6-4-2010

# Multi-objective optimization of the molecular structure of refrigerants

Suvrat Bhargava
*Florida International University*

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

# MULTI-OBJECTIVE OPTIMIZATION OF THE MOLECULAR STRUCTURE OF REFRIGERANTS

A thesis submitted in partial fulfillment of the

requirements for the degree of

MASTER OF SCIENCE

in

MECHANICAL ENGINEERING

by

Suvrat Bhargava

2010

To: Dean Amir Mirmiran
College of Engineering and Computing

This thesis, written by Suvrat Bhargava, and entitled Multi-Objective Optimization of the Molecular Structure of Refrigerants, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this thesis and recommend that it be approved.

Surendra K. Saxena

Arvind Agarwal, Co-Major Professor

George S. Dulikravich, Co-Major Professor

Date of Defense: June 4, 2010

The thesis of Suvrat Bhargava is approved.

Dean Amir Mirmiran
College of Engineering and Computing

Interim Dean Kevin O'Shea
University Graduate School

Florida International University, 2010

DEDICATION

I dedicate this thesis to my parents, Late Professor Sanjeev Bhargava and Mrs. Sarita Bhargava, and my sister, Ms. Akanksha Bhargava. Without their nurturing, guidance, inspiration, unwavering support, encouragement and love, the completion of this work would not have been possible.

ACKNOWLEDGMENTS

I would first like to express my gratitude to Dr. George S. Dulikravich, my co-major advisor, for his support in times when I needed it the most. During my stay in Miami, he has been one of the two professors who have supported me in the toughest of times. His confidence in my abilities, constant encouragement and positive attitude will always serve as a source of motivation to do better. I consider myself fortunate to have had so many opportunities to learn from him, not just about the research, but also about working as a professional.

I am also indebted to Dr. Arvind Agarwal, my co-major advisor for all his support. Along with Dr. Dulikravich, he has been the other professor on whom I have relied for timely suggestions and advice. I will always admire the way he supports students and makes them learn the value of hard work and perseverance.

I would also like to thank Dr. Surendra K. Saxena, for being a part of my thesis committee and being a source of encouragement. My discussions with him at the time of the inception of the project provided vital ideas.

My deepest gratitude is due to Dr. Igor N. Egorov who performed actual multi-objective constrained optimization tasks using his IOSO software.

I also thank Dr. Igor Tsukanov for his timely support and brief discussions during my stay in Florida International University as a graduate student.

I am also thankful to Mr. Octavio Oliva, Mr. Holger Paas, Mr. Byron Gaskin and Mr. Philip Clarke for helping me in managing the datasets and preparing a number of graphs. My friends and fellow researchers in the lab; Mr. Naveen Singamsetti, Mr. Ricardo Ardila, Mr. Carlos Velez, Mr. Stephen Wood, Mr. Souma Chowdhury, Mr. Flavio Vianna and Ms. Amanda Vianna, have all contributed to the whole learning experience, in many ways. I would like to extend my sincere thanks to all of them. I also thank fellow researcher Dr. Srinivasa Rao Bakshi for his intellectual discussions with me.

I am also grateful to the Department of Mechanical and Materials Engineering for providing me a platform to conduct my research and also for supporting me financially.

I am also indebted to Ms. Akanksha Bhargava and Mrs. Sarita Bhargava, for not only supporting me as family members, but also for being close friends and helping me with my work in the final stages. Finally, I express profound gratitude to my parents who gave me the ability to think rationally.

ABSTRACT OF THE THESIS

MULTI-OBJECTIVE OPTIMIZATION OF THE MOLECULAR STRUCTURE OF REFRIGERANTS

by

Suvrat Bhargava

Florida International University, 2010

Miami, Florida

Professor Arvind Agarwal, Co-Major Professor

Professor George S. Dulikravich, Co-Major Professor

The aim of this work was to develop a new methodology, which can be used to design new refrigerants that are better than the currently used refrigerants. The methodology draws some parallels with the general approach of computer aided molecular design. However, the mathematical way of representing the molecular structure of an organic compound and the use of meta models during the optimization process make it different. In essence, this approach aimed to generate molecules that conform to various property requirements that are known and specified *a priori*. A modified way of mathematically representing the molecular structure of an organic compound having up to four carbon atoms, along with atoms of other elements such as hydrogen, oxygen, fluorine, chlorine and bromine, was developed. The normal boiling temperature, enthalpy of vaporization, vapor pressure, tropospheric lifetime and biodegradability of 295 different organic compounds, were collected from open literature and data bases or estimated. Surrogate models linking the previously mentioned quantities with the molecular structure were developed. Constraints ensuring the generation of structurally feasible molecules were formulated and used in commercially available optimization algorithms to generate molecular structures of promising new refrigerants. This study was intended to serve as a proof-of-concept of designing refrigerants using the newly developed methodology.

TABLE OF CONTENTS

# LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

# LIST OF SYMBOLS

a, b, c - Coefficients of the Gaussian curve used for fitting the vapor pressure-temperature data

A, B, C – Constants used in the Antoine equation

$\Delta G$ – Gibb's free energy, kcal

h – Specific enthalpy, cal $mol^{-1}$ (or kcal/kg)

$h$ – Planck's constant ($6.626 \times 10^{-34} J \cdot s$)

$\Delta H_v$ – Enthalpy of vaporization, kcal/ mol

$\dot{m}$ - Mass flow rate, kg/s

s – Specific entropy, $cal/mol \cdot K$ (or $kcal/kg \cdot K$)

$Q_H$ - Heat transferred from the refrigerant to the warm region, kcal

$Q_L$ – Heat transferred from the cooled region to the refrigerant, kcal

T – Temperature, °C

$T_H$ - Condensation temperature, °C

$T_L$ – Evaporating temperature, °C

$T_{mean}$ – Mean temperature of the reference refrigeration cycle, °C

$T_{sc}$ – Scaled temperature used for fitting the vapor pressure-temperature data

$T_{superheat}$ – Superheat temperature of the reference refrigeration cycle, °C

$W_{input}$ – Work done by the compressor in the refrigeration cycle

$\nu$ – Frequency of the wave, $s^{-1}$

# LIST OF ACRONYMS

| | |
|---|---|
| AOPWIN | Atmospheric Oxidation Program for Microsoft Windows |
| BIOWIN | Biodegradation Probability Program |
| CAMD | Computer Aided Molecular Design |
| COP | Coefficient of Performance |
| DOE | Design of Experiments |
| GCA | Group Contribution Approach |
| GWP | Global Warming Potential |
| IOSO | Indirect Optimization on the basis of Self-Organization |
| MITI | Japanese Ministry of International Trade and Industry |
| NIST | National Institute of Standards and Technology |
| ODP | Ozone Depletion Potential |
| QSAR | Quantity Structure-Activity Relationship |
| QSPR | Quantitative Structure-Property Relationship |
| RMS | Root Mean Squared |
| SMILES | Simplified Molecular Input Line Entry System |
| SPARC | Sparc Performs Automated Reasoning in Chemistry |

# 1. INTRODUCTION

## 1.1. Problem Statement

As new environmental regulations are enforced and the designs of refrigeration systems improve, the need to design and produce better refrigerants becomes more profound. These newly designed refrigerants would gradually replace a number of existing refrigerants that would be deemed unsuitable for commercial use for various economic and environmental reasons (Calm 2008), (Brown 2009).

Traditional ways of designing, producing and testing a new chemical compound generally involve design heuristics coupled with direct experimental and computational studies. The performance of a refrigeration system can be evaluated from the thermodynamic and transport properties of the working fluid. Some of the most important thermodynamic properties are the freezing point, boiling temperature, vapor pressure, enthalpy of vaporization and liquid specific heat. Montreal and Kyoto protocols enforce certain restrictions and limit the use and production of many refrigerants due to their harmful effects on the atmosphere. Consequently, other properties are also desirable such as low *Ozone Depletion Potential* (ODP), very low *Global Warming Potential* (GWP), short atmospheric lifetime, low *Green House Gas* (GHG) emission and biodegradability.

In the design of a new refrigerant, all the above performance criteria must be looked at and addressed in a unified and simultaneous way. Little effort has been made in the past to address these issues simultaneously. Furthermore, since *Computer Aided Molecular Design* (CAMD) approach relies heavily on the use of group contribution theory and the availability of property prediction models, little or no effort has been made to investigate the usability of response surface methods or meta-models. The motivation of this research is to analyze a

different approach based on the creation of response surface methods and the optimization of the molecular structure to find potential candidates for the replacement of the existing refrigerants.

