INTEGRATING SAFETY ISSUES IN OPTIMIZING SOLVENT SELECTION AND PROCESS DESIGN

A Dissertation

by

SUHANI JITENDRA PATEL

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2010

Major Subject: Chemical Engineering

Integrating Safety Issues in Optimizing Solvent Selection and Process Design

INTEGRATING SAFETY ISSUES IN OPTIMIZING SOLVENT SELECTION AND PROCESS DESIGN

A Dissertation

by

SUHANI JITENDRA PATEL

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	M. Sam Mannan
Committee Members,	Mahmoud El-Halwagi
	Carl Laird
	Sergiy Butenko
Head of Department,	Michael Pishko

August 2010

Major Subject: Chemical Engineering

ABSTRACT

Integrating Safety Issues in Optimizing Solvent Selection

and Process Design. (August 2010)

Suhani Jitendra Patel, B.S., University of Mumbai

Chair of Advisory Committee: Dr. M. Sam Mannan

Incorporating consideration for safety issues while designing solvent processes has become crucial in light of the chemical process incidents involving solvents that have taken place in recent years. The implementation of inherently safer design concepts is considered beneficial to avoid hazards during early stages of design. The application of existing process design and modeling techniques that aid the concepts of 'substitution', 'intensification' and 'attenuation' has been shown in this work. For 'substitution', computer aided molecular design (CAMD) technique has been applied to select inherently safer solvents for a solvent operation. For 'intensification' and 'attenuation', consequence models and regulatory guidance from EPA RMP have been integrated into process simulation. Combining existing techniques provides a design team with a higher level of information to make decisions based on process safety.

CAMD is a methodology used for designing compounds with desired target properties. An important aspect of this methodology concerns the prediction of properties given the structure of the molecule. This work also investigates the applicability of Quantitative Structure Property Relationship (QSPR) and topological

indices to CAMD. The evaluation was based on models developed to predict flash point properties of different classes of solvents. Multiple linear regression and neural network analysis were used to develop QSPR models, but there are certain limitations associated with using QSPR in CAMD which have been discussed and need further work.

Practical application of molecular design and process design techniques have been demonstrated in a case study on liquid-liquid extraction of acetic acid-water mixture. Suitable inherently safer solvents were identified using ICAS-ProCAMD, and consequence models were integrated into Aspen Plus simulator using a calculator sheet. Upon integrating flammable and toxic hazard modeling, solvents such as 5-nonanone, 2-nonanone and 5-methyl-2-hexanone provide inherently safer options, while conventionally-used solvent, ethyl acetate, provides higher degree of separation capability. A conclusive decision regarding feasible solvents and operating conditions would depend on design requirements, regulatory guidance, and safety criteria specified for the process.

Inherent safety has always been an important consideration to be implemented during early design steps, and this research presents a methodology to incorporate the principles and obtain inherently safer alternatives.

DEDICATION

This dissertation and research is dedicated to my family for always being there for me and for spoiling me beyond measure;

my grandmother, Sushila Jashbhai Patel

my parents, Jitendra Patel and Malini Patel

and my brother, Tanmay Patel

ACKNOWLEDGEMENTS

While writing this part of my dissertation, mixed emotions and memories run through my head. Attending graduate school at Texas A&M University has been a wonderful experience of my life, and the credit for that goes to many individuals who have made this journey exciting and memorable.

There is a lot that I have learned from my advisor, Dr. M. Sam Mannan. He has continually shared his wisdom and knowledge for both academic and post-academic life. I am in awe of Dr. Mannan for being able to maneuver and inspire students from their confused, initial stages of research and lead them towards becoming professional individuals. He has been instrumental in converting my mindset to a 'can-do' attitude. I am also grateful to him for managing the Mary Kay O'Connor Process Safety Center with so much vigor. The center is definitely a significant contribution towards process safety.

I would also like to thank my committee members, Dr. Mahmoud El-Halwagi, Dr. Carl Laird and D. Sergiy Butenko, for providing guidance and valuable insight on my research. From my committee members, I have been able to learn the value of being kind and humble in all social interactions, as well as adding humor and gusto to research or other work.

I also want to thank my friends for countless days and nights of fun and entertainment. I have learned so much from their experiences as well. Special appreciation goes out to Katherine, Sara, Anisa, and Mahdiyati for adding bouts of

silliness and laughter in an otherwise serious environment. I am also grateful to all members of the MKOPSC, the old and the new, for always being ready to help me on many occasions, and celebrating many happy moments together.

I want to extend my gratitude to Valerie Green, Donna Startz and Mary Cass for arranging the center's activities and thereby, giving the students' the best experience possible. Thanks also go to my friends, colleagues, the department faculty and staff for making my time at Texas A&M University a great experience.

There is not enough I can put in writing to express appreciation for my family's contribution towards shaping my personality. They have instilled in me the significance of education and individuality, given me many comforts, and helped me in every step I took towards achieving my goals. I thank my mother for the discipline, my father for the humor, and my brother for being my pillar of strength.

Asato Maa, Sadgamaya

Tamaso Maa, Jyotirgamaya

Lead me from the unreal to the real

Lead me from darkness to light

- Brihadaranyaka Upanishad: I.iii.28

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Page

Page

Page

LIST OF TABLES

Page

Page

1. INTRODUCTION[*]

1.1 Solvent hazards

Solvents are widely used in chemical industries in several different processes and millions of industrial workers are exposed to solvents on a daily basis. Solvents are used in industries, such as construction, maritime, retail, and general industry. Certain characteristics and physical/chemical properties of solvents make them useful, while other properties make them extremely hazardous. Solvents are generally organic chemicals that tend to be highly flammable and toxic. Issues such as human and ecological toxicity, process safety hazards and waste/pollution management are a concern for solvent processes.

The main health hazards typically associated with organic solvent exposure include nervous system damage, kidney and liver damage, adverse reproductive effects, skin lesions, and cancer. Exposures to very high concentrations of certain solvents may even cause death. A review of OSHA records has shown that eight worker deaths between 1975 and 1992 occurred from over-exposure to a single solvent, trichloroethylene.[1] Some commonly used solvents and the various health effects associated with them have been shown in Table 1.

This dissertation follows the style of *Industrial and Engineering Chemistry Research*.

[*] Part of this section is reprinted with permission from "Inherently safer design of solvent processes at the conceptual stage: Practical application for substitution" by S. Patel, D. Ng, M. S. Mannan, 2010. *Journal of Loss Prevention in the Process Industries*, doi:10.1016/j.jlp.2010.03.002. Copyright 2010 by Elsevier Ltd.

Table 1. Health effects of commonly used solvents

Health effect	Examples of solvents
Damage to nervous system	n-hexane, perchloroethylene, n-butyl mercaptan
Damage to liver or kidney	Toluene, carbon tetrachloride, 1,1,2,2 – tetrachloroethane, chloroform
Reproductive hazards	2-methoxyethanol, 2-ethoxyethanol, methyl chloride
Suspected or known carcinogens	Carbon tetrachloride, trichloroethylene, 1,1,2,2-tetrachloroethane, perchloroethylene, perchloroethylene, methylene chloride

The main physical hazard of solvents is associated with its flammability. Solvents are organic chemicals with fairly low flash point, explosive (or flammability) limits, and autoignition temperature. These properties are indicative of the tendency of solvents to cause fires or explosions. Flammability related properties for some commonly used solvents have been shown in Table 2.

Table 2. Flammability related properties of commonly used solvents

Solvent	Flash point (K)	Flammability limits		Autoignition temperature (K)
		Lower (vol %)	Upper (vol %)	
Formaldehyde	220	7	73	697.15
Ethyl ether	228.15	1.9	48	433.15
Acetaldehyde	235	4	60	403.15
Carbon disulphide	243.15	1.3	50	363.15
Hexane	251.5	1.1	7.7	498
Acetone	255	2.6	12.8	738.15
Benzene	262	1.4	7.1	835
Methyl ethyl ketone	267	1.8	10	789
Ethyl acetate	269	2.2	11.4	700
Toluene	278	1.2	7.1	809
Methanol	284	7.3	36	737
1,4-dioxane	285	2	22	453.15
1,2 dichloroethane	286	6.2	16	686
Ethanol	286	4.3	19	696
Xylene	298.15	1.1	7	802
N,n,-dimethylformamide	331	2.2	15.2	718.15
Ethylene glycol	384	3.2	21.6	673.15

Thus, hazards and risk associated with solvent processes need to be assessed and mitigated at early stages of process design. Chemical industries reduce risk by placing emphasis on proper storage and handling procedures, operator training, proper ventilation systems, and minimization of ignition sources. Presently, the general areas of effort for diminishing solvent hazards are: hazard evaluation and monitoring, following

exposure related standards, application of control measures, engineering controls, administrative controls, personal protective equipment, storage, signs, labeling, education and training.[2] Despite abundant precautions and safety systems in chemical plants, many incidents and accidents have taken place in recent years. The incidents at CAI/Arnel facility in Massachusetts (November 2006),[3] and the Barton Solvents facilities in Kansas and Iowa (July 2007 & October 2007),[4, 5] involving fire and explosions, were intensified because of hazardous solvents being used at the facilities.

The CAI, Inc. and Arnel Company, Inc. incident on November 22, 2006 involved a confined vapor cloud explosion followed by fire that burned for 17 hours. The incident resulted in 10 injuries, destruction of 24 houses and 6 businesses, and evacuation of more than 300 residents. Such large scale of damage can be mainly attributed to the type and amount of fuel (i.e. solvents) being stored or utilized at the facility. The solvent mixture being used at the facility consisted of heptane, isopropyl alcohol, and normal propyl alcohol, and around 2000 gallons of the mixture participated in the explosion. One of the recommendations made by the U.S. Chemical Safety Board was inadequacy in ensuring a safe design for the flammable liquids processes.

Similarly, incidents at two facilities of Barton Solvents were investigated by the U.S. Chemical Safety Board. The incident at the Kansas facility experienced a fire and explosion incident on July 17, 2007. The consequences were 12 injuries, evacuation of 6000 residents, and complete destruction of the tank farm. The tanks stored the solvent, Varnish Makers' and Painters (VM&P) naphtha which is a highly flammable liquid. Although the main recommendation was about gaining better understanding of static

electricity ignition hazards, but using an alternative less hazardous solvent could have also lessened the consequences. Another incident occurred at the Barton Solvents facility in Iowa on October 29, 2007. The fire and series of explosions resulted in 2 injuries and evacuation of nearby businesses. This incident involved another commonly used solvent, ethyl acetate, which was being filled into a 300-gallon portable steel tank. Issues related to ethyl acetate application in solvent processes have been discussed in the case study in Section 6.

Apart from fire and explosion incidents, there have also been innumerable reports of workers being exposed to solvent vapors with toxic and adverse health effects. Such incidents result in loss of property and, at times, they result in loss of human lives as well. Thus, it is imperative to consider safety in solvent processes not only during the operating phase but also during the conceptual phase of process design.

1.2 Inherently safer design

Integrating safety in solvent processes during the design stage can be achieved by exploring Inherently Safer Design (ISD) concepts.[6] Inherently safer design (ISD) of chemical processes strives to achieve a higher level of safety by placing emphasis on eliminating or avoiding the hazards from the manufacturing process rather than relying on controlling the hazards.[7] The inherent safety of a process essentially lies in the fundamental characteristics of the materials, operations and conditions of the process. These are the characteristics which can be considered as inseparable from or inherent to

the process.[8] ISD remains a fairly undeveloped area of study compared to its counterpart in process safety: Quantitative Risk Assessment (QRA). QRA provides a detailed framework for estimating risk in terms of the frequency of occurrence of the hazardous event and severity of consequence associated with it. Having assumed the existence of a core process design, the QRA framework guides engineers in a step wise process to identify failure scenarios, estimate the consequences and likelihood, and thereby estimate the risk. The next step in the framework determines the tolerability of the estimated risk based on regulatory or company based risk criteria. In case of intolerable risk, it is required to apply proper risk reduction measures via several design alternatives.[9] Doing so can be a concern when the design of the process is extensively developed and agreed upon by an engineering team; and implementing changes to the design can become cumbersome. Changing the design to enhance safety in terms of general industry practices, internal company standards, and external codes/standards/regulations would require further evaluations for technical feasibility, cost benefit analysis, and evaluation of additional new hazards. Such an iterative approach for evaluation of risk (Figure 1) can become tedious and often overlooked or compromised. Thus, there is a need to incorporate safety during the design phase of the process by providing guidance for tolerable risk and safety parameters. This can be achieved by following the concepts of inherently safer design.

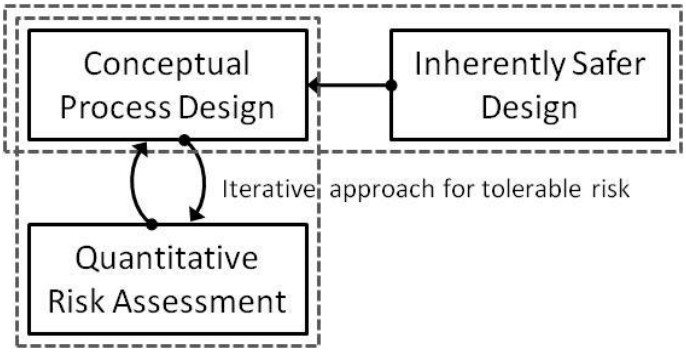


Figure 1. Comparing QRA and ISD for safer design approaches

There are four main concepts of inherently safer design:

1. Intensification/minimization

The amount of hazardous material involved in the process should be minimized as much as possible. The amount of hazardous material consisting of raw materials, or intermediates, could be reduced within unit operations, hazardous product storage, and pipelines. This is an extensive study of research because not only does intensification of processes benefit from safety perspective, but also from an economic perspective. Smaller sized equipments are being designed such that they are lower cost and easier control options. Some motivating factors and areas of research in the field of process intensification have been reviewed by Stankiewicz et al.[10]

2. Attenuation/moderation

Occasionally, changing process conditions (such as operating conditions or material phase) are able to render the substance/process less hazardous. For example,

when a hazardous material is diluted, then the partial pressure will be lower and the concentration of the material above the spill will be lowered.

3. Substitution

Inherent safety can be improved by considering replacing a more hazardous material or a more hazardous chemical synthesis route with less hazardous options. 'Green chemistry' technologies fall under this concept because they are able to eliminate the use of toxic and flammable materials. Some examples of substitution are water base latex paints that can replace organic solvent base paints, and aqueous systems replacing toxic chlorinated solvents.

4. Simplification

This concept refers to reduce unnecessary complexity in the plant and opportunities for human error. The most beneficial application of this concept would be to eliminate hazards by prohibiting hazardous operations to be conducted. An example is to remove unnecessary piece of piping which may become plugged, the valves may be shut or the gaskets may be degraded. In a recent CSB investigation, a massive fire was caused by a freeze related failure of some piping which had not been in service for about 15 years.[11] Removal of the pipe (or simplification of the process) could have avoided such a disaster.

Within the broad scope of process safety, strategies for risk management fall under four categories: inherent (eliminate or significantly reduce the hazard), passive (reduce the consequence or probability of an incident through devices not requiring activation or detection mechanisms), active (reduce the risk by means of devices that

detect and activate operations that interrupt the incident sequence of events), and procedural (reduce the risk by implementing procedures or human-process interaction).[8] If implemented properly, inherently safer design can achieve higher risk reduction benefits and prove to be more reliable and robust compared to active, passive and procedural safety systems.[12] More recently, ISD has also been considered as an inspiring philosophy which could be the basis of new trends in sustainability.[13] Thus, it is important to review and improve the techniques for implementing inherent safety.

1.3 Literature review

1.3.1 Hazard indices

In spite of having many advantages, previous work in the area of inherent safety implementation has been limited. Research has primarily focused on the development of inherently safer design indices,[14] integration of indices in process design and life cycle approach.[15, 16] Previous research in the area of inherent safety has been towards developing some sort of measure of inherent safety. Since there are no rules or methods to make a process inherently safer to date, many approaches could be followed. Previously, the main criteria for choosing between alternatives have been technical feasibility and economic viability, but recently safety and environmental concerns have become a substantial part of the decision making because of regulatory requirements. To aid such decision making, many inherently safer indices have been developed, such as the ones shown in Table 3.

Table 3. List of previously developed inherent safety indices[14]

Prototype index of inherent safety (Edwards and Lawrence)	Integrated inherent safety index (Khan and Amyotte)
Dow Fire & Explosion index and Mond index	Gentile et al. index
Heikkila and Hurme index	INSET toolkit
Palaniappan, et al. index	Gupta and Edwards index

Main parameters that are used in the indices are inventory, temperature, pressure, conversion, yield, toxicity, flammability, explosiveness, corrosiveness, side reactions, waste and co-products, reaction rate, heat of reaction, phase change etc. Some future work identified in the area has been to reduce the amount of information and data being used in the evaluation, and application to several practical examples.

Another avenue of inherent safety research is to comprehend the application of the ISD concepts at all phases in the lifecycle, such as process conception, thorough laboratory development, pilot plant, preliminary and detailed design, construction, operation and abandonment. At the stage of conceptual process research and development the following need to be assessed: selection of basic process technology, raw materials, intermediate products, by-products and chemical synthesis routes. At the process research and development stage the following need to be considered: selection of specific unit operations, types of reactors and other processing equipment, selection of operating conditions, recycle, and product purification. Similarly, other life cycle stages also have specific areas of evaluation.[8] The main limitations associated with implementing the inherent safety techniques is that the designer generally views

achieving process safety by ‘add-on’ safety features, and that ISD concepts are not part of the thinking of all engineers and chemists.

1.3.2 Limitations of previous work

Inherently safer design indices provide a simple approach to compare hazards posed by available alternatives, and aid in decision making. Nevertheless, the disadvantages associated with such index values are also widely accepted. Index calculations are generally used for comparison purposes only to choose a safer option among alternative process methods, and often require calibration against actual risk values.[14] Calculation of an index may also occasionally require extensive design data that may defeat the purpose of integrating safety at the design stage where limited information exists.[17] In addition, an index is an aggregate of many factors and different types of hazards, and does not provide enough information about individual effects of the design parameters on inherent safety of the process.[18] There is also lack of studies showing economic benefits of ISD, and lack of a tried and tested yet simple methodology for ISD implementation.[19] Methods to systematically integrate safety into process design have been approached by a few researchers by using optimization-based techniques,[20-22] and by developing integrated risk estimation tools.[23] Method developed to integrate Dow fire and explosion index (F&EI) into process optimization[20] is beneficial, but requires the user to develop separate functions for F&EI based on material inventory and operating pressure for each separate design using a sensitivity

analysis feature on an F&EI program. This is cumbersome because it would be repeated for all equipments and materials in the process. Also, it does not give a combined effect of the different process parameters on the safety, since it considers pressure and material inventory separately in the equations developed. Another work involves the development of an integrated risk estimation tool (iRET).[23] iRET comprised of the use of HYSYS for the process design simulation, and Microsoft Excel for developing the risk model. This configuration enabled the use of HYSYS features such as thermodynamic properties and readily available design templates. On the one hand, the integrated risk estimation tool is able to provide a means of including consequence estimation during conceptual design stage, but does not provide enough information about achieving inherently safer design of the process based on parameter modification. Thus, there is a need to incorporate safety considerations within the design procedure, and apply methods that provide a quantitative estimate of the hazard. Moreover, the inherently safer design concept of 'substitution' is not addressed by previously developed methods; thus available techniques in the area of 'substitution' need to be implemented suitably.

1.4 Dissertation outline

Having described within this section certain fundamental ideas behind this research, Section 2 explains the objectives of the research and a brief methodology. Methodology consists of two parts, molecular design and process design. Thereafter, Section 3 describes how an alternative technique can be applied to molecular design to improve its accuracy in some context, although certain limitations associated with the

newer approach have also been outlined. Section 4 describes the traditional approach for molecular design and its application to select inherently safer options. Also, database selection approach versus molecular design approach is compared. The next part of the research regarding integration of safety in process design is shown in Section 5. Sensitivity analysis of design parameters and their effect on process safety as well as avenues for optimizing results from process design are described in Section 5. Finally, techniques for molecular design and process design are implemented in a case study for acetic acid-water mixture extraction using solvent in Section 6. A brief summary of the conclusions and recommendations for future work can be found in Section 7.

2. PROBLEM STATEMENT

2.1 Objectives of research

Inherent safety can be incorporated in hazardous processes related to solvent operations by considering both molecular and process design as shown in Figure 2. Overall, the objective of simultaneously integrating inherent safety concepts into the design stage has been approached in this work, by means of novel application of existing techniques.

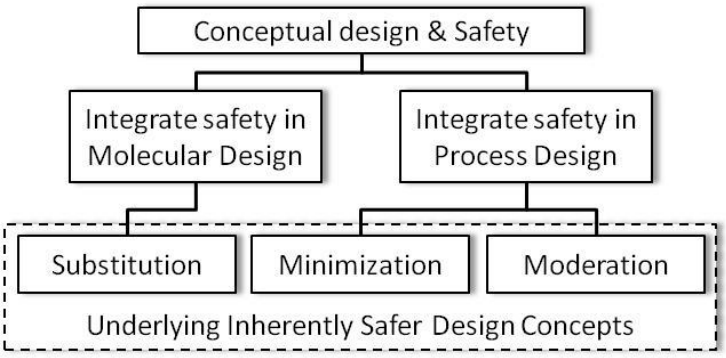


Figure 2. Conceptual design and inherent safety: 2 major approaches

Integrating safety in molecular design enables one to select inherently safer solvents based on their hazardous properties. Whereas, integrating safety in process design enables one to select inherently safer process parameters and conditions such that they satisfy requirements for both process efficiency and process safety. Methods that can be used to select safer chemicals, i.e. solvents, and to choose safer design parameters

along with guidelines and design constraints are described in Sections 3, 4, and 5. Application of this work, not only to solvent selection but also to other material and parameter selection will be extremely beneficial in early conceptual design for greater impact of inherent safety.

2.2 Methodology of research and integration of methods

Molecular design and process design can be achieved by using various tools and methods available. The application of these principles and techniques as well as safety considerations requires a sequential approach in order to integrate the various tools as described in Sections 3, 4, and 5, and this step wise approach has been shown in Figure 3. Firstly, in order to develop a database of inherently safer solvents, the ICAS software will be employed. Thereafter, suitable solvents will be carried forward to the process simulation step within which the hazardous scenario modeling guidelines of EPA will be incorporated by making use of existing features in the process simulator software.

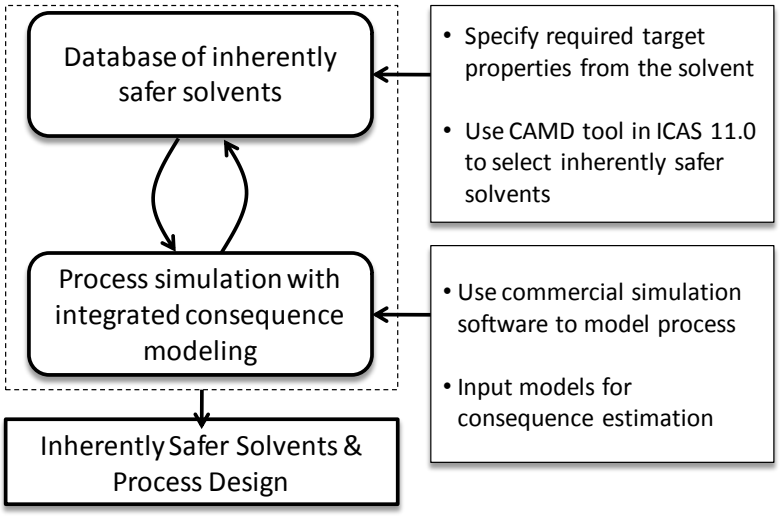


Figure 3. Sequential approach to Inherently Safer Design

In this work a sequential approach for solving the two problems has been applied, and further work needs to be done for obtaining a simultaneous solution for inherently safer design. It is to be noted that such a sequential approach may lead to a suboptimal design as stated by previous researchers who have worked towards obtaining a simultaneous solution for molecular design and process design.[24, 25] But the methods for obtaining a simultaneous solution are also limited in the number of properties that can be considered at a time for selecting solvent substitutes.

Apart from the methodology shown in Figure 3, certain considerations needed to be evaluated at the molecular design stage for developing better understanding of the methods available. These major considerations have been shown in Figure 4 and have been addressed in this research. In the area of molecular design, property prediction models needed further evaluation in order to obtain better accuracy (Section 3), as well

as conventional database approach and molecular design methodology needed to be compared (Section 4). Upon obtaining insight on issues related to molecular design, an appropriate method was selected.

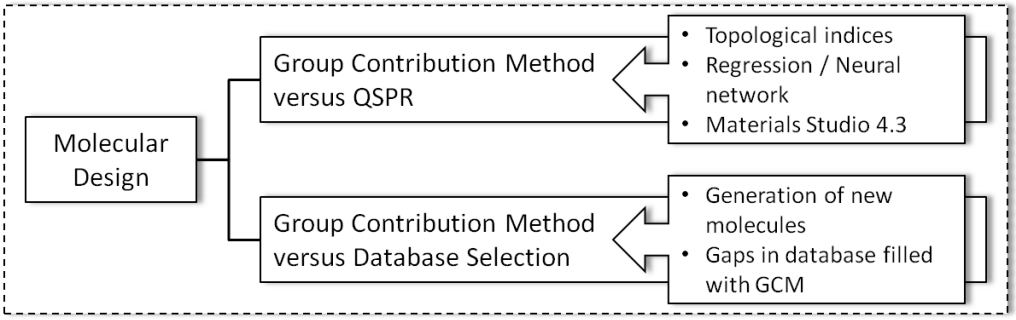


Figure 4. Major considerations for the molecular design stage

Overall, the proposed research will lead to the following principal outcomes:

1. Assessing the applicability of QSPR property prediction technique to molecular design
2. Using traditional molecular design method within ICAS to select inherently safer solvents
3. Developing process safety constraints based on credible release scenarios for flammable and toxic hazards
4. Assessing the influence of process parameters and conditions on safety measures and determining optimal design variables for a solvent process based on simultaneous safety considerations

3. MOLECULAR DESIGN: CAMD AND QSPR APPLICATION*

3.1 Introduction to Computer Aided Molecular Design (CAMD)

Molecular design is a methodology used to find a chemical or product that shows certain desirable behavior or that matches a desired set of target properties. When appropriate property models are not available, an empirical trial and error approach based on experimentation can be applied for molecular design. But in the case of available efficient property models, computer aided methodologies can be readily applied. As with most techniques used at the screening stage of material selection, results of molecular design may or may not work as replacements for specific applications, and experimental verification is strongly favored before implementation. Within Computer Aided Molecular Design (CAMD) the structures of molecules are represented using appropriate descriptors along with an algorithm that identifies the descriptors. Moreover the property evaluation models are also functions of the same descriptors. A general CAMD problem can be formulated as a mixed integer non-linear program (MINLP) which consists of property constraints and models.[26] A typical CAMD problem takes the form of an MINLP formulation as shown below in equation set (1).

