentene / 2	0.006568	0.1068	0.006983
trans-2-pentene / 3	0.006997	0.1054	0.004780
2-methyl-2-butene / 4	0.1158	0.1094	0.1307
trans-piperylene / 5	0.3663	0.1157	0.3613
cis-piperylene / 6	0.2387	0.1201	0.2294
2,2-dimethylbutane / 7	0.005914	0.1088	0.005259
cyclopentene / 8	0.2362	0.1120	0.2330
cyclopentane / 9	0.01574	0.1118	0.01245

*: Corresponds to elution order of components in GC of **Figure 5.1.a.**



Component number:	Component:	Peak area:
1	n-pentane	7392202
2	cis-2-pentene	6396260
3	trans-2-pentene	6828878
4	2-methyl-2-butene	117973068
5	trans-piperylene	354192967
6	cis-piperylene	53258056
7	2,2-dimethylbutane	1958284
8	cyclopentene	253945938
9	cyclopentane	13005456

Figure 5.1.b: Gas chromatogram and tabulated areas of nine-component mixture standard (G1).



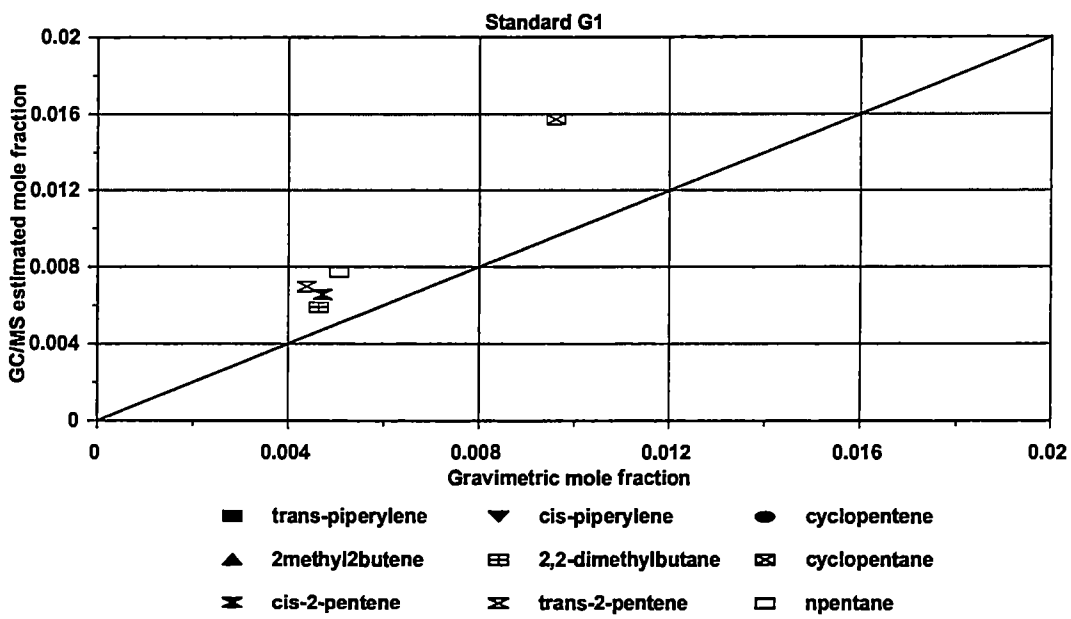


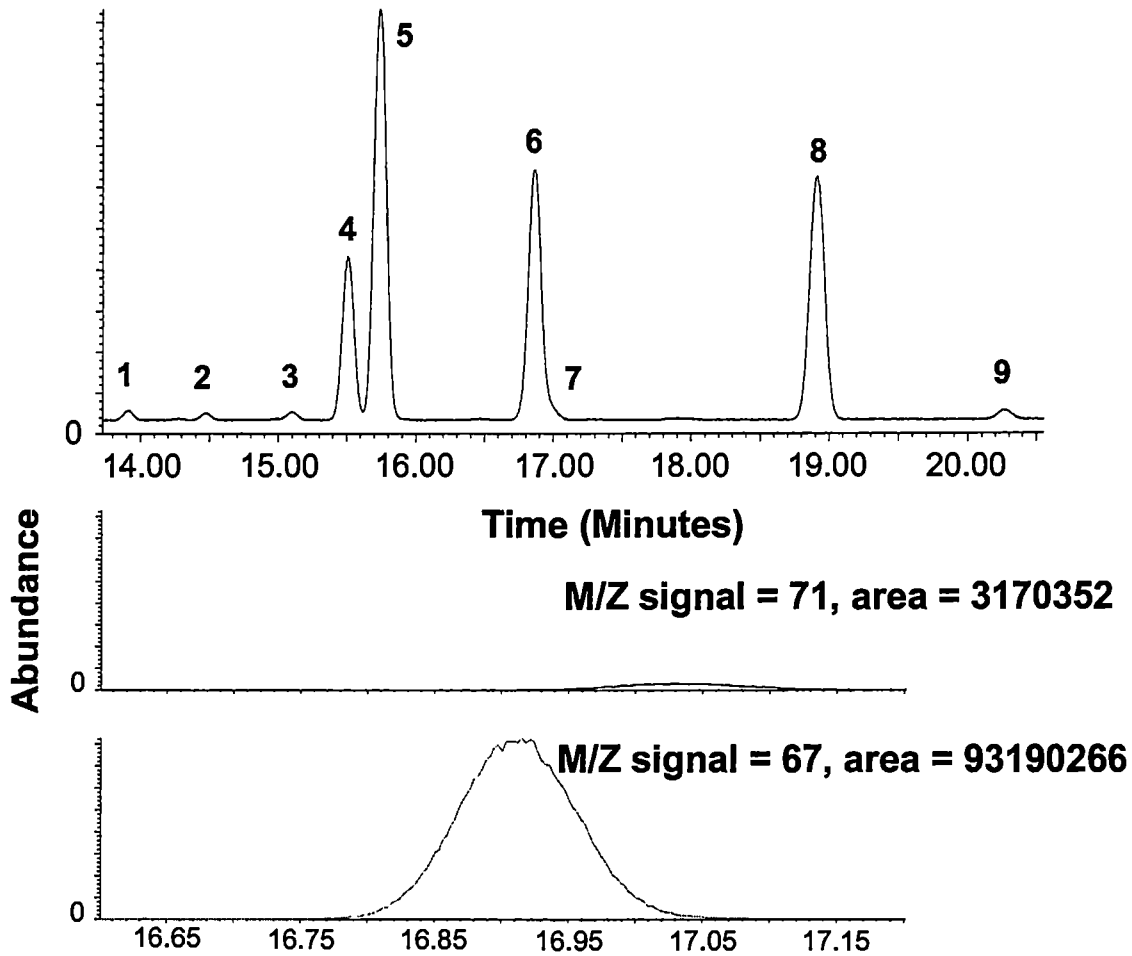
Figure 5.1.c: Plot of GC-MS vs. gravimetric mole fractions from GC-MS analysis of piperylene standard (G1) and lower concentration zoom.

which simulated the piperylene stream, the analysis of the piperylene *grab* sample was undertaken utilizing the techniques described above. **Figure 5.1.d** illustrates the GC obtained from the piperylene grab sample.

It can be seen in **Figure 5.1.d** that similar results are obtained as compared to the G1 standard in **Figure 5.1.b**. **Figure 5.1.e**, a plot of the results in **Table 5.1.a** for the piperylene grab sample vs. those in **Table 3.1.a**, indicates that the results obtained from the cryogenic GC-MS (collaborators laboratory) analysis are similar to those obtained using the 80 meter GC-column (non-cryogenic) in our laboratory. Deviations from the Y = X line in **Figure 5.1.e** may be due to issues including slight differences in the piperylene stream (grab sample) compositions and/or differences in sample handling. The results given in **Table 3.1.a** were obtained from the collaborator's laboratory in 1997 whereas the piperylene grab sample tested in this work was obtained in 1998. Hence, there may be slight differences in grab sample composition. Overall, reasonable resolution of the piperylene grab sample components via non-cryogenic GC is possible using the GC method executed in our laboratory.

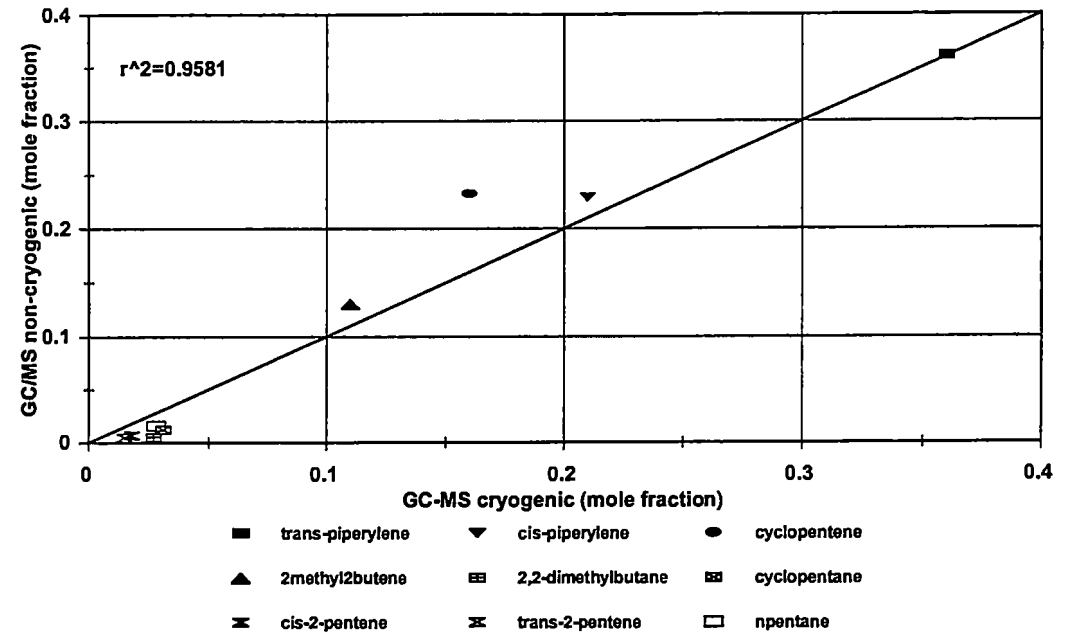
5.2: ANALYSIS OF 9-COMPONENT HYDROCARBON ISOMER MIXTURES BY PrMS

The GC standards (G1 and G2) and piperylene grab sample that were tested by GC-MS in the previous section were subjected to analysis by PrMS using the injection inlet described in Chapter 2. Upon applying the intuitive approach of parameterization to the nine-component mixtures, negative concentration estimates and a poor correlation



Component number:	Component:	Peak area:
1	n-pentane	27997389
2	cis-2-pentene	12380867
3	trans-2-pentene	8493165
4	2-methyl-2-butene	242453981
5	trans-piperylene	636045015
6	cis-piperylene	93190266
7	2,2-dimethylbutane	3170352
8	cyclopentene	455951552
9	cyclopentane	18728329

Figure 5.1.d: Gas chromatogram of piperylene grab sample (TIC) and tabulated GC peak areas. Selected ions m/z = 67 and 71 were used to determine the mole fractions for cis-piperylene and 2,2-dimethylbutane respectively.