## 1.2. Fundamentals of vapor compression refrigeration

As shown in Figure 1.1, the refrigerant in an ideal single stage vapor compression refrigeration cycle essentially undergoes the following four processes (Moran and Shapiro 2000):

a) Isobaric heat transfer to the refrigerant in the evaporator (process 1-2)

b) Isentropic compression of the refrigerant in the compressor (process 2-3)

c) Isobaric heat removal from the refrigerant in the condenser (process 3-4)

d) Adiabatic expansion of the refrigerant through the throttle valve/ turbine (process 4-1)

The corresponding T-s diagram is as shown in Figure 1.2. It should be noted that for successful operation of the refrigerator, the refrigerant temperature in the evaporator, $T_L$, should be less than

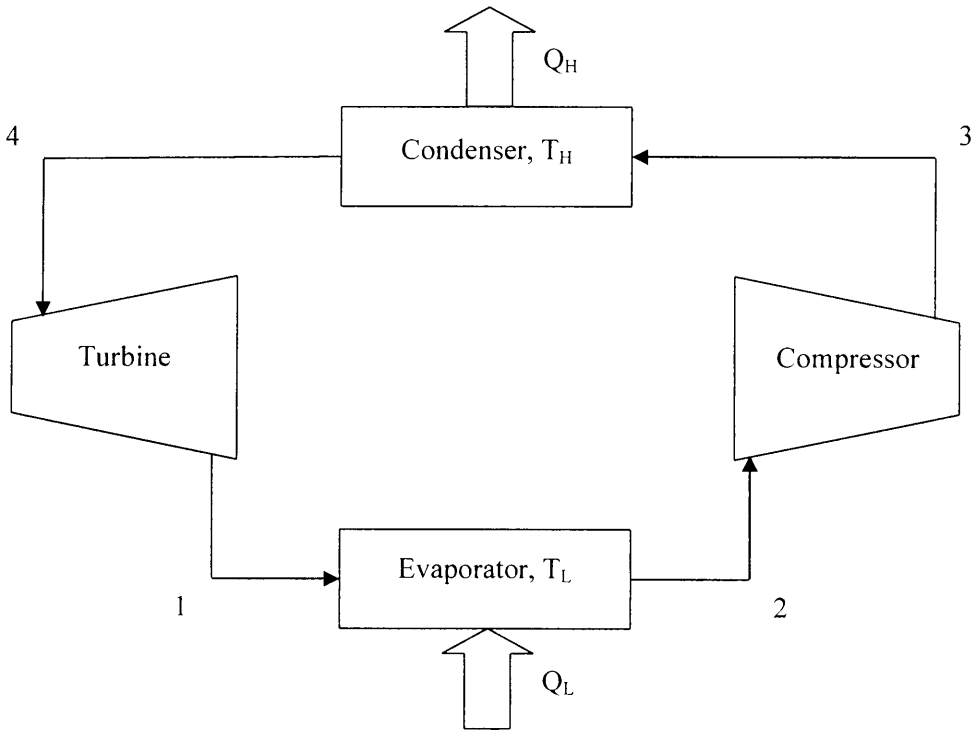



**Figure 1.1:** An illustration showing the components of an ideal single stage vapor compression refrigeration cycle

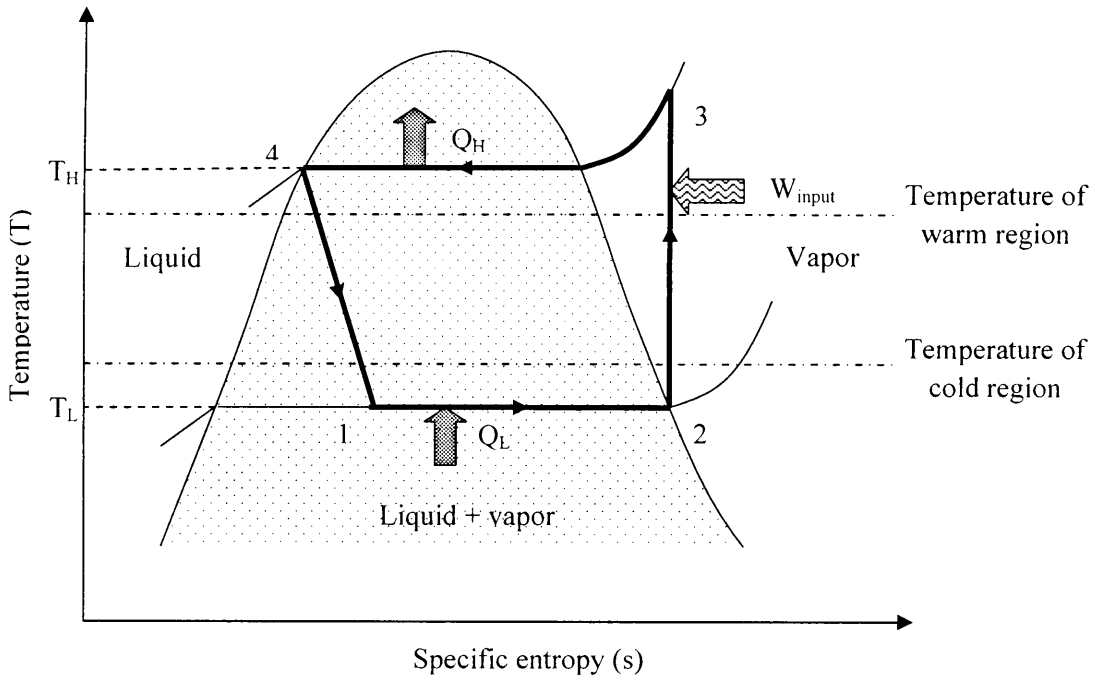



**Figure 1.2:** T-s diagram of a single stage ideal vapor compression refrigeration cycle

the cold region temperature and the refrigerant temperature in the evaporator, $T_H$, should be greater than the warm region temperature. Heat transfer from the refrigerated space to the refrigerant causes the vaporization of the refrigerant. This removal of heat causes the refrigeration of the colder region. The rate of this heat transfer $\dot{Q}_L$ is known as the refrigeration capacity and it is defined by the following equation:

$$\dot{Q}_L = \dot{m}(h_2 - h_1) \tag{1.1}$$

After the refrigerant exits the evaporator, it is compressed to a relatively higher pressure and temperature by the compressor. The rate of work done by the compressor is given by

$$\dot{W}_{input} = \dot{m}(h_3 - h_2) \tag{1.2}$$

The rate of heat removal from the refrigerant in the condenser is given by the following equation:

$$\dot{Q}_H = \dot{m}(h_4 - h_3) \tag{1.3}$$

The COP of the refrigeration cycle, β, is then defined by the following relation:

$$\beta = \frac{\dot{Q}_L/\dot{m}}{\dot{W}_{input}/\dot{m}} = \frac{(h_2 - h_1)}{(h_4 - h_2)} \tag{1.4}$$

The equations 1.1 through 1.4 remain valid under the following assumptions:

a) Any irreversibilities and stray heat transfer to the surroundings, within the evaporator and the condenser are ignored.

b) No pressure drops, due to friction effects, exist within the system.

## 1.3. Findings of past searches for new refrigerants

Calm and Didion (Calm and Didion 1998) summarized the results and observed trends of many past attempts to search for new refrigerants. The major findings are summarized below:

a) Increasing the number of carbon atoms in the refrigerant molecules not only increases the molecular weight, it generally also results in an increase in the normal boiling temperature and the heat capacity.

b) Presence of a large number of hydrogen atoms in the molecule generally decreases the atmospheric lifetime of the organic compound. Compounds with lower atmospheric lifetimes are known to have lower ODP and GWP. However, compounds with large number of hydrogen atoms tend to have higher flammability, especially when the number of hydrogen atoms in the refrigerant molecule becomes greater than the number of halogen atoms in the molecule.

c) An increase in the number of oxygen atoms results in a decrease in the ODP and GWP because it results in reduced stability of the particular compound in the atmosphere. However, it has also been associated with increased reactivity, flammability and toxicity.

d) The increase in the number of fluorine atoms increases the GWP. Perfluorinated compounds are examples which support this observation. Perfluorinated compounds belong to the family of organic compounds in which all the hydrogen atoms on a carbon

chain are replaced with fluorine atoms and the molecule contains at least one different functional group.

e) An increase in the chlorine content generally results in an increased ODP and toxicity. It also generally results in an increased lubricant miscibility.

f) An increase in the bromine content also generally increases the ODP of the organic compound. However, increased bromine content in the molecule has also been linked with reduced flammability of the compound. In general, a larger number of halogen atoms in the molecule results in an increase in the atmospheric lifetime, ODP and GWP. This trend is especially observed for perhalogenated compounds. Perhalogenated compounds belong to the family of organic compounds in which all the hydrogen atoms on a carbon chain are replaced with a particular halogen atom and the molecule contains at least one different functional group.

g) Presence of a large number of nitrogen atoms in the molecule has been associated with an increased reactivity of the compound, thereby reducing stability of the molecule. Increased nitrogen content is also linked with an increase in the toxicity. Presence of sulfur also shows similar effects on the toxicity and the stability of the compound.

h) Presence of boron in the molecule makes the compound more reactive and generally toxic. Iodine atoms in the molecule also increase the reactivity of the compound and some volatile iodine containing compounds have been rejected in the past because of their high toxicity. Because of their higher reactivity, organic compounds having iodine generally have lower atmospheric lifetime. Consequently, this also reduces their potential use as refrigerants because of similar issues related to their long term stability inside the refrigeration system.