---

* Part of this section is reprinted with permission from "QSPR Flash Point Prediction of Solvents Using Topological Indices for Application in Computer Aided Molecular Design" by S. Patel, D. Ng, M. S. Mannan, 2009. *Industrial and Engineering Chemistry Research, 48* (15), pp 7378-7387. Copyright 2009 by American Chemical Society.

$$\begin{aligned} \min \quad & F(\pi(y,x)) \\ s.t. \quad & \pi^L \leq \pi(y,x) \leq \pi^U \\ & g(y) \leq 0 \\ & h(y) = 0 \end{aligned} \tag{1}$$

where π is the vector of properties of the compound (also expressed as property evaluation models), y is the vector of integer variables that determine the molecular structure or binary variables for unit operation identity, descriptor identity, compound identity, x is the vector of continuous variables of relevant process conditions (such as flowrates, mixture compositions, condition of operation, design variables), π^U and π^L are upper and lower bounds on the property values, F is the performance criterion to be optimized, and g and h are vectors of constraints associated with structural feasibility requirements as well as other design specifications. CAMD approaches have been used successfully to develop novel materials given their specified target properties for materials such as refrigerants,[27] solvents and extractants,[28] catalysts,[29] polymers[30] and so on.

Property evaluation models play an important role in CAMD problems. The prediction ability of the models for pure/mixture properties governs the success of the CAMD methodology. There are many types of property estimation methods available as shown in Figure 5.[26]

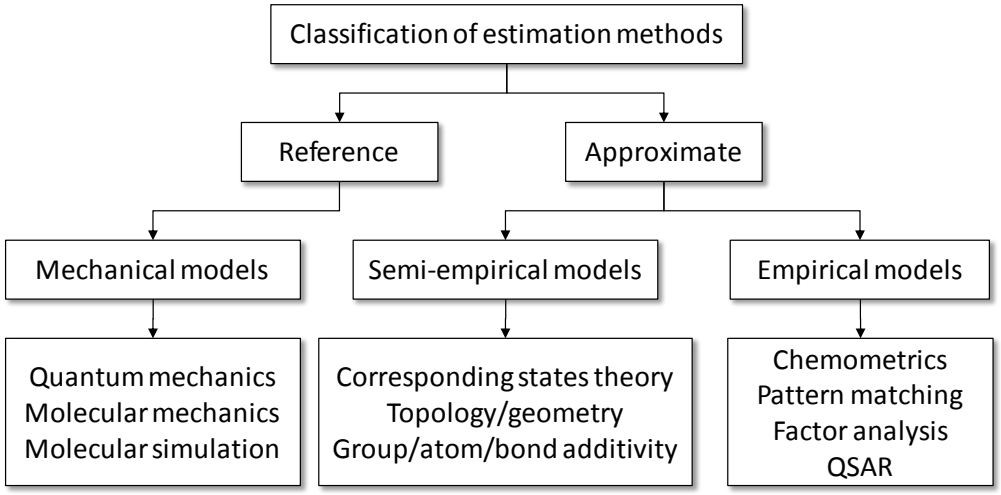


Figure 5. Classification of property estimation methods[26]

Not all property estimation methods are applicable to CAMD. One particular property prediction method expresses the property in terms of functions of the number of occurrences of predefined fragments/groups in the molecule. This class of property prediction method is known as Group Contribution Approach (GCA) and can be applied to CAMD. GCA based techniques are generally additive functions expressed as equation (2) shown below.

$$F(p) = w_1 \sum N_i C_i + w_2 \sum M_j D_j + w_3 \sum O_k E_k + \ldots \quad (2)$$

where C_i is the contribution of atom, bond or first-order group i; N_i is the number of occurrences of atom, bond or first-order group i; D_j is the contribution of atom, bond or second-order group j; M_j is the number of occurrences of atom, bond or second-order group j, and so on for O_k and E_k. Adding more number of higher order

terms to the equation denotes in principle, that the equation could possibly be highly accurate with large application range. But from a practical point of view, higher order terms are not feasible, and the most that is utilized currently is third order terms. Second and third order additive methods are able to distinguish between some isomers. At the same time, group contribution based methods have some limitations associated with them such as accuracy and ability to handle complex molecular structures. Thus, other lucrative property prediction methods (such as Quantitative Structure Property Relationship, which is an emerging technique) need to be investigated for application in CAMD as described in the next section.

3.2 Quantitative Structure Property Relationship (QSPR)

Quantitative Structure Property Relationship (QSPR) method is used to relate properties of substances/chemicals with entities obtained from the molecular structure. The relationships are generally linear correlations that use molecular descriptors as the inputs to the model. Molecular descriptors are the result of mathematical procedures that transform chemical information encoded within a symbolic representation of the molecule. There are various types of descriptors such as constitutional, geometrical, topological, electrostatic, quantum chemical and many others. Amongst the different types of predictive models described before, the group contribution approach for property estimation is employed by most studies in CAMD,[26, 31-34] while few studies have explored the applicability of QSPR/Quantitative structure-activity relationship

(QSAR) in CAMD.[35-37] Computer Aided Molecular Design using QSPR/QSAR consists of two parts- 1. Forward problem: A method to predict properties given the molecular structure, 2. Inverse problem: Applying the forward problem solution to obtain molecular structures that satisfy given target properties. Among the models described above, Quantitative Structure Property Relationship (QSPR) remains the choice of method for its predictive and new molecular design purposes. A schematic of the approach for applying QSPR in CAMD has been depicted in Figure 6.

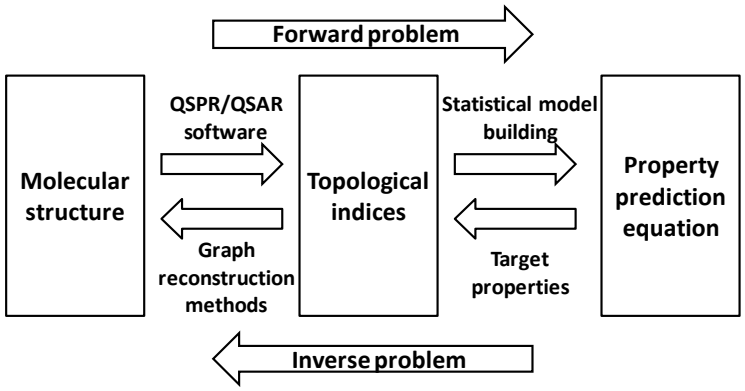


Figure 6. Schematic for QSPR application in CAMD

Topological indices prove to be more suitable for CAMD since they are calculated using information obtained from atomic constitution and bond characteristics of a molecule. Molecular descriptors based on topological information provide a higher level of molecular representation compared to functional groups or molecular fragment counts. They are able to differentiate molecules according to their size, degree of

branching, flexibility, and overall shape. In calculating topological indices, graph theory is used to evaluate information about the constituent atoms and the connecting bonds between them by means of adjacency, distance matrices or others. This makes it easier to visualize the molecular structure and simultaneously evaluate properties for the molecule. The possibility of solving the inverse problem for certain topological indices has been explored using graph reconstruction methods to obtain the exact molecular structure based on the index values.[38, 39] In recent years, more work has been done to resolve the inverse problem for QSPR, but few researchers have approached solving the forward problem using topological indices. Topological indices encode information on molecular connectivity which, in principle, would yield more accurate correlations than simpler group contribution methods. Some properties other than flash point have been modeled using topological indices (such as boiling point, molar volume, heat of vaporization for alkanes[40, 41]) but only for distinct groups of chemicals. While previous attempts did not address heteroatoms and multiple bonds, the necessity of extending this approach to make it more inclusive of different types of chemicals and variation in properties has been advocated in their work.

In relation with inherent safety principles, QSPR and CAMD can be used to aid the concept of ‘substitution’ of a more-hazardous compound with a less-hazardous one. To incorporate inherently safer substitution into chemical processes, consideration for hazardous properties (such as flash point, flammability limits, and toxicity) needs to be embedded into the solvent-selection process. Among the properties that describe a material’s flammability, flash point provides a stronger indication of flammability.

Flash point is the minimum temperature at which the liquid (or solid) emits sufficient vapor to form an ignitable mixture with air. NFPA ratings for flammability, as described in NFPA 704,[42] are also based on the flash point values for chemicals. Thus, the emphasis of this work is on predicting flash points for solvents using QSPR technique.

Models for prediction of flash point have been developed in the past. Boiling points of organic compounds have been correlated to flash point using quadratic relationships by Hshieh et al.,[43] and an exponential relationship by Satyanarayana and Rao.[44] Molecular structure information was also used by some researchers to predict flash point. Prugh[45] incorporated stoichiometric concentration and boiling point for flash point prediction. Suzuki et al.[46] used a combination of structural factors such as molecular connectivity index and group contributions to predict the flash point. A particular group-contribution-based model has also been developed by Stefanis et al.[47] Quantitative Structure Property Relationship (QSPR) has been applied for flash point prediction by Katritzky et al.[48, 49] using molecular descriptors which were of the types geometrical, electrostatic, quantum mechanical, and constitutional descriptors.[50] The QSPR method relies on predicting properties based on computable molecular descriptors which in turn are evaluated from information derived from the molecular structure. It has been used as a technique for prediction of properties such as critical temperature, boiling point, refractive index, octanol-water partition coefficient, and many others, with higher level of accuracy.[51-53]

In this work, the forward problem has been examined using topological descriptors (indices) to predict the flash point of solvents that are diverse in terms of

chemical constitution, bond saturation/unsaturation and cyclic/straight-chain/branch characteristics. Application of such predictive models to CAMD will aid in selecting solvents that are inherently safer.

3.3 Input data development

3.3.1 Experimental flash point

The experimental flash point data for 236 solvents were collected from the Acros[54] and Aldrich catalogs[55], and the Industrial Solvents handbook.[56] Major classes of solvents were used to form the dataset, such as monohydric alcohols, polyhydric alcohols, amines, ethers, and aliphatic and aromatic hydrocarbons. These solvents were selected in this study because of their frequent usage in the petrochemical industries, where safety concerns are escalated due to solvent processes. The distribution of the molecular weights and flash points for the entire data set has been shown in Figure 7 and Figure 8, respectively. The solvents used for this study show wide variability in composition and experimental flash point values (range of flash points = 513.15 K-157.15 K = 356 K, standard deviation = 64.81 K). The complete data set is shown in Table 4 along with the experimental and calculated flash point values.

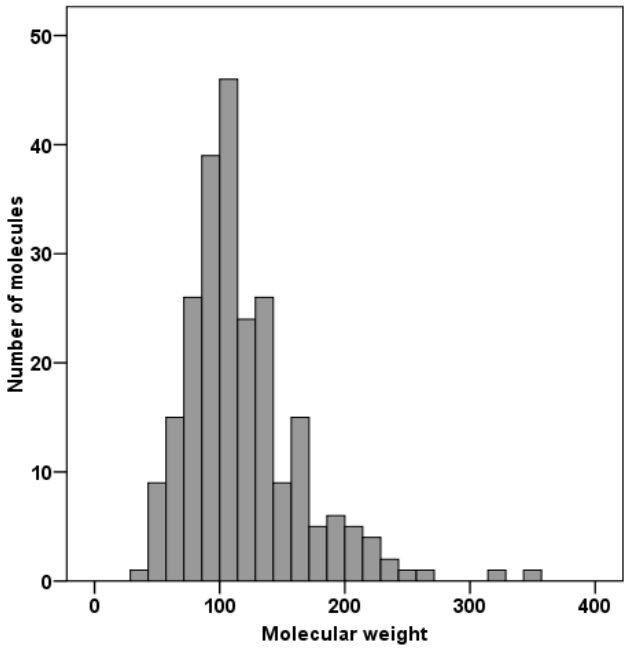


Figure 7. Distribution of molecular weights in dataset (n=236)

`

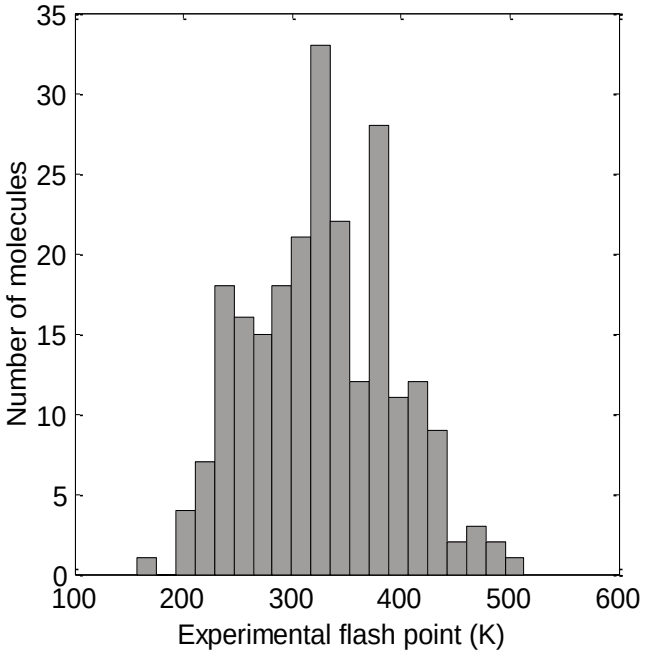


Figure 8. Distribution of flash point values in dataset (n=236)

Table 4. Solvent dataset with experimental and predicted (MLR and ANN) flash point

##	IUPAC CAS Name	T_f (exptl. K)	T_f (predicted K)	
			MLR	ANN
	Class -Monohydric alcohols			
1	Methanol	283.706	286.053	285.091
2	Ethanol	286.483	295.302	281.885
3	propan-1-ol	298.15	304.55	274.462
4	butan-2-ol	295.372	308.776	301.044
5	2-methylpropan-1-ol	303.15	308.776	291.057
6	2-methylpropan-2-ol	284.261	271.635	258.303
7	pentan-2-ol	313.706	318.024	308.542
8	2-methylbutan-2-ol	294.261	296.24	288.576
9	hexan-1-ol	347.039	332.294	321.632
10	2-ethylbutan-1-ol	331.483	334.407	325.531
11	octan-2-ol	358.15	345.769	347.779
12	nonan-1-ol	353.15	360.038	357.757
13	decan-1-ol	377.594	369.287	371.028
14	phenylmethanol	373.706	362.152	354.812
15	4-hydroxy-4-methylpentan-2-one	325.372	309.714	348.755
16	2-furylmethanol	356.483	352.904	350.476
17	2-methylpropanoic acid (3-hydroxy-2,2,4-trimethylpentyl) ester	393.15	396.931	378.539
18	1-chloropropan-2-ol	324.15	308.776	307.247
19	1,4-dibromobutan-2-ol	386.15	331.452	385.71
20	heptan-2-ol	332.15	336.521	333.879
21	heptan-1-ol	346.15	341.542	332.95
22	hexan-2-ol	314.15	327.272	321.81
23	1-methoxypropan-2-ol	306.15	318.024	333.533

Table 4. Continued

#	IUPAC CAS Name	T_f (exptl. K)	T_f (predicted K)	
			MLR	ANN
24	octan-1-ol	354.15	350.79	346.454
25	(2R)-2,4-dihydroxy-N-(3-hydroxypropyl)-3,3-dimethylbutanamide	386.15	392.489	472.342
26	pentan-3-ol	313.15	322.204	305.993
27	oct-1-en-3-ol	341.15	349.948	367.998
28	2,2,2-trifluoroethanol	302.15	296.24	306.357
29	2,3,4-trimethylpentan-1-ol	333.15	351.918	330.423
30	4-methylpentan-2-ol	314.15	322.251	316.174
31	2-methylbutan-2-ol	293.15	296.24	288.576
32	3,3-dimethylbutan-1-ol	302.15	305.488	311.749
33	2-ethylbutan-1-ol	330.15	334.407	325.531
34	2-ethylhexan-1-ol	350.15	352.904	347.905
35	2,2,4-trimethylpentan-1-ol	333.15	331.32	325.312
36	3-methylheptan-3-ol	327.15	336.342	307.242
37	tetradecan-1-ol	418.15	406.279	416.163
38	5-methylheptan-3-ol	327.15	349.106	332.281
39	8-methylnonan-1-ol	377.15	364.265	369.771
40	tridecan-1-ol	389.15	397.031	404.323
41	propan-2-ol *	284.817	293.617	278.532
42	butan-1-ol *	309.817	313.798	297.732
43	pentan-1-ol *	330.372	323.046	308.322
44	3-methylbutan-1-ol *	324.817	318.024	317.046
45	Cyclohexanol *	340.928	349.948	311.449
46	octan-1-ol *	363.706	350.79	346.454

Table 4. Continued

#	IUPAC CAS Name	T_f (exptl. K)	T_f (predicted K)	
			MLR	ANN
47	prop-2-en-1-ol *	295.372	304.55	326.194
48	2-tetrahydrofuranylmethanol *	347.039	352.904	352.845
49	(1R)-1-(2-furyl)ethanol *	383.15	361.058	378.618
50	5-methylheptan-1-ol *	--	349.948	408.592
	Class- Polyhydric alcohols			
51	ethylene glycol	392.594	389.86	375.36
52	2-(2-hydroxyethoxy)ethanol	411.483	411.722	407.13
53	2-[57]ethanol	435.928	423.681	445.617
54	2-[2-[2-(2-hydroxyethoxy)ethoxy]ethoxy]ethanol	477.594	438.116	478.926
55	propane-1,2-diol	377.594	380.062	373.765
56	3-(3-hydroxypropoxy)propan-1-ol	397.039	420.52	422.148
57	2-[2-(2-hydroxypropoxy)propoxy]propan-1-ol	416.483	409.9	412.671
58	Glycerol	433.15	401.138	424.783
59	butane-1,3-diol	382.039	380.919	392.755
60	butane-1,4-diol	394.261	404.435	394.196
61	pentane-1,5-diol	402.594	411.722	406.614
62	propane-1,1-diol	372.15	380.062	369.324
63	propane-1,3-diol	352.15	402.925	382.459
64	butane-1,2-diol	366.15	401.138	385.339
65	butane-2,3-diol	358.15	388.364	366.991
66	(E)-but-2-ene-1,4-diol	401.15	404.435	383.484
67	but-2-yne-1,4-diol	425.15	404.435	377.609
68	hexane-1,6-diol	420.15	414.388	417.903
69	hexane-2,5-diol	374.15	374.379	395.676

Table 4. Continued

#	IUPAC CAS Name	T_f (exptl. K)	T_f (predicted K)	
			MLR	ANN
70	2,2-diethylpropane-1,3-diol	380.15	419.21	383.197
71	2,5-dimethylhex-3-yne-2,5-diol	347.05	344.495	342.359
72	benzoic acid [4-[(oxo-phenylmethoxy)methyl]cyclohexyl]methyl ester	434.15	471.191	445.312
73	[4-(hydroxymethyl)phenyl]methanol	460.95	417.78	455.097
74	2-butyl-2-ethylpropane-1,3-diol	386.15	430.521	398.383
75	3,6-dimethyloct-4-yne-3,6-diol	382.15	395.317	383.748
76	2-ethyl-2-(hydroxymethyl)propane-1,3-diol	445.15	419.21	430
77	2,2-bis(hydroxymethyl)propane-1,3-diol	513.15	419.21	437.612
78	(2R,3R,4R,5S)-hexane-1,2,3,4,5,6-hexol	422.05	428.185	420.714
79	2-(hydroxymethyl)-2-methylpropane-1,3-diol	433.15	406.717	444.377
80	2,2-dimethylpropane-1,3-diol *	424.817	392.84	408.592
81	2-methylpentane-2,4-diol *	374.817	355.86	339.105
82	(2S)-butane-1,2,4-triol *	385.15	406.488	427.459
83	2,3-dimethylbutane-2,3-diol *	350.15	388.935	328.351
84	2-ethylhexane-1,3-diol *	409.15	428.496	400.474
85	2-(2-hydroxyethylthio)ethanol *	433.15	411.722	424.956
86	hexane-1,2,6-triol *	471.15	414.417	463.775
	Class- Hydrocarbons			
87	Pentane	224.15	240.353	244.722
88	Hexane	250.15	255.318	258.673
89	Heptanes	272.15	270.283	275.413
90	Octane	289.15	285.248	289.989
91	Nonane	304.15	300.212	306.757

Table 4. Continued

#	IUPAC CAS Name	T_f (exptl. K)	T_f (predicted K)	
			MLR	ANN
92	Decane	319.15	315.177	321.133
93	Dodecane	344.15	345.107	351.052
94	Cyclohexane	255.15	257.885	257.711
95	Cyclohexene	243.15	257.885	247.574
96	Benzene	262.15	257.885	228.335
97	Methylbenzene	280.15	269.673	284.393
98	1,2-dimethylbenzene	305.15	281.965	334.563
99	1,4-dimethylbenzene	300.15	281.461	328.904
100	1,3,5-trimethylbenzene	317.15	293.249	355.192
101	Isopropylbenzene	319.15	296.93	349.884
102	Tridecane	352.15	360.072	366.468
103	Tetradecane	372.15	375.037	378.976
104	Hexadecane	408.15	404.967	404.401
105	Heptadecane	421.15	419.932	417.163
106	Nonadecane	441.15	449.862	438.092
107	Isopentane	222.15	236.038	235.877
108	Isohexane	250.15	251.003	248.665
109	3-methylpentane	241.15	252.141	247.182
110	2,3-dimethylbutane	244.15	247.192	241.401
111	buta-1,3-diene; vinylbenzene	197.05	285.775	313.615
112	but-2-ene	199.85	225.388	213.151
113	(Z)-but-2-ene	200.15	225.388	213.151
114	but-1-ene	193.15	225.388	230.56
115	2-methylprop-1-ene	157.15	219.936	257.358

Table 4. Continued

#	IUPAC CAS Name	T_f (exptl. K)	T_f (predicted K)	
			MLR	ANN
116	(Z)-pent-2-ene	255.15	240.353	211.277
117	pent-1-ene	222.15	240.353	239.464
118	dec-1-ene	320.15	315.177	307.973
119	hept-1-ene	264.15	270.283	267.451
120	Cyclooctane	301.15	287.815	292.725
121	cyclopenta-1,3-diene	273.15	242.92	229.461
122	Cyclopentane	236.15	242.92	242.668
123	Cyclopentene	243.15	242.92	235.476
124	2-methylheptane	277.15	280.933	277.254
125	4-vinylcyclohexene	294.15	285.775	302.944
126	4-methylpent-1-yne	269.15	251.003	267.136
127	Ethylbenzene *	295.15	285.775	300.509
128	Undecane *	335.15	330.142	337.442
129	Pentadecane *	405.15	390.002	393.184
130	Octadecane *	439.15	434.897	426.962
131	Tricosane *	386.15	509.721	470.086
132	2,2-dimethylbutane *	225.15	244.736	223.052
133	2-methylhexane *	270.15	265.968	263.382
134	but-2-yne *	260.15	225.388	206.812
135	2,4,4-trimethylpent-1-ene *	267.15	270.351	262.589
136	2,3,4-trimethylpent-2-ene *	275.15	274.449	284.713
	Class – Ether			
137	Methoxymethane	232.039	228.675	262.678
138	Ethoxyethane	233.15	258.72	237.91

Table 4. Continued

#	IUPAC CAS Name	T_f (exptl. K)	T_f (predicted K)	
			MLR	ANN
139	2-methoxy-2-methylpropane	238.706	263.12	254.253
140	2-isopropoxypropane	245.372	280.103	268.041
141	1-butoxybutane	304.261	318.811	291.177
142	1-pentoxypentane	330.372	348.856	327.25
143	1-ethenoxybutane	263.706	288.765	274.498
144	2-methyloxirane	235.928	243.085	242.002
145	2-ethyloxirane	260.928	259.25	258.008
146	1,4-dioxane	291.483	276.32	298.632
147	(2,2-dimethyl-1,3-dioxolan-4-yl)methanol	267.039	310.541	320.874
148	2-methylfuran	243.15	273.131	280.25
149	Tetrahydrofuran	247.594	261.298	266.445
150	Tetrahydropyran	253.15	276.32	280.536
151	Methoxybenzene	324.817	304.318	321.482
152	1,2-bis(2-methoxyethoxy)ethane	386.15	363.879	364.855
153	2-(phenoxymethyl)oxirane	388.15	349.916	379.468
154	2-methoxy-2-methylbutane	262.15	279.965	278.489
155	2-(2-hydroxyethoxy)ethanol	416.15	288.765	407.13
156	Ethoxyethylene	228.15	258.72	259.356
157	Allyloxybenzene	335.15	334.363	336.911
158	2-methoxy-2-methylpropane	245.15	263.12	254.253
159	phenylmethoxymethylbenzene	408.15	410.007	422.252
160	1-chloro-2-(2-chloroethoxy)ethane	328.15	288.765	328.89
161	1-chloro-2-methoxyethane	288.15	258.72	278.95
162	1-[2-(2-butoxyethoxy)ethoxy]butane	374.15	408.947	373.549