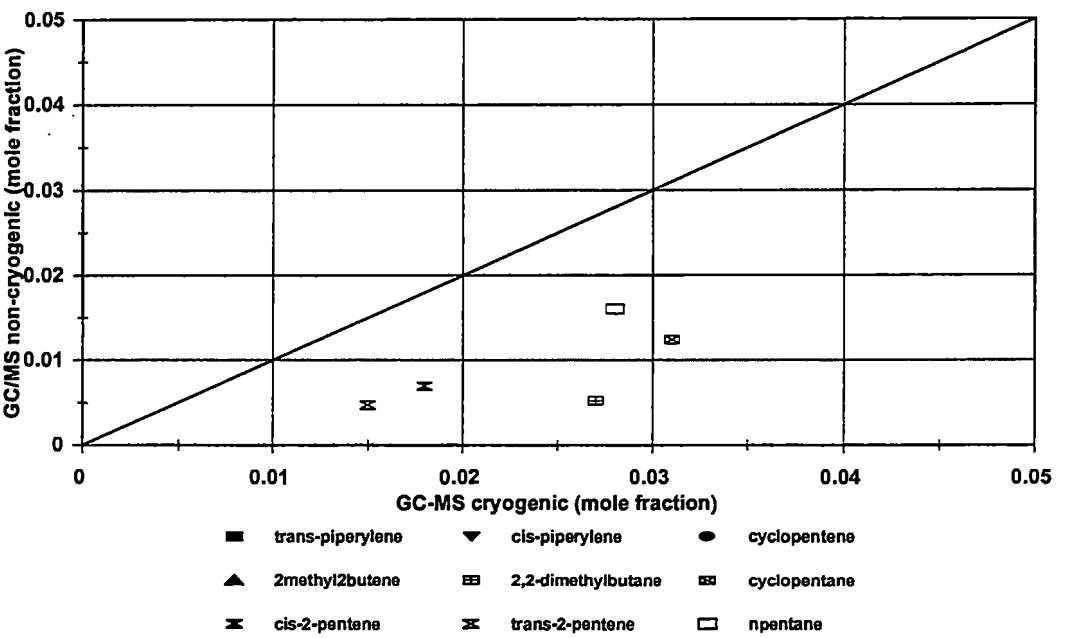


Figure 5.1.e: Plot of non-cryogenic GC-MS vs. cryogenic GC-MS for piperylene grab sample. Bottom figure is a magnification at the lower concentrations.

coefficient result were obtained. This was not surprising since the intuitive parameterization applied to the piperylene binary mixtures afforded poor results as well. Thus, the brute force method of parameterization was applied to the G1 standard sample using G2 as the calibration sample.

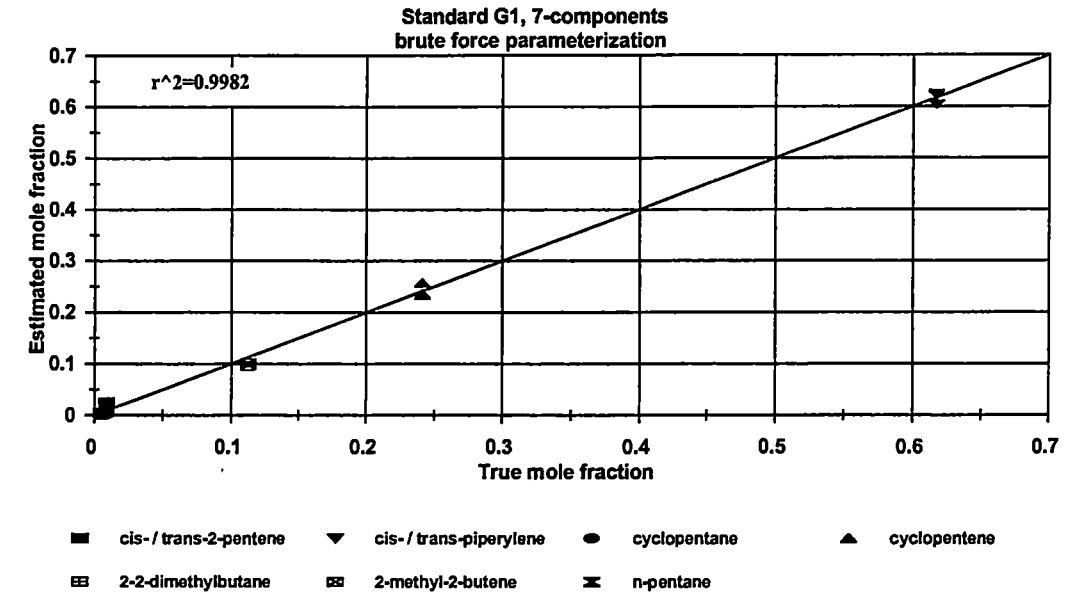
The results of the brute force parameterization method indicated that there were no parameterizations that afforded all non-negative concentration estimates for the 9-component system. In response to these results, and the results obtained from the binary mixtures of piperylene, it was concluded that the 9-component system was too complex (too many components) for accurate deconvolution by PrMS and thus simplification (reduction of the number of components) of the system was necessary. Simplification was performed by combining (averaging) both pairs of the stereoisomer reference spectra to afford 2 quasi-components. The cis- / trans-piperylene reference spectra were averaged with respect to the individual m/z signals to afford one piperylene reference spectrum. In addition, the individual cis- / trans-piperylene mole fractions were also combined to afford a total concentration of piperylenes. The same averaging and combining procedure was used for the cis- / trans-2-pentene components. Thus, a 7-component system was the result of combining the two pairs of stereoisomers from the 9-component system. The stereoisomers were combined since they were the problem components (negative concentration estimates) in the 9-component brute force analysis and because they afforded the highest SI in the previous chapter.

Upon applying a brute force parameterization approach to standard G1, utilizing G2 as the calibration standard, and simplifying the stereoisomer components as described, results for the 7-component piperylene system are illustrated in the validation

plot of **Figure 5.2.a**. The upper portion of the figure represents the success achieved in deconvolving the 9-component G1 standard by combining the stereoisomers (r^2 = 0.9982). The bottom portion of the figure is a magnification of the lower concentrations. What is evident here is that the cyclopentane and cis- / trans-2-pentene components demonstrate a compromised precision as compared to the saturated components (*n*-pentane and 2,2-dimethylbutane). This evidence supports the same theme that is apparent in the other data presented in this work with regard to the differences in the relative reactivity of the components tested. That is, the saturated hydrocarbons (*n*-pentane and 2,2-dimethylbutane) are measured with higher accuracy and precision than the unsaturated hydrocarbons. This is also supported by the requirement of combining the stereoisomers into one quasicomponent before the accurate and precise deconvolution of their components could be achieved.

Also, in observing the lower validation plot in **Figure 5.2.a**, it is evident that a horizontal line can be drawn through the replicate points of the 2-pentene and cyclopentane components. This implies that these components may exist at a level that is below the limit of detection. It is also possible that the same limit of detection issue applies to the remaining components in this validation plot (*n*-pentane and 2,2-dimethylbutane). It is recommended that for future work this issue regarding the limit of detection be investigated in more detail. For purposes of this work, these results indicate that it is feasible to use PrMS to rapidly monitor the three major components (piperylenes, cyclopentene and 2-methyl-2-butene) with reasonable accuracy and precision.

Finally, the piperylene grab sample was analyzed utilizing the same approach and



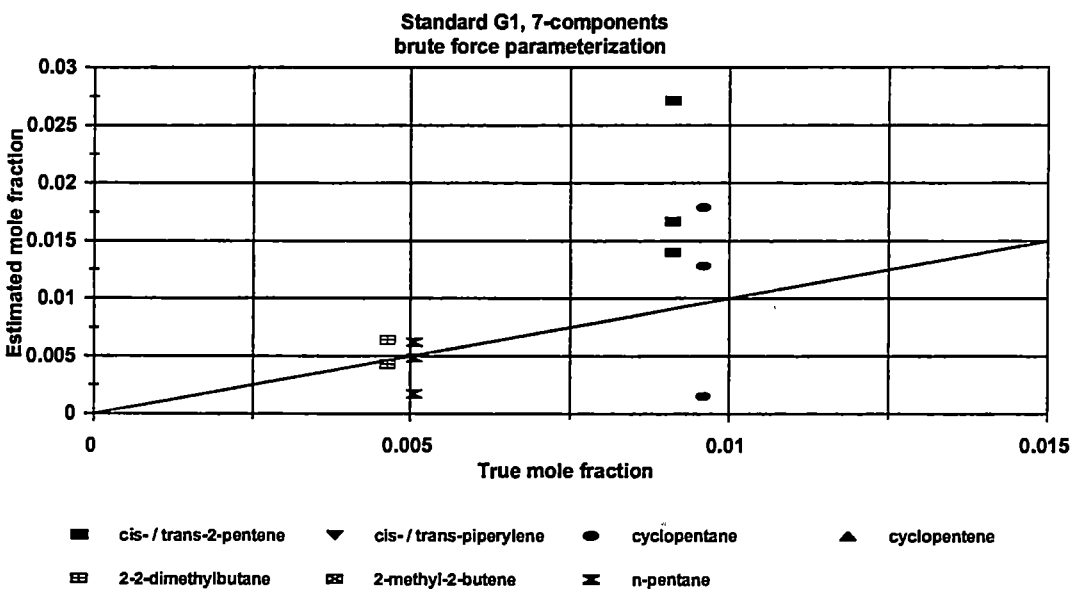


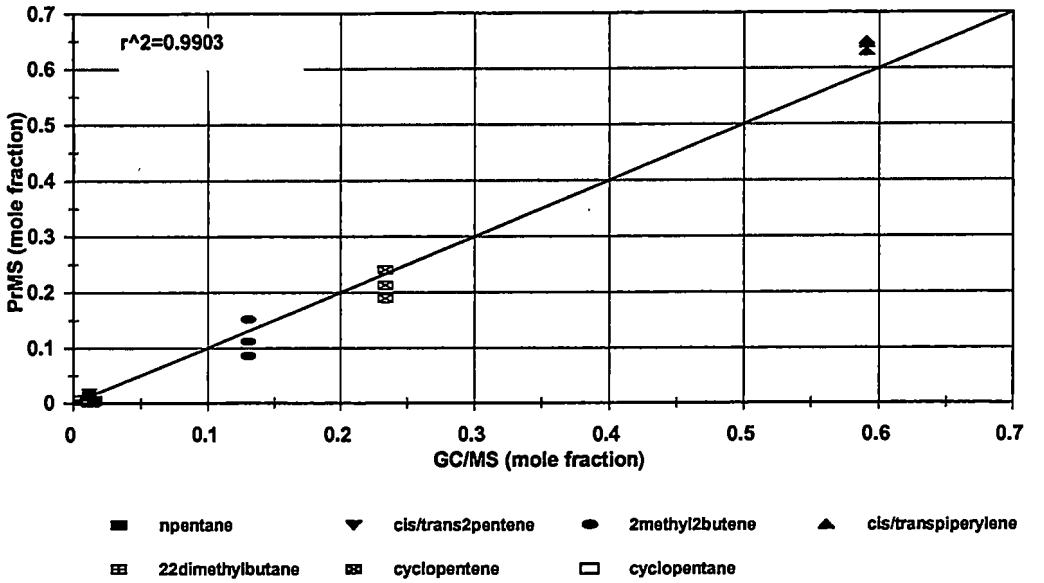
Figure 5.2.a: Validation plot for 7-component piperylene stream standard (G1) calculated from brute force parameterization: 29, 40, 41, 43, 51, 56, 57, 65, 67, 71.

brute force parameterization that was applied to the G1 standard. The resulting accuracy and precision was consistent with that observed in **Figure 5.2.a**; these results are illustrated in the validation plot of **Figure 5.2.b** ($r^2 = 0.9903$). In this case the estimated concentrations predicted by the PrMS method were validated with the GC-MS results for the grab sample given in **Table 5.1.a**. Similar to the case with the G1 standard at low concentrations, upon observing the lower concentrations in the bottom half of **Figure 5.2.b**, it is apparent that the *n*-pentane and cyclopentane components may be present at the limit of detection. Specifically, a horizontal line could be drawn through the estimated concentrations implying that these components are present at the limit of detection.

5.3: SUMMARY AND CONCLUSIONS

Using the ABB-Extrel process mass spectrometer and the injection inlet described in Chapter 2, it was illustrated that at least three of the major components of the piperylene stream could be rapidly (< 5 seconds) measured with good accuracy and precision as illustrated in **Figure 5.2.a**. This turns out to be a much more rapid method of analysis than the GC analysis method which required more than 20 minutes. This near real-time information is useful in a process analysis setting. The analysis time (< 5 seconds) is determined by considering the parameterization utilized in the deconvolution calculation (10 m/z's) and the time required to analyze each selected ion (~ 300 milliseconds).