## 1.4. Research objectives

This research aims to test and evaluate the performance of a different approach, which incorporates some fundamentals of CAMD, in designing new refrigerants. The performance of the newly designed refrigerants will be evaluated on the basis of the normal boiling temperature, latent heat of vaporization, vapor pressure variation over the operating range, half life in troposphere and biodegradability.

This research represents an endeavor of finding an alternative strategy for designing new functional molecules. Freon-12 will be used as the reference refrigerant and the new refrigerants will be designed to be equal or better than Freon-12 with respect to all the objectives defined previously. This new strategy does not rely on the availability of group contribution models. Instead, it depends on the availability of data for those classes of compounds which are considered suitable for being used as refrigerants.

## 1.5. Methodology

The rationale behind the newly developed methodology is very similar to the basis of a group/ atom/ bond contribution approach (Achenie, Gani and Venkatasubramanian 2003). It is the fact that intermolecular forces in a molecule are determined by the nature and the number of the elements constituting the particular molecule, types and locations of bonds between a particular pair of atoms and the arrangement of atoms themselves. Thus, a given set of structural variables describing a topology of a molecule should be linked to a unique set of property values for such a molecule.

Figure 1.3 shows the sequence of operations which together constitute the newly developed methodology for performing multi-objective optimization of the molecular structure of refrigerants. An initial pool of organic compounds was constituted. The compounds included in this pool were chosen while considering currently used refrigerants, the availability of

experimental data and/ or accurate property estimation models and stability criteria required by an organic compound to be used as a refrigerant. This initial pool of organic compounds comprised of the general family of alkanes, halogen substituted alkanes, alkenes, halogen substituted alkenes, ethers and halogen substituted ethers. These chemicals were chosen in such a way that none of them had more than four constituting carbon atoms.

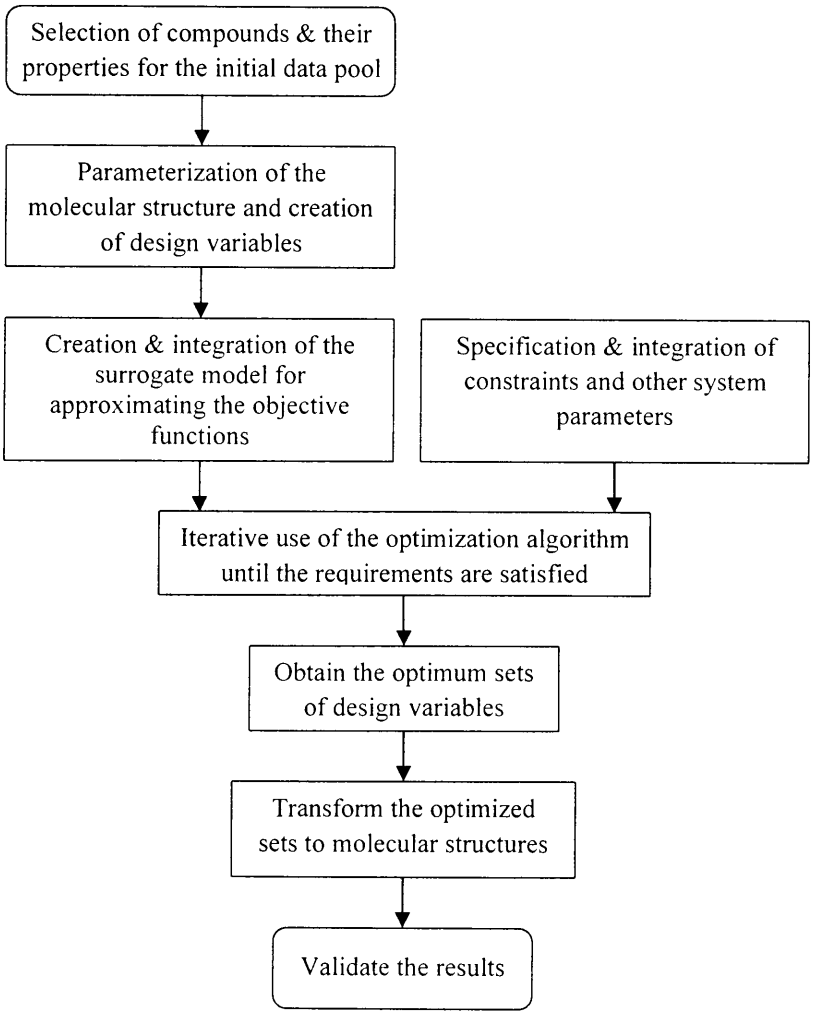



**Figure 1.3:** A flowchart explaining the methodology used in performing multi-objective design optimization of the molecular structure of refrigerants

A methodology to mathematically represent the molecular structures of those compounds was developed. Using this methodology, the molecular structures of all the compounds in the initial data set were represented by a unique set of twenty eight structural design variables. This approach of mathematically representing the molecular structure was also found capable of differentiating between isomers of a particular compound.

Using the available data, multi-dimensional surrogate models/ response surfaces were developed to approximate the objective functions. The weighted approximation method (Sigma-Technology Company 2009) and modified radial basis functions based method (Sigma-Technology Company 2009) were considered for creating the response surfaces. The results were analyzed, and it was found that estimates from the weighted approximation method were more accurate than the ones which were produced from the radial basis functions based method.

Constraints for governing the individual value of all the design variables, enforcing limits on the combination of constraints and setting the lower and/ or upper bounds of the desired property values, were formulated. Constraints were specifically developed and enforced to ensure the chemical feasibility of a generated molecular structure.

Along with the constraints, the information regarding the meta-models was used in the process of optimization. The optimization was performed using the commercially available multi-objective optimization software package, IOSO NM version 1.0 (Sigma Technology).

The parameterization of the molecular structure of an organic compound is described in section 1.6. Section 3.3 discusses the basic optimization algorithm used in IOSO and advantages of IOSO over other currently used optimization algorithms. Chapter 4 discusses the selection of compounds for the data pool, selection of their properties and compilation of the property related data. Significance of the constraints and their importance is explained in detail in chapter 5. Chapter 6 discusses the creation of response surfaces, their analysis and comparison of the results from the two methods used for the construction of the meta-models.

## 1.6. Mathematical representation of the molecular structure

The developed methodology is based on the creation and evaluation of 28 structural design variables, used for representing the molecular structure of any compound considered in this study. The methodology has been designed to handle molecules which have up to four carbon atoms. Apart from carbon atoms, the other elements which might be present in the molecule are restricted to hydrogen, chlorine, fluorine, bromine and oxygen. The compound is only allowed to have any of the following types of seven bonds:

a) Carbon-carbon single bond

b) Carbon-carbon double bond

c) Carbon-hydrogen single bond

d) Carbon-chlorine single bond

e) Carbon-fluorine single bond

f) Carbon-bromine single bond

g) Carbon-oxygen single bond

Each of the four carbon atoms is associated directly with seven out of a total of twenty eight structural design variables. Each of these seven structural design variables physically represents the number of a particular type of bond associated with the particular carbon atom.

Table 1.1 shows the twenty eight structural design variables ($X_1$, $X_2$, $X_3$..., $X_{27}$, $X_{28}$). The first row of the design variables, $X_1$, $X_2$, $X_3$ and $X_4$, represents the number of carbon-carbon single bonds associated with each of the carbon atoms designated by $C_1$, $C_2$, $C_3$ and $C_4$, respectively.

Similarly, the second row of design variables, $X_5$, $X_6$, $X_7$ and $X_8$, represents the number of carbon-carbon double bonds associated with each of the four carbon atoms designated by $C_1$, $C_2$, $C_3$ and $C_4$, respectively. In a similar manner, $X_9$, $X_{10}$, $X_{11}$ and $X_{12}$ represent the number of carbon-hydrogen single bonds associated with $C_1$, $C_2$, $C_3$ and $C_4$, respectively; $X_{13}$, $X_{14}$, $X_{15}$ and

$X_{16}$ represent the number of carbon-chlorine single bonds associated with $C_1$, $C_2$, $C_3$ and $C_4$, respectively; $X_{17}$, $X_{18}$, $X_{19}$ and $X_{20}$ represent the number of carbon-fluorine single bonds associated with $C_1$, $C_2$, $C_3$ and $C_4$, respectively; $X_{21}$, $X_{22}$, $X_{23}$ and $X_{24}$ represent the number of carbon-bromine single bonds associated with $C_1$, $C_2$, $C_3$ and $C_4$, respectively and $X_{25}$, $X_{26}$, $X_{27}$ and $X_{28}$ represent the number of carbon-oxygen single bonds associated with $C_1$, $C_2$, $C_3$ and $C_4$, respectively.