Table 4. Continued

#	IUPAC CAS Name	T_f (exptl. K)	T_f (predicted K)	
			MLR	ANN
163	3-allyloxyprop-1-ene	266.15	288.765	295.111
164	1-chloro-4-(phenoxy)benzene	386.15	391.794	415.391
165	1-isopentyloxy-3-methylbutane	319.15	340.193	318.73
166	2-(allyloxymethyl)oxirane	321.15	304.318	298.392
167	chloro-(chloromethoxy)methane	292.15	258.72	300.365
168	2-(butoxymethyl)oxirane	314.15	319.341	330.12
169	1-chloro-1-(1-chloroethoxy)ethane	328.15	280.103	305.909
170	1-butoxybutane	298.15	318.811	291.177
171	1,2-dimethoxybenzene	345.15	332.822	363.198
172	1,2-dimethoxyethane	271.15	273.743	292.028
173	(3S,3aR,6R,6aR)-3,6-dimethoxy-2,3,3a,5,6,6a-hexahydrofuro[3,2-b]furan	376.15	362.361	355.042
174	Phenoxybenzene	388.15	379.961	391.754
175	1,3,5-trioxane	318.15	276.32	314.384
176	Ethoxyethylene	227.15	258.72	259.356
177	1-hexoxyhexane *	349.817	378.901	363.131
178	Ethoxyethylene *	227.594	258.72	259.356
179	1-ethenoxy-2-methylpropane *	263.706	284.434	285.693
180	Furan *	237.594	261.297	238.432
181	Phenoxybenzene *	388.15	379.961	391.754
182	4-(4-aminophenoxy)aniline *	491.15	403.628	464.648
183	chloro-methoxymethane *	289.15	243.697	268.983
184	4,7,7-trimethyl-8-oxabicyclo[2.2.2]octane *	322.15	336.993	322.116
185	1-pentoxypentane *	330.15	348.856	327.25

Table 4. Continued

#	IUPAC CAS Name	T_f (exptl. K)	T_f (predicted K)	
			MLR	ANN
186	1-ethenoxypropane *	247.15	273.743	253.605
	Class- Amines			
187	Acetamide	315.15	278.371	310.272
188	1-(2-pyridyl)ethanone	349.15	370.67	389.519
189	2-(bis(2-hydroxyethyl)amino)ethanol	452.15	390.196	436.342
190	prop-2-en-1-amine	245.15	290.279	300.604
191	6-methyl-2-pyridinamine	376.15	346.839	351.516
192	2-(2-aminoethoxy)ethanol	400.15	346.534	385.27
193	2-pyridinamine	236.15	341.367	318.273
194	Aniline	343.15	336.39	310.158
195	Phenylmethanamine	345.15	354.535	337.317
196	3-bromopyridine	324.15	314.827	298.997
197	butan-2-amine	254.15	280.705	283.293
198	2-methylpropan-2-amine	235.15	235.866	257.271
199	3-chloroaniline	391.15	337.668	352.07
200	2-chloroaniline	371.15	340.514	365.247
201	2-chloropyridine	337.15	332.276	306.731
202	2,4,6-trimethylpyridine	330.15	343.568	358.223
203	2-pyridinecarbonitrile	362.15	366.108	341.904
204	3-pyridinecarbonitrile	357.15	365.256	344.993
205	4-pyridinecarbonitrile	361.15	365.441	347.087
206	1-cyclohexyl-2-pyrrolidinone	418.15	379.197	404.158
207	Cyclohexanamine	300.15	304.905	295.267
208	N-pentylpentan-1-amine	277.15	382.028	329.366

Table 4. Continued

#	IUPAC CAS Name	T_f (exptl. K)	T_f (predicted K)	
			MLR	ANN
209	2,6-ditert-butylpyridine	345.15	334.552	331.44
210	ethane-1,2-diamine	330.35	288.59	327.026
211	N-cyclohexylcyclohexanamine	376.15	380.74	376.24
212	N,N-diethylacetamide	343.15	343.773	326.557
213	N-ethylethanamine	245.15	301.613	237.992
214	2-diethylaminoethanol	324.65	348.285	336.194
215	Diethylcyanamide	342.15	338.401	316.907
216	N-(2-aminoethyl)ethane-1,2-diamine	363.15	340.071	358.463
217	N,N-diethylformamide	333.15	336.669	316.305
218	1-(2-hydroxypropylamino)propan-2-ol	399.15	343.034	405.999
219	N-isopropylpropan-2-amine	266.15	288.95	263.146
220	N-methylmethanamine	255.15	265.924	244.246
221	N,N-dimethylaniline	336.15	356.04	353.182
222	2-dimethylaminoethanol	313.15	301.33	323.81
223	N,N-dimethylformamide	330.15	290.798	293.096
224	3,5-dimethylpiperidine	294.65	302.183	303.333
225	heptan-2-amine	327.15	321.424	316.324
226	heptan-1-amine	317.15	337.285	316.826
227	3-methyl-2-pyridinamine *	384.15	347.631	371.888
228	propan-1-amine *	243.15	281.851	259.021
229	2-bromopyridine *	327.15	318.717	303.486
230	butan-1-amine *	261.15	295.71	279.44
231	4-chloroaniline *	461.15	337.778	356.749
232	N-allylprop-2-en-1-amine *	280.15	342.997	286.198

Table 4. Continued

#	IUPAC CAS Name	T_f (exptl. K)	T_f (predicted K)	
			MLR	ANN
233	N-butylbutan-1-amine *	312.15	354.311	294.626
234	2-(2-hydroxyethylamino)ethanol *	411.15	344.221	405.273
235	N,N-dimethylacetamide *	336.15	298.548	328.412
236	2-aminoethanol *	358.15	290.664	351.183

*The solvents belonging to the test set for MLR calculations

The data set is divided into a training set (80%) and test set (20%), and cross-validation is performed by omitting each of 3 groups in turn. The total range of flash point values was divided into n smaller ranges. From each range a proportional number of solvents were used to form the test set. The box plots of the training set and test set for the entire data set are shown in Figure 9. The test set and training set are representative of the data set.

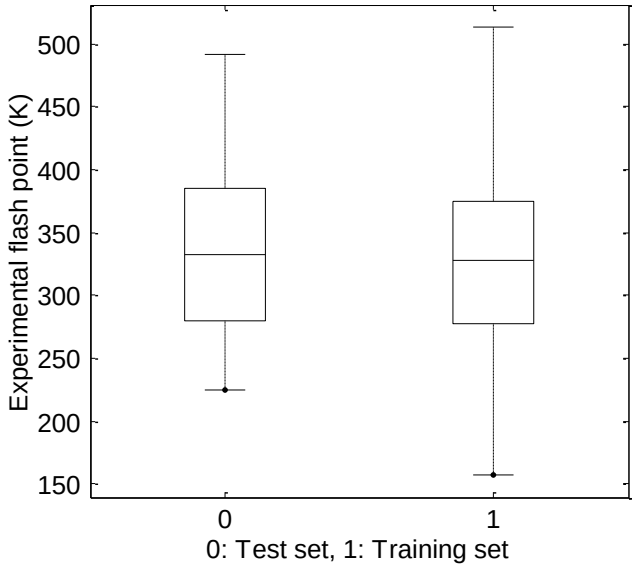


Figure 9. Box plots of test set and training set

3.3.2 Molecular structures and topological indices

Structures for the chemical compounds were obtained from the PubChem Database[58] in the standard data format (SDF). The molecular structure information was obtained for the 2D structure (i.e. without optimizing the geometry of the molecule for lowest energy state). This format is able to provide sufficient information to calculate topological indices using molecular graphs and related matrices (such as distance and adjacency matrices). The PubChem compound database contains chemical structures that have been validated. It also provides query and search tools for chemicals based on their names, molecular weights, and other criteria. Within PubChem, the structure of the compound can be found in the ASN, XML or SDF formats. An example of molecular structure stored in the SDF format has been shown in Figure 10.

The molecular descriptors (29 topological indices) were then calculated for all molecular structures using the Materials Studio 4.3 software (Accelrys Software Inc.), and the correlations were obtained using the Materials Studio Software package.[59]

```
702
  -OEChem-03301010282D

  9  8  0     0  0  0  0  0  0999 V2000
    2.5369   -0.2500    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
    3.4030    0.2500    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    4.2690   -0.2500    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    3.8015    0.7249    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    3.0044    0.7249    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    3.9590   -0.7869    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    4.8059   -0.5600    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    4.5790    0.2869    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    2.0000    0.0600    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0  0  0  0
  1  9  1  0  0  0  0
  2  3  1  0  0  0  0
  2  4  1  0  0  0  0
  2  5  1  0  0  0  0
  3  6  1  0  0  0  0
  3  7  1  0  0  0  0
  3  8  1  0  0  0  0
M  END
> <PUBCHEM_COMPOUND_CID>
702

> <PUBCHEM_IUPAC_CAS_NAME>
ethanol

> <PUBCHEM_NIST_INCHI>
InChI=1S/C2H6O/c1-2-3/h3H,2H2,1H3

> <PUBCHEM_MOLECULAR_FORMULA>
C2H6O

> <PUBCHEM_MOLECULAR_WEIGHT>
46.06844

> <PUBCHEM_OPENEYE_ISO_SMILES>
CCO

$$$$
```

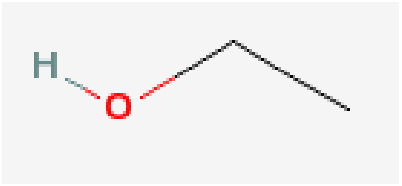


Figure 10. Example of molecular structure information stored in the SDF format

3.4 Statistical methods

Both multiple linear regression (MLR) analysis and artificial neural network (ANN) were used to evaluate the models and corresponding accuracy. Multiple regression models can be depicted using equation (3).

$$y = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n + c \tag{3}$$

where a_1, a_2, etc. and *c* are constants chosen to give the smallest possible sum of least squares difference between true y values and the y' values predicted using this equation. Neural network is a model-building technique that may better represent non-linear functions. ANN typically consists of three layers, an input layer, a hidden layer, and an output layer. Each layer is connected to the next layer, and the connections are associated with certain "weights". The connection weights are generally adjusted through a training method. Back-propagation has been used here for training the model. The algorithm comprises of the forward pass initially, wherein the input layer propagates a component of the input vector to each node in the middle hidden layer. Consequently, the middle layer computes output values, which become inputs to the nodes in the output layer. The output layer computes the network output for a particular input vector. These steps comprise the forward pass which is based on the current state of the network weights. The network weights are initially given as random values; thus, prior to training the weights it is unlikely that reasonable outputs will be obtained. Hence the weights are adjusted to reduce the error by backward propagation through the network. This process is known as the backward pass. The error values are computed for each node, based on

the known desired output. The error for the middle-layer nodes is then calculated by assigning a portion of the error at the output layer node to the middle node. The amount of error attributed depends on the magnitude of the connection weight.[60] Furthermore, the weight values are adjusted to improve the network performance according to the BFGS (Broyden-Fletcher-Goldfarb-Shanno) method or steepest descent algorithm.[59] This method enables the network to model complex non-linear functions for engineering applications.[61] The equation (4) best describes a neural network.[62]

$$y_k = f\left(\sum_{j=1}^{n_H} w_{kj} f\left(\sum_{i=1}^{d} w_{ji} x_i + w_{j0}\right) + w_{k0}\right) \tag{4}$$

where y_k denotes the output, n_H is the number of hidden nodes, w_{kj} is the hidden-to-output layer weights at the output layer k, w_{ji} is the input-to-hidden layer weights at the hidden unit j, x_i is the i^{th} input of total d inputs, and w_{jo} and w_{ko} are known as the bias. Also, f is the non-linear transfer function which calculates the output at a node. The transfer function used in Materials Studio is an s-shaped sigmoid function. This function is chosen because it is smooth and easily differentiable which makes it easier to train the network. The s-sigmoid function is depicted by the following underlying equation.[63]

$$f(z) = \frac{1}{1 + e^{-z}} \tag{5}$$

3.5 Results and discussion

3.5.1 Multiple linear regression

The calculated values for the flash points of solvents using multiple linear regression and back-propagation neural network analysis are shown in Table 4. Upon using the entire data set for multiple linear regression analysis, poor accuracy was obtained (R^2 = 0.479, r = 0.692). This is partially due to the large variability in chemical constitution and structure. Thus, the data set was divided into different classes of solvents (Figure 11 shows the distribution of flash point for each class) which have been analyzed using MLR, and the results are as shown in Table 5, along with the correlation coefficient, r, R-squared value, R^2, R^2 (CV) for cross validation and F value, being indicative of their predictive capability.

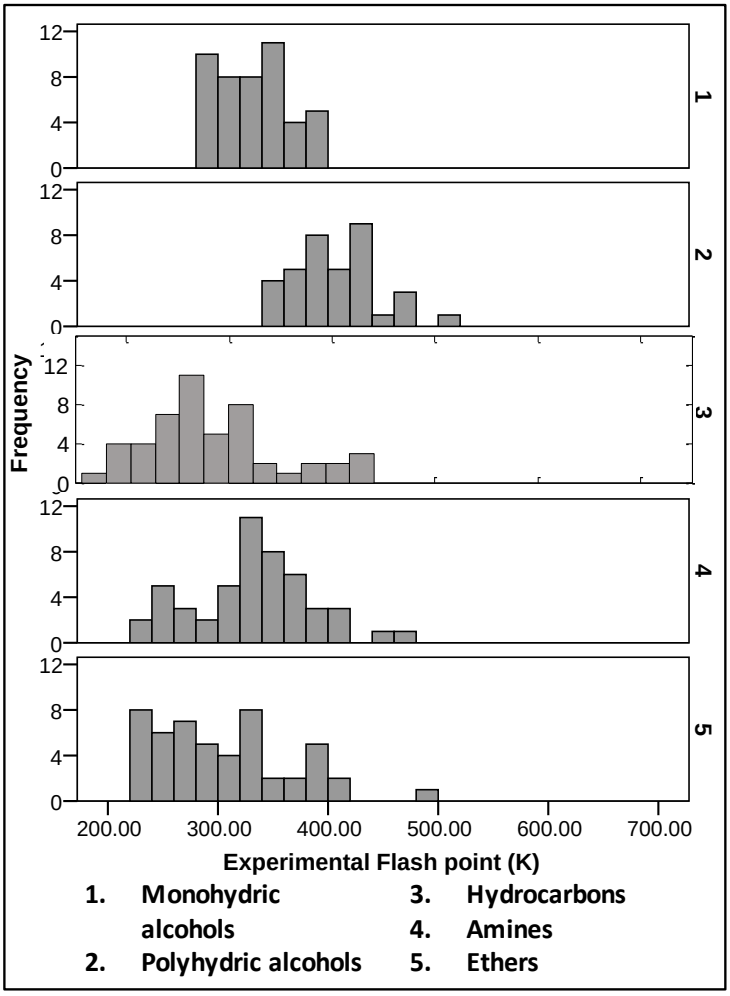


Figure 11. Distribution of flash point values for each class

Table 5. Results of multiple linear regression on different classes of solvents for flash point prediction

Class of solvents	Equation	Training set				Test set	
		r	R^2	R^2(CV)	F	r	R^2
Monohydric alcohols	$T_f = 9.248 \times SC:2 - 34.953 \times {}^{3}\chi_{cluster} + 286.053$	0.925	0.855	0.613	109.06	0.933	0.870
Polyhydric alcohols	$T_f = 13.064 \times {}^{2}\kappa - 8.665 \times {}^{3}\kappa + 385.328$	0.608	0.370	-	7.62	0.500	0.251
Hydrocarbons	$T_f = 29.929 \times {}^{1}\chi - 105.054$	0.881	0.776	0.696	132.15	0.792	0.628
Amines	$T_f = 50.264 \times {}^{1}\chi - 31.887 \times {}^{2}\chi_v + 210.783$	0.691	0.477	0.22	16.89	0.396	0.157
Ethers	$T_f = 30.045 \times {}^{1}\chi + 186.184$	0.825	0.680	0.600	80.75	0.875	0.770

R^2 = coefficient of determination; r = correlation coefficient; R^2 (CV) = R^2 for the cross validation set, F = Fisher test statistic; SC: 2 = Subgraph counts (second order): path; ${}^{3}\chi_{cluster}$ = Chi(3): cluster; ${}^{n}\kappa$ = Kappa-n; ${}^{n}\chi$ = Chi (n); ${}^{2}\chi_v$ = Chi(2) (Valence modified).

An overall depiction of MLR results is shown in Figure 12 as a plot of calculated versus experimental values.

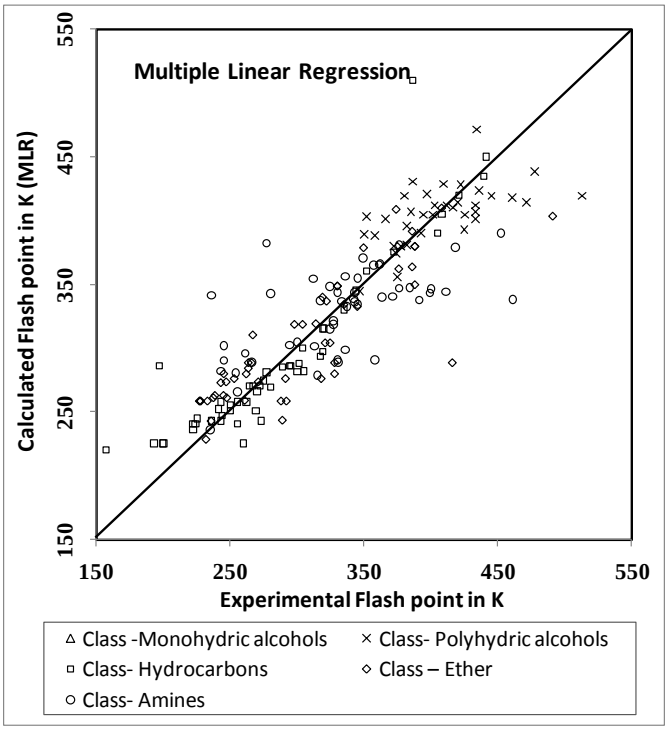


Figure 12. Plot of calculated versus experimental values of flash point using MLR (graph depicts correlations from Table 1 for all classes)

A statistical evaluation of error and deviation in calculated-versus-experimental values is found using the following definitions for the entire set of solvents.

Average absolute deviation = $\frac{1}{n}\sum_{i=1}^{n}\left|T_{\exp} - T_{calc}\right|$

Average absolute relative deviation = $\frac{1}{n}\sum_{i=1}^{n}\left|\frac{T_{\exp}-T_{calc}}{T_{\exp}}\right|$

Average percent bias = $\frac{1}{n}\sum_{i=1}^{n}\frac{T_{\exp}-T_{calc}}{T_{\exp}}\times 100$

The average absolute deviation is 20.819 K, the average absolute relative deviation is 6.57%, and the average bias is -0.21% for the data set using MLR.

From Table 1 monohydric alcohols and ethers show a consistent and higher accuracy in prediction (as seen from the results of the training set and test set), whereas hydrocarbons show higher accuracy for the training set and do not perform well for the test set. Models for polyhydric alcohols and amines do not perform better overall. Alcohols and amines have additional molecular phenomena that dictate the physical properties associated with the chemicals. Hydrogen bonding that occurs between molecules where a hydrogen atom is attached to one of the electronegative elements - fluorine, oxygen, or nitrogen - is one of the contributing factors for such chemical types. Vapor pressure is an influencing factor for flash point determination[64] which in turn is governed by attractions among molecules, unevenly distributed electron densities, and bonded hydrogen atoms.[65]

Molecular connectivity indices χ show good correlation for most types of solvents as seen in Table 1. Previous work has also shown that this particular type of topological index has been used successfully to predict properties for normal and branched alkanes.[40] The topological indices applied here have some physical meaning. The molecular connectivity index χ provides a quantitative assessment of the degree of

branching of molecules and the valence modified connectivity index χ_v provides information on the chemical nature of the atoms. The shape of a molecule – as determined by the different degrees and location of branching, is described by Kier's shape indices κ. The subgraph count index (second order) is a measure of the number of pairs of connected edges (i.e. number of paths of length 2). These indices collectively can provide information on the size of a molecule (volume occupied by the molecule), and the shape of the molecule (distribution of the molecular volume in space).[66] The topological indices are able to provide some information on the interactions among molecules, but do not give sufficient information on hydrogen bonding ability. As aforesaid for certain classes of compounds, the properties would depend on the patterns in intermolecular attractions. Thus, other types of molecular descriptors may perform better in predicting flash point for more complex compounds.

3.5.2 Artificial neural network

To enhance the predictive power, neural network analysis using topological indices was performed on the entire data set of 236 solvents. The training set consisted of 189 (~80%) compounds and the test set consisted of 47 compounds. A 16:6:1 network (consisting of 16 input nodes as given in Table 6, one output node, and one hidden layer with 6 nodes) gave higher accuracy for prediction of flash point as shown in Table 7. It is to be noted though that such a network configuration points to a highly complex network with a large number of coefficients and variables.

Table 6. Input variables to the neural network.

16 Input (predictive) variables for ANN		
Kier Shape indices[67]	$^{1}\kappa$	Kappa-1
	$^{2}\kappa$	Kappa-2
	$^{3}\kappa$	Kappa-3
	$^{1}\kappa$	Kappa-1 (alpha modified)
	$^{2}\kappa$	Kappa-2 (alpha modified)
Kier and Hall Subgraph count indices	SC: 0	Subgraph counts (0): path
	SC: 1	Subgraph counts (1): path
	SC: 2	Subgraph counts (2): path
Kier and Hall Molecular connectivity indices[68]	$^{0}\chi$	Chi (0)
	$^{1}\chi$	Chi (1)
	$^{2}\chi$	Chi (2)
	$^{3}\chi$	Chi (3): path
	$^{0}\chi_{v}$	Chi (0) (valence modified)
	$^{1}\chi_{v}$	Chi (1) (valence modified)
	$^{2}\chi_{v}$	Chi (2) (valence modified)
	$^{3}\chi_{v}$	Chi (3): path (valence modified)

Table 7. Results of neural network analysis for flash point prediction

n = 236, Network configuration = 16:6:1			
	r	R^2	R^2(CV)
Training set	0.940	0.883	0.638
Test set	0.878	0.772	0.664

The input parameters are selected based on the best set of descriptors found in the MLR models. Neural networks are good at fitting functions, but could occasionally result in over-fitting. Thus, a test set is needed to verify the predictive power of the neural network and, as expected, the accuracy for the test set is lower than the training set. Figure 13 shows the plot of calculated values against experimental values of flash point. The average absolute relative deviation is 5.35%, the average absolute deviation is 16.08 K, and the average percent bias is -0.22% for the complete data set using ANN.

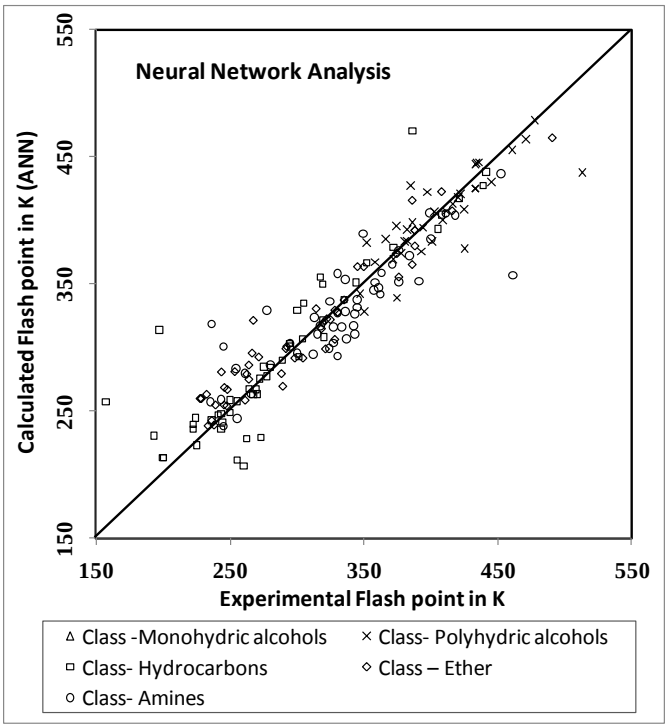


Figure 13. Plot of calculated versus experimental values of flash point using ANN

The results from ANN show higher accuracy than MLR although the complexity and non-linearity of the ANN model makes it difficult to directly apply it to the CAMD problem (since CAMD requires simpler, preferably linear, relationships for an MINLP based formulation). Thus, a suitable methodology that incorporates neural network models in the solution of an inverse problem would be a significant development in this approach. Overall, neural network analysis gives better prediction of flash points for solvents.

3.6 Conclusions

Computer Aided Molecular Design is one means for finding inherently safer chemical substitutes for solvents. Supplementing the group contribution methods by newly developed methods is a promising venture for CAMD in future. This section discusses an approach to achieve this objective. Efficient QSPR approaches have become an attractive option in recent years for property estimation in general. Thus, it is also important to identify its applicability in CAMD for solvent substitution. One of the hazardous properties, flash point, was evaluated using QSPR for different classes of solvents. Topological indices in particular have been used in this work to facilitate future application in CAMD as explained earlier. Although the results proved to be promising, some aspects can be explored further with regards to CAMD:

1. An ANN model for flash point gave higher accuracy for the entire data set than MLR, but the application of this complex model in CAMD has not yet been

investigated. The non-linear equation that defines a neural network will prove to be difficult to implement in CAMD. Previous studies have used linear relationships preferably, to reduce the CPU time.

2. Further justification for using QSPR instead of group contribution methods in CAMD is needed. The established group contribution methods are applicable to a wider range of chemical species, whereas QSPR methods are occasionally specific to a particular class of chemical compounds. In this paper, the flash point prediction proved to be of higher accuracy for certain classes of solvents (monohydric alcohols and ethers) as compared to other classes (amines, polyhydric alcohols). In order to obtain higher accuracy in the QSPR model, sub-classes for each of the classes investigated here need to be assessed. Subsequent sub-division defeats the purpose of CAMD to choose amongst a larger range of chemicals.
3. More properties which need to be considered during selection of solvents such as solubility parameter, boiling point, surface tension, and vapor pressure for diverse chemical data sets need to be predicted and assessed using topological indices. Other safety related properties that need to be evaluated are toxicity levels (e.g. LC_{50}), and reactivity or stability related properties.

Thus, there is a pressing need to explore the applicability of QSPR (and topological indices) in future CAMD studies. QSPR has been successfully applied in quantifying certain biological responses and polymer behaviors in the past, but analyzing its suitability for CAMD remains to be accomplished.