Mixtures of isomeric hydrocarbons as complex as nine-components and



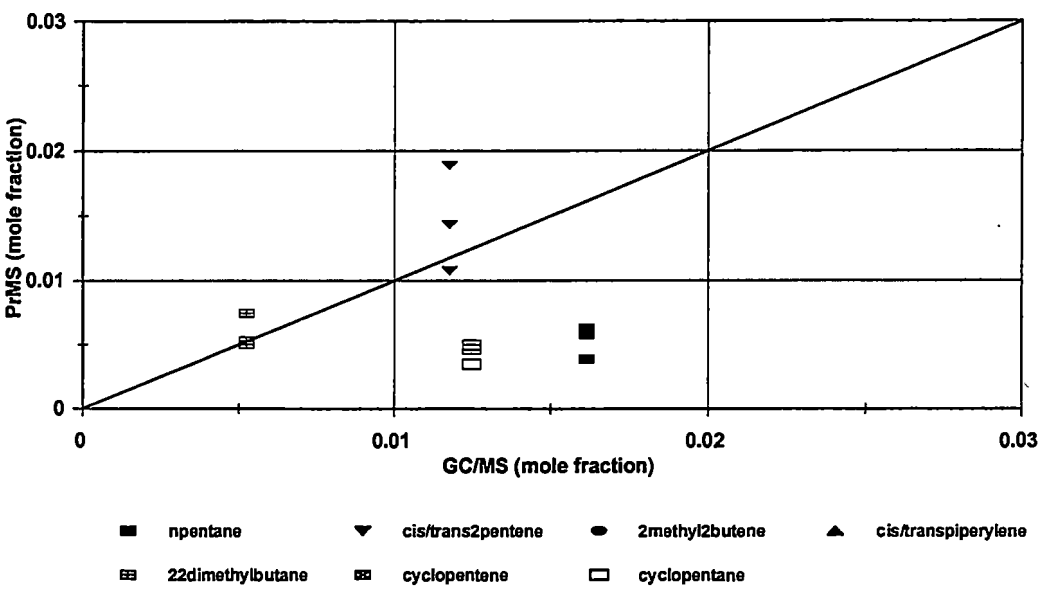


Figure 5.2.b: Validation plots for piperylene stream grab sample calculated from brute force parameterization: 29, 40, 41, 43, 51, 56, 57, 65, 67, 71.

containing stereoisomers require simplification (combination of the stereoisomers) for accurate and precise analysis. However, this assessment is based upon the analysis of one particular sample with component concentrations targeted toward the piperylene stream. Other combinations of stereoisomers should be tested and compared. A brute force analysis is required to identify a parameterization that affords the most accurate concentration estimates for mixtures as complex as nine-components. However, a caveat of the brute force parameterization method is the possibility of overfitting the data. Isomeric components for which the SI approaches the high nineties will likely have to be combined as a quasicomponent for accurate and precise analysis via PrMS if they are to be tested among other stereoisomers.

The EA described in this work requires extension to additional components. This could be done by taking all possible pairs of components and summing the EA score. Developments and improvements in the EA such as this are currently taking place in our laboratory.

REFERENCES

REFERENCES

1. *Chemical and Engineering News*, March 15, **1999**, 21.

2. Puzniak, T. J. "On-line Analysis for Process Optimization and Control." *CEP*, August, **1984**, 29 - 32.

3. Holzworth, R. E. "On-line Plant Performance Monitoring." *ISA Transactions*, 26, No. 1, **1987**, 41 - 45.

4. Clarke, J. R. P. "A Review of On-line Analysis." *Anal. Chim. Acta.*, 190, **1986**, 1 - 11.

5. Devins, D. W. "Economic Benefits of On-line Analysis." *Journal of Coal Quality*, January, 7, No. 1, **1988**, 11 - 14.

6. Parnell, D.; Saltzman, R. S. "Energy Savings through use of On-line Analysis Instrumentation in Alkanol Amine Processes." *Gas Conditioning Conference*, **1984**, 1 - 12.

7. Loeblein, C.; Perkins, J. D.; Srinivasan, B.; Bonvin, D. "Economic Performance Analysis in the Design of On-line Batch Optimization Systems." *J. Process Control*, 9, **1999**, 61 - 78.

8. Marlin T. E. *Process Control, Designing Process and Control Systems for Dynamic Performance*, McGraw Hill Publishers: New York, NY, **1995**.

9. Morris W. E. "In-line Blending and On-line Analysis Offer Advantages." *Oil and Gas Journal*, 84, No. 36, **1986**, 112 - 114.

10. Marco, E.; Pena, J. A.; Santamaria, J. "Early Detection of Runaway Reactions in Systems with Gas Evolution using On-line Mass Spectrometry." *Chem. Eng. Sci.*, 52, No. 18, **1997**, 3107 - 3115.

11. Personal communication. Measurement and Control Engineering Center industrial member. University of Tennessee-Knoxville. October, **1997**.

12. Robinson, C.F. "Process Control by Mass Spectrometer." *Chem. Eng.*, 58, **1951**, 136 - 137.

13. Walker, J. K.; Gifford, A. P.; Nelson, R. H. "Process-Monitor Mass Spectrometer." *Ind. Chem. Eng.*, 46, **1954**, 1400 - 1403.

14. Thomas, B. W. "Future Holds Great Promise for Continuous Analysis." *Petroleum Refiner*, 30, **1951**, 81 - 84.

15. Spracklen, S. B. "Continuous Analysis Sampling in Petroleum and Petroleum-Chemical Processing. I. Objective: A Measurable and Representative Sample." *Oil and Gas J.*, 52, No. 31, **1953**, 94 - 96.

16. Thomas, B. W. "How Analyzers Fit into System." *Oil and Gas J.*, 52, No. 31, **1953**, 96 - 98.

17. Barr, F. L. "Petroleum-Chemical Systems Must Be Tailor Made." *Oil and Gas J.*, 52, No. 31, **1953**, 98 - 100.

18. Thomas, B. W.; Martin, R. L. "Sampling Systems for Plant Analyzers." *J. Instrum. Soc. Am.*, 1, No. 6, **1954**, 29 - 31.

19. Schaafsma, J. G. "Use of the Mass Spectrometer in the Routine Analysis of Refinery Gas Samples." *Natl. Petroleum News*, 36, **1944**, R196 - R200.

20. Schaafsma, J. G. "Use of the Mass Spectrometer in the Routine Analysis of Refinery Gas Samples." *Petroleum Refiner*, 23, **1944**, 123 - 127.

21. Mawson, J. "The On-line Analysis of Process Streams." *Automatic Control in the Chemical Process and Allied Industries Conference*, April 14-16, **1964**, 15-28.

22. Nicholas P. "Comprehensive Gas Analysis for Processing." *Spectroscopy*, 6, No. 9, **1991**, 36 - 40, 42 - 45.

23. Schugerl K. "New Trends in Measurement and Control of Biotechnological Processes." *Pure Appl. Chem.*, 64, No. 2, **1992**, 205 - 212.

24. Carr-Brion, K. G. "Sampling System for On-line Analysis of Materials that are Difficult to Handle." *Anal. Chim. Acta*, 238, **1990**, 11 - 16.

25. Callis, J. B.; Illman, D. L.; Kowalski, B. R. "Process Analytical." *Anal. Chem.*, 59, No. 9, **1987**, 624A - 637A.

26. Converse, J. G. "Taking the Analysis to the Sample." *Process Control Quality*, 5, **1993**, 131 - 136.

27. Cornish, D. C.; Jepson, G.; Smurthwaite, M. J. *Sampling Systems for Process analyzers* Butterworths: London, **1981**.

28. Crandall, J. A. "On-line Process Analyzers." *Chemical Engineering*, April **1993**, 94 - 98.

29. Hall, J.; Lundgren, S.; Keck, K. E.; Kasemo, B. "A Gas Inlet System for Quantitative Mass Spectrometry." *Int. J. Mass Spectrom. and Ion Proc.*, 108, **1991**, 1 - 21.

30. Scrivens, J. H.; Ramage, J. C. "Current Developments in On-line Mass Spectrometry." *Int. J. Mass Spectrom. and Ion Proc.*, 60, **1984**, 299 - 306.

31. Hoeil, C.; Joon-Sik, L.; Min-Sik, K. "Feasibility of Simultaneous Measurement of Xylene Isomers and other Hydrocarbons in p-Xylene Production Processes using Near-Infrared Spectroscopy." *Appl. Spectrosc.*, 52, No. 6, **1998**, 885 - 889.

32. Townshend, A.; Green, J. D.; Dunn, W. B. "The Potential of On-line Headspace Analysis of Liquid Streams using Direct Mass Spectrometry Detection." *Anal. Proc. Including Anal. Commun.*, 32, September, **1995**, 361 - 353.

33. Dunn, W. B.; Townshend, A.; Green, J. D. "Comparison of Total Vaporization and Dynamic Headspace Techniques Combined with Direct Mass Spectrometric Detection for the On-line Analysis of Liquid Process Streams." *Analyst*, 123, February **1998**, 343 - 348.

34. Brodbelt, J. S. "Inverse Sampling Valve Interface for On-line Process Monitoring with a Mass Spectrometer." *Anal. Chem.*, 64, **1992**, 827 - 829.

35. Didden, C.; Duisings, J. "On-line Measurement of a Liquid Reactor Feed with a Mass Spectrometer." *Process Control and Qual.*, 3, **1992**, 263 - 271.

36. Lewis, G. "Mass Spectrometers move On-line." *InTech*, 36, No. 2, **1989**, 45 - 47.

37. Klemp, M.; Sacks, R. "Sample Decomposition in an Electrically Heated Cold-Trap Inlet System for High-Speed Gas-Chromatography." *HRC-J. High Res. Chromatogr.*, 14, No. 4, **1991**, 235 - 240.

38. Cusworth, W. H.; Tureček, F. "Electrospray Ionization Techniques for Process Mass Spectrometry." *J. Process Anal. Chem.*, 1, No. 2, **1995**, 110 - 114.

39. Villalobos, R. "Process Gas Chromatography." *Anal. Chem.*, 47, No. 11, **1975**, 983A - 1004A.

40. Chopra, L. K.; Kapoor, V. B.; Bhagat S. D. "Aromatised Product of n-Heptane." *Erdeol Kohle, Erdgas, Petrochemie.*, 48, No. 4 - 5, **1995**, 186 - 190.

41. Scharfenberger, G. "Methods of Gas Sampling and Solvent Determination in Waste Gases." *Chem. Labor Biotech.*, **1992**, 43, No. 8, 434 - 437.

42. Villalobos, R.; Annino, R. "Methods of On-line Analysis of Volatile Organics in Wastewater by Process Gas Chromatography." *Adv. Instrum. Control*, **1991**, 46,

No. 1, 707 - 718.

43. Guillemin, C. L. "Progress and Prospects in On-line Process Gas and Liquid Chromatography." *Anal. Chim. Acta*, **1990**, 238, No. 1, 17 - 33.

44. Klemp, M.; Peters, A.; Sacks, R. "High Speed GC Analysis of VOC's: Sample Collection and Inlet Systems." *Environ. Sci. Technol.*, 28, No.8, **1994**, 369A - 428A.

45. Sacks, R.; Akard, M. "High-speed GC Analysis of VOC's: Tunable Selectivity and Column Selection." *Environ.Sci.Technol.*, No. 9, 28, **1994**, 428A - 433A.

46. Ke, H. Q.; Levine, S. P.; Berkley, R. "Analysis of Complex-Mixtures of Vapors in Ambient Air by Fast-Gas Chromatography." *J. Air and Waste Management Assoc.*, 42, No. 11, **1992**, 1446 - 1452.