**Table 1.1:** Structural design variables used in the design methodology

| | $C_1$ | $C_2$ | $C_3$ | $C_4$ |
|---|---|---|---|---|
| **Number of C-C bonds** | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
| **Number of C=C bonds** | $X_5$ | $X_6$ | $X_7$ | $X_8$ |
| **Number of C-H bonds** | $X_9$ | $X_{10}$ | $X_{11}$ | $X_{12}$ |
| **Number of C-Cl bonds** | $X_{13}$ | $X_{14}$ | $X_{15}$ | $X_{16}$ |
| **Number of C-F bonds** | $X_{17}$ | $X_{18}$ | $X_{19}$ | $X_{20}$ |
| **Number of C-Br bonds** | $X_{21}$ | $X_{22}$ | $X_{23}$ | $X_{24}$ |
| **Number of C-O bonds** | $X_{25}$ | $X_{26}$ | $X_{27}$ | $X_{28}$ |

### 1.6.1. Distinguishing structural isomers using the twenty eight structural design variables

As mentioned earlier, the developed design methodology is capable of differentiating between structural isomers of an organic compound. In addition differentiates between different structural isomers by generating unique sets of twenty eight structural design variables for each of the structural isomers.

Isomers are compounds that have the same molecular formula, but differ in the arrangement of individual atoms in the molecule. For example, the molecular formula $C_4H_9Cl$ may correspond to any of the following organic compounds:

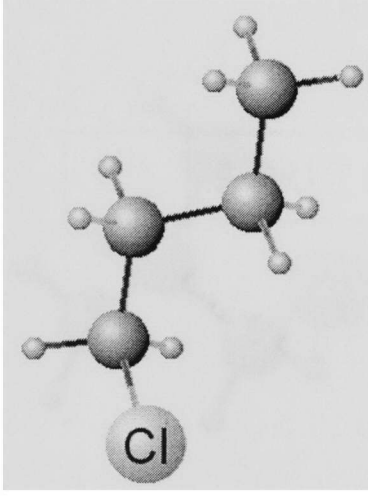
a) 1-chlorobutane

b) 2-chlorobutane

c) 1-chloro-2-methylpropane

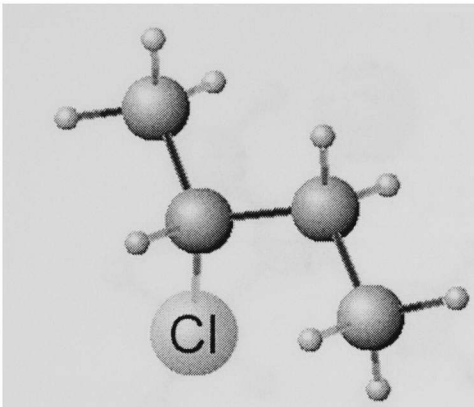
d) 2-chloro-2-methylpropane

Figure 1.4, Figure 1.5, Figure 1.6 and Figure 1.7 shows the molecular structure of 1-chlorobutane, 2-chlorobutane, 1-chloro-2-methylpropane and 2-chloro-2-methylpropane, respectively. Even though they are represented by the same molecular formula, distinct structural features can be clearly seen in their molecular structures. Values of the corresponding sets of twenty eight structural design variables show that the proposed design methodology represents the molecular structure of all the four compounds in a unique manner. Hence, it successfully differentiates between the four isomers of the organic compound having the formula $C_4H_9Cl$.



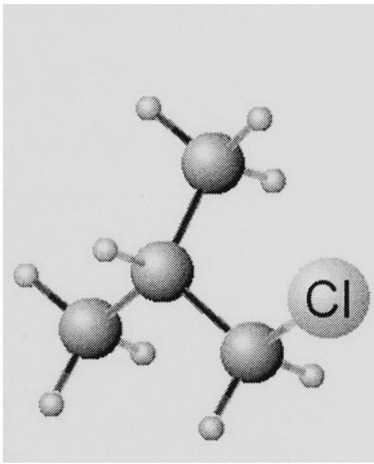

| Number of | $C_1$ | $C_2$ | $C_3$ | $C_4$ |
|---|---|---|---|---|
| C-C bonds | 1 | 2 | 2 | 1 |
| C=C bonds | 0 | 0 | 0 | 0 |
| C-H bonds | 2 | 2 | 2 | 3 |
| C-Cl bonds | 1 | 0 | 0 | 0 |
| C-F bonds | 0 | 0 | 0 | 0 |
| C-Br bonds | 0 | 0 | 0 | 0 |
| C-O bonds | 0 | 0 | 0 | 0 |

**Figure 1.4:** Molecular structure of 1-chlorobutane and the corresponding set of structural design variables



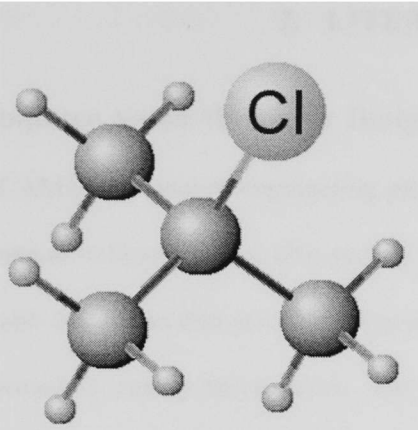

| Number of | $C_1$ | $C_2$ | $C_3$ | $C_4$ |
|---|---|---|---|---|
| C-C bonds | 1 | 2 | 2 | 1 |
| C=C bonds | 0 | 0 | 0 | 0 |
| C-H bonds | 3 | 1 | 2 | 3 |
| C-Cl bonds | 0 | 1 | 0 | 0 |
| C-F bonds | 0 | 0 | 0 | 0 |
| C-Br bonds | 0 | 0 | 0 | 0 |
| C-O bonds | 0 | 0 | 0 | 0 |

**Figure 1.5:** Molecular structure of 2-chlorobutane and the corresponding set of structural design variables




| Number of | $C_1$ | $C_2$ | $C_3$ | $C_4$ |
|---|---|---|---|---|
| C-C bonds | 1 | 3 | 1 | 1 |
| C=C bonds | 0 | 0 | 0 | 0 |
| C-H bonds | 2 | 1 | 3 | 3 |
| C-Cl bonds | 1 | 0 | 0 | 0 |
| C-F bonds | 0 | 0 | 0 | 0 |
| C-Br bonds | 0 | 0 | 0 | 0 |
| C-O bonds | 0 | 0 | 0 | 0 |

**Figure 1.6:** Molecular structure of 1-chloro-2-methylpropane and the corresponding set of structural design variables




| Number of | $C_1$ | $C_2$ | $C_3$ | $C_4$ |
|---|---|---|---|---|
| C-C bonds | 1 | 3 | 1 | 1 |
| C=C bonds | 0 | 0 | 0 | 0 |
| C-H bonds | 3 | 0 | 3 | 3 |
| C-Cl bonds | 0 | 1 | 0 | 0 |
| C-F bonds | 0 | 0 | 0 | 0 |
| C-Br bonds | 0 | 0 | 0 | 0 |
| C-O bonds | 0 | 0 | 0 | 0 |

**Figure 1.7:** Molecular structure of 2-chloro-2-methylpropane and the corresponding set of the structural design variables

# 2. LITERATURE REVIEW

## 2.1. Computer Aided Molecular Design (CAMD) in the prediction of properties

CAMD is a reverse engineering procedure which not only reduces the time to develop a new chemical compound, but also proves to be a more systematic methodology in generating several new molecules that are promising enough to be considered for experimental verification. This approach is usually based on the idea that any organic molecule can be constructed from a pre-specified set of building blocks (descriptors of the molecular structure). The value of a property for the compound can be estimated by summing the contributions for the specific property from each chemical group constituting the molecular structure of the compound. However, this usually requires the availability of structure based property models, which evaluate the contributions for the specific property of each of the chemical groups used in describing the molecule. In essence, this approach aims to generate molecules that conform to various property requirements that are known and specified *a priori*. However, experimentation, based on empirical trial and error, becomes the only viable option in most of the cases in which property models do not exist. Achenie *et al.* described the methods, tools and the applications of CAMD methodology in details (Achenie, Gani and Venkatasubramanian 2003).

According to Poling *et al.* (Poling, Prausnitz and O'Connell 2004), most of the CAMD methodologies used for the estimation of properties are based on additive methods like the Group Contribution Approach (GCA) or bond/ atom contribution approach. In any of these methodologies, the properties of a chemical compound are expressed in terms of the number of occurrences of certain groups of atoms arranged in a specified manner (or individual bonds/ individual atoms), in the molecule. Although these methodologies allow the flexibility to analyze structurally diverse set of chemicals, their use in certain problems becomes restricted due to their dependence on the number and types of the chemical groups (or bonds/ atoms) selected *a priori*.

Nevertheless, in the past, a number of efforts have been made to predict different properties like the boiling temperature (Walters, Myrdal and Yalkowsky 1995), (Cholakov, Wakeham and Stateva 1999), (Nannoolal, *et al.* 2004), (Li, Higashi and Tamura 2006), (Wang, *et al.* 2009); critical volumes (Sastri, Mohanty and Rao 1997); acute toxicity (96-h $LC_{50}$) of organic compounds to the Fathead Minnow aquatic species (Martin and Young 2001); enthalpy of vaporization at different temperatures (Basarova and Svoboda 1995), (Li, *et al.* 1997); probability and rate of aerobic biodegradation (Boethling, *et al.* 1994); by using the group contribution approach.