4. MOLECULAR DESIGN: CAMD AND ICAS

4.1 ICAS information

The Integrated Computer Aided System, i.e. ICAS, Version 11.0 developed at the CAPEC group in Technical University of Denmark provides a tool (ProCAMD) for the design and selection of solvents and process fluids.[57] The feasible solvent candidates generated through the toolbox satisfy desired property constraints based on group contribution models. ProCAMD is based on a hybrid methodology for CAMD[69] employing group contribution based property prediction methods. ProCAMD is linked to two other modules in ICAS, ProPred and CAPEC Database. This enables all generated and tested structures to be further analyzed through ProPred and cross-checked for the molecule's existence in the CAPEC database. Some of the properties that can be included in the selection process are as shown in Table 8.

Table 8. Properties for solvent selection in ICAS 11.0

Non-temperature dependent properties	
Critical temperature, pressure, volume	Normal boiling point
Normal melting point	Gibbs energy of formation
Enthalpy of formation	Enthalpy of vaporization
Liquid molar volume	Open cup flashpoint
Total solubility parameter	Surface tension
Log P (Octanol water partition coefficient)	Refractive index
Temperature dependent properties	
Liquid density	Viscosity
Diffusion coefficient in water	Vapor pressure
Thermal conductivity	Liquid heat capacity
Mixture properties	
Selectivity	Solvent power
Solvent loss	Distribution coefficient
Separation factor	Solute loss
Solvent capacity	Feed selectivity

These properties are estimated using group contribution based approaches, which are derived from molecular structure information. The desired property of the molecule is expressed as a function (generally additive) of the number of groups of type i present in a molecule and the contribution of that group towards the final property. Individual contributions of functional/ structural groups are predetermined using empirical data and property prediction modeling techniques. Some group contribution methods developed

for property prediction have been published previously.[70-72] Definitions of mixture properties as given in the documentation are as shown in Table 9.[57]

Table 9. Definition of solvent-mixture properties relevant to solvent performance

Solute distribution coefficient, m	$m = \frac{\gamma^{\infty}_{A,B}}{\gamma^{\infty}_{A,S}} \frac{MW_B}{MW_S}$
Solvent loss, S_l	$S_l = \frac{1}{\gamma^{\infty}_{B,S}} \frac{MW_S}{MW_B}$
Solvent power, S_p	$S_P = \frac{1}{\gamma^{\infty}_{A,S}} \frac{MW_A}{MW_B}$
Solvent selectivity, β	$\beta = \frac{\gamma^{\infty}_{B,S}}{\gamma^{\infty}_{A,S}} \frac{MW_A}{MW_B}$
Solute loss	Amount of solute leaving with the raffinate
Solvent capacity	This is similar to solvent power. It is expressed as mass of solute dissolved per mass of solvent
Separation factor	The ratio of equilibrium constants for A and B with respect to S corresponding to the feed selectivity calculations

MW is the molecular weight, γ^{∞} is the activity coefficient at infinite dilution, S as subscript indicates solvent, A as subscript indicates solute A that is to be recovered, B as subscript indicates solute B that is to be recovered, {A,S} {B,S} indicate binary mixtures where the first compound is in infinite dilution.

In order to select inherently safer solvents, hazardous properties of the solvent should be investigated, such as flash point, toxicity levels etc. For assessing flammability hazard, the indicative properties are lower flammability limit and its vapor pressure at the prevailing temperature. Flash point is the lowest temperature at which the liquid gives off enough flammable vapors to form an ignitable mixture with air. Thus, flash point accounts for the phenomenon of high-enough vapor pressure such that the concentration approximately corresponds to the lower flammability limit. Thus, the flash point is the main parameter to assess flammability of material.[73] For assessing toxicity hazard, the commonly used values are threshold limit value (TLV), LD_{50} and LC_{50} values. TLVs are better suited for emergency planning because they provide exposure limits for workers over 8-hour working days, whereas legislations generally make use of the LD_{50} or LC_{50} values which can provide better estimates to protect a broader population. Models used to predict open cup flash point and LC_{50} toxicity values in the ProCAMD module are as shown in Table 10.

Table 10. Prediction models for flash point and toxicity value

Property	Group contribution method of prediction[57]	Model accuracy
Open cup flash point (K)	$T_f(K) = 3.63 \times \sum_i N_i T_{fi} + 0.409 \times T_b(K) + 88.43$ where $T_b(K) = 204.359 \times \log\left(\sum_i N_i T_{bi} + \sum_j M_j T_{bj}\right)$	$T_b(K)$ Avg. absolute error = 5.35 K Avg. relative error = 1.42%
Open cup flash point (K) - Marrero & Gani	$T_f(K) = 155.192 + \sum_i N_i C_i + \sum_j M_j D_j + \sum_k O_k E_k$	Avg. absolute error = 8.56 K Avg. relative error = 2.66% r = 0.9692
LC_{50} toxicity values (mol/lt)	$-\log\ LC_{50} = \sum_i N_i \alpha_i$	R^2 = 0.91 & RMSE= 0.37

4.2 Integration of safety aspects

Apart from providing property prediction models, suitable limits or corresponding requirements for the hazardous properties need to be specified. Such information can be obtained from external standards and guidance; National Fire Protection Association rating,[42] OSHA HCS - Globally Harmonized System of hazard classification,[74] EU directive,[75] Dow Fire and Explosion index,[76] and Heikkila Inherent Safety Index.[73] Table 11 shows the flammability classification obtained from different sources given above. It can be deduced that for the purposes of selecting inherently less-

flammable solvents, the flash point Tf should be greater than ~300 - 330 K. Similarly, Table 12 indicates that in order to select an inherently less-toxic solvent, the LC_{50} toxicity level should be greater than ~2 mg/lt for a 4-hr exposure. The ProCAMD module enables us to provide such limiting values for open cup flash point and toxicity levels prior to generating molecules using CAMD. For toxicity level of LC_{50}, an indirect constraint is placed on log P (octanol/water partition coefficient), because it has been shown that log LC_{50} (in moles/lt) is almost linearly related to log P.[77, 78]

Table 11. Flammability classification criteria according to different guiding documents

Hazard category	Flash point (K) and boiling point (K) criteria		
	NFPA rating and Dow F&EI	OSHA GHS	EU Directive 67/548 and Heikkila ISI
Flammability decreases from top to bottom	$T_f < 295.8$; $T_b < 310.8$	$T_f < 296$; $T_b < 308$	$T_f < 273$; $T_b < 308$
	$T_f < 295.8$; $T_b > 310.8$ or $T_f > 295.8$; $T_f < 310.8$	$T_f < 296$; $T_b > 308$	$T_f < 294$; $T_f > 273$
	$T_f > 310.8$; $T_f < 366.3$	$T_f > 296$; $T_f < 333$	$T_f > 294$; $T_f < 328$
	$T_f > 366.3$	$T_f > 333$; $T_b < 366$	$T_f > 328$

Table 12. Toxicity classification criteria according to different guiding documents

Hazard category	LC_{50} toxicity level criteria (mg/lt) for 4 hr exposure OSHA GHS inhalation vapors, EU Directive 67/548
Toxicity decreases from top to bottom	$LC_{50} < 0.5$
	$LC_{50} > 0.5$; $LC_{50} < 2.0$
	$LC_{50} > 2.0$; $LC_{50} < 10.0$
	$LC_{50} > 10.0$; $LC_{50} < 20.0$

The inherent safety of the selected solvent will also depend on reactivity characteristics (if any) of the chemical. But in this work, flammability and toxicity levels have been considered for the selection of inherently safer options for solvents. Thus, by integrating Computer Aided Molecular Design and inherent safety guidance based on hazard classification, a solvent selection strategy can be developed that accounts for 'substitution' during the early stages of design. The application is further shown in the case study in Section 6.

4.3 Comparison of database selection and CAMD

Another logical approach for selection of safer solvents is to investigate a database of solvents. ICAS 11.0 includes ProCAMD module as described in previous sections, and the CAPEC Database that includes extensive data for various types of mixtures, pure component properties, and solvent properties. The type of solvents selected by both methods can be compared by carrying out a simple case study. Results

from a representative separation case study for phenol-water mixture for both CapecDB Manager and ICAS-ProCAMD have been assessed. The following guidelines were considered as initial first-estimate criteria for solvent selection:

- Easier separation in the solvent recovery unit is achieved when the normal boiling point of the solvent and vapour pressure is not close to that of phenol
- The solvent should have high feed selectivity and separation factor, and low solvent losses
- Solvent should possess favourable characteristics that make it inherently safer such as high flash point and low toxicity value. The flash point is estimated directly, while the toxicity level, LC_{50} can be estimated based on octanol-water partition coefficient (P) value as shown in equation (6).[78]

$$\log LC_{50} = -0.94 \log P + 0.94 \log(6.8E^{-05}P + 1) - 1.25 \quad (6)$$

Toluene is traditionally used for phenol-water separation. Thus, the selection criteria can be quantitatively inferred from toluene's physical properties as shown in Table 13.

Table 13. Selection criteria for phenol-water mixture

Normal boiling point (K)	Max: 450
Open cup flash point (K)	Min: 320
Log P (Octanol/water partition coefficient)	Min: 1.5
Liquid density (g/cm^3)	Max: 0.9 at 298 K
Vapor pressure (bar)	Min: 0.003 at 360 K

The ProCAMD module enables the user to include constraints for mixture properties and solvent performance such as solvent loss (max: 0.001), separation factor (min: 80), solvent capacity (min: 2), and feed selectivity (min: 8).

The results of both search approaches have been shown in Table 14. In terms of overall applicability of the two approaches, certain strengths and weaknesses can be deduced as given in Table 15.

Table 14. Comparison of results from ProCAMD & database search

	ICAS ProCAMD	Database search
Number of solvents identified	58	23
Isomers	57	13
Maximum value of flash point (K)	335.1	430

Table 15. Strengths and weaknesses of CAMD & database screening methods

	ICAS ProCAMD	Database search
Strengths	• Capable of generating new molecular structures • Ability to select based on solvent performance characteristics as well	• Provides more accurate property values from credible data sources
Weaknesses	• Group contribution based property prediction is not very accurate for isomers or complex molecular structures	• Limited properties available for selection criteria • Solvent performance criteria cannot be specified

Another database has been developed for solvent selection, PARIS II (Program for Assisting the Replacement of Industrial Solvents).[79] It can be used to design single-chemical substitutes and mixture substitutes. The chemical families found in the database are normal hydrocarbons, ketones, alcohols, aromatics, and organic and aqueous mixtures. PARIS II takes into account performance and operational properties such as molecular mass, liquid density, boiling temperature, vapor pressure, surface tension, viscosity, thermal conductivity, and flash point. For pure components information is obtained from DIPPR correlations and data, and for mixtures the properties are calculated using weighted sum approach, UNIFAC method, and other correlations developed by previous researchers. In order to assess the interactions between the solvent and the solute, the approach used in PARIS II is that the solvents of interest are investigated for their interactions based on the infinite-dilution activity coefficient with a set of representative solutes from different chemical families. In order to assess human and environmental impact of solvent use, two indexes are used: environmental index and air index. When dealing with chemical mixtures, the PARIS II software deals with weighted additive functions. This is a drawback when assessing environmental impact of mixtures because the effect of two or more chemicals can be antagonistic or synergistic.

Another limitation of such a database searching method is that new molecular structures cannot be generated, and the selection can only be made with existing chemicals in the database. This is also seen in a solvent substitution example shown in Li et al.[79] where the solvent to be replaced is benzene and the suggested replacements

(single chemicals and mixtures) violate some property constraints for molecular mass, liquid density and flash point. Also, this database does not account for cost considerations for solvent substitutes.

4.4 Conclusions

This section discusses a tool, ICAS –ProCAMD that can be utilized to carry out CAMD studies using group contribution methods of property prediction. This tool works as a multi-level, test-and-generate software for CAMD purpose.

1. ProCAMD includes non-temperature dependent properties, temperature dependent properties, and solvent-solute mixture properties. It also includes estimations related to safety related properties such as flash point and LC_{50} (log P).
2. The limiting conditions for hazardous properties can be inferred from existing guidelines such as NFPA 704 hazardous classification system, OSHA GHS classification system, and the European Union directive based classification system.
3. Solvent substitutes can be designed to be single chemicals or mixtures of solvents. At the same time, solvents can be selected from an existing database with relevant properties, or by using the CAMD methodology. Both approaches have certain shortcomings and advantages. Database technique is unable to always provide solutions that match the property requirements whereas; CAMD can in principle

generate new molecular structures that match the target properties if suitable property prediction models are available.

4. Future research can be carried out in the area of solvent selection using database, such that it accounts for mixtures and pure components as well as cost considerations.

5. PROCESS DESIGN: SIMULATION WITH SENSITIVITY ANALYSIS AND OPTIMIZATION

5.1 Introduction

Along with 'substitution', other inherently safer design concepts that need to be accounted for are 'attenuation/ moderation' and 'intensification/ minimization'. These can be incorporated by placing suitable constraints on process design, such that the constraints provide a relationship linking the level of hazard/ risk and the design parameters. For example, on the one hand, considering flash point of solvents as shown in Section 4.2 provides an indication of the ease of igniting the material (as well as amount of heat energy required to enable ignition). On the other hand considering properties such as vapor pressure and heat of combustion gives an indication of how quickly the liquid evaporates and how severe the consequences could be. Consequence modeling for incidents provides a framework for the development of such relationships among variables.

---

[*] Part of this section is reprinted with permission from "Inherently safer design of solvent processes at the conceptual stage: Practical application for substitution" by S. Patel, D. Ng, M. S. Mannan, 2010. *Journal of Loss Prevention in the Process Industries*, doi:10.1016/j.jlp.2010.03.002. Copyright 2010 by Elsevier Ltd.

5.2 Integration of safety consequence modeling

5.2.1 General consequence modeling

There are many approaches and types of constraints that can be implemented during the design stage of the process as shown in Figure 14. The figure shows that at each step of consequence modeling, there is an opportunity to place a suitable limiting condition on the parameter being estimated. For example, if dispersion modeling equation is utilized, then a suitable limit on the concentration level can be imposed, such as 'Concentration should not exceed LC_{50} toxicity levels at a particular distance' or 'Concentration must not lie between the LFL (lower flammability limit) and UFL (upper flammability limit)'. Similarly, some examples of the equations to be used along with the limiting condition are shown in Figure 14.

5.2.2 EPA RMP worst case modeling

Apart from general consequence models, another approach as described below can be applied to integrate safety consideration during the design phase. This approach makes use of information obtained from the EPA Risk Management Program guidelines[80] and has been employed in the case study in Section 6. The regulation covers facilities that contain more than the threshold quantity of 140 regulated substances, which includes many commonly used solvents such as carbon disulfide, pentane, toluene etc.[81] An offsite consequence analysis provides information to the government and the public about the potential consequences of an accidental chemical release at the facility.

For flammable and toxic substances, the regulation states that the consequence analysis must consist of a worst-case release scenario, and alternative release scenarios.

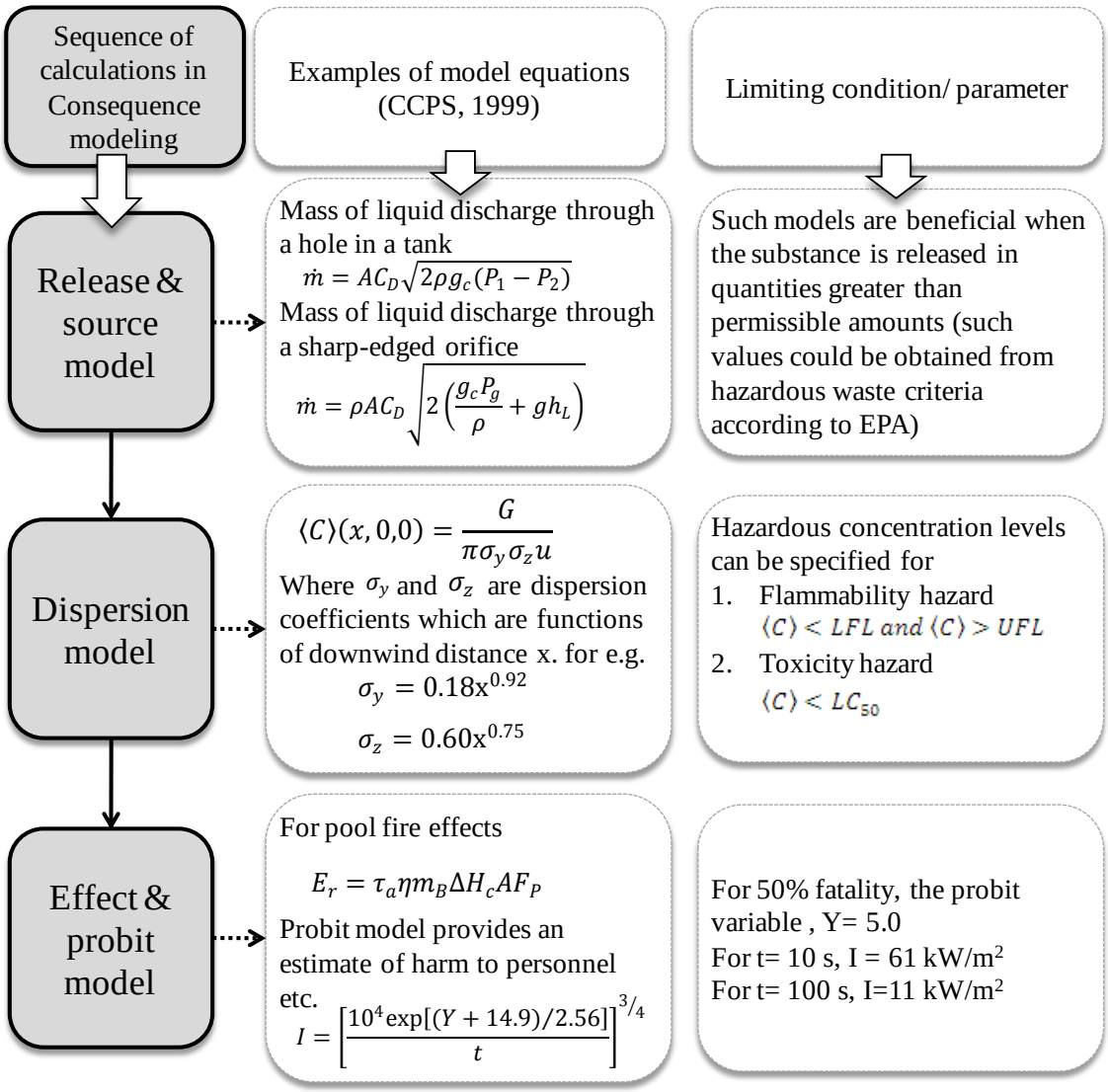


Figure 14. Forming safety constraints based on consequence modeling and other criteria

5.2.2.1 Flammable hazard

EPA RMP states the requirements for the worst-case scenario of flammable release, for which the endpoint has been defined as the distance to 1 psi overpressure resulting from a vapor cloud explosion. This endpoint has been determined to be the threshold for potential serious injuries to people as a result of property damage caused by an explosion (for e.g. shattering of glass windows, and partial demolition of houses). For liquid releases, the quantity participating in the vapor cloud explosion is the amount that volatilizes in 10 minutes from a pool formed by the entire quantity of the mixture. The evaporation rate can be estimated with equation (7) as shown below.[80]

$$QR = \frac{0.284 \times U^{0.78} \times MW^{2/3} \times A \times VP}{82.05 \times T} \tag{7}$$

where QR = evaporation rate (pounds per minute), U = wind speed (meters per second) ~1.5m/s, MW = molecular weight, A = surface area of pool formed by the entire quantity of mixture (square feet), VP = vapor pressure (mmHg), T = temperature of released substance (K).

Thereafter, the total amount vaporized in 10 minutes can be found by multiplying QR with 10. This amount can then be further incorporated into the explosion overpressure-distance calculation using equation (8) given below.

$$D_{flam} = 0.0081 \times \left(0.1 \times W_f \times \frac{HC_f}{HC_{TNT}} \right)^{1/3} \tag{8}$$

where D_{flam} = distance to overpressure of 1 psi (miles), W_f = weight of flammable substance (pounds) $\cong QR \times 10$, HC_f = heat of combustion of flammable

substance (kilojoules per kilogram), HC_{TNT} = heat of explosion of trinitrotoluene. The factor 0.0081 is a constant for damages associated with 1 psi overpressure, and the factor 0.1 represents an explosion efficiency of 10 %.

5.2.2.2 Toxic hazard

Guidelines for the toxic worst case release quantity remain similar, that is the total amount of toxic material held in a single vessel is released. Having assumed the quantity being released, the next step is to calculate the amount being vaporized by using equation (7) given in flammable consequence modeling. Thereafter, the total time to evaporate can be calculated. The EPA RMP guidelines for worst case-scenario related with toxic exposure provide the toxic endpoints individually for all chemicals listed in the hazardous category list by EPA. The toxic endpoints correspond to the maximum airborne concentrations below which it is believed that nearly all individuals can be exposed for up to one hour without experiencing or developing irreversible or other serious health effects or symptoms which could impair an individual's ability to take protective action. This toxic endpoint is not the same as the one calculated by the ProCAMD model, which is for estimating LC_{50}: the aqueous concentration causing 50% mortality in fathead minnow after 96 h. Thus, the dispersion calculations cannot be based off of the EPA guidance. Rather the method applied here is based on simple Pasquill-Gifford dispersion model calculations. For the worst case scenario the stability class is F (stable conditions, nighttime, <50 % cloud cover and wind < 3 m/s). The main

equation used to relate concentration of the toxic plume with the distance is shown below.

$$C(x) = \frac{Q}{\pi \sigma_y \sigma_z u} \tag{9}$$

$$\sigma_y = 0.067(x)^{0.90} \tag{10}$$

$$\sigma_z = 0.057(x)^{0.80} \tag{11}$$

where C(x) is the vapor concentration, Q is the release rate, u is the wind speed and σ_y, σ_z are the dispersion coefficients in the y, z directions. The above equation needs to be manipulated such that distance to toxic endpoint becomes a function of the other parameters. The resulting equations that can be used to perform toxic consequence modeling are as shown below.

$$QR = \frac{0.284 \times U^{0.78} \times MW^{2/3} \times A \times VP}{82.05 \times T} \tag{12}$$

where QR = evaporation rate (pounds per minute), U = wind speed (meters per second) ~1.5m/s, MW = molecular weight, A = surface area of pool formed by the entire quantity of mixture (square feet), VP = vapor pressure (mmHg), T = temperature of released substance (K).

$$D_{tox} = \left(\frac{0.31505 \times QR}{MW \times LC_{50}} \right)^{1/1.7} \tag{13}$$

where D_{tox} = distance to toxic endpoint of LC_{50} (meter), QR is the vaporization rate (lb/min), LC_{50} is the lethal concentration causing 50% mortality in fathead minnow after 96 h (mol/lt).

5.3 Sensitivity analysis of single and multiple parameters

Sensitivity analysis provides a means of determining the effects of changes in input variables to desired output variables. It can aid in determining what factor needs better estimation, and identifying weak links of the assessment chain. Sensitivity can be measured using different approaches, such as local derivative based method, regression method, Morris method, variance based methods, and Monte Carlo filtering. In this work, sensitivity analysis has been performed using regression techniques to obtain standardized regression coefficients. This method can be applied to multiple parameter sensitivity analysis. Data for combinations of various parameters and their influence of process outputs can be measured using Aspen Plus simulator and the sensitivity feature described in Section 5.4. This collection of data is based on combinations of inputs obtained by varying multiple parameters, each taking discrete values as specified in the simulator. The data is standardized using the mean and the standard deviation and then the regression algorithm (such as ordinary least squares) is fed with model input and output values. A regressed meta-model is returned whereby the output Y is expressed as a linear combination of the input factors.

$$Y = \beta_s P_s + \beta_t P_t + \beta_j P_j \tag{14}$$

Y is the set of estimated/measured values of interest which has been standardized, and the parameter sets are P_s, P_t, and P_j, which are independent and standardized. The input and output data is transformed to its standardized state by first subtracting the mean value, and then dividing by the standard deviation. β_i s are called the standardized regression coefficients. The standardized regression coefficients provide a direct measure of the relative importance of the input variables.[82]

The effectiveness of the regression coefficients will be based on R_y^2 of the meta-model, which is defined as $R_y^2 = \sum_{i=1}^{N} (\hat{y}^{(i)} - \overline{y})^2 \; / \; \sum_{i=1}^{N} (y^{(i)} - \overline{y})^2$. For a measure of linearity, if the R_y^2 is larger than ~0.8 then the meta-model represents a large part of the variation in *Y*. The disadvantage of using such a regression based technique is that it is not altogether suitable for non-linear models, and can be misleading for non-monotonic models.

Thus, a measure of the sensitivity parameter S_i^{σ} can be obtained using equation (15) shown below. This is possible because standardized data has been applied to the regression analysis.[83]

$$S_i^{\sigma} = \frac{\partial Y}{\partial P_i} = \beta_i \tag{15}$$

The data collected from Aspen Plus is large in number because of the ease of obtaining simulation points. But the drawback associated with this assessment is that by simply performing data collection based on incremental values of the parameters, it does

not guarantee normal distribution in the data. This can result in some variation in the sensitivity parameters obtained using this assessment. Thus, the values should be subject to proper validation and judgment.

For single parameter sensitivity analysis, only one factor is varied and its effects can be observed by 2D graphical approach. For two parameters, the effects can be visualized by 3D graphs.

5.4 Aspen Plus® sensitivity feature and optimization feature

Sensitivity in Aspen Plus is a part of model analysis tools, and it is able to vary one or more flowsheet variables and study the effect of that variation on other flowsheet variables. Sensitivity analysis results are displayed in a table with the first n columns as the list of n variables to be varied, and the remaining columns are the variables to be estimated. The estimated variables of interest could be either internal flowsheet variables or valid Fortran expressions. Fortran expressions can be utilized to include the consequence modeling related equations as described in Section 5.2.2 The sheet for specifying variables to vary has been shown in Figure 15.