47. Peters, A.; Sacks, R. "Inlet Pressure Programming for High-Speed Gas-Chromatography." *J. Chromatogr. Sci.*, 30, No. 5, **1992**, 187 - 191.

48. Akard, M.; Sacks, R. "Column Bifurcation with Tunable Selectivity for High-Speed Gas-Chromatography." *J. Chromatogr. Sci.*, 31, No. 8, **1993**, 297 - 304.

49. Munchmeyer, W.; Walte, A.; Matz, G. "On-line Measurement of Polycyclic Aromatic Hydrocarbons by Fast GC/MS." *Polycyclic Aromatic Compounds*, 9, **1996**, 299 - 306.

50. Ke, H.; Levine, S. P.; Mouradian, R. F.; Berkley R. "Fast Gas Chromatography for Air Monitoring: Limits of Detection and Quantitation." *Am. Ind. Hyg. Assoc. J.*, 53, **1992**, 130 - 137.

51. Martin, A. J. P.; James, A. T. "Gas-Liquid Chromatography: The Gas Density Meter, A New Apparatus for the Detection of Vapors in Flowing Gas Streams." *Biochem. J.* (London), 63, **1956**, 138 - 143.

52. Schuy, K. D.; Reinhold B. "Process Gas Analysis in the Iron and Steel Industry by Mass Spectrometry." *Meas. and Control*, 4, No. 5, **1971**, Y84 - T89.

53. Thompson, K. R. "Process Mass Spectrometry." *DuPont Innovation, Wilmington, Del., E.I. duPont de Nemours and Co.*, 5, No. 3, **1974**, 18 - 20.

54. Gregori, A. M.; Kephart, L. A. "Mass Spectrometry Improves Polyethylene Process Control." *Intech*, 43, No. 9, **1995**, 68 - 70.

55. Sharkey, Jr., A. G. "Mass Spectrometric Analysis of Process Streams for Coal-Derived Fuels." *Carcinogenesis - A Comprehensive Survey.*, Freudenthal, R. I.; Jones, P. W., Raven Press, New York, 1, **1976**, 341 - 347.

56. Damoth, D. C.; Montgomery, K. L. "On-line Monitoring with Mass Spectrometers." *Chem. Eng. Prog.*, 68, No. 10, **1972**, 51 - 52.

57. McKeown M. "Some Rules of Thumb for using On-line Mass Spectrometry." *InTech*, 8, **1981**, 45 - 50.

58. Heinzle, E. "Present and Potential Applications of Mass Spectrometry for Bioprocess Research and Control." *J. Biotech.*, 25, **1992**, 81 - 114.

59. Colin, T. "Vacuum Dryer Endpoint Prediction using Process Mass Spectrometry." *AT-Process*, **1996**, 63 - 70.

60. White, G. "Process Gas Analysis by Mass Spectrometry." *Proc. Process Chemists' Conf.*, 28, **1975**, 83 - 93.

61. Scrivens, J. H. "Process Analysis using Mass Spectrometry." *Royal Society of Chemistry*, 154, No. 154, **1994**, 145 - 152.

62. Cessna, G. D. "On-line Analysis of Fuel Gas Streams in a Refinery using Process Mass Spectrometry." *ISA*, No. 90-455, **1990**, 383 - 389.

63. Nicholas, P. "Process and Environmental Monitoring using Mass Spectrometry." *Spectroscopy*, 6, No. 9, **1991**, 36 - 45.

64. Adams, V. H. "Process Mass Spectrometry." *ISA*, No. 89-0344, **1989**, 389 - 400.

65. Cook, K. D.; Bennett, K. H.; Haddix, M. L. "On-line Mass Spectrometry: A Faster Route to Process Monitoring and Control." *Ind. Eng. Chem. Res.*, 38, No. 4, **1999**, 1192 - 1204.

66. Norris, P. E.; Scrivens, J. H. "On-line Mass Spectrometry Evaluation and Implementation." *Dynamic Mass Spectrometry*, London, Philadelphia, Heyden, 6, Chapter 13, **1981**, 155 - 166.

67. Beebe, K. R.; Blaser, W. W.; Bredeweg, R. A.; Chauvel, Jr., J. P.; Harner, R. S.; LaPack, M.; Leugers, A.; Martin, D. P.; Wright, L. G.; Yalvac E. D. "Process Analytical Chemistry." *Anal. Chem.*, 65, **1993**, 199R - 216R.

68. Venkatesan, T. R. "Mass Spectrometer in Exploration of Venusian Atmosphere; A Review." *Indian Soc. for Mass Spec., India* | Counc. for Sci. and Ind. Res., Dep. Atomic Energy India | Geol. Surv. India., **1981**, 2.

69. Nier, A. O. "Mass Spectrometry in Planetary Research." *Int. J. Mass Spec. and Ion Proc.*, 66, No. 1, **1985**, 55 - 73.

70. Chernavskii, D.; Glianenko, A.; Ishikawa, Y.; Kaneko, T.; Kawasaki, Y.; Kobayashi, K.; Koike, J.; Kotov, Y.; Kuzitcheva, E.; Martynov, A.; Oshima, T.; Saito, T.; Tsarev, V.; Yammamoto, T.; Yanagawa, H. "Search for Bioorganic Compounds and Organisms on Mars." *J. Biol. Phys.*, 20, No. 1 - 4, **1994**, 55 - 59.

71. Wyatt, J. R.; Koslin, M. E. "Atmosphere Analysis On-board Nuclear Submarines." *Proc. 37th ASMS Conference on Mass Spectrometry and Allied Topics*, Miami Beach, FL, **1989**, 80 - 81.

72. Ronse, K.; Tritchkov, A.; Randall, J.; Jonckheere, R.; Ghandehari, K.; Van-den-hove, L. "Automated OPC for Application in Advanced Lithography." *Proceedings of SPIE the International Society for Optical Engineering*. 3096, **1997**, 138 - 144.

73. Webb, A. P. "Process Control for III - V Semiconductor Device Fabrication using Mass Spectroscopy." *Applied Surface Sci.*, 63, **1993**, 70 - 74.

74. Scrivens, J. H. "The Use of Mass Spectrometry as a Process Technique: An Evaluation." *Vacuum*, 32, No. 3, **1982**, 169 - 174.

75. Watson, J. T. *Introduction to Mass Spectrometry*, 3rd Ed.; Lippincott-Raven Publishers: Philadelphia, PA, **1997**.

76. Ponton, J. W. "Chemical Process Control in the 1990's." *Chem. Ind.*, **1993**, 315 - 318.

77. Walsh, M. R.; LaPack, M. A. "On-line Measurements using Mass Spectrometry." *ISA Transactions*, 34, **1995**, 67 - 85.

78. Colin, T. B. "Process Mass Spectrometry as a Tool for Process Understanding." *AT-Process*, **1995**, 410 - 415.

79. Walsh, M. R.; LaPack, M. A. "On-line Measurements using Mass Spectrometry." *Symposium on Instrumentation for the Process Industries*, Proceedings: Annual Symposium on Instrumentation for the Process Industries., 49, **1994**, 97 - 108.

80. Brereton, R. G. "Deconvolution of Mixtures by Factor Analysis." *Analyst*, 120, **1995**, 2313 - 2336.

81. Hindmarch, P.; Demir, C.; Brereton, R. G. "Deconvolution and Spectral Clean-up of Two-component Mixtures by Factor Analysis of Gas Chromatographic-Mass Spectrometric Data." *Analyst*, 121, **1996**, 993 - 1001.

82. Shadmehr, R.; Angell, D.; Chou, P. B.; Oehrlein, G. S.; Jaffe, R. S. "Principal Component Analysis of Optical Emission Spectroscopy and Mass Spectrometry: Application to Reactive Ion Etch Process Parameter Estimation Using Neural

Networks." *J. Electrochem. Soc.*, 139, No. 3, **1992**, 907 - 914.

83. Kourti, T.; MacGregor, J. F. "Process Analysis, Monitoring and Diagnosis, using Multivariate Projection Methods." *Chemomet. Intell. Lab. Sys.*, 28, **1995**, 3 - 21.

84. Sengupta, S. K.; Cheeseman, S. C.; Brown, S. D.; Foley, H. C. "Partial Least Squares for Mass Spectral Analysis of H_2 and D_2 Plasma Scrambling Products." *Ind. Eng. Chem. Res.*, 31, **1992**, 2003 - 2010.

85. McShane, M. J.; Cote, G. L.; Spiegelman, C. H. "Assessment of Partial Least-Squares Calibration and Wavelength Selection for Complex Near-Infrared Spectra." *Applied Spectrosc.*, 52, No. 6, **1998**, 878 - 884.

86. Fuller, M. P.; Ritter, G. L.; Draper, C. S. "Partial Least-Squares Quantitative Analysis of Infrared Spectroscopic Data. Part I: Algorithm Implementation." *Applied Spectrosc.*, 42, No. 2, **1988**, 217 - 227.

87. Thomas, E. V. "A Primer on Multivariate Calibration." *Anal. Chem.*, 66, No. 15, **1994**, 795A - 804A.

88. Kaspar, M. H.; Ray, W. H. "Chemometric Methods for Process Monitoring and High-Performance Controller Design." *AIChE Journal*, 38, No. 10, **1992**, 1593 - 1608.

89. Dahlberg, D. B; Lee, S. M.; Wenger, S. J.; Vargo, J. A. "Classification of Vegetable Oils by FT-IR." *Applied Spectrosc.*, 51, No. 8, **1997**, 1118 - 1124.

90. McShane, M. J.; Cote, G. L.; Spiegelman, C. H. "Assessment of Partial Least-Squares Calibration and Wavelength Selection for Complex Near-Infrared Spectra." *Applied Spectrosc.*, 52, No. 6, **1998**, 878 - 884.

91. Edye, L. A.; Clarke, M. A. "Near Infrared Analysis of Sugarbeet Juices and Factory Products." *J. Sugar Beet Res.*, 30, No. 1 - 2, **1993**, 90.

92. Brown, P. J.; Spiegelman, C. H.; Denham, M. C. "Chemometrics and Spectral Frequency Selection." *Phil. Trans. R. Soc. Lond. A.*, 337, **1991**, 311 - 322.

93. Personal communication; Dr. D. W. Noid, MCEC academic collaborator, **1995**.

94. Cook, K. D.; Haddix, M. L.; Noid, D. W.; Sumpter, B. G.; Keator, E. A. University of Tennessee, unpublished results.

95. Kang, S. G.; Lee, D.H.; Ward, A. C.; Lee, K. J. "Rapid and Quantitative Analysis of Clavulanic Acid Production by the Combination of Pyrolysis Mass Spectrometry and Artificial Neural Network." *J. Microbiol. and Biotech.*, 8, No. 5, **1998**, 523 - 530.

96. Jimenez, C. R.; Li, K. W.; Dreisewerd, K.; Spijker, S.; Kingston, R.; Bateman, R. H.; Burlingame, A. L.; Smit, A. B.; Van-Minnen, J.; Geraerts, W. P. M. "Direct Mass Spectrometric Peptide Profiling and Sequencing of Single Neurons Reveals Differential Peptide Patterns in a Small Neuronal Network." *Biochemistry*, 37, No. 7, **1998**, 2070 - 2076.

97. Kenyon, R. G. W.; Ferguson, E. V.; Ward, A. C. "Application of Neural Networks to the Analysis of Pyrolysis Mass Spectra." *Z. Bakteriologie*, 285, No. 2, **1997**, 267 - 277.

98. Wise, B. M.; Gallagher, N. B. "The Process Chemometrics Approach to Process Monitoring and Fault Detection." *J. Proc. Cont.*, 6:, No. 6, **1996**, 329 - 348.

99. Tauler, R.; Kowalski, B. "Multivariate Curve Resolution Applied to Spectral Data from Multiple Runs of an Industrial Process." *Anal. Chem.*, 65, **1993**, 2040 - 2047.