In many applications a conventional group/ bond/ atom contribution based approach cannot be applied due to the absence of suitable property prediction models and the complexity of the generated molecules. Most of the alternatives to these methodologies are either *Quantitative Structure-Property Relationship* (QSPR) based approaches that use molecular descriptors such as local charge densities, molecular surface area, *etc.* or are based on the addition of the weighted contributions of additive properties such as molecular weight, boiling temperature, *etc.* Harper *et al.* (Harper, *et al.* 1999) proposed a modified CAMD approach based on an atomic/ molecular structure generation algorithm and molecular modeling tools. However, this methodology still relied on the conventionally used generation and test approach. This modified methodology was then applied to the designing of industrial solvents. It was reported that this methodology lacked the capability to address many of the issues related with conformers. It should be pointed out that this design methodology is different from the methodology explained in this work which utilizes multi-dimensional response surfaces and multi-objective constrained optimization algorithms.

Marrero and Gani further tried to enhance the accuracy, reliability and the range of application of the CAMD methodology (Marrero and Gani 2001). This methodology relied heavily on the group contribution approach. However, the researchers proposed a three-level scheme that was developed to estimate the normal boiling temperature, critical temperature,

critical pressure, critical volume, standard enthalpy of formation, standard enthalpy of vaporization, standard Gibb's energy, normal melting temperature and standard enthalpy of fusion of a wide variety of organic compounds. Estimation was performed successively at all three levels where the first level provided the initial approximation of the property values (only first order groups were used in the first level), which was consequently improved and refined in the second (both first order and second order groups were used in the second level) and the third level (groups belonging to first order, second order and third order, were used in the third level), respectively.

Fermeglia *et al.* (Fermeglia, Ferrone and Pricl 2003) proposed a general all atom force field model for predicting the phase equilibria of low molecular weight hydrofluorocarbons that can be used as refrigerants. This model was based on the ab initio calculations and molecular dynamics simulations. Firstly, the molecular structures of all the eight hydrocarbons (each hydrocarbon only consisted of atoms of carbon, hydrogen and fluorine) considered in this study was optimized using density functional theory based calculations. Secondly, varied configurations of the molecules used in this study were obtained to generate the potential energy surface sampling. The results were consequently fitted using least-square fitting routines. The calculated intramolecular energetical properties (for molecules in isolation) such as absolute energy, enthalpy of formation, ionization potential, dipole moment, *etc.* were reported to be in good agreement with the experimental data.

Commercially available software packages such as the *Estimation Program Interface (EPI)* suite have also been developed for the estimation of a wide variety of chemical, physical and environment related properties. For example, the atmospheric oxidation (U.S. Environmental Protection Agency 2008) prediction and biodegradability probability (U.S. Environmental Protection Agency 2009) of an organic compound may be performed by inputting its *Simplified Molecular Input Line Entry System* (SMILES) notation in the user interface of the *Atmospheric*

*Oxidation Program for Microsoft Windows* (AOPWIN) and *Biodegradation Probability Program* (BIOWIN).

## 2.2. Development of replacement refrigerants using computational approaches

One of the earliest reported works on the use of computer aided molecular design in designing replacement refrigerants was published in the late 1980's (Joback and Stephanopoulos 1989). Some other research efforts such as the one reported by Bhattacharjee (Bhattacharjee 1995), to predict the molecular structure of organic compounds that can be used as viable refrigerants, were limited to the study of compounds belonging particularly to haloethanes.

With the enforcement of newer and stricter environmental laws and regulations, the design methodologies developed in the 1990's started incorporating objectives that were specifically addressing the impact of refrigerants on the environment. For example, Churi and Achenie (Churi and Achenie 1996) restricted the ODP of the new refrigerant to be less than or equal to 0.2 (The Clean Air Act of 1995 required the refrigerants to have ODP's below 0.2 by the year 2000). This work was aimed at optimizing many performance objectives such as the specific heat and the enthalpy of vaporization, *via* the formation of a single performance objective. Mixed integer non-linear programming approach was used in their work. Duvedi and Achenie (Duvedi and Achenie 1996) also incorporated the ODP of compounds, however, using a different method. An analytical expression to find the ODP of chlorofluorocarbons, with one or two carbon atoms and a certain number of chlorine atoms, was developed and utilized. However, again all the individual performance objectives were linearly combined to form a single objective, which was optimized using an algorithm designed to solve a finite sequence of non-linear programming sub problems and mixed integer linear programming master problems.

In 1997, the same authors also reported results of a proof-of-concept study to design environmentally benign refrigerant mixtures (Duvedi and Achenie 1997). Due to the limited

availability of experimental data, the basis set consisted of only 21 of the most commonly used refrigerants. Moreover, only binary refrigerant mixtures were considered due to the relative ease of vapor-liquid equilibrium calculations and the unavailability of accurate experimental data related to the normal boiling temperature, pressure, volume, critical temperature, *etc*. The objective function comprised of the compressor displacement and the ODP.

Sahinidis and Tawarmalani (Sahinidis and Tawarmalani 2000) demonstrated the importance of identifying the global optima of a mixed integer non-linear programming formulation when they attempted to find a replacement for Freon 12. The use of a modified branch-and-bound algorithm which included range reduction techniques, finite branching schemes and a theory of convex enclosures for non-linear functions of binary variables, provided new solutions. However, all the performance objectives were linearly combined to formulate a single objective optimization problem.

Khetib *et al.* (Khetib, Meniai and Lallemand 2009) attempted to design CFC and HCFC substitutes by using the group contribution approach. This approach incorporated functional groups that specifically excluded chlorine atoms and it explored the idea of generating binary mixtures between certain generated refrigerants themselves or with other selected compounds. However, the applications of the method were limited because only five functional groups were chosen and those groups contained only three different elements: carbon, hydrogen and fluorine.

# 3. MULTI-OBJECTIVE OPTIMIZATION

## 3.1. Basic terms and definitions

This section explains the basic terms necessary for understanding the rest of the chapter. The terms are commonly used in current literature found on the general topic of multi-objective optimization. Here, we will relate this terminology to specific applications in design optimization of molecules.

In our particular application there will be several simultaneous objectives that are often contradictory For example, we would like to maximize the enthalpy of vapor liquid phase change while at the same time minimizing the ODP and GWP. Each of these objectives depends on the number of elements constituting a given refrigerant, the number of atoms of each of these elements and the number and type of bonds among each of these elements. All of these parameters thus, represent design variables that could be optimized in order to create an entire set of new refrigerants that will have the best trade-off among the desired multiple objectives as defined earlier. In general, we could have $m$ such objective functions each depending on $n$ design variables. It will be seen at a later stage that the optimization problem studied in this work, deals with a mixture of discrete and continuous design variables. However, all the objective functions have real values.

A set of inequality and equality constraints must be specified which limit the values that individual design variables and/ or a combinations of design variables (and/ or objective functions) may achieve in order to guarantee the physical relevance of the numerical results obtained from the optimization algorithm.

The region of search space in which the candidate new refrigerants (designs) satisfy all the constraints is known as the feasible search space. The region of the search space in which the designs do not satisfy at least one constraint is called the infeasible search space.

Multi-objective optimization algorithms are conceptually different from most single-objective optimization algorithms. Specifically, they do not utilize gradients of the objective functions with respect to each of the design variables. Instead, they use essentially stochastic procedures in their search for the best trade-off solutions among several desired objectives. This approach requires a pool (population) of candidate solutions (population members) to be evaluated for their multiple properties (objective functions) Consider an example in which $m$ objectives are to be minimized. A solution $\overrightarrow{x_1}$ is considered to be a dominated solution if there exists a feasible solution $\vec{x}$ which satisfies the following two conditions:

a) $\vec{x}$ is at least as good as $\overrightarrow{x_1}$ for all objectives $f_i$ (i = 1, …, m), i.e.,

$$f_i(\vec{x}) \leq f_i(\overrightarrow{x_1}) \; \forall \; 1 \leq i \leq m \tag{3.1}$$

b) $\vec{x}$ is strictly better than $\overrightarrow{x_1}$ on at least one objective i, i.e.,

$$f_i(\vec{x}) < f_i(\overrightarrow{x_1}) \text{ for at least one } 1 \leq i \leq m \tag{3.2}$$

If equation 3.1 and equation 3.2 are satisfied, then it is said that $\vec{x}$ weakly dominates $\overrightarrow{x_1}$. For example, consider an optimization problem in which two objectives, objective 1 and objective 2

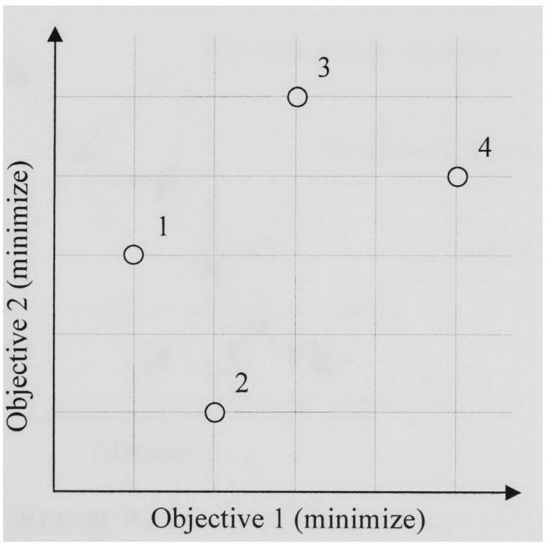



**Figure 3.1:** An example explaining the concept of dominance

that need to be minimized. There are four solutions to this problem. These solutions are represented in the objective function space (Figure 3.1), by small circles numbered 1, 2, 3 and 4. It can be seen that solution 1 performs better than all the other solutions with respect to objective 1. Considering the conditions for dominance, it can be concluded that solution 1 dominates solution 3 and solution 4 with respect to both objectives. Solution 2 does better as compared to solution 1, with respect to only objective 2. Hence, neither solution 1 is completely dominated by solution 2 nor solution 2 is completely dominated by solution 1. Thus, these two solutions (1 and 2) represent the best trade-off solutions among the four candidate solutions presented.