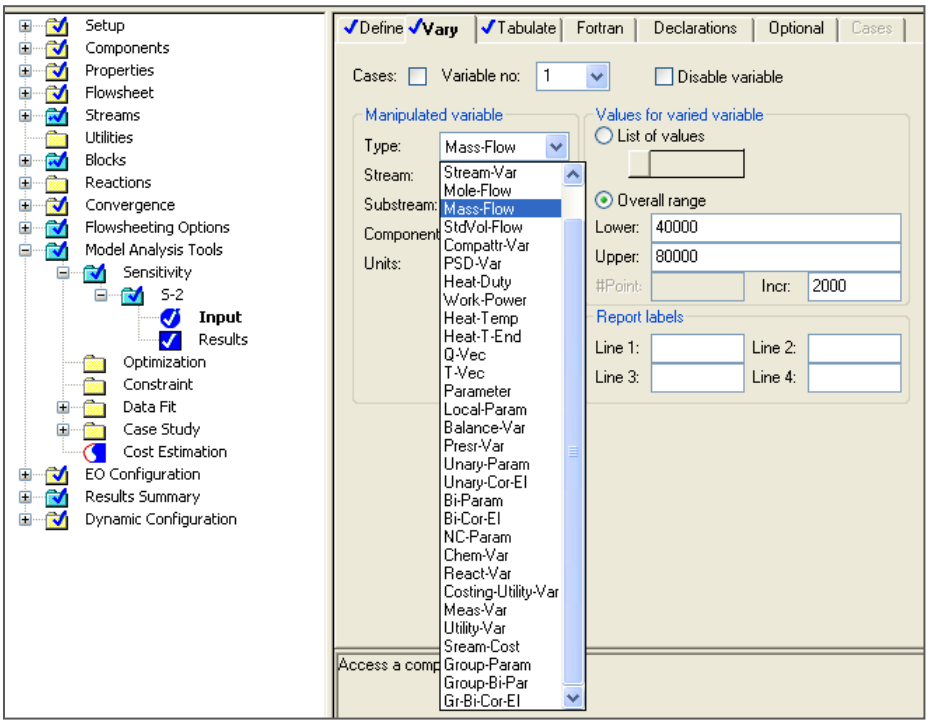


Figure 15. Snapshot of Aspen Plus sensitivity tool

Along with sensitivity option, Aspen Plus also provides optimization feature. This tool can be utilized to obtain enhanced solutions for different solvents by changing process parameters such that they optimize a user-specified objective function. The objective function can be any valid Fortran expression involving one or more flowsheet quantities. Equality or inequality constraints can be imposed, which can be functions of flowsheet variables using Fortran expressions.

Aspen Plus uses an iterative approach to solve the problems. The variables in the stream or block inputs are provided as initial estimates. The results of the optimization block are the value of the objective function and the convergence status of the

constraints. It is recommended to include sensitivity analysis before optimization to find appropriate decision variables and their ranges. Also, the results of optimization can be evaluated to find out if the optimum is broad or narrow.

Two optimization algorithms are available in Aspen Plus. The COMPLEX method uses the well-known Complex algorithm, and can handle inequality constraints and bounds on decision variables. Equality constraints must be handled as design specifications. The COMPLEX method frequently takes many iterations to converge, but does not require numerical derivatives. The SQP method is a quasi-Newton nonlinear programming algorithm. The SQP method usually converges in fewer iterations but requires numerical derivatives for all decision and tear variables at each iteration. The default optimization convergence procedure in Aspen Plus converges the tear streams and the optimization problem simultaneously, using the SQP method.

5.5 Conclusions

This section describes the approach used to integrate safety considerations at the design stage using available tools and features in commercially available process simulators, such as Aspen Plus. This approach benefits from the vast extent of information and process models available in Aspen Plus, thus eliminating the need to include process models separately when integrating safety aspects. This would be the case if a purely optimization based method was selected. The code for the optimization problem would include governing process models (such as mass balance, energy

balance, equilibrium models, rate equations, models to compute physical properties) and consequence models (such as release model, dispersion model, probit model). Simultaneously, much information for properties and other parameters would be needed and that would have to be manually inputted to the optimization program. Using Aspen Plus gives access to the in-built database of properties, process models and convergence tools. The inequality constraints (such as material flow limits, pressure, temperature and concentration upper and lower bounds, environmental constraints, and safety constraints) can be included easily using simulator options. [84]

The conclusions of this section are as shown below.

1. In order to integrate consequence modeling, guidance can be obtained from general consequence modeling equations or regulatory based requirements. General consequence equations require the user to make assumptions regarding certain physical and environmental parameters, whereas regulatory requirements contain more specific guidelines that eliminate the need for assumptions. But the drawback of regulatory guidance is that it is more simplified and cannot be applied to all types of hazards or chemicals.
2. Aspen Plus can be used to assess the inherent safety of the process based on consequence modeling by simply providing input for consequence models in terms of an excel spreadsheet or Fortran code. Two toolboxes within Aspen Plus can be used for further analysis, the sensitivity toolbox and optimization toolbox within model analysis.

3. Sensitivity analysis can be performed for single or multiple parameters. The data from the simulation can be graphed and further analyzed to obtain the sensitivity parameters. Analyzing the outputs of the models both in terms of process efficiency and process safety provides the user more information to choose better options in terms of feasible solvent and process conditions for further evaluation.
4. Optimization tool provides a means of placing constraints on the safety requirements of the process and thereby, assess the more attractive options available. Further description of benefits associated with these features has been included in the case study shown in Section 6.

6. CASE STUDY: LIQUID-LIQUID EXTRACTION OF ACETIC ACID-WATER MIXTURE*

The case study used to demonstrate this method is a liquid-liquid extraction process for an aqueous solution of acetic acid. The separation of acetic acid and water by simple rectification is very difficult and costly, requiring a column with many stages and a high reflux ratio. Thus, generally extraction is the chosen method to separate acetic acid and water. Usually, ethyl acetate and methyl isobutyl ketone are preferred solvents for this separation because of their separation power and lower boiling points. Having a lower boiling point may in turn reduce the energy costs associated with the distillation stage.

The liquid extraction process is simulated using Aspen Plus, which is able to facilitate rigorous calculation of the number of theoretical stages required provided that an accurate liquid-liquid equilibrium model is employed. The program is not able to sufficiently provide information about mass-transfer performance in terms of stage efficiency or extraction column height requirements, throughput and flooding characteristics, which can be evaluated using other methods and software.

---

* Part of this section is reprinted with permission from "Inherently safer design of solvent processes at the conceptual stage: Practical application for substitution" by S. Patel, D. Ng, M. S. Mannan, 2010. *Journal of Loss Prevention in the Process Industries,* doi:10.1016/j.jlp.2010.03.002. Copyright 2010 by Elsevier Ltd.

Also, another limitation associated with the use of a simulator is that it is dependent on the quality of the liquid-liquid equilibrium model programmed into the simulation. In most cases, an experimentally validated model is needed, and UNIFAC or other estimation methods are not sufficient.[85] But for the purpose of this study, the UNIF-LL property method is employed, which uses interaction parameters developed for liquid-liquid equilibrium applications.

An advantage of using the process simulator is that it facilitates process optimization by allowing rapid evaluation of numerous design cases. The process simulator is suitable because it contains the model of the process which is the bulk of the constraints in an optimization problem. These equality constraints in the simulator include all the mathematical relations such as material balance, energy balance, rate equations, phase relations, and methods of computing physical properties.

6.1 Molecular design and inherently safer solvents

Target property requirements are specified in the ProCAMD module of ICAS. The non-temperature dependent properties specified are boiling point greater than 400 K, open cup flash point greater than 300 K, and log P (octanol/water partition coefficient) less than 3.5. Mixture properties such as selectivity, solvent losses and solvent capacity are also specified. Upon running the module, total number of compounds selected or generated (along with isomeric structures) are 308. Figure 16 shows the flash point and octanol-water partition coefficient estimated for the 308

selected solvents, which includes solvents identified in the internal database as well. Solvents that were carried forward in the extraction simulation step, as shown in Table 16, were chosen such that they were also found in the ICAS internal chemical database. For these representative solvents, Figure 17 shows the estimated open cup flash point temperature versus estimated values of $-\log(LC_{50})$. The identified solvents satisfy requirements for both types of hazardous properties. Some solvents have lower toxicity levels (5-methyl-2-hexanone, and 3-heptanone), while some solvents possess higher flash points (5-nonanone, and 2-nonanone). For the purpose of this case study, 10 solvents which are identified in the database for acetic acid-water separation will be carried forward for further analysis because these solvents can be easily characterized in the simulator software used in Step 2. The performance of ethyl acetate (which is a commonly used solvent for acetic acid-water separation) is also assessed and compared with the solvents identified using ProCAMD. Ethyl acetate has a flash point of 280.2 K and LC_{50} of 45 mg/lt. Solvents with lower flash point would indicate higher tendency to ignite, thus increasing probability of occurrence of fires or explosions.

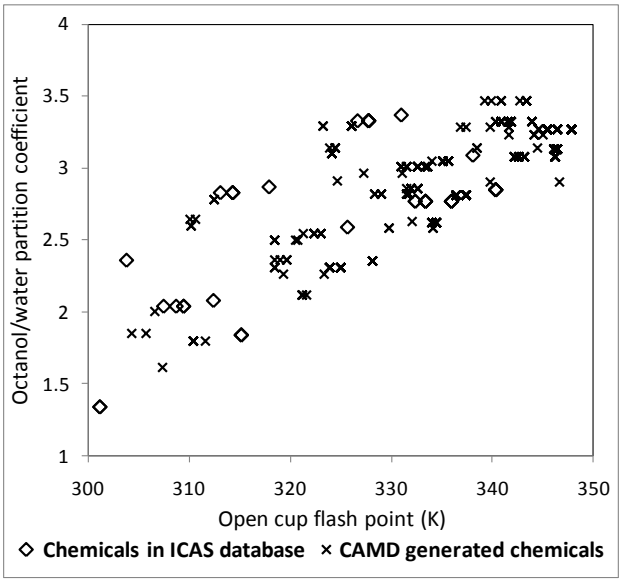


Figure 16. Hazardous properties for solvents selected using ProCAMD

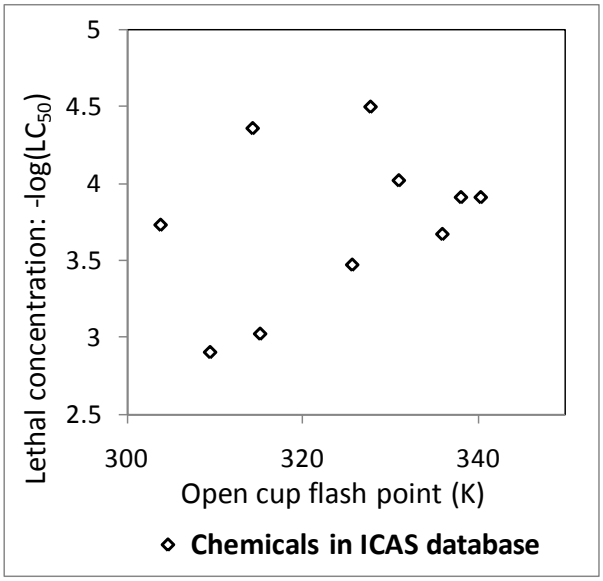
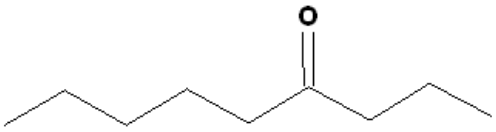
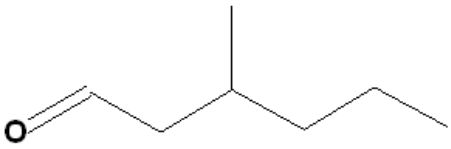
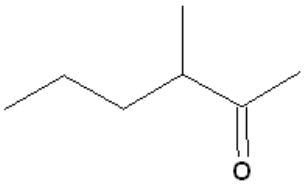
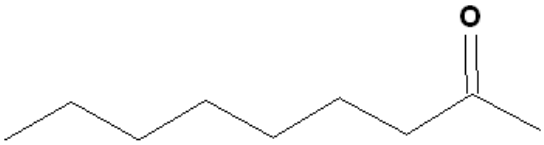
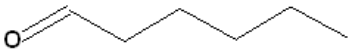
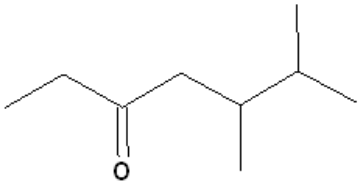
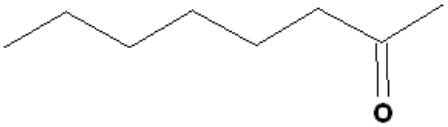
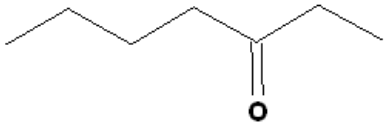
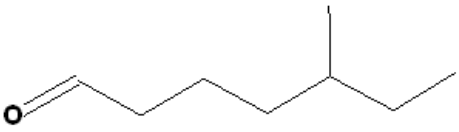


Figure 17. Hazardous properties for solvents listed in Table 16

Table 16. Solvents selected using ProCAMD module in ICAS 11.0

5-nonanone	
2-methyl hexanal	
5-methyl-2-hexanone	
2-nonanone	
1-hexanal	
diisobutyl ketone	
1-octanal	
2-octanone	
3-heptanone	
2-ethyl hexanal	

6.2 Process design with consequence modeling

A simple flow diagram used for simulating the extraction and solvent recovery section has been shown in Figure 18. This flow diagram is intended only for depicting the initial conceptual design stage, primarily for screening solvents based on separation capability and safety aspects (related to explosion overpressures). The input to the simulation for feed streams and column characteristics has been shown in Table 17. The process simulator used for this case study is Aspen Plus, Version 2006®.

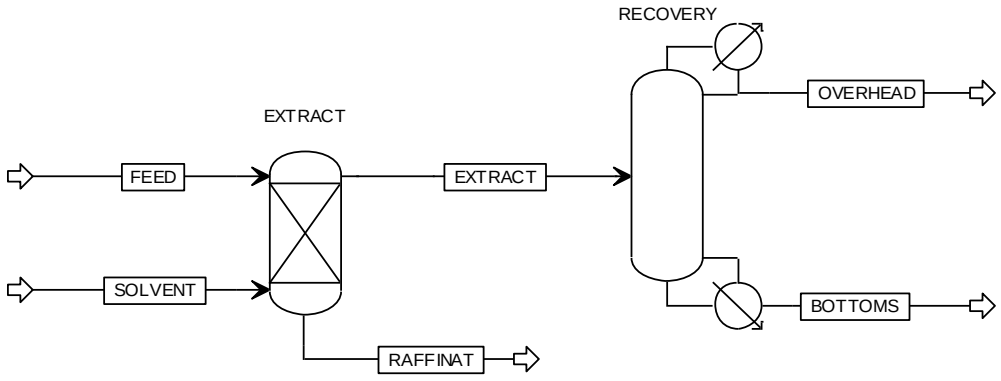


Figure 18. Flow diagram for extraction and solvent recovery

Table 17. Simulation inputs for extraction and solvent recovery

Feed stream	Acetic acid – 6,660 lb/hr; Water – 23,600 lb/hr; Temperature – 100˚F; Pressure – 100 psi
Extraction column	Stages – 6; Pressure – 100 psi
Solvent recovery column	Recovery: Light key component – 0.95; Heavy key component – 0.05
Solvent stream	Vary flowrate to assess effects on separation and safety

In order to integrate safety modeling into the process simulation, models as specified in Section 5.2.2 can be used, for example in this case study equations (7) and (8) are simultaneously calculated using properties and parameters obtained from the database within the simulator. For flammable hazards, the solvent-related parameters of interest for this calculation are molecular weight, volumetric flowrate (that is used to calculate the area of the liquid pool), vapor pressure, temperature, and heat of combustion. These parameters are provided as inputs to the calculator block within flowsheeting options and corresponding calculations are performed using an excel spreadsheet (Figure 19). Similarly, spreadsheet evaluation can be included for the calculation of toxic hazards as described in section 5.2.2.2. Aspen Plus also allows for the inclusion of simple Fortran code for flammable and toxic consequence modeling and this has been shown in Sections 6.3.4 and 6.4.4

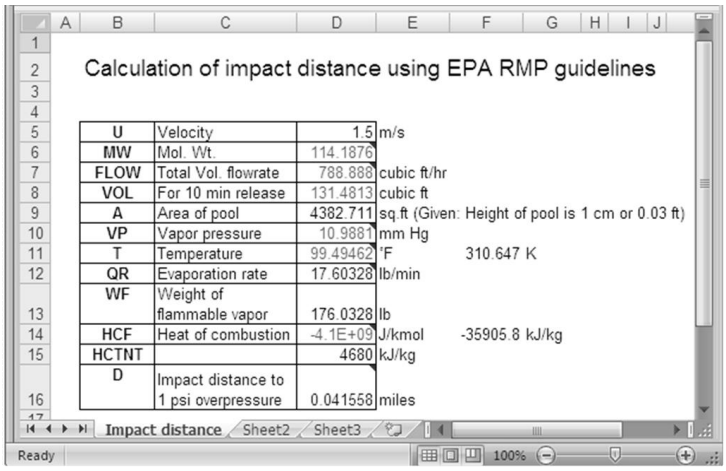


Figure 19. Microsoft Excel spreadsheet linked with calculator option in Aspen Plus®

6.3 Results and discussion for flammable hazards

6.3.1 Single parameter sensitivity analysis

Furthermore, in order to assess how changing a single parameter, such as flowrate of the solvents affects the separation characteristics and inherent safety of the process, the sensitivity block within the model analysis tools of the simulator is used. The flowrate is generally varied from 40,000 lb/hr to 80,000 lb/hr for all solvents identified. The calculator and sensitivity analysis blocks are coupled to obtain simultaneous results for process simulation and consequence modeling, thereby reducing the time to perform the assessment.

The results of the simulation assessment are plotted to observe and analyze the viable options for separation. Figure 20 shows the extent of separation in terms of acetic acid in the extract phase versus the impact distances calculated for 1 psi overpressure. 11 solvents were evaluated and their individual trends consisted of multiple datapoints that resulted from modifying the solvent flowrate from 40,000 to 80,000 lb/hr. It can be observed that solvents such as 5-nonanone, 2-nonanone have shorter impact distances, whereas 5-methyl-2-hexanone and ethyl acetate have better separating power. Decision for appropriate solvent to be used would need to be based on the trade-offs and design requirements. Figure 21 shows the reboiler heat duties in the solvent recovery column versus the impact distances. The energy requirements of solvent recovery greatly affect the economic viability of the overall extraction process. Thus, it is important to assess the reboiler heat duty to make an estimation of the associated costs. Again, 5-nonanone

and 2-nonanone have low requirements for heat duty, but 5-methyl-2-hexanone shows a trend of greatly increasing heat loads. In order to achieve the same amount of separation the required flowrate of solvents with inherently safer characteristics is more than the flowrate of ethyl acetate, but the consequence associated with the inherently safer solvents remain lower because of the combined effects of other intrinsic properties such as vapor pressure and heat of combustion. Thus, application of such a preliminary screening approach identifies viable candidates for the separation process taking into account trade-offs between process efficiency and inherent safety.

6.3.2 Dual parameter sensitivity analysis

Figure 20 and Figure 21 show the variation of one variable and its effects on the parameters of interest. Similarly, the effects of two variables can be visualized by using 3D graphs or surface plots. Thus, having changed the solvent flowrate and the temperature of the solvent simultaneously, the associated effects on acetic acid in extract phase and the impact distance for 1 psi overpressure can be measured and plotted. For the conventional solvent ethyl acetate the effects on acetic acid in extract phase and distance to overpressure are shown in Figure 22 and Figure 23. Similarly, plots for the inherently safer solvent, 5-nonanone, are shown in Figure 24 and Figure 25. It can be seen that the conventional solvent, ethyl acetate, is better for separation but does not perform well in terms of the safety measure.

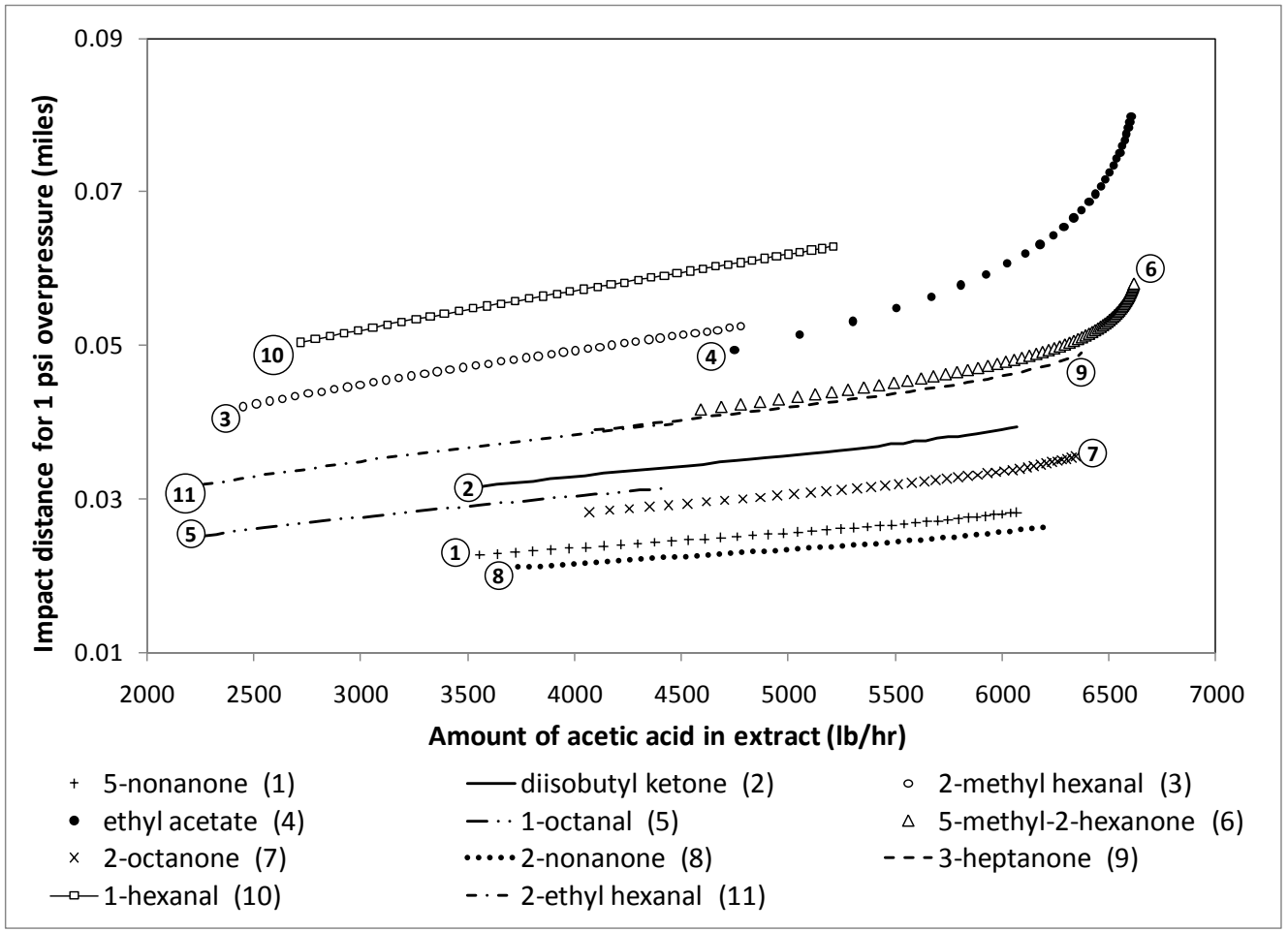


Figure 20. Amount of acetic acid separated versus impact distance for 1 psi overpressure