100. Questor IV Process Mass Spectrometer Operations and Maintenance Manual. ABB-Extrel, 575 Epsilon Drive, Pittsburgh, PA 15238-2838

101. McLafferty, F. W.; Zhang, M. Y. "The Best Algorithm and Data Base for Matching Unknown Mass Spectra." Proceedings of the 45th ASMS Conference on Mass Spectrometry and Allied Topics, Palm Springs, CA, **1997**, 361.

102. Leardi, R.; Boggia, R.; Terrile, M. "Genetic Algorithms as a Strategy for Feature Selection." *J. Chemometrics*, 6, **1992**, 267 - 281.

103. Leardi, R. "Application of a Genetic Algorithm to Feature Selection Under Full Validation Conditions and to Outlier Detection." *J. Chemometrics*, 8, **1994**, 65 - 79.

104. Van den Broek, W. H. A. M.; Wienke, D.; Melssen, W. J.; Buydens M. C. "Optimal Wavelength Range Selection by a Genetic Algorithm for Discrimination Purposes in Spectroscopic Infrared Imaging." *Appl. Spectrosc.*, 51, No. 8, **1997**, 1210 - 1217.

105. Broadhurst, D.; Goodacre, R.; Jones, A.; Rowland J. J.; Kell, D. B. "Genetic Algorithms as a Method for Variable Selection in Multiple Linear Regression and Partial Least Squares Regression, with Applications to Pyrolysis Mass Spectrometry." *Anal. Chim. Acta*, 348, **1997**, 71 - 86.

106. Li, T. H.; Lucasius, C. B.; Kateman, G. "Optimization of Calibration Data with the Dynamic Genetic Algorithm." *Anal. Chim. Acta*, 268, **1992**, 123 - 134.

107. Guo, A.; Lee, P. W. N.; Katz, J.; Bergner, A. E. "Quantitation of Positional Isomers of Deuterium-Labeled Glucose by Gas Chromatography/Mass

Spectrometry." *Anal. Biochem.*, 204, **1992**, 273 - 282.

108. Suzuki, M. "Gas Chromatography-Mass Spectrometry Selected Ion Monitoring Method for Quantitative Analysis of 1,2,3,4,5,6-Hexachlorocyclohexane Isomers in Environmental Samples." *Chromatogr. Mass Spectrom. in Biomed. Sci.*, 14, No. 2, **1983**, 485 - 496.

109. Hect, S. S.; Bondinell, W. E.; Hoffman, D. "Chrysene and Methylchrysene. XXIX. Presence in Tobacco Smoke and Carcinogenicity." *J. Natl. Cancer Inst.*, 53, No. 4, **1974**, 1121 - 1133.

110. Hecht, S. S.; Loy, M.; Hoffman, D. *"Carcinogenesis - On the Structure and Carcinogenicity of the Methylchrysenes."* R. Freudenthal and P. W. Jones, Eds., Raven Press, New York, 1, **1976**, 325 - 340.

111. Mullin, M. D.; Pochini, C. M. "High-Resolution PCB Analysis: Synthesis and Chromatographic Properties of all 209 PCB Congeners." *Environ. Sci. Technol.*, 18, **1984**, 468 - 476.

112. Silvon, L. E.; Gebhart, J. E.; Hayes, T. L. "Automated Procedures for Mass Spectrometric Determination of Polychlorinated Biphenyls as Isomer Groups." *Anal. Chem.*, 57, **1985**, 2464 - 2469.

113. Gebhart, J. E.; Hayes, T. L. "Mass Spectrometric Determination of Polychlorinated Biphenyls as Isomer Groups." *Anal. Chem.*, 57, **1985**, 2458 - 2463.

114. Hold, K. M.; deBoer, D.; Bos, K. L.; Van Ooijen, R. D.; Zuidema J.; Maes, R. A. A. "Enantioselective Quantitation of (R)- and (S)-Alprenolol by Gas Chromatography-Mass Spectrometry in Human Saliva and Plasma." *J. Chromatogr. Sci.*, 34, **1996**, 13 - 19.

115. Beck, H.; Eckart, K.; Mathar, W.; Ruhl, Ch.-S.; Wittkowski, R. "Isomer-specific Determination of 2,3,7,8-tetrachlorodibenzo-p-dioxin and Related Compounds in Human Fat and Food." *Biomed. Environ. Mass Spectrom.*, 16, No. 1 - 12, **1988**, 161 - 165.

116. Wilson, R.; Smith, R.; Wilson, P.; Shepherd, M. J.; Riemersma, R. A. "Quantitative Gas Chromatography-Mass Spectrometry Isomer-Specific Measurement of Hydroxy Fatty Acids in Biological Samples and Food as a Marker of Lipid Peroxidation." *Anal. Biochem.*, 248, **1997**, 76 - 85.

117. Christie, W. W.; Brechany, E. Y. "Mass Spectrometry of Derivatives of Isomeric Allenic Fatty Acids." *Biolog. Mass Spectrom.*, 21, **1992**, 267 - 270.

118. Husain, S.; Krishnamurthy, A. S. R.; Sarma, P. N. "Separation and Identification of Chloropyridine Isomers and Quantitative Determination of Pentachloropyridine in Chlorinated Pyridine Residues by Gas Chromatography/Mass Spectrometry." *J. Chromatogr.*, 285, **1984**, 509 - 514.

119. Hasler, E.; Rettinghaus, G. "Mass Spectrometric Contamination Control of Process Gases." *Vacuum*, 38, No. 8 - 10, **1988**, 777 - 781.

120. Buckley, M. E. "Process Control in the Semiconductor Manufacturing Environment using a High Pressure Quadrupole Mass Spectrometer." *Vacuum*, 44, No. 5 - 7, **1993**, 665 - 668.

121. Smyrl, N. R.; Hembree, D. M. Jr.; Davis, W. E.; Williams, D. M.; Vance, J. C. "Simultaneous GC-FT-IR/GC-MS Analysis for Isomer-Specific Identification and Quantitation of Complex Mixture Components." *Applied Spectrom.*, 46, No. 2, **1992**, 277 - 282.

122. Schneider, J. F.; Reedy, G. T.; Ettinger, D. G. "GC/Matrix Isolation/FTIR Applications: Analysis of PCB's." *J. Chromatogr. Sci.*, 23, No. 2, **1985**, 49 - 53.

123. Dass C. "Ion-Molecule Reactions of $[Ketene]^+$ as a Diagnostic Probe for Distinguishing Isomeric Alkenes, Alkynes and Dienes: A Study of the C_4H_8 and C_5H_8 Isomeric Hydrocarbons." *Org. Mass Spectrom.*, 28, **1993**, 940 - 944.

124. Guilhaus, M.; Kingston, R. G.; Brenton, A. G.; Beynon, J. H. "Mass Spectrometric Methods of Differentiating Between Structures of Isomeric Ions." *Org. Mass Spectrom.*, 20, No. 6, **1985**, 424 - 425.

125. Zhang, M. Y.; McLafferty, F. W. "Quantitative Analysis of Isomeric Ion Mixtures." *Org. Mass Spectrom.*, 27, **1992**, 991 - 994.

126. Grange, A. H.; Sovocool, G. W.; Donnelly, J. R.; Genicola, F. A.; Gurka, D. F. "Identification of Pollutants in a Municipal Well using High-Resolution Mass Spectrometry." *Rapid Commun. Mass Spectrom.*, 12., No. 17, **1998**, 1161 - 1169.

127. GC parameters obtained from MCEC industrial collaborator. November, **1998**.

128. Cook, K. D.; Bennett, K. H.; Haddix, M. L.; Keator, E. A.; Seebach, G. L.; Falconer J. L. "Process Mass Spectrometric Monitoring of Isomeric Hydrocarbon Gases." *J. Process Anal. Chem.*, 3, No. 3, 4, **1997**, 115 - 124.

129. Cook, K. D.; Haddix, M. L.; Seebach, G. L.; Bennett, K. H. "Parameterization for Mass Spectrometric Monitoring of Multicomponent Streams." *J. Process Anal. Chem.*, 3, No. 1, 2, **1997**, 53 - 62.

130. Personal communication: Dr. Paul Gemperline, Eastern Carolina University, **1997.**

131. Drahos, L.; Vekey, K "Quantification of Isomeric Differences in Mass Spectra." *Rapid Comm. Mass Spectrom.*, 10, **1996**, 1309 - 1315.

132. Swain, D.; Winans, R. E.; Dunn, W. J. "Positional Isomer Differentiation of Monoalkylated Naphthalenes using Principal Components Analysis and Mass Spectrometry." *Anal. Chem.*, Vol. 68, No. 18, **1996**, 3244 - 3249.

133. Breci, L.; Smith, L; Tsaprailis, G.; Vaisar, T.; Wysocki, V. "ESI / SID Fragmentation Spectra of Peptides Containing D- or L- Stereoisomers." Presented at the 47th ASMS Conference on Mass Spectrometry and Allied Topics, Houston, TX, **1999**.

134. Cook, K. D.; Haddix, M. L.; Seebach, G. L. "Isomer Resolution using Process Mass Spectrometry." *J. Process Anal. Chem.*, 2, No. 2, **1996**, 219 - 223.

135. Cook, K. D.; Haddix, M. L.; Seebach G. L. "Impact of Parameterization on the Least Squares Deconvolution of Mass Spectral Data for Simultaneous Determination of Isomeric Alkanes." *Adv. Instrum. Control*, 50, No. 3, **1995**, 1169 - 1178.

136. Personal communication: ABB-Extrel research scientist, **1998.**

137. Lay, J. O.; Gross, M. L.; Zwinselman J. J.; Nibbering, N. M. M.; "A Field Ionization and Collisionally Activated Dissociation / Charge Stripping Study of Some $[C_9H_{10}]^+$ Ions." *Org. Mass Spectrom.*, 18, No. 1, **1983**, 16 - 21.

138. Stein, S. E.; Scott, D. R. "Optimization and Testing of Mass Spectral Library Search Algorithms for Compound Identification." *J. Amer. Soc. Mass Spectrom.* 5, **1994**, 859 - 863.

139. Personal communication: Bennett, K. H., Cook, K. D., University of Tennessee-Knoxville, **1999**.

APPENDICES

APPENDIX 1

WHAT *IS* MASS SPECTROMETRY ?

Mass spectrometry (MS) is a powerful analytical technique that is used to provide valuable information to many different kinds of professionals such as physicians, law enforcement officials, chemists, astronomers and especially (in the context of this thesis) process control scientists. It can be used to identify unknown compounds, quantify known materials and to help define the structural and chemical properties of molecules. A mass spectrometer is an instrument that is used to "weigh" molecules. The manner in which it performs this task involves converting the molecules to ions which are molecules that have been electrically charged. Thus, the mass spectrometer does not actually measure the mass directly (as does a conventional scale or balance) but rather via the mass-to-charge ratio of ions formed from the molecules.

Molecules are so small that the use of typical units of weight (such as kilograms or pounds) is not convenient. For example, the mass of a hydrogen atom is approximately 1×10^{-24} grams and thus the use of a smaller unit of mass such as the Dalton is more favorable. A Dalton is a unit of mass used in mass spectrometry that is equivalent to 1/12th of the mass of a single atom of the isotope carbon-12 (^{12}C). Thus, the carbon-12 isotope has a mass of exactly twelve mass units or twelve Daltons. Very minute quantities of sample are typically used (on the order of 10^{-12} grams). For a compound with mass 1000 Daltons this translates into 10^{-15} moles and implies that compounds can be detected at concentrations on the order of one part in 10^{12}.

MS is a destructive technique in the sense that the sample is consumed during the analysis. In all cases of MS, energy is imparted to the analyte to invoke ionization of the constituent molecules. In effect, ions (positive and negative) representative of the analyte are formed and, in most cases, the molecule fragments into smaller pieces. In the case where the analyte is of a reasonably high purity, the resulting fragmentation is characteristic. Thus, when the mass of each fragment is analyzed by the mass spectrometer, the resulting mass spectrum can be considered a "fingerprint" that can be used to characterize the sample. Typically a mass spectrum is a bar-graph plot of intensity vs. m/z, where m/z is mass-to-charge ratio and is equal to the mass in Daltons when z (the charge) is (+) or (-) one.