The non-dominated set of solutions refers to those solutions which are not dominated by any other solution. A non-dominated solution is also called *Pareto-optimal*, and the non-dominated solution set among all feasible solutions is termed the *Pareto-optimal* set. In the case of a two-objective problem, the *Pareto-optimal* set of solutions manifests as a two-dimensional curve.

For example, consider the case of an optimization problem with two competing objectives and both objectives need to be maximized. Figure 3.2 illustrates the concept of non-

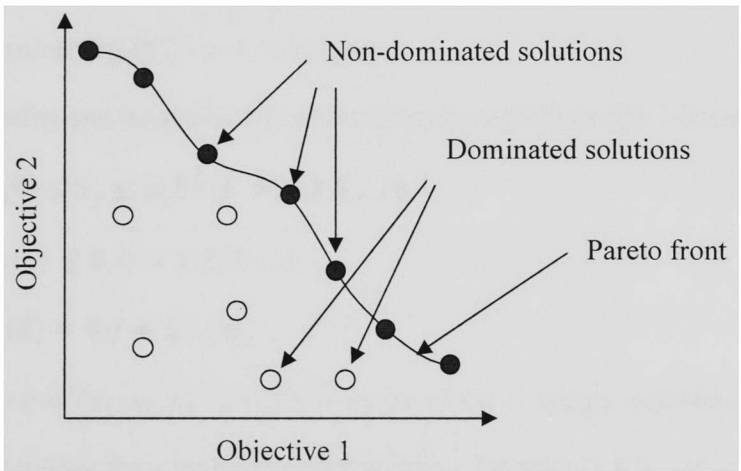



**Figure 3.2:** A schematic illustrating the concept of non-dominated solution, dominated solution and Pareto front for a two-objective problem (adapted from (Zitzler 2002))

dominated solution, dominated solution and the Pareto front, for such a problem.

For a problem involving three objectives, the Pareto front is generally a surface. In general, for a problem involving $m$ competing objectives, the Pareto front will manifest itself as a hyper surface in an $m$ dimensional space.

It should be pointed out that a commonly used approach of minimizing the linear combination (a weighted sum) of the objective functions will result in finding only a single point on the Pareto front. Thus, creating a Pareto set of optimal solutions using this approach would be computationally prohibitively expensive.

### 3.2. Overview of multi-objective optimization

Consider a multi-objective optimization problem which has $n$ decision variables and $m$ objective functions defined by $f_1, f_2, f_3 \ldots, f_m$. Without any loss of generality, it can be assumed that each of the $m$ components of the vector of objective functions $f(\vec{x})$ is to be minimized. Such a minimization problem could be converted to a maximization problem by multiplying the objective functions by -1.

This multi-objective optimization problem can be mathematically defined as

$$minimize\ f_i(\vec{x}), i = 1, 2, 3 \ldots, m \tag{3.3}$$

subject to the equality and inequality constraints given by equations 3.4, 3.5 and 3.6.

$$x_p{}^{(l)} \leq x_p \leq x_p{}^{(u)}, p = 1, 2, 3 \ldots, n \tag{3.4}$$

$$g_k(\vec{x}) \leq 0, k = 1, 2, 3 \ldots, l \tag{3.5}$$

$$h_j(\vec{x}) = 0, j = 1, \ldots, q \tag{3.6}$$

A solution vector $x = (x_1, x_2, x_3 \ldots, x_n)$ is a vector of the $n$ design variables and, as mentioned previously, is not unique for a multi-objective problem. Equation 3.4 represents a constraint that limits the value of the decision variable $x_p$ between an upper bound ($x_p{}^{(u)}$) and a lower bound ($x_p{}^{(l)}$). Equation 3.5 is a generic representation of $l$ inequality constraints which constrict the

objective space. Similarly, equation 3.6 is a generic representation of $q$ equality constraints, which also limit the objective space.

## 3.3. Multi-objective optimization using Indirect Optimization on the basis of Self Organization (IOSO) software

The capabilities of IOSO technology algorithms have been documented in the past in detail (Egorov, *et al.* 2007), (Egorov-Yegorov, *et al.* 2003), (Sigma Technology). This algorithm is based on the application of response surface methodology and the adaptive use of global and middle range multi point approximation. These techniques allow for a more accurate response surface fit in the vicinity of the optimum of the most recent optimization cycle. Another advantage of using these techniques is that they are capable of generating accurate response surfaces, even when very small amount of data (candidate solutions) is available. IOSO also achieves this by implementing a number of independent highly efficient evolutionary self organizing algorithms. The selection of a particular algorithm is done internally and adaptively and it depends on the specific problem.

Any iteration of IOSO essentially comprises of the following two steps, in the same order in which they are listed (Egorov, *et al.* 2007):

1) Creation of surrogate models that analytically approximate the objective functions.
2) Optimization of the approximation function, in the current search region.

Figure 3.3 shows a schematic diagram of the sequence of steps constituting the basic optimization algorithm used in IOSO. Between two successive iterations the experiment plan is modified; the current search space is adaptively selected and modified; a global or a middle range response function is chosen and the response surface is transformed; the parameters and the structure of the optimization algorithms are modified and if necessary, new potentially promising points are selected within the current search region. Because of the small number of points in the