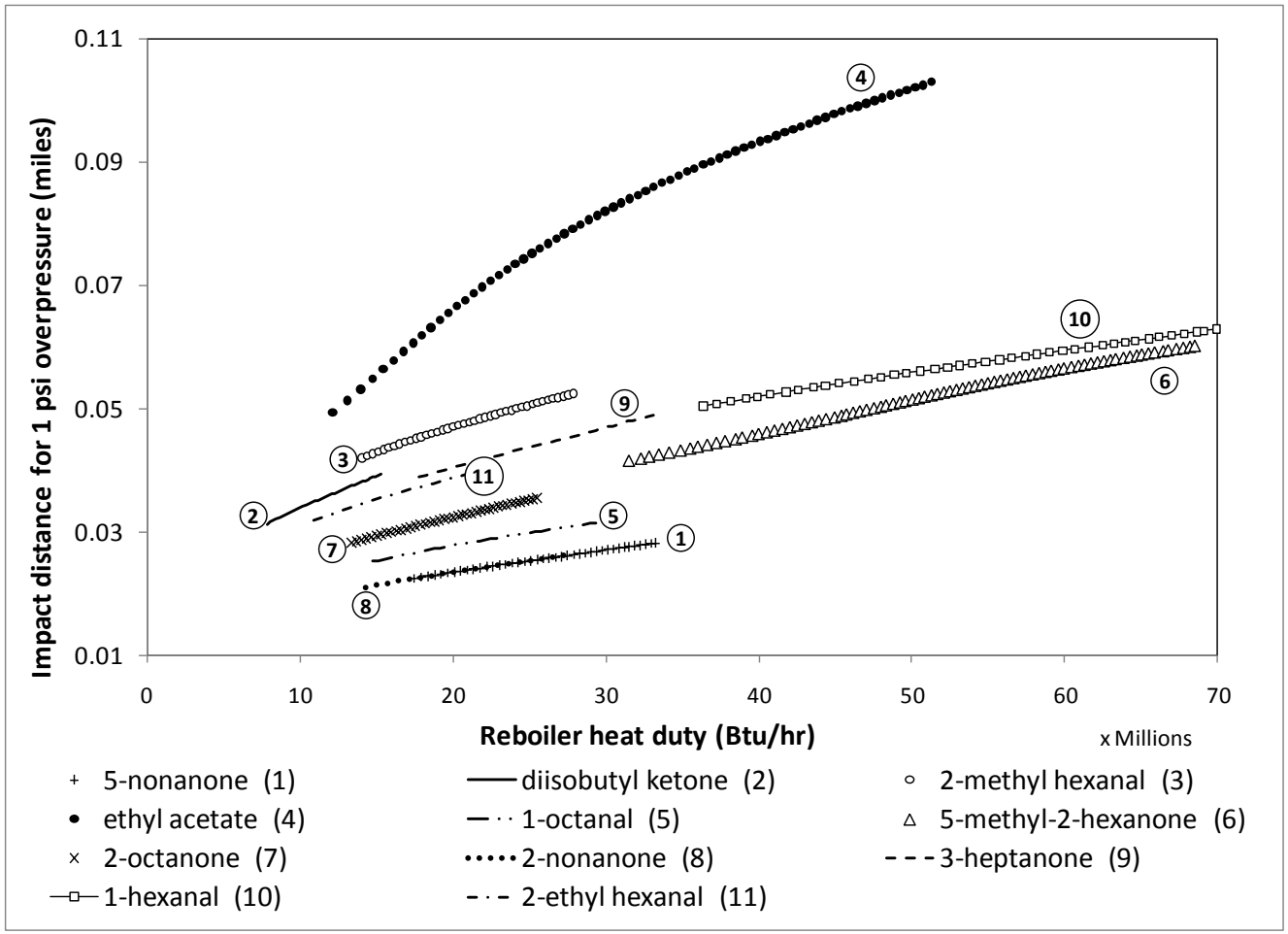


Figure 21. Reboiler heat duty versus impact distance for 1 psi overpressure

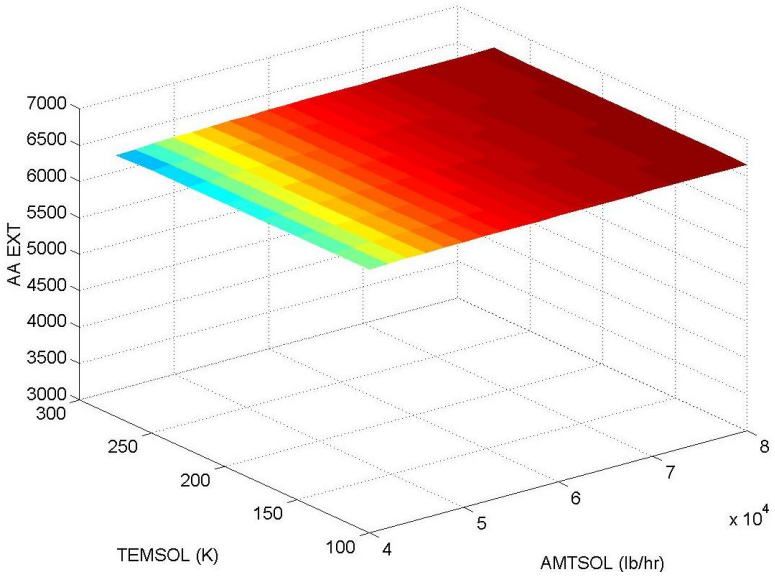


Figure 22. Ethyl acetate: Amount of acetic acid in extract phase upon varying 2 parameters

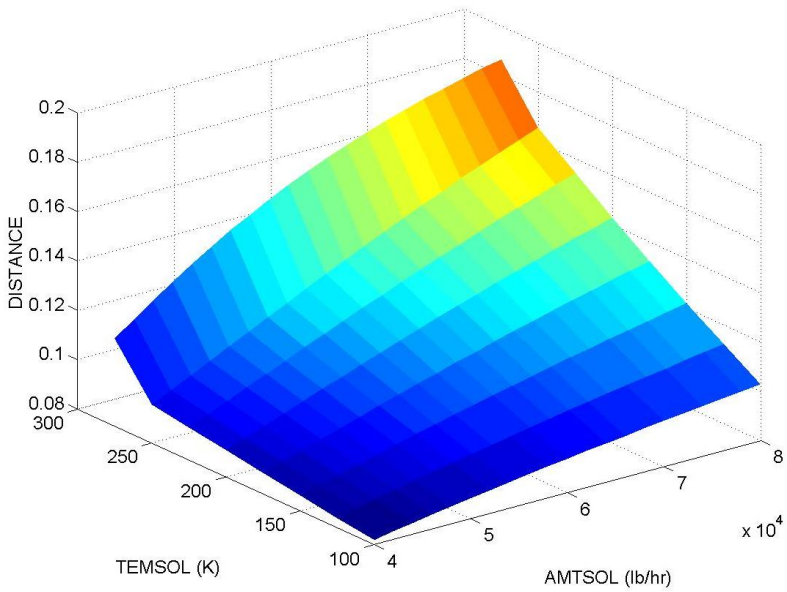


Figure 23. Ethyl acetate: Distance to 1 psi overpressure upon varying 2 parameters

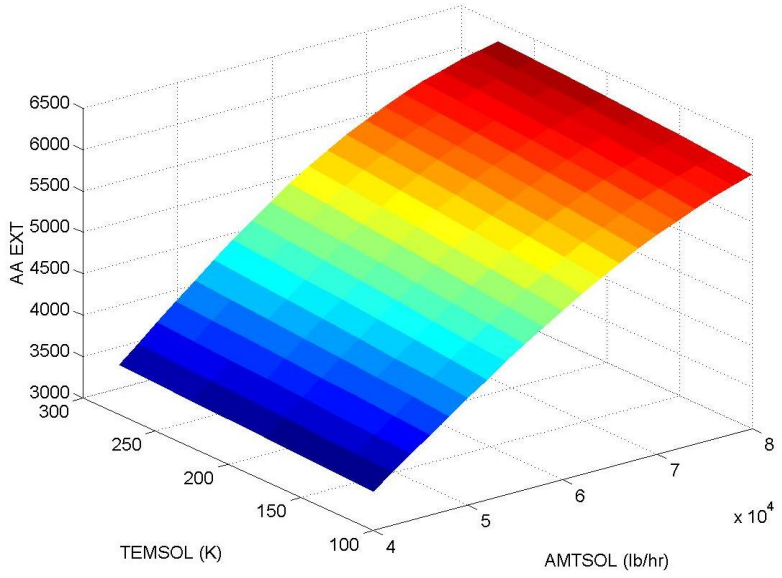


Figure 24. 5-nonanone: Amount of acetic acid in extract phase upon varying 2 parameters

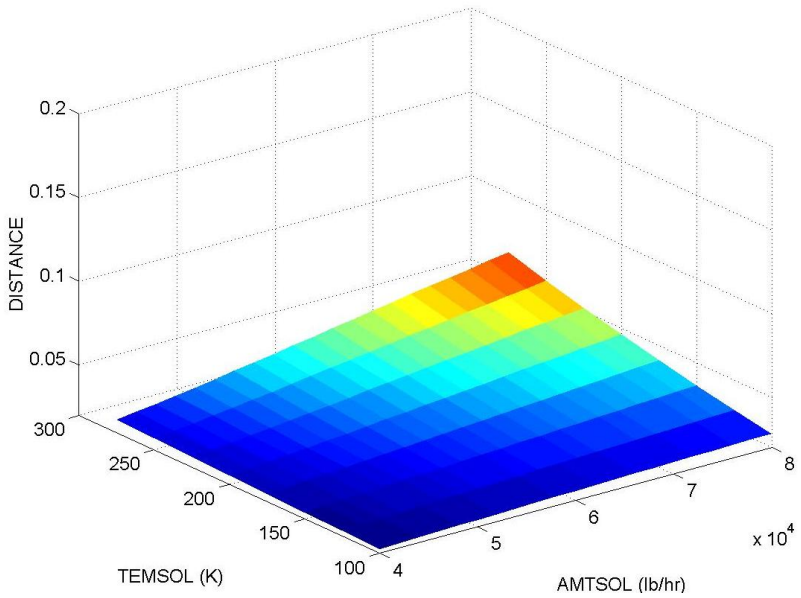


Figure 25. 5-nonanone: Distance to 1 psi overpressure upon varying 2 parameters

Also, temperature plays an important role for distance calculations associated with ethyl acetate, but effect of temperature on measuring acetic acid separation is nominal. Assessing such visual aids provides the user with a crude measure of suitable operating conditions that can be employed for safer operation. In the case of 5-nonanone, safer operating conditions that yield higher separation are amount of flowrate: 70,000 to 80,000 lb/hr and temperature of solvent: 250 to 300 K.

6.3.3 Multiple parameter sensitivity analysis

Using the same process for acetic acid- water separation in Aspen Plus and varying multiple parameters, further analysis can be performed to judge the effect of multiple parameters on process efficiency and safety. In this case study, three parameters were changed: solvent flowrate, temperature and pressure. The flowrate was changed from 40,000 lb/hr to 80,000 lb/hr, the temperature of the solvent feed was changed from 100 K to 300 K, and the pressure of the solvent feed was changed from 100 to 300 psi. Data points were obtained for all combinations of discrete values and thereafter, the results were interpreted using regression based sensitivity analysis.

The total number of data points collected through the sensitivity analysis are

$$\begin{aligned} 18700 = 11 \text{ solvents} \times 21 \text{ increments in amount of solvent} \\ \times 9 \text{ increments in pressure} \times 9 \text{ increments in temperature} \end{aligned}$$

Regression model was developed for a set of variables, which includes the ones used as input to consequence modeling for calculation of distance to overpressure. The input data includes properties of solvents such as molecular weight MW, vapor pressure VP, heat of combustion of solvent HCF, as well as process conditions such as temperature of the stream TEM and amount of solvent in the stream AMTSOL. The regression model developed is as shown in Table 18 and the fit of the regressed data with the actual data can be seen in Figure 26.

Table 18. Regression model #1 for distance to overpressure

$$\text{Distance to overpressure}^* = 0.101 \times MW^* + 0.177 \times VP^* + 0.431 \times TEM^* + 0.136 \times AMTSOL^* + 0.753 \times HCF^*$$

R^2	0.9618,	F	94240,	Standard error for y estimate	0.195

Note: Distance to overpressure*, MW*, VP*, TEM*, AMTSOL*, HCF* are standardized data obtained from original values

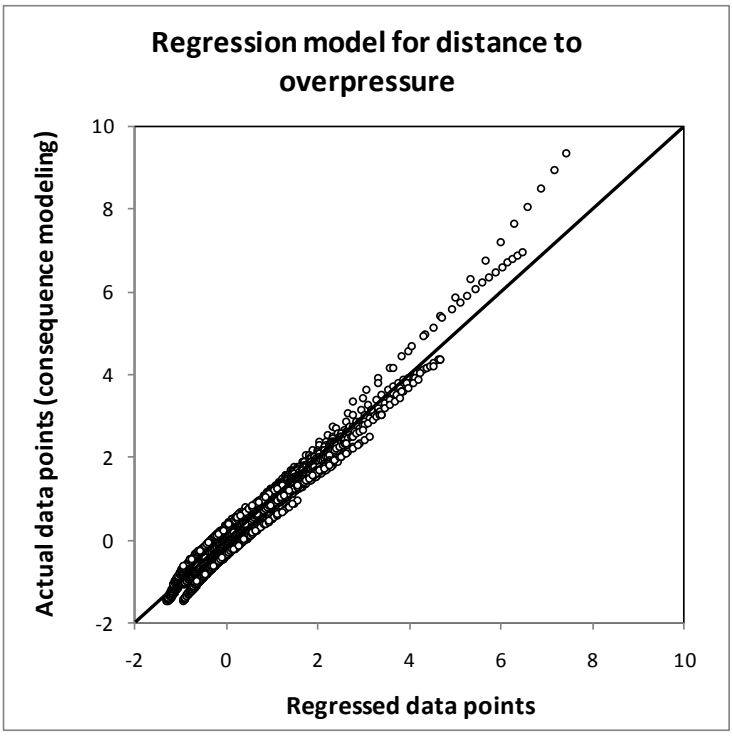


Figure 26. Actual data points from consequence modeling versus regressed data points for model #1

Thus, using the standardized regression coefficients $S_i^{\sigma} = \beta_i$ the sensitivity parameters can be found and plotted as shown in Figure 27. Heat of combustion when expressed in the units of J/kmol has a direct effect on the safety measure as shown below, and when expressed in J/kg has an inverse effect. It is to be noted that the heat of combustion considered here is negative because of exothermic reaction.

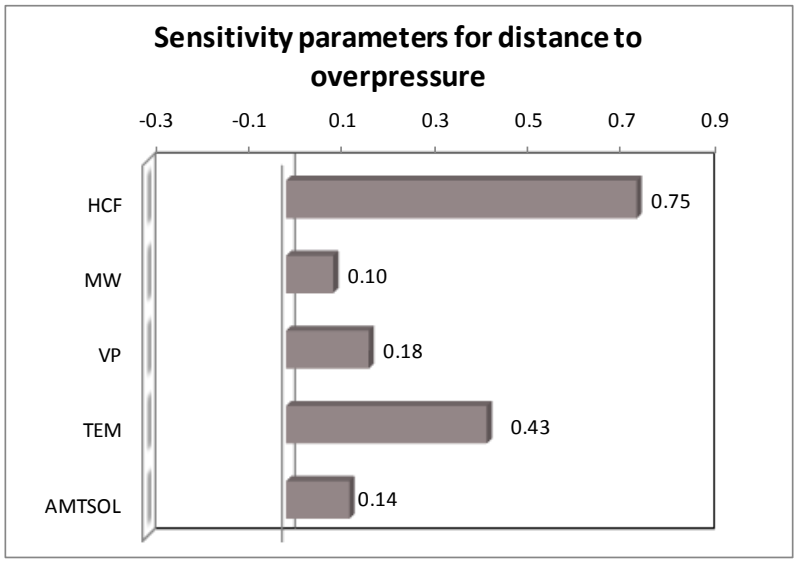


Figure 27. Sensitivity parameters for variables in model #1

Results in Figure 27 show that the distance to 1 psi overpressure is affected strongly by heat of combustion, and temperature of the stream. But parameters such as molecular weight, vapor pressure and amount of solvent have little effect on the output variable. This is contrary to popular belief that the amount of solvent in the stream has greater impact for higher consequence. Also, such results show that physical properties other than flash point, and flammability limits need to be incorporated into the evaluation of inherently safer processes and the associated index calculations.

6.3.4 Optimization within simulator

Another type of assessment that can be carried out to investigate the different solvents identified is optimizing the process with safety constraints. If the process can be

optimized such that it maximizes process efficiency or minimizes cost, but at the same time also satisfies constraints on safety, then valuable information can be obtained in order to make an informed decision regarding solvents to be used in the process and the operating conditions.

The basic formulation of the optimization problem is as shown below.

$$\max \quad Acetic\,acid\,in\,extract\,phase$$

$$\text{subject to} \quad \left\{ \begin{array}{c} governing\,process\,model\,for\,countercurrent\,extraction\,of \\ a\,liquid\,stream\,using\,a\,solvent \end{array} \right\}$$

$$QR = \frac{0.284 \times U^{0.78} \times MW^{2/3} \times A \times VP}{82.05 \times T}$$

$$D_{flam} = 0.0081 \times \left(QR \times \frac{HC_f}{HC_{TNT}} \right)^{1/3}$$

$$D_{flam} < 0.1 \text{miles}$$

$$40000 < AMTSOL < 80000$$

$$100 < TEMSOL < 300$$

$$100 < PRESSOL < 300$$

The Fortran code for the consequence modeling to be included in Aspen Plus is

```
F       AREA=(VFLOW*10.0)/60.0/0.03
F       QR=0.3896*(MW**0.67)*AREA*VP/(82.05*(TEM+459.67)*5/9)
F       DIST=0.0081*(QR*(-1.0*HCF/(MW*1000.0)/4680))**(1.0/3.0)
```

Where VFLOW is the solvent flowrate in extract stream, MW is the molecular weight of the solvent, VP is the vapor pressure of the extract stream, TEM is temperature of extract stream, and HCF is the heat of combustion of the solvent. User defined parameters are specified in the 'define' sheet: DIST is the distance in miles to 1 psi overpressure, AREA is the area of the pool formed by quantity released in 10 minutes, and QR is the amount of solvent vaporized. In the EPA RMP guidelines it is specified that the entire quantity of material is released and forms a pool. But in this work it has been assumed that a representative amount to consider for formation of the pool is the release that takes place for 10 minutes at the flowrate in the extract stream. This is based on judgment which can be appropriate for evaluating worst case type scenarios.

Having performed the same assessment on 11 solvents identified in Table 16, unique solutions are found for the optimized values. The input variables and output variables from the optimization are shown in Table 19.

Table 19. Optimization results upon considering flammability hazard

Solvent name	Input flowsheet variables/ decision variables								Output variables		Objectives and constraint values		
	Amount of solvent used (AMTSOL)	Temperature of solvent (TEMSOL)	Pressure of solvent (PRESSOL)	Solvent molecular weight (MW)	Volumetric flowrate (VFLOW)	Solvent vapor pressure (VP)	Extract phase temperature (TEM)	Solvent Heat of combustion (HCF)	Area of pool (AREA)	Vaporization rate (QR)	Flash point (Tf)	Amount of acetic acid in extract phase (AAEXT)	Distance to overpressure (DIST)
	lb/hr	F	psi	g/mol	ft^3/hr	mm Hg	F	J/kmol	sqft	lb/min	K	lb/hr	mile
5-nonanone	80000	300	294	142	1567	27	195	-5.32E+09	8706	85	333	6295	0.0712
diisobutyl ketone	70000	300	156	142	1398	41	180	-5.31E+09	7767	118	322	5976	0.0794
2-methyl hexanal	80000	263	101	114	1600	85	174	-4.13E+09	8891	243	305	5582	0.0999
ethyl acetate	73051	100	196	88	1298	170	100	-2.06E+09	7212	377	269	6658	0.1001
1-octanal	70000	300	100	128	1369	47	196	-4.74E+09	7608	120	324	4946	0.0795

Solvent name	Input flowsheet variables/ decision variables								Output variables		Objectives and constraint values		
	Amount of solvent used (AMTSOL)	Temperature of solvent (TEMSOL)	Pressure of solvent (PRESSOL)	Solvent molecular weight (MW)	Volumetric flowrate (VFLOW)	Solvent vapor pressure (VP)	Extract phase temperature (TEM)	Solvent Heat of combustion (HCF)	Area of pool (AREA)	Vaporization rate (QR)	Flash point (Tf)	Amount of acetic acid in extract phase (AAEXT)	Distance to overpressure (DIST)
	lb/hr	F	psi	g/mol	ft^3/hr	mm Hg	F	J/kmol	sqft	lb/min	K	lb/hr	mile
5-methyl-2-hexanone	70000	298	122	114	1382	100	183	-4.10E+09	7680	245	309	6396	0.1000
2-octanone	80000	300	100	128	1570	54	200	-4.70E+09	8721	158	344	6461	0.0870
2-nonanone	80000	300	101	142	1566	26	203	-5.32E+09	8701	82	337	6387	0.0703
3-heptanone	80000	268	100	114	1571	88	183	-4.10E+09	8729	245	310	6470	0.1000
1-hexanal	79999	218	100	100	1578	94	154	-3.52E+09	8766	250	300	5763	0.1000
2-ethylhexanal	80000	300	256	128	1566	64	192	-4.73E+09	8701	189	317	5434	0.0926

The results from Table 19 can be plotted to represent multiple objectives by using a contour plot (colored cell) technique. One objective is maximizing amount of acetic acid extracted and the other objective is to minimize hazard. Figure 28 shows the plot of hazards of the individual solvents such as flash point (indicating the tendency of the solvent to ignite) and distance to overpressure (indicating the severity of worst case scenario for the solvent). The color variation in the plot mimics the variation in the objective function (i.e. acetic acid in the extract phase). Thus, the zones formed by the intersection of higher objective function values, higher flash point and lower impact distances are preferred, and solvents within these zones are preferred over others in terms of efficiency and safety.

The figure shows that the solvents in the preferable zones on the plot are 1 (5-nonanone), 7 (2-octanone), and 8 (2-nonanone). Even though solvent 4 (ethyl acetate) has greater separation power; it is not favorable in terms of the flammability hazard measures.

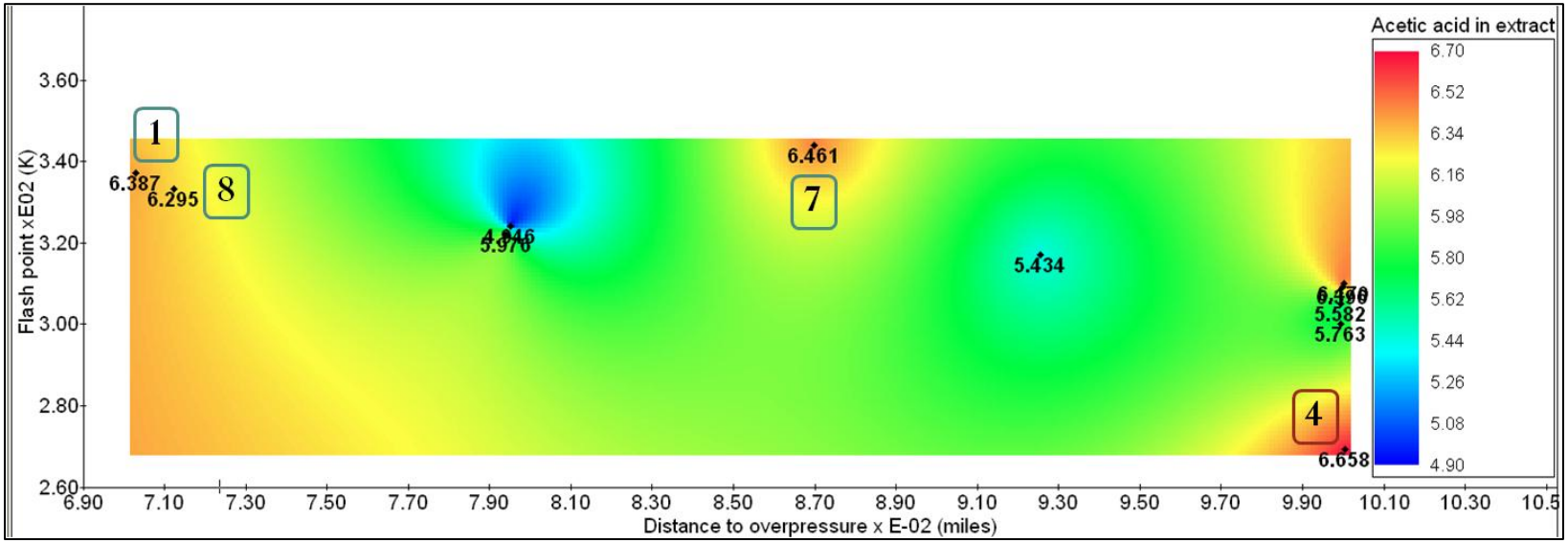


Figure 28. Contour colored cell plot of optimization results considering flammability hazards (1: 5-nonanone, 7: 2-octanone, 8: 2-nonanone, 4: ethyl acetate)

6.3.5 Validation of EPA model with available PHAST models

The EPA RMP worst case consequence models used in this work are simplified both in terms of equations used and in terms of assumptions followed. The major assumptions that govern this consequence modeling are that the pool is formed from the entire contents of the vessel or pipe, and that the amount of material participating in the vapor cloud explosion is the amount of material being evaporated in 10 minutes. The equations employed for estimating the effects are also simplified. The equations have been developed based on the TNT equivalence method. This method relies on calculating the mass of TNT that is equivalent to the effective flammable mass in the cloud, and thereafter the overpressure is calculated as a function of the TNT mass and explosion radius. Fundamentally, this method assumes that all VCEs are detonations that behave like a condensed-phase high explosive.

Thus, an enhanced consequence model that can be applied to such work is the Baker-Strehlow-Tang method that uses a continuum of numerically determined pressure and impulse curves that are based on the Mach number of the VCE flame front. The Baker-Strehlow-Tang method provides guidance on selecting a flame speed based on broad categories of congestion (high-medium-low) and fuel reactivity (high-medium-low).[86] Unlike the TNT methods where only one graph is available for scaled overpressure against scaled distance, the Baker-Strehlow-Tang method provides graphs for eleven different values of flame speed.

Both types of consequence modeling techniques described above can be implemented using PHAST 6.53.1 (Process Hazard Analysis Software Tool) by Det Norske Veritas, which is a comprehensive consequence analysis tool. It is capable of examining the process of a complete hazardous scenario from the initial release to far field dispersion, including modeling of pool vaporization and evaporation, and flammable and toxic effects.

The total material participating in the vapor cloud explosion was specified for both ethyl acetate and 5-nonanone. These values were taken from the optimization results. Also, for the materials, ethyl acetate and 5-nonanone, various properties such as critical temperature, pressure, boiling point, flammability properties, DIPPR relations for surface tension, thermal conductivity, viscosity, heat capacity, vapor pressure, and density, were needed to be given as input to the software.

The results from the consequence models used in this work versus the TNT and Baker-Strehlow-Tang have been shown in Table 20. The results for EPA RMP model and the TNT model from PHAST show good agreement because both models are similar in form. On the other hand, the results for Baker-Strehlow-Tang method are higher than the TNT method. This may be preferable because the method can provide a more realistic estimation of the effects based on phenomena such as congestion due to obstacle density, and effect of ground reflectance on the explosion overpressure. The graphical results from PHAST are as shown in Figure 29 .