Mass spectrometers come in a variety of sizes ranging from the size of a small box which can be easily handled by one or more persons to large research instruments that can dominate an entire laboratory.

The most common ionization process in MS is electron ionization (EI) which is typically performed using an electron energy of about 70 electron volts (eV). Generally, EI involves the partial transfer of energy (*e.g.,* 70 eV) to a neutral analyte molecule in the vapor state. Typically, a neutral molecule or atom in this state will require less than 24 eV (the ionization potential (IP) of helium, the element in the periodic table with the highest IP) to acquire sufficient energy to eject one of its own electrons and become a positively charged molecule or atom. The relatively large amount of left-over energy is dissipated via fragmentation of chemical bonds to afford an average of different fragment ions. In other words, not all of the molecules that are ionized break down to their elemental ions. In some cases a strong molecular ion (little or no fragmentation) is

afforded while in other cases there is no molecular ion. The masses of the fragment ions which are formed is equal to the sum of the atomic masses of the group of atoms which retained the positive charge during the fragmentation process.

Mass spectrometry has been used to determine how drugs are utilized by the body, analyze biopolymers such as proteins and oligosaccharides and identify the structures of biomolecules including carbohydrates, nucleic acids and steroids. It has also been used as a breath monitoring system for patients of anesthesiologists during surgery, to analyze the composition of molecular species in outer-space, to detect dioxins in contaminated fish and to locate oil deposits by monitoring petroleum precursors in rocks.

APPENDIX 2

MATLAB CODE FOR EXECUTING BRUTE FORCE METHOD OF PARAMETERIZATION

This appendix lists computer code for the brute force Matlab program with line numbers such that the following detailed description can be followed. The following example describes a 2 component system with 5 total samples and 3 replicate analyses of all samples and reference spectra. However, any number of components and m/z ratios can readily be applied to this program with minor modifications. In this particular application sample number 2 is used for the calculation of the sensitivity factor. The use of this factor compensates for differences between components due to variations in vapor pressure, ionization cross-section and ionization potential. This calibration sample is selected based upon its component composition falling closest to the average composition of all other samples. The remaining samples are used as test samples.

```
1   for xxx=1:M-2;
2   for yyy=xxx+1:M-1;
3   for zzz=yyy+1:M;
4   idx=[xxx yyy zzz];
5   negchkgen=0;%Global negatives checking variable
6   ee=refdataa(idx,:)\kcda(idx,:); % Begin calc. of sens. factor for run 1
7   ff=ee./consena;
8   sena=diag(ff); % sensitivity factor for run 1
9   ee=refdatab(idx,:)\kcdb(idx,:); % Begin calc. of sens. factor for run 2
10  ff=ee./consenb;
11  senb=diag(ff); % sensitivity factor for run 2
12  ee=refdatac(idx,:)\kcdc(idx,:); % Begin calc. of sens. factor for run 3
13  ff=ee./consenc;
14  senc=diag(ff); % sensitivity factor for run 3
15  x1a1=[(refdataa(idx,:)*sena)\bdataX1a(idx,:)];%est.conc.run1,unnrmlzd
16  x2a1=sum(x1a1);
17  xa1=x1a1./x2a1;%est.conc.run1,normalized
18  negchk=0;negsum=0;
19  negchk=xa1<0;
20  negsum=sum(negchk);
21  if negsum >0, negchkgen=negchkgen+1;end
22  x1a2=[(refdatab(idx,:)*senb)\bdataX1b(idx,:)];%est.conc.run2,unnrmlzd
23  x2a2=sum(x1a2);
24  xa2=x1a2./x2a2;%est.conc.run2,normalized
25  negchk=0;negsum=0;
26  negchk=xa2<0;
27  negsum=sum(negchk);
28  if negsum >0, negchkgen=negchkgen+1;end
29  x1a3=[(refdatac(idx,:)*senc)\bdataX1c(idx,:)];%est.conc.run3,unnrmlzd
30  x2a3=sum(x1a3);
31  xa3=x1a3./x2a3;%est.conc.run3,normalized
32  negchk=0;negsum=0;
```

```
negchk=xa3<0;
negsum=sum(negchk);
if negsum >0, negchkgen=negchkgen+1;end
xall=[xa1 xa2 xa3]; xallrot=rot90(xall); %Collect the 3 [norm]'s, rotate
xa=mean(xallrot);%Find average
xdeva=xa - condatx1;%Deviations between ave.norm.est.conc. and condatx1
xdevsuma=sum((xdeva).^2); %Sum squared deviations
x1a1=[(refdataa(idx,:)*sena)\bdataX3a(idx,:)];%est.conc.run1,unnrmlzd
x2a1=sum(x1a1);
xa1=x1a1./x2a1;%est.conc.run1,normalized
negchk=0;negsum=0;
negchk=xa1<0;
negsum=sum(negchk);
if negsum >0, negchkgen=negchkgen+1;end
x1a2=[(refdatab(idx,:)*senb)\bdataX3b(idx,:)];%est.conc.run2,unnrmlzd
x2a2=sum(x1a2);
xa2=x1a2./x2a2;%est.conc.run2,normalized
negchk=0;negsum=0;
negchk=xa2<0;
negsum=sum(negchk);
if negsum >0, negchkgen=negchkgen+1;end
x1a3=[(refdatac(idx,:)*senc)\bdataX3c(idx,:)];%est.conc.run3,unnrmlzd
x2a3=sum(x1a3);
xa3=x1a3./x2a3;%est.conc.run3,normalized
negchk=0;negsum=0;
negchk=xa3<0;
negsum=sum(negchk);
if negsum >0, negchkgen=negchkgen+1;end
xall=[xa1 xa2 xa3]; xallrot=rot90(xall); %Collect the 3 [norm]'s, rotate
xa=mean(xallrot);%Find average
xdeva=xa - condatx3;%Deviations between ave.norm.est.conc. and condatx1
xdevsumb=sum((xdeva).^2); %Sum squared deviations
x1a1=[(refdataa(idx,:)*sena)\bdataX4a(idx,:)];%est.conc.run1,unnrmlzd
x2a1=sum(x1a1);
xa1=x1a1./x2a1;%est.conc.run1,normalized
negchk=0;negsum=0;
negchk=xa1<0;
negsum=sum(negchk);
if negsum >0, negchkgen=negchkgen+1;end
x1a2=[(refdatab(idx,:)*senb)\bdataX4b(idx,:)];%est.conc.run2,unnrmlzd
x2a2=sum(x1a2);
xa2=x1a2./x2a2;%est.conc.run2,normalized
negchk=0;negsum=0;
negchk=xa2<0;
negsum=sum(negchk);
if negsum >0, negchkgen=negchkgen+1;end
x1a3=[(refdatac(idx,:)*senc)\bdataX4c(idx,:)];%est.conc.run3,unnrmlzd
x2a3=sum(x1a3);
xa3=x1a3./x2a3;%est.conc.run3,normalized
negchk=0;negsum=0;
negchk=xa3<0;
negsum=sum(negchk);
if negsum >0, negchkgen=negchkgen+1;end
xall=[xa1 xa2 xa3]; xallrot=rot90(xall); %Collect the 3 [norm]'s, rotate
xa=mean(xallrot);%Find average
xdeva=xa - condatx4;%Deviations between ave.norm.est.conc. and condatx1
xdevsumc=sum((xdeva).^2); %Sum squared deviations
x1a1=[(refdataa(idx,:)*sena)\bdataX5a(idx,:)];%est.conc.run1,unnrmlzd
x2a1=sum(x1a1);
xa1=x1a1./x2a1;%est.conc.run1,normalized
negchk=0;negsum=0;
negchk=xa1<0;
negsum=sum(negchk);
if negsum >0, negchkgen=negchkgen+1;end
x1a2=[(refdatab(idx,:)*senb)\bdataX5b(idx,:)];%est.conc.run2,unnrmlzd
x2a2=sum(x1a2);
xa2=x1a2./x2a2;%est.conc.run2,normalized
negchk=0;negsum=0;
negchk=xa2<0;
negsum=sum(negchk);
if negsum >0, negchkgen=negchkgen+1;end
x1a3=[(refdatac(idx,:)*senc)\bdataX5c(idx,:)];%est.conc.run3
x2a3=sum(x1a3);
xa3=x1a3./x2a3;%est.conc.run3,normalized
negchk=0;negsum=0;
negchk=xa3<0;
```

```
109 negsum=sum(negchk);
110 if negsum >0, negchkgen=negchkgen+1;end
111 xall=[xa1 xa2 xa3]; xallrot=rot90(xall); %Collect the 3 norm.concs
112 xa=mean(xallrot);%Find average
113 xdeva=xa - condatx5;%Deviations between ave.norm.est.conc. and condatx1
114 xdevsumd=sum((xdeva).^2); %Sum squared deviations
115 xdevsumtot=xdevsuma+xdevsumb+xdevsumc+xdevsumd;
116 negchk=0;negsum=0;
117 negchk=sena<0;%Check sensitivities for negatives
118 negsum=sum(negchk);negsum2=sum(negsum);
119 if negsum2 >0, negchkgen=negchkgen+1;end
120 negchk=0;negsum=0;
121 negchk=senb<0;%Check sensitivities for negatives
122 negsum=sum(negchk);negsum2=sum(negsum);
123 if negsum2 >0, negchkgen=negchkgen+1;end
124 negchk=0;negsum=0;
125 negchk=senc<0;%Check sensitivities for negatives
126 negsum=sum(negchk);negsum2=sum(negsum);
127 if negsum2 >0, negchkgen=negchkgen+1;end
128 if negchkgen > 0, xdevsumtot = -xdevsumtot;end%If negs anywhere set
129 index=index+1; %Increment pmtzn counter
130 if xdevsumtot > 0, fprintf(fid1,'%.9f %.0f\n',xdevsumtot,index');, else  131
fprintf(fid2,'%.9f%.0f\n',xdevsumtot,index');end
132 end
133 end
134 end
```

Lines 1 - 4 and 132 - 134 correspond to letter "A" and "Q" in the brute force flow chart diagramed in **Figure A.2** and set-up a series of nested loops that iterate through all combinations of 3 m/z ratios. Line 5 resets a global variable to zero which is used to check for negative sensitivity factors or estimated component concentrations in any of the 4 samples or 3 replicates. Lines 6 through 14 serve to calculate the sensitivity factor for all 3 runs of data (step "B" in the flow chart). Note that for each run of data there is a different set of reference and mixture spectra. Line 15 calculates the three un-normalized estimated concentrations for mixture 1 (step "C"). Lines 16 and 17 normalize the estimated component concentrations calculated in line 15 (step "D"). This is performed by summing the estimated component concentrations in line 15 and dividing each by the sum. Line 18 uses two variables to check for negative estimated component concentrations (step "E"). Line 19 checks for negative estimated component concentrations by creating a vector of 1's and 0's where 1's are negative estimated component concentrations (step "F"). Line 20 sums the vector created in line 19. If any