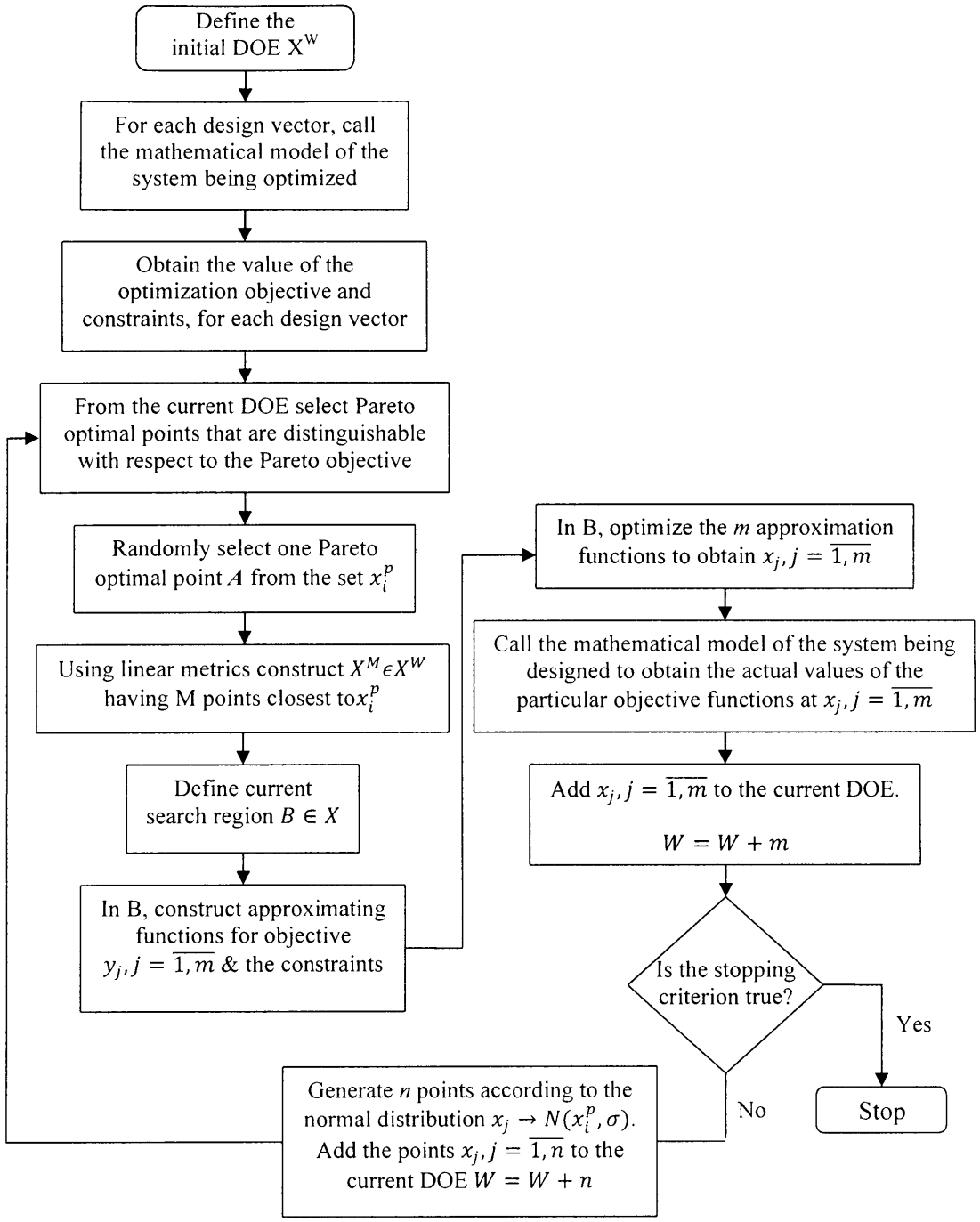



**Figure 3.3:** An illustration of the basic optimization algorithm employed in IOSO

initial *Design of Experiments* (DOE) and a large search space, the accuracy of the process during initial stages could be poor. During each iteration, the behavior of the objective functions, inside

the current search space, near the extremum, is stored and the response function is made more accurate. The number of points in the search region is increased and the search space itself is progressively reduced. This results in an increase in the approximation function accuracy, which in turn increases the efficiency of the whole optimization process.

IOSO NM provides many advantages over the commonly used techniques in solving multi-objective optimization problems (Sigma Technology). These advantages are the following.

1. The algorithm can solve practical tasks which have non-convex, non-differentiable and stochastic objective functions and constraints with mixed variables.
2. The algorithm does not need to significantly adapt the mathematical models of the system being designed.
3. For tasks involving multiple extrema, the algorithm finds the global optimum with a relatively high probability.
4. The *Pareto–optimal* solutions are generally obtained by making a relatively small number of direct calls to the mathematical models of the system, which is being designed.
5. The algorithms used in IOSO find desired number of Pareto points, distributed uniformly in the objective function space.
6. Convolution methods are not employed to solve multi-objective problems (Egorov, *et al.* 2007)
7. Parallelization of the whole computational process is possible without much additional complexity.

# 4. PREPARATION OF THE DATA SET FOR OPTIMIZATION

## 4.1. Introduction

Figure 4.1 shows the process flow used in the preparation of the data set. As shown, the process starts with the identification of the different classes of organic compounds that are represented in the pool of data. Based on the structures of existing refrigerants, availability of thermo-physical properties and stability related issues, compounds constituting the data set were identified. Section 4.2 explains this selection in more detail.

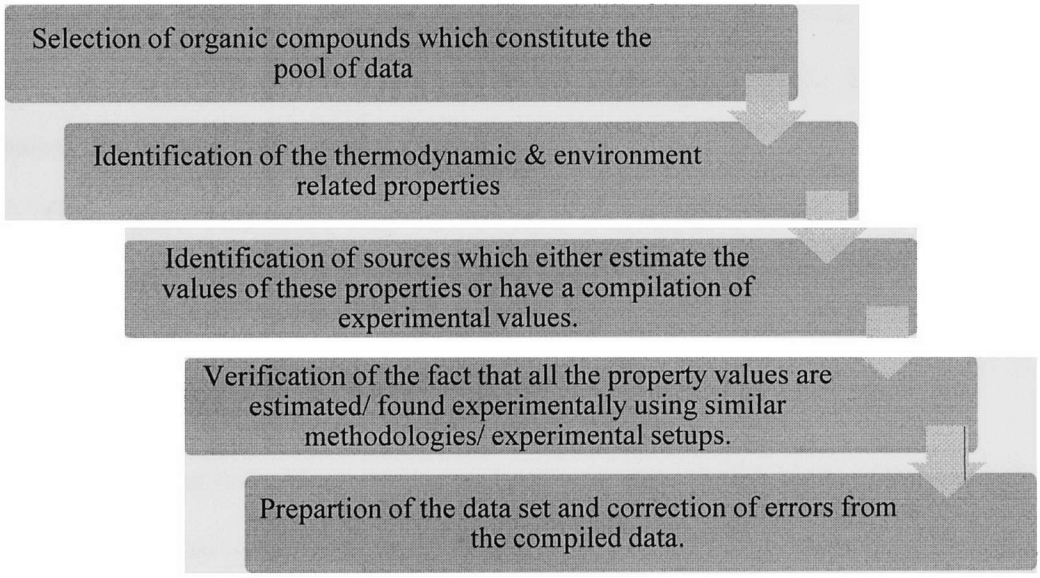



**Figure 4.1:** An illustration showing the process flow used in the preparation of the data set

Many thermodynamic and environment related properties are important for evaluating the performance of a given refrigerant and its suitability for a particular application. The thermodynamic properties may include normal boiling temperature, enthalpy of vaporization, liquid specific heat, vapor pressure, *etc*. Whereas, the properties required for estimating its suitability for a given application may include the ODP, GWP, half-life in troposphere, biodegradability, flammability and toxicity among others. Because of the non-availability of experimental/ estimated values of many of such properties in the public domain, this study is

limited to the use of normal boiling temperature, enthalpy of vaporization, vapor pressure, half-life in the atmosphere and biodegradability. However, with the availability of more data, this study can be easily expanded to incorporate any number of properties. Different sources were identified and used to extract the information about each compound selected in this study.

## 4.2. Selection of organic compounds for the data pool

Currently used refrigerants involve an extensive use of chlorofluorocarbon (CFC), hydrochlorofluorocarbon (HCFC), perfluorocarbon (PFC) and other halogenated hydrocarbon based compounds. Non-fluorochemical, fluorochemical and hydrofluoroethers based compounds are also used as refrigerants in some applications (Sekiya and Misaki 2000), (Calm 2008), (Mohanraj, Jayaraj and Muraleedharan 2009).

Only those compounds which have four or fewer carbon atoms were selected in this study. Duvedi *et al.* (Duvedi and Achenie 1996) reported that molecules with larger number of carbon atoms do not exhibit vapor pressures in the range desired for refrigerants. Churi *et al.* (Churi and Achenie 1996) also reported similar observations.

It is observed that compounds having double or higher bonds show stability problems by spontaneously decomposing or by polymerizing in a short period of time. In many situations stability problems can be avoided by simply ensuring that the new refrigerant does not have double or triple bonds. Therefore, compounds containing any triple bonds are not considered. However, this study uses compounds with carbon-carbon double bonds to prevent any elimination of potentially good molecules.

Compounds containing nitrogen atoms are not considered, because of primarily two reasons. They are:

1. When both halogen and nitrogen groups exist in a low boiling compound, the compound shows strong tendencies to explode (Duvedi and Achenie 1996).

2. We could not find a sufficiently large number of nitrogen containing compounds, with their properties available. Consequently, such compounds would not be represented well in the data pool.

Compounds containing other elements like sulphur, phosphorous, iodine, *etc.* are also not considered. Again, this is because of the fact that the number of such compounds for which either experimental or estimated property data could be found, is not sufficient to be incorporated in the data pool. An insufficient representation in the data pool would result in a decrease in the accuracy of the response surface. Because of the above concerns, compounds belonging to the generic class of alkanes, alkenes, alkyl halides, ethers and halogenated ethers only were considered for selection in the present effort.

## 4.3. Selection of properties for the refrigerant problem

Each of the sections numbered between 4.3.1 and 4.3.5 explains the necessity of choosing a particular property, its relevance to the refrigerant design problem and in certain cases, the comparison of different models considered for the estimation of that specific property.

### 4.3.1. Normal boiling temperature

In general, refrigerants with normal boiling temperatures (boiling temperatures at one standard atmosphere pressure) lower than the room temperature have been found to have wider applications. This requirement might be different for compounds which are intended to be used as refrigerants in industrial applications. However, in most cases, the replacement refrigerant should be such that it has its normal boiling temperature close to the normal boiling temperature of the refrigerant which it is designed to replace.

### 4.3.2. Vapor pressure

The vapor pressure of the refrigerant fluid is considered to be an important property for the refrigerant design. In order to reduce the chances of air/ moisture leaking into the system and to reduce the chances of any explosions if a hydrocarbon is used as a refrigerant, the lowest pressure in the cycle should be more than the atmospheric pressure. Whereas, a high value of system pressure increases the size of the system, thereby, indirectly increasing the cost of the system. It has been reported in the current literature that the capacity of a refrigeration system is almost directly proportional to the evaporating pressure inside the system (Granryd 2001). Therefore, the replacement refrigerant should have similar vapor pressure as that of the refrigerant it is designed to replace. It has also been observed that those refrigerants that are required to work in heat pumps above 90°C should possess vapor pressures which vary with temperature at a moderate rate (Pfeiffenberger 1982).