Table 20. Comparing results for distance to overpressure with PHAST models

Distance to 1 psi overpressure (miles)	EPA RMP models	TNT model	Baker-Strehlow-Tang model
Ethyl acetate	0.1001	0.1012	0.1967
5-nonanone	0.0712	0.0790	0.1550
Assumptions		• 10% explosion efficiency • Ground burst	• Low material reactivity • 2D flame expansion • Medium obstacle density • Ground reflection factor -2

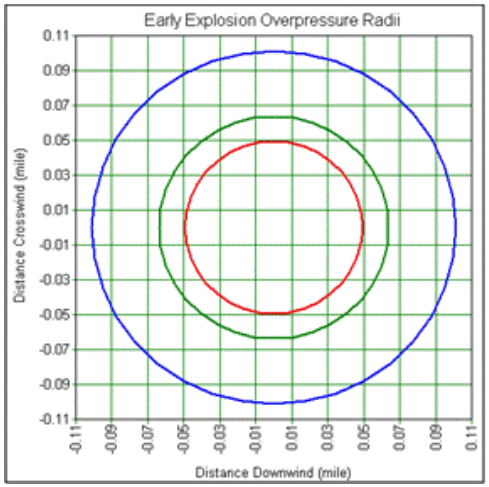
TNT model

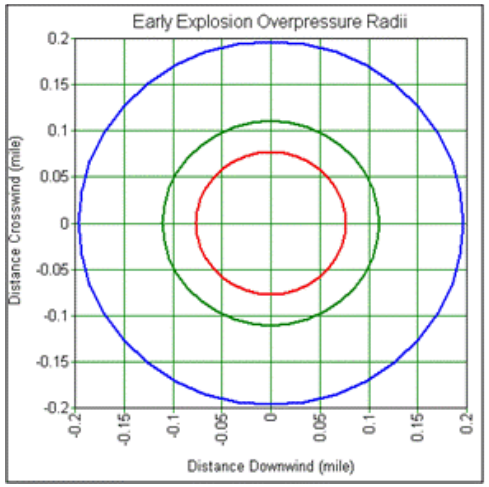
Baker-Strehlow-Tang model

Ethyl acetate

Weather: Category 1.5/F

1 psi 2 psi 3 psi

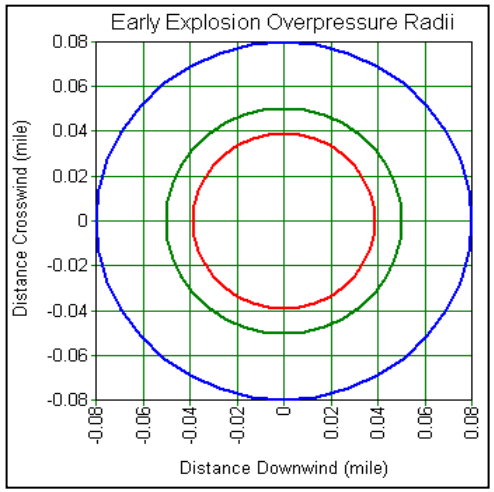




5-nonanone

Weather: Category 1.5/F

1 psi 2 psi 3 psi



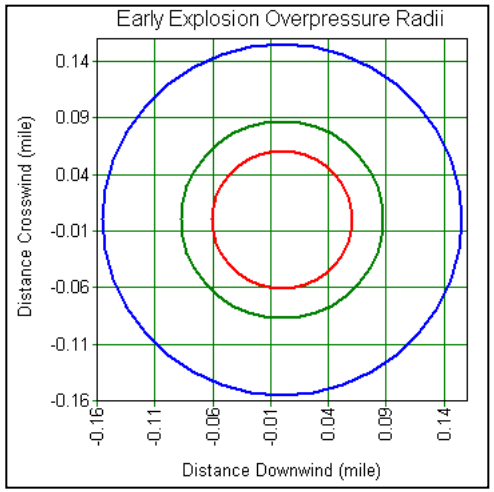


Figure 29. PHAST graphs for TNT and Baker-Strehlow-Tang models

6.4 Results and discussion for toxic hazards

6.4.1 Single parameter sensitivity analysis

Similar to the assessment for flammable hazards, within this section the toxic hazard has been assessed. A single parameter is varied, such as flowrate of the solvents to assess the effects on separation characteristics and inherent safety of the process. The sensitivity block within the model analysis tools of the simulator is used and the flowrate is varied from 40,000 lb/hr to 80,000 lb/hr for all solvents identified. The results of the simulation assessment are plotted to observe and analyze the viable options for separation. Figure 30 shows the extent of separation in terms of acetic acid in the extract phase versus the LC_{50} toxic endpoint distances calculated. It can be observed that solvents such as 2-nonanone, 2-octanone, 5-methyl-2-hexanone have shorter distances to the toxic endpoint, whereas ethyl acetate has better separating power. Similar to flammable hazards, in order to achieve the same amount of separation the required flowrate of solvents with inherently safer characteristics is more than the flowrate of ethyl acetate, but the consequence associated with the inherently safer solvents remain lower because of the combined effects of other intrinsic properties such as vapor pressure and LC_{50}.

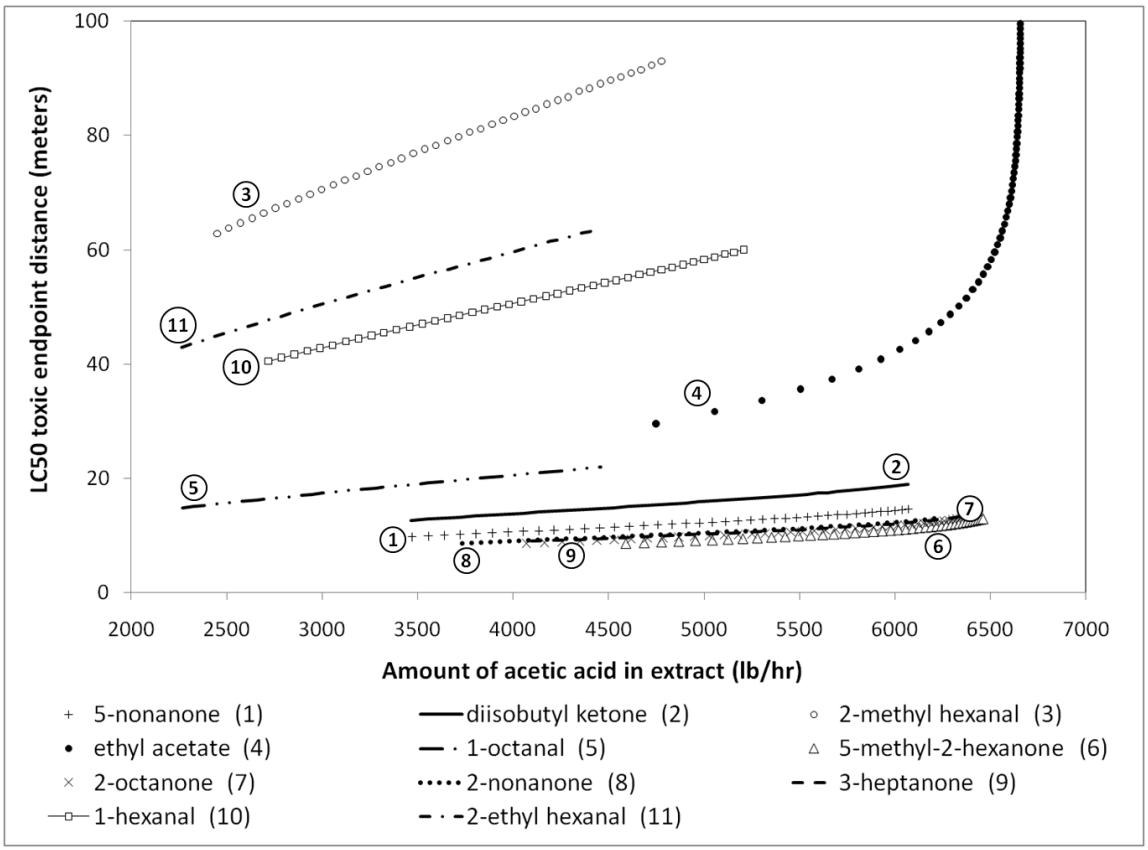


Figure 30. Amount of acetic acid separated versus distance to toxic endpoint

6.4.2 Dual parameter sensitivity analysis

Two variables such as temperature and amount of solvent can be changed and the corresponding effects on amount of acetic acid in extract phase and the distance to toxic endpoint can be measured and visualized. Figure 31 and Figure 32 show the results for conventional solvent, ethyl acetate. Figure 33 and Figure 34 show results for the inherently safer solvent, 5-methyl-2-hexanone. There are certain regions on the graphs where the performance of the inherently less-toxic solvent is comparable to the conventional solvent, ethyl acetate. Thus, suitable operating conditions for the inherently less toxic solvent, 5-methyl-2-hexanone are a temperature range from 250 to 300 K and solvent flowrate from 70,000 to 80,000 lb/hr.

Although the toxic, health related consequences associated with the less toxic solvent are mild in comparison to ethyl acetate, other hazards for storage and transportation would have to be evaluated. For larger quantities being used in the process, storage requirements would increase possibly leading to higher risk. Other properties of the material such as corrosivity should also be considered in the final selection.

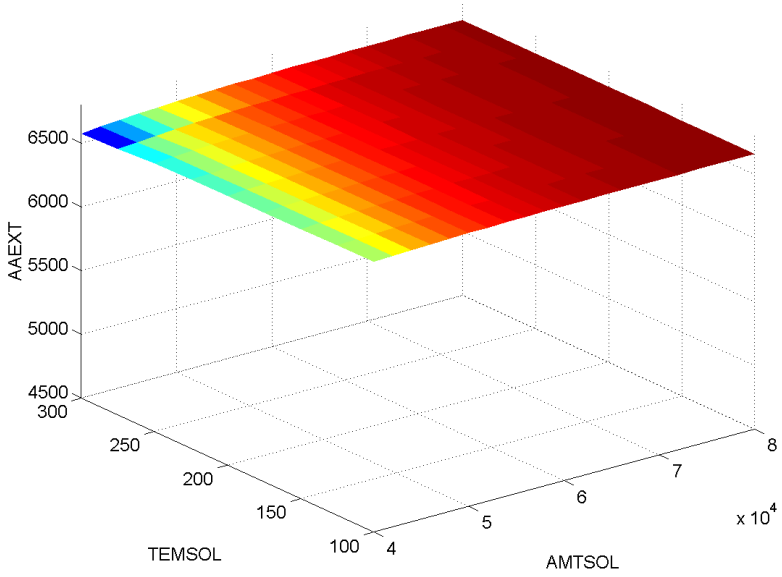


Figure 31. Ethyl acetate: Amount of acetic acid in extract phase upon varying 2 parameters to compare toxicity modeling results

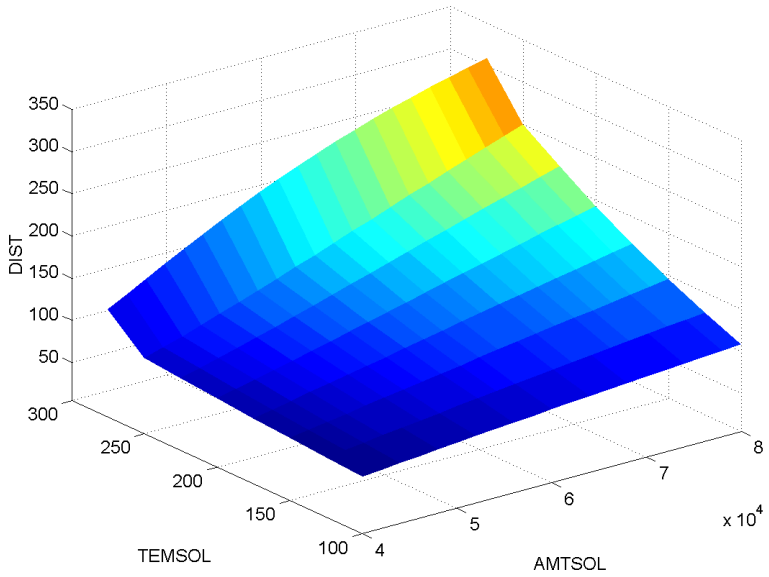


Figure 32. Ethyl acetate: Distance to toxic endpoint upon varying 2 parameters

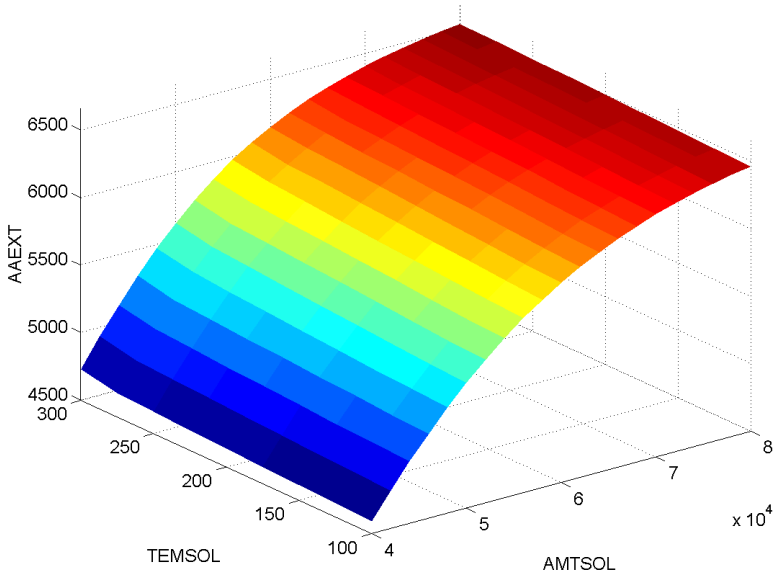


Figure 33. 5-methyl-2-hexanone: Amount of acetic acid in extract phase upon varying 2 parameters

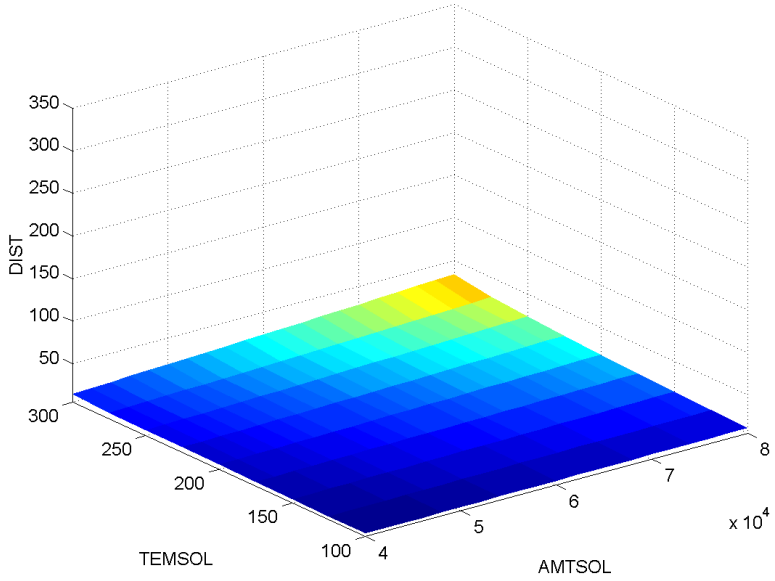


Figure 34. 5-methyl-2-hexanone: Distance to toxic endpoint upon varying 2 parameters

6.4.3 Multiple parameter sensitivity analysis

Similar to the flammable hazard assessment, three parameters were changed: solvent flowrate, temperature and pressure. The flowrate was changed from 40,000 lb/hr to 80,000 lb/hr, the temperature of the solvent feed was changed from 100 K to 300 K, and the pressure of the solvent feed was changed from 100 to 300 psi.

The total number of data points collected through the sensitivity analysis is 18700. Regression model was developed for a set of variables that includes the ones used as input to consequence modeling for calculation of distance to toxic endpoint, which are properties of solvents such as molecular weight MW, vapor pressure VP, lethal concentration LC_{50}, as well as process condition such as temperature of the stream TEM that is released and amount of solvent in the stream AMTSOL. The regression model is as shown in Table 21 and the fit of the regressed data with the actual data can be seen in Figure 35.

Table 21. Regression model #2 for distance to toxic endpoint

$$\text{Distance to toxic endpoint}^* = -0.460 \times LC_{50}{}^* + 0.066 \times AMTSOL^* - 0.557 \times MW^* + 0.045 \times VP^* + 0.191 \times TEM^*$$

R^2 0.8553, F 22123, Standard error for y estimate 0.38

Note: Distance to toxic endpoint*, MW*, VP*, TEM*, AMTSOL*, LC_{50}* are standardized data obtained from original values

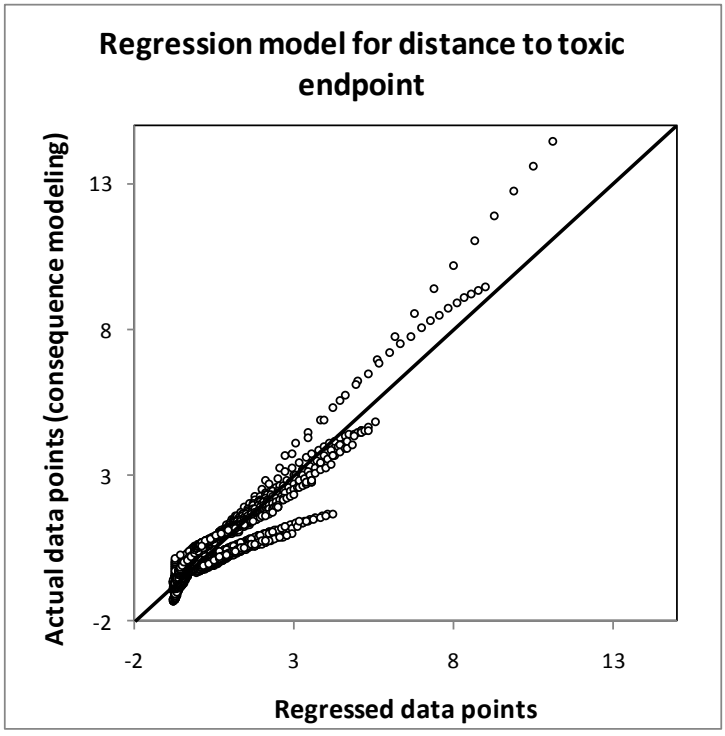


Figure 35. Actual data points from consequence modeling versus regressed data points for model #2

The sensitivity parameters can be found and plotted as shown in Figure 36. Results in Figure 36 show that the dependence of the distance to toxic endpoint is affected strongly by molecular weight, vapor pressure and LC_{50}. The sensitivity parameters indicate that for evaluation of the extent of toxic effects of materials and any associated index calculations, the properties of interest are vapor pressure, LC_{50} and molecular weight of the material.

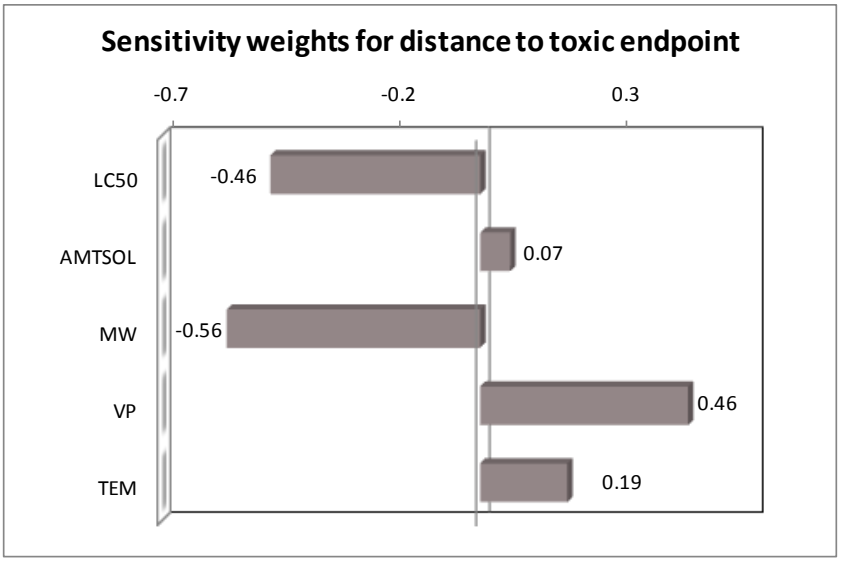


Figure 36. Sensitivity parameters for variables in model #2

6.4.4 Optimization within simulator

Optimization of the process using Aspen Plus was performed following similar approach as given in Section 6.3.4. The basic formulation of the optimization problem is as shown.

$$\max \; Acetic\,acid\,in\,extract\,phase$$

$$\text{subject to} \quad \left\{ \begin{array}{c} governing\,process\,model\,for\,countercurrent\,extraction\,of \\ a\,liquid\,stream\,using\,a\,solvent \end{array} \right\}$$

$$QR = \frac{0.284 \times U^{0.78} \times MW^{2/3} \times A \times VP}{82.05 \times T}$$

$$D_{tox} = \left(\frac{0.31505 \times Q}{MW \times LC_{50}} \right)^{1/1.7}$$

$$D_{tox} < 100\,\text{meters}$$

$$40000 < AMTSOL < 80000$$

$$100 < TEMSOL < 300$$

$$100 < PRESSOL < 300$$

The Fortran code for the consequence modeling to be included in Aspen Plus is

```
F       AREA=(VFLOW*10.0)/60.0/0.03
F       QR=0.3896*(EAMW**0.67)*AREA*VP/(82.05*(TEM+459.67)*5/9)
F       DIST=(0.31505*QR/(EAMW*LC50))**(1.0/1.7)
```

where VFLOW is the solvent flowrate in extract stream, MW is the molecular weight of the solvent, VP is the vapor pressure of the extract stream, TEM is temperature of extract stream, and LC50 is the lethal concentration for 50% fatality probability. User defined parameters are specified in the ‘define’ sheet: DIST is the distance in meters to reach toxic endpoint (LC50), AREA is the area of the pool formed by quantity released in 10 minutes, and QR is the amount of solvent vaporized.

Having performed the same assessment on 11 solvents, unique solutions are found for the optimized values of the input variables and output variables as shown in Table 22.

Table 22. Optimization results upon considering toxicity hazard

Solvent name	Input flowsheet variables/ decision variables							Output variables			Objectives and constraint values		
	Amount of solvent used (AMTSOL)	Temperature of solvent (TEMSOL)	Pressure of solvent (PRESSOL)	Solvent molecular weight (MW)	Volumetric flowrate (VFLOW)	Solvent Vapor pressure (VP)	Extract phase temperature (TEM)	LC_{50} (LC50)	Area of pool (AREA)	Vaporization rate (QR)	LC_{50} (LC50)	Amount of acetic acid in extract phase (AAEXT)	Distance to toxic endpoint (DIST)
	lb/hr	F	psi	g/mol	ft^3/hr	mm Hg	F	mol/lt	sqft	lb/min	mol/lt	lb/hr	meter
5-nonanone	80000	300	174	142	1567	27	194	0.000123	8706	85	333	6295	74.49
diisobutyl ketone	70000	300	100	142	1398	60	196	0.000214	7766	167	322	6014	80.37
2-methyl hexanal	80000	109	241	114	1601	12	103	0.0000437	8895	40	305	4833	99.99
ethyl acetate	69426	100	208	88	1233	171	101	0.000511	6851	359	269	6657	100.07
1-octanal	70000	300	204	128	1370	31	180	0.0000955	7608	81	324	4812	89.86

Solvent name	Input flowsheet variables/ decision variables							Output variables			Objectives and constraint values		
	Amount of solvent used (AMTSOL)	Temperature of solvent (TEMSOL)	Pressure of solvent (PRESSOL)	Solvent molecular weight (MW)	Volumetric flowrate (VFLOW)	Solvent Vapor pressure (VP)	Extract phase temperature (TEM)	LC_{50} (LC50)	Area of pool (AREA)	Vaporization rate (QR)	LC_{50} (LC50)	Amount of acetic acid in extract phase (AAEXT)	Distance to toxic endpoint (DIST)
	lb/hr	F	psi	g/mol	ft^3/hr	mm Hg	F	mol/lt	sqft	lb/min	mol/lt	lb/hr	meter
5-methyl-2-hexanone	70000	300	100	114	1382	121	191	0.001259	7679	291	309	6403	44.64
2-octanone	80000	300	183	128	1570	47	195	0.000339	8722	140	344	6457	58.56
2-nonanone	80000	300	296	142	1566	21	195	0.000123	8701	68	337	6379	65.32
3-heptanone	80000	100	100	114	1573	9	99	0.000955	8736	29	310	6374	13.47
1-hexanal	80000	171	157	100	1578	53	131	0.000186	8769	149	300	5574	100.01
2-ethylhexanal	80000	157	113	128	1567	10	123	0.0000316	8703	32	317	4772	99.90

The results from Table 22 can be plotted to represent multiple objectives by using a contour plot (colored cell) technique. Figure 37 shows the plot of hazards of the individual solvents such as LC_{50} (indicating the toxic property of the material) and distance to toxic endpoint (indicating the severity of worst case scenario for the solvent). The color variation in the plot shows the variation in the objective function. Thus, the solvents within the zones formed by the intersection of higher objective function values, higher LC_{50} and lower impact distances are preferred.

The figure shows that there are no solvents that satisfy all requirements sufficiently, but the solvents present in the preferable zones on the plot are 6 (5-methyl-2-hexanone), 9 (3-heptanone), and 4 (ethyl acetate). The solvents that proved to be better for flammable hazards (5-nonanone, 2-octanone, 2-nonanone) display intermediate ranges for distance, but lower values for LC_{50}, thus, proving to be not as effective in the case of toxic hazard assessment.