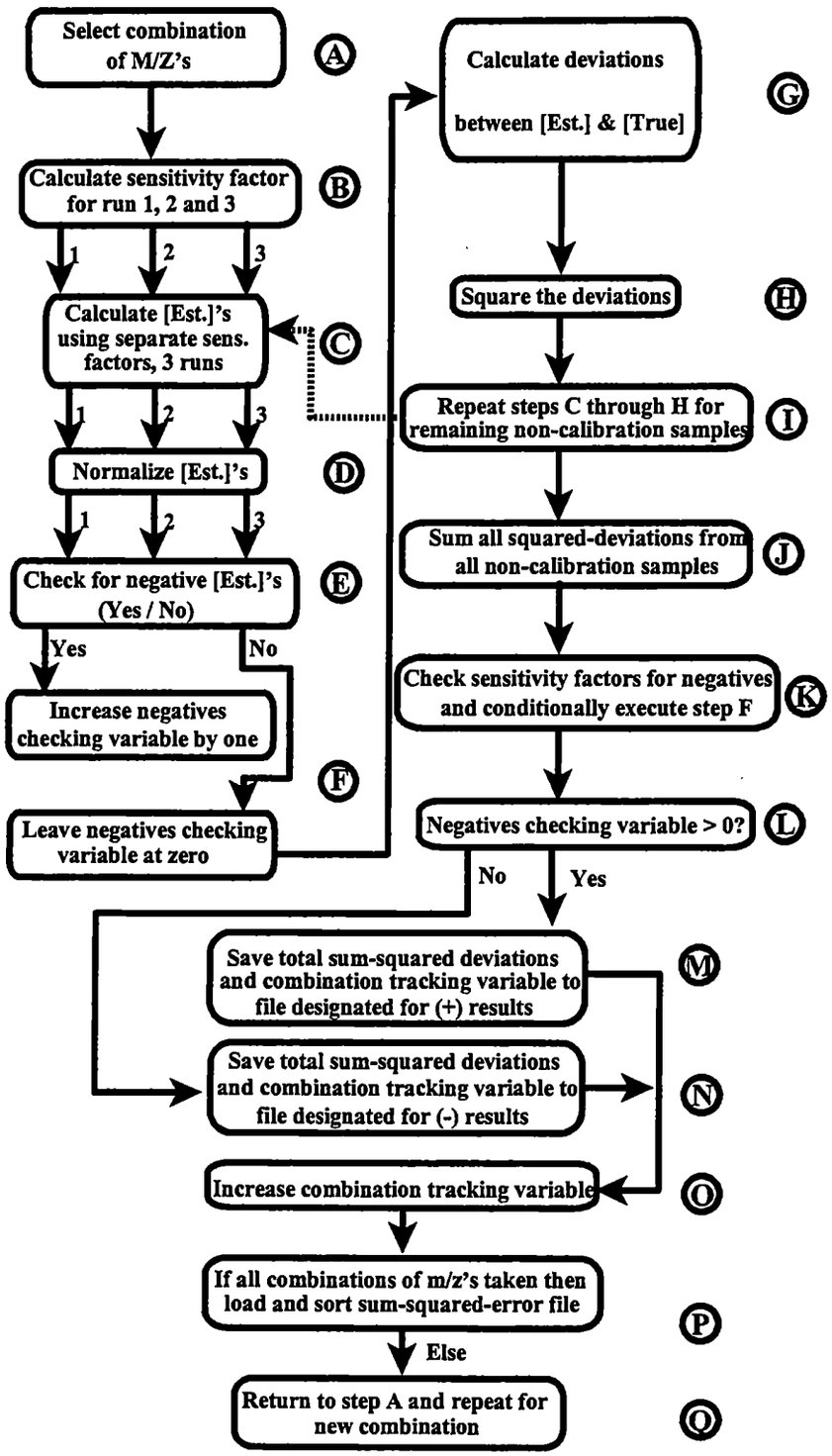


Figure A.2: Flow chart for brute force parameterization method.

negatives exist in this vector then the sum will be greater than zero. Line 21 analyzes the sum created in line 20, if any negatives exist then the global negatives checking variable is incremented by 1. Lines 22 - 28 repeat the calculations described for lines 15 - 21 using the second run of mixture spectrum number 1, note that the different sensitivities for each particular replicate are used in the corresponding replicates for the sample. In other words, the same calibration data from "day 1" is utilized for "day 1" sample data. Lines 29 - 35 repeat the calculations described for lines 15 - 21 using the third run of mixture spectrum number 1. Line 36 collects all estimated component concentrations into one variable and rotates the resulting matrix such that average estimated component concentrations can be calculated among the three replicates. Line 37 calculates the average estimated component concentrations among the three replicates (step "G"). Line 38 calculates the deviations between the average estimated component concentrations and the true concentrations (on a component by component basis) for mixture sample number 1 (step "H"). Line 39 squares the deviations and sums them for mixture sample number 1 (step "I"). Lines 40 - 64 repeat the calculations described for lines 15 - 39 using mixture sample number 3 (recall mixture number 2 was used in the sensitivity (calibration) calculations (begin step "J"). Lines 65 - 89 repeat the calculations described for lines 15 - 39 using mixture sample number 4. Lines 90 - 114 repeat the calculations described for lines 15 - 39 using mixture sample number 5. Line 115 sums the sum-squared deviations calculated among all 4 mixture samples and all 3 replicates, this sum represents the performance of the particular combination of ion signals (the variable "idx" in line 4) for all samples, all components and all replicates (step "K"). Lines 116 - 127 examine the sensitivities calculated in each replicate and identify negative sensitivities in a manner

analogous to the process of identifying negative estimated component concentrations (step "L"). Line 128 analyzes the global negatives checking variable, if the value is greater than zero then there was a negative estimated concentration or sensitivity calculated somewhere for the associated combination of signals (parameterization) and the variable created in line 115 is thus assigned a negative value (step "M"). Line 129 serves as a variable for which the combination of ion signals utilized in all of the calculations can be identified. This variable is increased for each new combination (parameterization) (step "P"). Line 130 - 131 saves the total-sum-squared deviations and the parameterization index (the variable "index" from line 129) to one of two files. If the total-sum-squared deviations is tagged negative then the result is saved to a separate file designated for negative results (step "O"). Conversely, if the total-sum-squared deviations result is positive, the value and the parameterization index is saved to a file designated for positive results (step "N"). Step "Q" is initiated upon execution of all possible combinations (when all the loops have been completed). In step "Q" the large files that were formed in the previously mentioned program are loaded, combined and sorted according to the total sum-squared-error. The following illustrates a sample of the Matlab program that performs the task of loading, combining and sorting the output files that are generated upon execution of the brute force program.

```
1  clear
2  %Clear all variables:
3  a=0;
4  %Set precision:
5  format long e
6  %Preallocate variables
7  loop2=0;loop3=0;loop4=0;loop5=0;loop6=0;loop7=0;loop8=0;loop9=0; ...
8  loop10=0;loop11=0;loop12=0;loop13=0;loop14=0;loop15=0;loop16=0; ...
9  loop17=0;loop18=0;loop19=0;loop20=0;loop21=0;
10 fid1 = fopen('/usr/local/scrtch/haddix/cmbd1.out','wt');
11 fid2 = fopen('/usr/local/scrtch/haddix/cmbd2.out','wt');
12 fid3 = fopen('/usr/local/scrtch/haddix/cmbd3.out','wt');
13 load('/usr/local/scrtch/haddix/loop2.out');
14 fprintf(fid1,'%.9f %.0f\n',loop2');
15 loop2=0;
```

```
load('/usr/local/scrtch/haddix/loop3.out');
fprintf(fid1,'%.9f %.0f\n',loop3');
loop3=0;
load('/usr/local/scrtch/haddix/loop4.out');
fprintf(fid1,'%.9f %.0f\n',loop4');
loop4=0;
load('/usr/local/scrtch/haddix/loop5.out');
fprintf(fid1,'%.9f %.0f\n',loop5');
loop5=0;
load('/usr/local/scrtch/haddix/loop6.out');
fprintf(fid1,'%.9f %.0f\n',loop6');
loop6=0;
load('/usr/local/scrtch/haddix/loop7.out');
fprintf(fid1,'%.9f %.0f\n',loop7');
loop7=0;
load('/usr/local/scrtch/haddix/loop8.out');
fprintf(fid1,'%.9f %.0f\n',loop8');
loop8=0;
load('/usr/local/scrtch/haddix/loop9.out');
fprintf(fid1,'%.9f %.0f\n',loop9');
loop9=0;
load('/usr/local/scrtch/haddix/loop10.out');
fprintf(fid1,'%.9f %.0f\n',loop10');
loop10=0;
load('/usr/local/scrtch/haddix/loop11.out');
fprintf(fid2,'%.9f %.0f\n',loop11');
loop11=0;
load('/usr/local/scrtch/haddix/loop12.out');
fprintf(fid2,'%.9f %.0f\n',loop12');
loop12=0;
load('/usr/local/scrtch/haddix/loop13.out');
fprintf(fid3,'%.9f %.0f\n',loop13');
loop13=0;
load('/usr/local/scrtch/haddix/loop14.out');
fprintf(fid3,'%.9f %.0f\n',loop14');
loop14=0;
load('/usr/local/scrtch/haddix/loop15.out');
fprintf(fid3,'%.9f %.0f\n',loop15');
loop15=0;
load('/usr/local/scrtch/haddix/loop16.out');
fprintf(fid3,'%.9f %.0f\n',loop16');
loop16=0;
load('/usr/local/scrtch/haddix/loop17.out');
fprintf(fid3,'%.9f %.0f\n',loop17');
loop17=0;
load('/usr/local/scrtch/haddix/loop18.out');
fprintf(fid3,'%.9f %.0f\n',loop18');
loop18=0;
load('/usr/local/scrtch/haddix/loop19.out');
fprintf(fid3,'%.9f %.0f\n',loop19');
loop19=0;
load('/usr/local/scrtch/haddix/loop20.out');
fprintf(fid3,'%.9f %.0f\n',loop20');
loop20=0;
load('/usr/local/scrtch/haddix/loop21.out');
fprintf(fid3,'%.9f %.0f\n',loop21');
loop21=0;
fclose(fid1);
fclose(fid2);
fclose(fid3);
fid4 = fopen('/usr/local/scrtch/haddix/cmbsrtd1.out','wt');
fid5 = fopen('/usr/local/scrtch/haddix/cmbsrtd2.out','wt');
fid6 = fopen('/usr/local/scrtch/haddix/cmbsrtd3.out','wt');
load('/usr/local/scrtch/haddix/cmbd1.out');
a=sortrows(cmbd1,1);
fprintf(fid4,'%.9f %.0f\n',a');
a=0;
load('/usr/local/scrtch/haddix/cmbd2.out');
a=sortrows(cmbd2,1);
fprintf(fid5,'%.9f %.0f\n',a');
a=0;
load('/usr/local/scrtch/haddix/cmbd3.out');
a=sortrows(cmbd3,1);
fprintf(fid6,'%.9f %.0f\n',a');
a=0;
fclose(fid4);
```

```
92 fclose(fid5);
93 fclose(fid6);
```

Lines 1 - 9 clear any variables that might be left over in the Matlab memory, set the precision of the calculations to scientific notation and preallocates the variables " loop3" through " loop 21". Line 10 - 12 open files that will be used for the output of the sorted and combined results of loops 3 through 21. Three files are created that contain all of the brute force data. The entire brute force generated dataset combined often consumes over 150 megabytes of disk space and it is frequently the case that the University UNIX computer system is not able to load the entire file. Hence, a portion of this program creates 3 separate files containing approximately 50 megabytes each such that the UNIX system is successful in handling the files. In order to do this, line 13 loads the brute force generated file "loop 2.out" while lines 14, 17, 20, 23, 26, 29, 32, 35 and 38 save the contents to the file which will contain the loops 2 through loop 10. Line 15, 18, 21, 24, 27, 30, 33 and 39 reset the variable "loop2, 3, 4, ..." back to zero to conserve computer memory. Lines 40 - 72 repeat similar actions mentioned previously for the remaining loops. Lines 73 - 75 close the files generated in lines 10 - 12. Lines 76 - 78 open new files that will contain the sorted, combined output files. Lines 79, 83, 87 load the combined results obtained from the execution of lines 1 - 75. Lines 80, 84, 88 utilize a subroutine called "sortrows.m" which sorts the loaded files "cmbd1.out, cmbd2.out and cmbd3.out" according to the total-sum-squared error. Lines 81, 85, 89 save the sorted results to the files generated in lines 76 - 78. Lines 82, 86, 90 conserve UNIX computer memory by clearing the variable "a". Lines 91 - 93 finish the process of loading, combining and sorting the brute force output by closing the file generated in line 76 - 78 and ending the program.

APPENDIX 3

MATLAB PROGRAM FOR EXECUTING THE EMPIRICAL ALGORITHM

METHOD OF PARAMETERIZATION

This appendix lists the Matlab program used to execute the Empirical Algorithm.

```
%Empirical algorithm based upon differences between reference spectra
clear
format long e
mzs=[m/z's...]; %order of m/z's placed in matrices
comps=2; %Enter the number of components here
M=30; %Enter the number of m/z's in matrices
refdataa=[data...];%Run 1
refdatab=[data...];%Run 2
refdatac=[data...];%Run 3
bdataa=[data...]; %50/50 calibration mixture, Run 1
bdatab=[data...]; %50/50 calibration mixture, Run 2
bdatac=[data...]; %50/50 calibration mixture, Run 3
condat=[]; %"True" mole fractions of Calibration Sample
ee=refdataa(:,:)\bdataa(:,:); % Begin calc. of sens. factor for run 1
ff=ee./condat;
sena=diag(ff); % sensitivity factor for run 1
ee=refdatab(:,:)\bdatab(:,:); % Begin calc. of sens. factor for run 2
ff=ee./condat;
senb=diag(ff); % sensitivity factor for run 2
ee=refdatac(:,:)\bdatac(:,:); % Begin calc. of sens. factor for run 3
ff=ee./condat;
senc=diag(ff); % sensitivity factor for run 3
data1=(refdataa(:,:))*sena;
data2=(refdatab(:,:))*senb;
data3=(refdatac(:,:))*senc;
ave1=rot90(data1(:,1));
ave2=rot90(data2(:,1));
ave3=rot90(data3(:,1));
compa=[ave1;ave2;ave3];
compavg=mean(compa);
comp1=rot90(compavg,-1);
ave1=rot90(data1(:,2));
ave2=rot90(data2(:,2));
ave3=rot90(data3(:,2));
compa=[ave1;ave2;ave3];
compavg=mean(compa);
comp2=rot90(compavg,-1);
datavg=[comp1 comp2];%avged data in column format like refdata
printout=[mzs datavg];
holdintdiff=zeros(M,comps);
count=0;
for mass=1:M; %Loop through all m/z's
a=datavg(mass,1);
b=datavg(mass,2);
count=count+1;
intdiff=(a - b)^2;%Square of intensity difference at selected m/z
holdintdiff(mass,count)=intdiff; %Hold score in 1st column of variable
end
rotate=rot90(holdintdiff);%adjust data for summing
sumit=sum(rotate); %Sum intensity differences to give score for each m/z
unrotate=rot90(sumit,-1); %Place scores back in column orientation
massnscores=[mzs unrotate]; %Insert m/z labels, aligned w/ scores
ranked=sortrows(massnscores,2); %Sort, smallest on top
flipped=flipud(ranked); %Flip ranked scores; Largest on top, w/ m/z labels
norm1=max(unrotate);%Find maximum score for normalization
normd=(flipped(:,2) ./ norm1)*100; %Normalized final score to 100
final=[flipped(:,1) normd]; %Final EA score
```

APPENDIX 4

MATLAB PROGRAM FOR EXECUTING THE DRAHOS-VEKEY EQUATION FOR CALCULATION OF SPECTRAL SIMILARITY

The following Matlab program was used to perform the Drahos-Vekey similarity index analysis of sensitized reference spectra.

```
1  clear
2  format long e
3  refdataa=[1;2;3];% Sensitized Reference spectra Run 1
4  refdatab=[1;2;3];% Sensitized Reference spectra Run 2
5  refdatac=[1;2;3];% Sensitized Reference spectra Run 3
6  %Run 1:
7  deltacomp=(refdataa(:,1) - refdataa(:,2));%Numerator of DVSimilarity index
8  deltacompsqrd=deltacomp.^2;%Square to find abs. val. of diffs in spectra
9  deltasqrt=sqrt(deltacompsqrd);%Absolute value of differences
10 numeratorsum=sum(deltasqrt);%Sum of numerator
11 numerator=100*numeratorsum;
12 denominator=sum((refdataa(:,1) + refdataa(:,2)));%Denominator
13 SIlast=numerator / denominator;
14 S1=100 - SIlast; %Final DHSi ("S") in reference, run 1
15 %Run 2:
16 deltacomp=(refdatab(:,1) - refdatab(:,2));%Numerator of DVSimilarity index
17 deltacompsqrd=deltacomp.^2;%Square to find abs. val. of diffs in spectra
18 deltasqrt=sqrt(deltacompsqrd);%Absolute value of differences
19 numeratorsum=sum(deltasqrt);%Sum of numerator
20 numerator=100*numeratorsum;
21 denominator=sum((refdatab(:,1) + refdatab(:,2)));%Denominator
22 SIlast=numerator / denominator;
23 S2=100 - SIlast; %Final DHSi ("S") in reference, run 2
24 %Run 3:
25 deltacomp=(refdatac(:,1) - refdatac(:,2));%Numerator of DVSimilarity index
26 deltacompsqrd=deltacomp.^2;%Square to find abs. val. of diffs in spectra
27 deltasqrt=sqrt(deltacompsqrd);%Absolute value of differences
28 numeratorsum=sum(deltasqrt);%Sum of numerator
29 numerator=100*numeratorsum;
30 denominator=sum((refdatac(:,1) + refdatac(:,2)));%Denominator
31 SIlast=numerator / denominator;
32 S3=100 - SIlast; %Final DHSi ("S") in reference, run 3
```

Lines 1 - 5 clear any left-over variables in the Matlab memory, set the precision of the calculations and contain the variables "refdata" of the sensitized reference spectra. Line 6 is a comment for lines 7 - 14 which calculate the SI for replicate number one of the data. Line 7 calculates the difference between each m/z signal intensity in the sensitized reference spectra. Line 8 calculates the square of the differences calculated in Line 7. Line 9 calculates the square-root of the values calculated in Line 8 thus effectively Lines 8 and 9 calculate the absolute values of the differences calculated in Line 7. Line 10

calculates the sum of the values calculated in line 9 while line 11 multiplies the resulting sum by 100 which is used as the numerator in **Equation 2.5.5.a**. Line 12 calculates the denominator of **Equation 2.5.5.a** by summing the intensities of the sensitized reference spectra. Line 13 calculates the quotient of **Equation 2.5.5.a** by dividing the result of line 11 by the result from line 12. The final similarity index value is calculated in line 14 by subtracting the ratio calculated in line 13 from 100. Lines 15 - 23 and 25 - 32 repeat the calculations performed in lines 7 - 14 utilizing spectra from replicates 2 and 3.

APPENDIX 5

MATLAB PROGRAM FOR CALCULATING THE CORRELATION COEFFICIENT (R^2) TO THE LINE Y = X

The following Matlab program was used to calculate the correlation coefficient with respect to the Y = X line.

```
1  trumestsq=(alltruex - allconcy).^2;
2  avest=mean(allconcy);
3  estmavestsq=(allconcy - avest).^2;
4  sum1=sum(trumestsq);
5  sum2=sum(estmavestsq);
6  diff=sum2 - sum1;
7  rsqyx=diff/sum2;
```

The variable alltruex contains the mole fractions, determined gravimetrically, for each sample tested. The variable allconcy contains the estimated (calculated) mole fractions. Line 1 calculates the differences between the estimated (allconcy) and true (alltruex) mole fractions and squares the resulting difference. Line 2 calculates the average of the estimated mole fractions. Line 3 calculates the differences between the estimated mole fractions and the average calculated in Line 2 and then squares the differences. Line 4 sums the squared differences calculated in line 1. Line 5 sums the squared differences calculated in line 3. Line 6 calculates the difference between the two sums calculated in lines 4 and 5. Finally, the correlation coefficient with respect to the line Y = X is calculated in line 7 by taking the ratio of line 6 and line 5.

VITA

Martin L. Haddix, son of Leo L. and Coralee A. Haddix, was born in Kearney, Nebraska on November 12, 1968. He graduated from Kearney High School in May of 1987. In the fall of 1987 he began studies at The University of Nebraska at Lincoln. He received the Sheldon-Coleman Charitable and Educational Trust Scholarship for the academic year 1987/1988 and again for the 1988/1989 academic year. He graduated with a Bachelors of Science degree in chemistry in May of 1992. In the fall of 1992 he began graduate studies at the University of Tennessee at Knoxville and began to conduct research under the direction of professor Kelsey D. Cook.

As a graduate student, Martin received funding from graduate teaching assistantships, a Science Alliance Center of Excellence Fellowship, an American Society for Mass Spectrometry Graduate Student Award and an East Tennessee Mass Spectrometry Discussion Group Award. In April of 1995 he participated in the Union Carbide Kenan Analytical Award Symposium. In November of 1996 he received an award from the University of Tennessee-Knoxville University Programs Council. He also worked as a research assistant for Professor Cook which was supported by the University of Tennessee-Knoxville's Measurement and Control Engineering Center.

Martin is a member of the American Chemical Society, the American Society for Mass Spectrometry, The East Tennessee Mass Spectrometry Discussion Group, as well as an academic member of The University of Tennessee's Measurement and Control Engineering Center.

Martin has co-authored the following publications and presentations:

1. K.D. Cook, K.H. Bennett, M.L. Haddix, "On-line Mass Spectrometry: A Faster Route to Process Monitoring and Control." *Indust. Eng. Chem. Res.*, **1999**, 38(4), 1192-1204.

2. K.D. Cook, M.L. Haddix, G.L. Seebach, and K.H. Bennett, "Parameterization for Mass Spectrometric Monitoring of Multi-Component Streams." *J. Process Anal. Chem.*, **1997**, 3 (1,2), 53-62.

3. K. D. Cook, M.L. Haddix,G.L. Seebach, "Isomer Resolution Using Process Mass Spectrometry." *J. Process Anal.Chem.*, **1996**, 2, 2, 219-223.

4. K.D. Cook, M.L. Haddix, G.L. Seebach, "Impact of Parameterization on the Least Squares Deconvolution of Mass Spectral Data for Simultaneous Determination of Isomeric Alkanes." *Adv. Instrum. Control*, **1995**, 50 (Part 3), 1169-1178.

5. Cook, K. D.; Bennett, K. H.; Haddix, M. L.; Keator, E. A.; Seebach, G. L.; Falconer J. L. "Process Mass Spectrometric Monitoring of Isomeric Hydrocarbon Gases." *J. Process Anal. Chem.*, **1997**, *3* (3,4), 115-24.

6. M.L. Haddix, K.D. Cook, American Society of Mass Spectrometry (ASMS) poster presentation, "Spectral resolution of isomeric alkanes: A process mass spectrometry application." Atlanta, GA, June, 1995.

7. M.L. Haddix, Instrument Society of America International Conference, oral presentation, "Spectral resolution of isomeric alkanes: A process mass spectrometry application." New Orleans, LA, October, 1995.

8. M.L. Haddix, K.D. Cook, K.H. Bennett, ASMS poster presentation, "Spectral resolution of isomeric hydrocarbon mixtures: A process mass spectrometry application." Portland, OR, May, 1996.

9. M.L. Haddix, East Tennessee Mass Spectrometry Discussion Group Fellowship, oral presentation, "Process Mass Spectrometry." Knoxville, TN, May, 1997.

10. M.L. Haddix, K.D. Cook, ASMS poster presentation, "Optimum parameterization for deconvolution of hydrocarbon mixtures using process mass spectrometry." Palm Springs, CA, June, 1997.

11. M.L. Haddix, K.D. Cook, University of Tennessee-Knoxville Statistics Department, oral presentation, "Process Mass Spectrometry." Knoxville, TN, March, 1998.

12. M.L. Haddix, K.D. Cook, American Chemical Society Local Section Meeting, oral presentation, "Process Mass Spectrometry at UTK." Wilmington, DE, July, 1998.