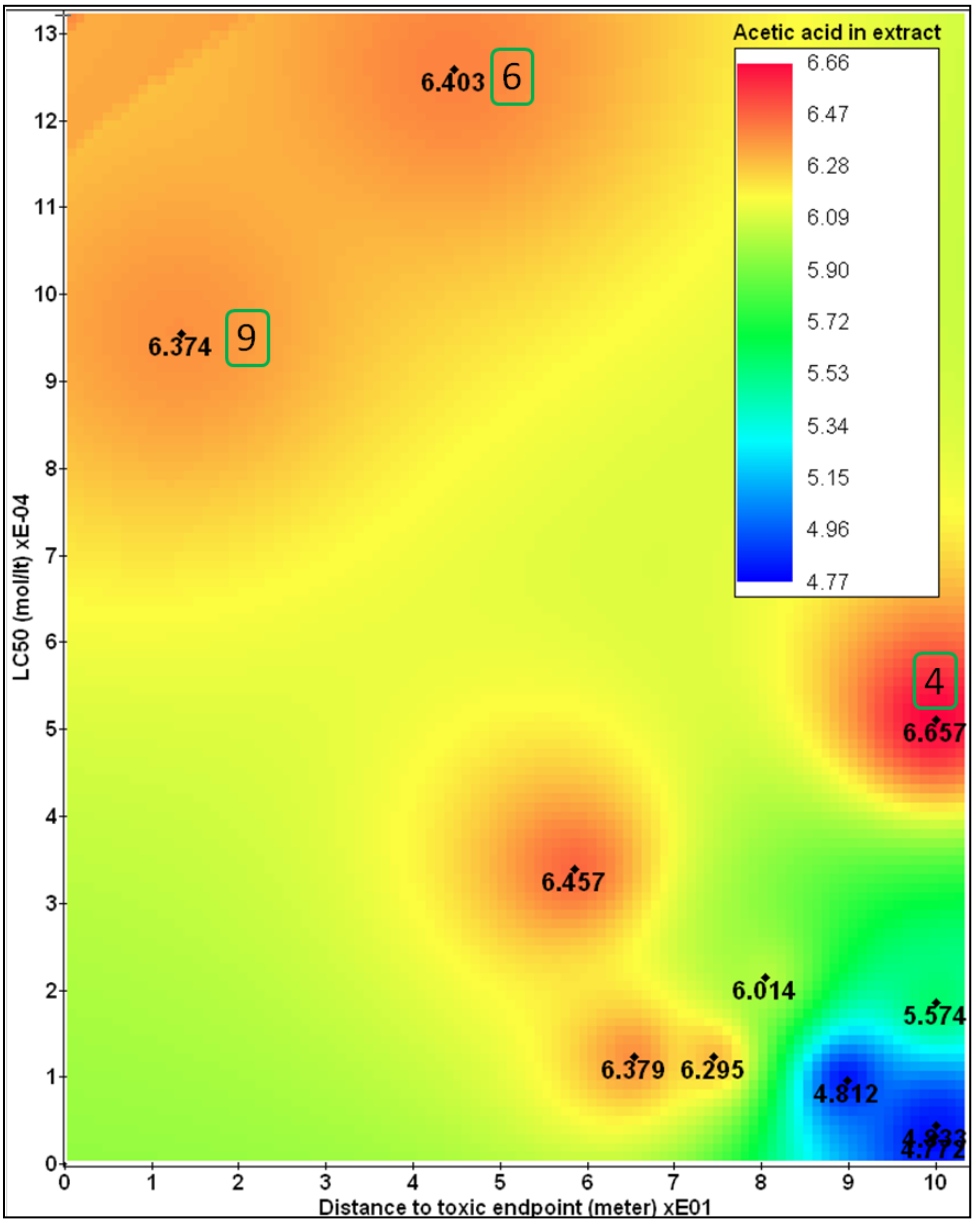


Figure 37. Contour colored cell plot of optimization results considering toxicity hazards (6: 5-methyl-2-hexanone, 9: 3-heptanone, 4: ethyl acetate)

6.5 Conclusions

Integration of safety in molecular design and process design has been implemented on a case study: liquid-liquid extraction of acetic acid–water mixture. The results provide a detailed understanding of inherently safer characteristics at the conceptual design stage. Traditionally, acetic acid-water separation is carried out using ethyl acetate solvent, but based on the results of the case study it would be beneficial to further assess safer solvent substitutes and operating parameters using experimental studies or dynamic simulation methods.

1. ICAS ProCAMD generated several solvents that matched requirements for solvent performance and safety (such as flash point and LC_{50}). Most structures identified were isomeric structures, and only 10 of the solvents were also found in the internal database. These 10 solvents were further evaluated using Aspen Plus, along with a conventional solvent (ethyl acetate) for comparison purposes.
2. When flammable hazard modeling was integrated with Aspen Plus, solvents such as 5-nonanone, 2-nonanone and 2-octanone were identified as inherently safer in the case study. Also, ranges for process operating conditions (temperature and solvent flowrate) were identified by using multiparameter sensitivity analysis and optimization within Aspen Plus.
3. Similarly, in the case of reducing toxic hazards the solvent substitutes identified for further evaluation are 5-methyl-2-hexanone and 3-heptanone.
4. Simpler EPA RMP consequence models were employed which were compared with more stringent models such as Baker-Strehlow-Tang for validation.

7. SUMMARY AND CONCLUSIONS

7.1 Summary and conclusions

This research focuses on practical applications of tools and methods to inherently safer design. This work also focuses on solvent related processes, so that the methods can be utilized for ‘substitution’, ‘attenuation’ and ‘intensification’. It has been identified that knowledge and integration of safety considerations is required at early stages of any project. The early stages of process design are generally material selection and process operating range or parameter selection. Thus, considering molecular design and process design and simultaneous integration of inherent safety concepts in both areas can be greatly beneficial. The conclusions of this research can be summarized as shown below.

1. Molecular design and QSPR

Molecular design can be used to select molecules for desired purpose using property prediction models. QSPR is an emerging, and efficient method for property prediction. Thus, the applicability of QSPR to molecular design has been assessed in this work. Although the molecular descriptors within QSPR provide a higher level of molecular representation compared to group contribution method, the models developed in this work show that there are certain limitations that need to be overcome. These limitations are that the QSPR model is not universally applicable to all classes of solvents, and that the more complex, non-linear model based on neural network is not suitable for the inverse problem in molecular design.

2. Integration of safety in molecular and process design

The method proposed in this research demonstrates a systematic technique to incorporate principles of inherently safer design into the early stages of conceptual process design. This work deals with solvent processes primarily, in order to effectively demonstrate the application of CAMD tools to select inherently safer solvent substitutes. Thereafter, an approach to integrate basic consequence modeling for hazardous events in the initial design steps has been identified in this work. This can be used in assessing cost effectiveness and regulatory requirements.

The case study on acetic acid-water separation utilizes CAMD method and hazard classification guidance for selection of inherently safer solvents, as well as Aspen Plus simulator and basic consequence modeling for selection of inherently safer systems. Multiple assessments such as sensitivity analysis and optimization provide additional information and insight on factors affecting decisive choices related to the solvent process.

7.2 Future work and recommendations

This research area, as any other study, can be expanded and improved in many ways. There are several possible avenues of future work that could be assessed for obtaining better solutions to the problem. This work is a preliminary attempt to incorporate safety into the conceptual design of solvent processes, which has not yet been extensively approached by previous studies. Some aspects for development in future studies are as shown below.

1. QSPR application in CAMD is a promising field of study, but QSPR has limitations in its present form. A potential improvement would be to develop novel methods or molecular descriptors to be appropriate for CAMD purposes in future.
2. For CAMD and inherent safety, similar molecular design approach can be applied to other material types which are considered hazardous. This would depend on the applicability of CAMD techniques and models used within CAMD to the materials of interest. Database search techniques can be developed for selection of materials, but numerous gaps in data related to hazardous properties as well as other physical properties need to be accounted for. Databases could incorporate property estimation techniques like the ones in CAMD to be more effective as a screening tool in early process design.
3. Solvents find application in many other unit operations as well, such as crystallization, extractive distillation, absorption, acting as a medium for reactions, and for diluting paints and varnishes. Recently, a prominent area of research is carbon capture using physical and chemical solvents. Thus, a similar approach can be applied to other operations to assess inherent safety.
4. Different types of hazard assessment, such as pool/jet fire hazards, or reactive chemicals hazard can also be included at the process design stage. Apart from EPA RMP guidelines and simpler consequence models, more realistic and comprehensive models need to be incorporated at the process design stage.
5. An optimization approach with an objective to reduce cost (that factors both process and safety-related cost estimates), subject to constraints on feasible operating

conditions and limiting values on hazardous consequences should be developed. An approach to simultaneously solve the molecular and process design problems needs to be implemented. This can be computationally challenging, but will provide the design team with an advanced optimal solution to the problem.

The concept of inherently safer design has existed since decades, but effective techniques to include these concepts at the design stage have been limited. There is a pressing need to explore the applicability of existing techniques and models to achieve inherent safety of processes. A comprehensive tool or methodology that accounts for both safer material selection and process design will greatly enhance the area of inherently safer design.

LITERATURE CITED

1. Ford, E. S.; Rhodes, S.; McDiarmid, M.; Schwartz, S. L.; Brown, J., Deaths from Acute Exposure to Trichloroethylene. *Journal of Occupational and Environmental Medicine* **1995,** *37*, (6), 749-754.

2. NOHSC *Industrial Organic Solvents*; National Occupational Health and Safety Commission: Canberra, Australia, 1990.

3. CSB *Confined Vapor Cloud Explosion*; No. 2007-03-I-MA; U.S. Chemical Safety and Hazard Investigation Board: 2008.

4. CSB *Static Spark Ignites Explosion Inside Flammable Liquid Storage Tank*; No. 2007-06-I-KS; U.S. Chemical Safety and Hazard Investigation Board: 2007.

5. CSB *Static Sparks Ignites Flammable Liquid During Portable Tank Filling Operation*; No. 2008-02-I-IA; U.S. Chemical Safety and Hazard Investigation Board: 2008.

6. Kletz, T. A., *Process Plants-A Handbook for Inherently Safer Design*. Taylor & Francis, Inc.: Philadelphia, PA, 1998.

7. Kletz, T. A., Inherently Safer Design: The Growth of an Idea. *Process Safety Progress* **1996,** *15*, (1), 5-8.

8. Mannan, M. S., *Lees' Loss Prevention in the Process Industries*. 3rd ed.; Elsevier: Burlington, MA, 2005; Vol. 2.

9. CCPS, *Guidelines for Design Solutions for Process Equipment Failures*. Center for Chemical Process Safety/AIChE: New York, NY, 1998.

10. Stankiewicz, A.; Moulijn, J. A., Process Intensification. *Industrial & Engineering Chemistry Research* **2002,** *41*, (8), 1920-1924.

11. CSB *LPG Fire at Valero - McKee Refinery*; U.S. Chemical Safety Board: 2008

12. Hendershot, D. C., Inherently Safer Chemical Process Design. *Journal of Loss Prevention in the Process Industries* **1997,** *10*, (3), 151-157.

13. García-Serna, J.; Pérez-Barrigón, L.; Cocero, M. J., New Trends for Design Towards Sustainability in Chemical Engineering: Green Engineering. *Chemical Engineering Journal* **2007,** *133*, (1-3), 7-30.

14. Khan, F. I.; Sadiq, R.; Amyotte, P. R., Evaluation of Available Indices for Inherently Safer Design Options. *Process Safety Progress* **2003,** *22*, (2), 83-97.

15. Bollinger, R. E.; Clark, D. G.; Dowell, A. M.; Rodger, M.; Hendershot, D. C.; Lutz, W. K.; Meszaros, S. I.; Park, D. E.; Wixom, E. D., *Inherently Safer Chemical Processes: A Life Cycle Approach*. Center for Chemical Process Safety/AIChE: New York, NY, 1996.

16. Hurme, M.; Rahman, M., Implementing Inherent Safety Throughout Process Lifecycle. *Journal of Loss Prevention in the Process Industries* **2005,** *18*, (4-6), 238-244.

17. Khan, F. I.; Amyotte, P. R., I2SI: A Comprehensive Quantitative Tool for Inherent Safety and Cost Evaluation. *Journal of Loss Prevention in the Process Industries* **2005,** *18*, (4-6), 310-326.

18. Gupta, J. P.; Edwards, D. W., A Simple Graphical Method for Measuring Inherent Safety. *Journal of Hazardous Materials* **2003,** *104*, 15-30.

19. Gupta, J. P.; Edwards, D. W., Inherently Safer Design-Present and Future. *Process Safety and Environmental Protection* **2002,** *80*, (3), 115-125.

20. Suardin, J.; Mannan, M. S.; El-Halwagi, M., The Integration of Dow's Fire and Explosion Index (F&EI) into Process Design and Optimization to Achieve Inherently Safer Design. *Journal of Loss Prevention in the Process Industries* **2007,** *20*, 79-90.

21. Medina, H.; Arnaldos, J.; Casal, J., Process Design Optimization and Risk Analysis. *Journal of Loss Prevention in the Process Industries* **2009,** *22*, 566-573.

22. Kim, D.; Kim, J.; Moon, I., Integration of Accident Scenario Generation and Multiobjective Optimization for Safety-cost Decision Making in Chemical Processes. *Journal of Loss Prevention in the Process Industries* **2006,** *19*, 705-713.

23. Shariff, A. M.; Rusli, R.; Leong, C. T.; Radhakrishnan, V. R.; Buang, A., Inherent Safety Tool for Explosion Consequences Study. *Journal of Loss Prevention in the Process Industries* **2006,** *19*, 409-418.

24. Eljack, F. T.; Eden, M. R.; Kazantzi, V.; Qin, X.; El-Halwagi, M., Simultaneous Process and Molecular Design- A Property Based Approach. *AIChE Journal* **2007,** *53*, (5), 1232-1239.

25. Eden, M. R.; Jorgensen, S. B.; Gani, R.; El-Halwagi, M., A Novel Framework for Simultaneous Separation Process and Product Design. *Chemical Engineering and Processing* **2004,** *43*, 595-608.

26. Achenie, L. E. K.; Gani, R.; Venkatasubramanian, V., *Computer Aided Molecular Design: Theory and Practice*. Elsevier: Amsterdam, The Netherlands, 2003.

27. Churi, N.; Achenie, L. E. K., Novel Mathematical Programming Model for Computer Aided Molecular Design *Industrial and Engineering Chemistry Research* **1996,** *35*, (10), 3788-3794.

28. Wang, Y.; Achenie, L. E. K., Computer Aided Solvent Design for Extractive Fermentation. *Fluid Phase Equilibria* **2002,** *201*, (1), 1-18.

29. Chavali, S.; Lin, B.; Miller, D. C.; Camarda, K. V., Environmentally-benign Transition Metal Catalyst Design Using Optimization Techniques. *Computers and Chemical Engineering* **2004,** *28*, (5), 605-611.

30. Vaidyanathan, R.; El-Halwagi, M., Computer Aided Synthesis of Polymers and Blends with Target Properties. *Industrial and Engineering Chemistry Research* **1996,** *35*, (2), 627-634.

31. Odele, O.; Macchietto, S., Computer Aided Molecular Design: A Novel Method for Optimal Solvent Selection. *Fluid Phase Equilibria* **1993,** *82*, 47-54.

32. Wang, Y.; Achenie, L. E. K., A Hybrid Global Optimization Approach for Solvent Design. *Computers & Chemical Engineering* **2002,** *26*, (10), 1415-1425.

33. Karunanithi, A. T.; Achenie, L. E. K.; Gani, R., A New Decomposition-based Computer-aided Molecular/Mixture Design Methodology for the Design of Optimal Solvents and Solvent Mixtures. *Industrial & Engineering Chemistry Research* **2005,** *44*, (13), 4785-4797.

34. Kazantzi, V.; Qin, X.; El-Halwagi, M.; Eljack, F.; Eden, M., Simultaneous Process and Molecular Design through Property Clustering Techniques: A Visualization Tool. *Industrial & Engineering Chemistry Research* **2007,** *46*, (10), 3400-3409.

35. Kier, L. B.; Hall, L. H.; Frazer, J. W., Design of Molecules from Quantitative Structure-Activity Relationship Models. 1. Information Transfer Between Path and Vertex Degree Counts. *Journal of Chemical Information and Computer Sciences* **1993,** *33*, (1), 143-147.

36. Visco, D. P.; Pophale, R. S.; Rintoul, M. D.; Faulon, J.-L., Developing a Methodology for an Inverse Quantitative Structure-activity Relationship Using the Signature Molecular Descriptor. *Journal of Molecular Graphics and Modeling* **2002,** *20*, (6), 429-438.

37. Brown, N.; McKay, B.; Gasteiger, J., A Novel Workflow for the Inverse QSPR Problem Using Multiobjective Optimization. *Journal of Computer Aided Molecular Design* **2006,** *20*, 333-341.

38. Skvortsova, M. I.; Baskin, I. I.; Slovokhotova, O. L.; Palyulin, V. A.; Zefirov, N. S., Inverse Problem in QSAR/QSPR Studies for the Case of Topological Indexes Characterizing Molecular Shape (Kier Indices). *Journal of Chemical Information and Computer Science* **1993,** *33*, (4), 630-634.

39. Skvortsova, M. I.; Fedyaev, K. S.; Palyulin, V. A.; Zefirov, N. S., Inverse Structure-property Relationship Problem for the Case of Correlation Equation Containing the Hosoya Index. *Doklady Chemistry* **2001,** *379*, (1-3), 191-195.

40. Needham, D. E.; Wei, I. C.; Seybold, P. G., Molecular Modeling of the Physical Properties of Alkanes. *Journal of the American Chemical Society* **1988,** *110*, (13), 4186-4194.

41. Raman, V. S.; Maranas, C. D., Optimization in Product Design with Properties Correlated with Topological Indices. *Computers & Chemical Engineering* **1998,** *22*, (6), 747-763.

42. NFPA, 704, Standard System for the Identification of the Hazards of Materials for Emergency Response. In National Fire Protection Agency: Quincy, MA, 2007.

43. Hshieh, F.-Y., Correlation of Closed-cup Flash Points with Normal Boiling Points for Silicone and General Organic Compounds. *Fire and Materials* **1997,** *21*, (6), 277-282.

44. Satyanarayana, K.; Rao, P. G., Improved Equation to Estimate Flash Points of Organic Compounds. *Journal of Hazardous materials* **1992,** *32*, 81-85.

45. Prugh, R. W., Estimation of Flash Point Temperature. *Journal of Chemical Education* **1973,** *50*, (2).

46. Suzuki, T.; Ohtaguchi, K.; Koide, K., A Method for Estimating Flash Points of Organic Compounds from Molecular Structures. *Journal of Chemical Engineering Japan* **1991,** *24*, (2).

47. Stefanis, E.; Constantinou, L.; Panayiotou, C., A Group-Contribution Method for Predicting Pure Component Properties of Biochemical and Safety Interest. *Industrial & Engineering Chemistry Research* **2004,** *43*, (19), 6253-6261.

48. Katritzky, A. R.; Petrukhin, R.; Jain, R.; Karelson, M., QSPR Analysis of Flash Points. *Journal of Chemical Information and Computer Sciences* **2001,** *41*, (6), 1521-1530.

49. Katritzky, A. R.; Stoyanova-Slavova, I. B.; Dobchev, D. A.; Karelson, M., QSPR Modeling of Flash Points: An Update. *Journal of Molecular Graphics and Modelling* **2007,** *26*, (2), 529-536.

50. Karelson, M., *Molecular Descriptors in QSAR/QSPR*. Wiley-Interscience: New York, NY, 2000.

51. Katritzky, A. R.; Maran, U.; Lobanov, V. S.; Karelson, M., Structurally Diverse Quantitative Structure-property Relationship Correlations of Technologically Relevant Physical Properties. *Journal of Chemical Information and Computer Science* **2000,** *40*, (1), 1-18.

52. Godavarthy, S. S.; Robinson, R. L.; Gasem, K. A. M., An Improved Structure-Property Model for Predicting Melting-point Temperatures. *Industrial & Engineering Chemistry Research* **2006,** *45*, (14), 5117-5126.

53. Sola, D.; Ferri, A.; Banchero, M.; Manna, L.; Sicardi, S., QSPR Prediction of n-Boiling Point and Critical Properties of Organic Compounds and Comparison with a Group-contribution Method. *Fluid Phase Equilibria* **2008,** *263*, (1), 33-42.

54. Acros Organics, Catalog of Organic and Fine Chemicals. http://www.acros.com/portal/alias__Rainbow/lang__en/tabID__34/DesktopDefault.aspx. (May 2008)

55. Aldrich Chemical Co., I., Catalog Handbook of Fine Chemicals. http://www.sigmaaldrich.com/safety-center.html (May 2008)

56. Archer, W. L., *Industrial Solvents Handbook*. Marcel Dekker, Inc: New York, NY, 1996.

57. *ICAS Documentation* CAPEC, Department of Chemical Engineering, Technical University of Denmark: Lyngby, Denmark, 2003.

58. NCBI PubChem Project. http://pubchem.ncbi.nlm.nih.gov/ (May 2008)

59. *Materials Studio Release Notes*, Release 4.3; Accelrys Software, Inc San Diego, CA 2006.

60. Shihab, K., A Backpropagation Neural Network for Computer Network Security. *Journal of Computer Science* **2006,** *2*, (9), 710-715.

61. Nash, S. G.; Sofer, A., *Linear and Nonlinear Programming*. Mc-Graw Hill: New York, NY, 1996.

62. Himmelblau, D. M., Accounts of Experiences in the Application of Artificial Neural Networks in Chemical Engineering. *Industrial & Engineering Chemistry Research* **2008,** *47*, (16), 5782-5796.

63. Duda, R. O.; Hart, P. E.; Stork, D. G., *Pattern Classification*. 2nd ed.; Wiley: New York, NY, 2001.

64. Vidal, M.; Rogers, W. J.; Mannan, M. S., Prediction of Minimum Flash Point Behaviour for Binary Mixtures. *Process Safety and Environmental Protection* **2006,** *84*, (1), 1-9.

65. Katritzky, A. R.; Slavov, S. H.; Dobchev, D. A.; Karelson, M., Rapid QSPR Model Development Technique for Prediction of Vapor Pressure of Organic Compounds. *Computers & Chemical Engineering* **2007,** *31*, (9), 1123-1130.

66. Rouvray, D. H., The Modeling of Chemical Phenomena Using Topological Indices. *Journal of Computational Chemistry* **1987,** *8*, (4), 470-480.

67. Kier, L. B., A Shape Index from Molecular Graphs. *Quantitative Structure-Activity Relationships* **1985,** *4*, (3), 109-116.

68. Kier, L. B.; Hall, L. H., Derivation and Significance of Valence Molecular Connectivity. *Journal of Pharmaceutical Sciences* **1981,** *70*, (6), 583-589.

69. Harper, P. M.; Gani, R., A Multi-step and Multi-level Approach for Computer Aided Molecular Design. *Computers & Chemical Engineering* **2000,** *24*, (2-7), 677-683.

70. Marrero, J.; Gani, R., Group Contribution Based Estimation of Pure Component Properties. *Fluid Phase Equilibria* **2001,** *183*, 183-208.

71. Constantinou, L.; Gani, R., New Group Contribution Method for Estimating Properties of Pure Compounds. *A.I.Ch.E. Journal* **1994,** *40*, (10), 1697-1710.

72. Joback, K. G.; Reid, R. C., Estimation of Pure Component Properties from Group Contributions. *Chemical Engineering Communications* **1987,** *57*, 233-243.

73. Heikkila, A.-M. Inherent Safety in Process Plant Design: An Index-based Approach. Helsinki University of Technology, Helsinki, Finland, 1999.

74. OSHA, Hazard Communication -71:53617-53627 (Proposed rule making). In Office of the Federal Register: Washington, D.C., 2006.

75. 67/548/EEC, Council Directive of 27 June 1967 on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labeling of dangerous substances, Off. J. P 196 (1967) 0001-0098 (16 August). . In.

76. Dow *Dow's Fire & Explosion Index Hazard Classification Guide*; AIChE: New York, NY, 1994.

77. Veith, G. D.; Call, D. J.; Brooke, L. T., Structure-toxicity Relationships for the Fathead Minnow, Pimephales Promelas: Narcotic Industrial Chemicals. *Canadian Journal of Fisheries and Aquatic Sciences* **1983,** *40*, (6), 743-748.

78. Veith, G. D.; Call, D. J.; Brooke, L. T. In *Estimating the Acute Toxicity of Narcotic Industrial Chemicals to Fathead Minnows*, Aquatic Toxicology and Hazard Assessment: Sixth Symposium Philadelphia, PA, 1983; Bishop, W. E.; Cardwell, R. D.; Heidolph, B. B., Eds. American Society for Testing and Materials: Philadelphia, PA, 1983; pp 90-97.

79. Li, M.; Harten, P. F.; Cabezas, H., Experiences in Designing Solvents for the Environment. *Industrial & Engineering Chemistry Research* **2002,** *41*, 5867-5877.

80. EPA, Risk Management Program Guidance for Offsite Consequence Analysis. In 2009; Vol. EPA 550-B-99-009, Washington, D.C.

81. EPA Regulated substances for accidental release prevention (40CFR68.130). Washington, D.C. http://frwebgate.access.gpo.gov/cgi-bin/get-cfr.cgi?TITLE=40&PART=68&SECTION=130&TYPE=TEXT

82. Diwekar, U. M.; Rubin, E. S., Stochastic Modeling of Chemical Processes. *Computers & Chemical Engineering* **1991,** *15*, (2), 105-114.

83. Saltelli, A.; Tarantola, S.; Campolongo, F.; Ratto, M., *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. John Wiley & Sons, Ltd.: West Sussex, England, 2004.

84. Edgar, T. F.; Himmelblau, D. M.; Lasdon, L. S., *Optimization of Chemical Processes*. 2nd ed.; McGraw-Hill: New York, NY, 2001.

85. Perry, R. H.; Green, D. W., *Perry's Chemical Engineers' Handbook*. 7th ed.; McGraw-Hill: New York, NY, 1997.

86. Plerorazio, A. J.; Thomas, J. K.; Baker, Q. A.; Ketchum, D. E., An Update to the Baker-Strehlow-Tang Vapor Cloud Explosion Prediction Methodology Flame Speed Ttable. *Process Safety Progress* **2005,** *24*, (1), 59-65.

VITA

Name: Suhani Jitendra Patel

TAMU Dept. address: 3122 TAMU Room 244 College Station, TX 77843-3122

Email Address: suhani.patel@gmail.com

Education: Ph.D., Chemical Engineering, Texas A&M University, College Station TX, August 2010

B.S., Chemical Engineering, Thadomal Shahani Engineering College, University of Mumbai, May 2006

Publications:

1. Patel, S. J.; Ng, D.; Mannan, M. S., QSPR Flash Point Prediction of Solvents Using Topological Indices for Application in Computer Aided Molecular Design. *Industrial & Engineering Chemistry Research* **2009**, *48*, (15), 7378-7387.

2. Patel, S. J.; Ng, D.; Mannan, M. S., Inherently Safer Design of Solvent Processes at the Conceptual Stage: Practical Application for 'Substitution', *Journal of Loss Prevention in the Process Industries* **2010**, doi:10.1016/j.jlp.2010.03.002

Conference talks:

1. Patel, S. J., Ng, D., Mannan, M. S., 'Integrating Safety Issues in Conceptual Design of Solvent Processes', AIChE Conference, November 2009, Nashville TN

2. Patel, S. J., Ng, D., Mannan, M. S., 'Integrating Safety Issues in Solvent Selection and Process Design', AIChE Conference, March 2010, San Antonio TX