### 4.3.3. Enthalpy of vaporization

The enthalpy of vaporization (latent heat of liquid/ vapor phase change) is directly linked to the volumetric flow necessary to produce a certain amount of cooling. A higher enthalpy of vaporization of the refrigerant fluid reduces the amount of the volumetric flow needed to produce the desired amount of cooling. A reduced volumetric flow requirement indirectly reduces the size of the refrigeration system, thereby reducing the cost of operation and installation.

### 4.3.4. Half-life time in troposphere

The atmospheric lifetime is one of the major factors in determining the environmental impact of a refrigerant. In some cases, it has even been observed that refrigerants with longer lifetimes can have a higher negative environmental impact than what can be expected by measuring their actual concentration in the atmosphere (Halimic, *et al.* 2003). Atmospheric

lifetime has been linked directly and indirectly with global warming. Because of this reason, the use of refrigerants with extremely large lifetimes should be avoided (Kopko 1990).

The hydroxyl radical, OH, possesses the ability to react with almost all the gases which are emitted into the atmosphere. This ability makes it the most important radical for driving the atmospheric chemistry (Carr, Heard and Blitz 2009). Therefore, hydroxyl radicals are often seen as oxidants that control the removal of pollutants from the atmosphere.

In troposphere, OH radicals are mainly produced through the photolysis of ozone, which results in the formation of excited O ($^1$D) species at wavelengths $\lambda \leq 320$nm. This is followed by the reaction of O ($^1$D) with a molecule of either $N_2$ or $O_2$. Although most of the atoms of the excited species get deactivated through collisions with other gases, a small percentage of them do cause the formation of OH radicals. The whole process is illustrated in the chemical reactions, which are listed below (Li, Matthews and Sinha 2008).

$$O_3 + h\nu(\leq 320\text{nm}) \rightarrow O(^1D) + O_2 \quad (4.1)$$

$$O(^1D) + M \rightarrow O(^3P) + M \text{ (where M is } N_2 \text{ or } O_2) \quad (4.2)$$

$$O(^1D) + H_2O \rightarrow 2OH \quad (4.3)$$

Under a steady state approximation, the production rate for OH ($R_{OH}$) can be calculated from the kinetics analysis of the above reaction sequence. $R_{OH}$ is given by equation

$$R_{OH} = \frac{2 \cdot j \cdot k_1 \cdot [H_2O] \cdot [O_3]}{(k_1[H_2O] + k_2[M])} \quad (4.4)$$

In the above equation, $j$ is the $O_3$ photolysis rate to form O ($^1$D), $k_1$ is the rate constant for the reaction of O ($^1$D) with water and $k_2$ is the rate constant for deactivation of O ($^1$D) by gases like $N_2$ and $O_2$. [$O_3$] and [$H_2O$] represent the concentration of ozone and water, respectively.

It can be safely assumed that the chemical, with which OH radical is reacting, is present in the troposphere at a very low concentration. Moreover, a steady state concentration of OH radical is maintained by its generation in the presence of sunlight. Under these assumptions the

reaction can be considered a pseudo first order reaction and the half-life in the troposphere can be calculated by using equation 4.5.

$$t_{1/2} = \frac{0.693}{k_{OH}[OH]} \tag{4.5}$$

Here,

$k_{OH}$ – hydroxyl radical rate constant (in $cm^3$/ molecule-second)

[OH] – hydroxyl radical concentration (in molecules (or radicals)/$cm^3$)

Hence, the half-life of a chemical compound can be calculated if an experimental value or an estimated value of the rate constant of the chemical is known as well as some average hydroxyl radical concentration. This average hydroxyl radical concentration depends on the time frame in which it is calculated. For chemicals which react faster, a twelve hour daylight time frame might be used to estimate the average OH radical concentration. However, for the chemicals which react slowly (the average life times in the atmosphere are expected to be more than a few days), a twenty four hour time frame might be more applicable.

It should be noted that the hydroxyl radical concentration varies with the temperature, time of the year, cloud cover, latitude, the hemisphere of the place where the OH radical concentration is found out, *etc*. However, global mean OH radical concentrations for the troposphere (up to a height of 14 km) have been reported previously as $9.2 \times 10^5$ mol/$cm^3$ (Bahm and Khalil 2004). The concentrations were averaged over a 24-hour time period. The zonally and monthly averaged OH concentrations have been reported previously. They are listed in Appendix B.

### 4.3.5. Biodegradability

Biodegradation is the process of the transformation of organic compounds by micro organisms through enzymatic reactions like oxidation, reduction and/ or hydrolysis. The biodegradability of organic compounds, which could be used as refrigerants, has been a subject of

many experimental studies (Sablijic and Peijnenburg 2001). For many of the existing substances, results of a ready biodegradability test may be available which help in determining if the substance is potentially persistent. However, for many other substances, such a preliminary assessment aid may not be available. QSAR based models have been used to overcome this difficulty. A number of models for predicting the probability and/or rate of aerobic/anaerobic biodegradation of chemical compounds have been proposed. Some of the most commonly used approaches to develop models for estimating the biodegradability of organic compounds have been the group contribution approach, the expert system approach and a combination of expert system and probabilistic modeling of pathways.

BIOWIN (U.S. Environmental Protection Agency 2009) has been used to estimate the biodegradability of all the 295 compounds considered in this study. This program estimates the probability of an organic compound for rapid aerobic and anaerobic biodegradation. It is assumed that the compound is subjected to an atmosphere in which mixed populations of environmental organisms is present. The program classifies the organic compounds in one of the two categories – the ones which biodegrade easily and the ones which biodegrade slowly. The models used in the program have been described in more details in current literature (Sablijic and Peijnenburg 2001), (Pavan and Worth 2006).

The biodegradability of any compound can be estimated based on the possible contribution of any of the 36 chemical groups accounted for in this software. For each of these 36 chemical groups, the contribution to biodegradability and a molecular weight chemical group constant is utilized. During the development of the program, 295 chemicals from the *Biodegradation Probability Program* (BIODEG) database were used to derive the chemical group probability values. These constants were developed by using multiple linear and non-linear regression analyses. This dataset consisted of 186 easily biodegradable chemicals and 109 slowly

biodegradable compounds. The version 4.10 of BIOWIN contains seven separate models. These models are listed below:

1. Biowin1 – linear probability model
2. Biowin2 – non-linear probability model
3. Biowin3 – expert survey ultimate biodegradation model
4. Biowin4 – expert survey primary biodegradation model
5. Biowin5 – *Japanese Ministry of International Trade and Industry* (MITI) linear model
6. Biowin6 – MITI nonlinear model
7. Biowin7 – anaerobic biodegradation model

Tests conducted for the validation of the models using a large set of consistent biodegradation data of 733 compounds, have previously shown that the overall performance of the BIODEG model is only 61.1% of correct predictions. However, the model correctly identified 91.1% of the chemical compounds that were classified as 'ready' biodegradable.

BIOWIN is a model which is based on the group contribution approach. Therefore, like many other models which are based on group contribution approach, it lacks the sophistication to consider the neighboring substituents and their position. For example, 2, 4, 6 – trichlorophenol which is considered 'ready' degradable in the MITI-1 test, is predicted to be 'not-ready' biodegradable by this model. When small molecules, containing only commonly found chemical groups that are present in small numbers, are presented to the model for prediction, it is generally found that the model produces reasonable estimates.

However, wrong predictions become more likely if larger molecules with relatively complicated structures are presented. For example, it is found that the model is also prone to make unreasonable predictions for isomers of larger molecules and other similar chemicals. Minor differences in the chemical structure may lead to differences in the ways micro organisms perform degradation of an organic compound. Due to this fact, these seemingly small structural

differences may also cause large differences in the rates of degradation. BIOWIN also ignores the role of steric factors or low solubility in predicting the biodegradability of an organic compound. For practical reasons, many chemical groups which are significant for biodegradation do not have sufficient representation in the training set. Due to this reason, the model lacks coefficients for many chemical groups including phosphonate (C-P bond), pyrimidine ring and cycloaliphatics. If a compound containing such features is presented for the estimation of biodegradability, such structural features are not considered for the prediction of biodegradability.

Table 4.1 and Table 4.2 summarize the key results of the evaluation of all the seven Biowin models used in BIOWIN v 4.10. It should be noted that among all the seven models, Biowin 5 and Biowin 6 were trained using the largest number of chemical compounds. The total number of compounds in their training set was 589.

**Table 4.1:** Results of the evaluation of Biowin 1, Biowin 2, Biowin 3, Biowin 4 and Biowin 7

| **Parameter** | **Biowin 1** Linear Model | **Biowin 2** Non-Linear Model | **Biowin 3** Ultimate Model | **Biowin 4** Primary Model | **Biowin 7** Anaerobic Model |
|---|---|---|---|---|---|
| Total Correct | 264 | 275 | 167 | 165 | 152 |
| Compounds in training set | 295 | 295 | 200 | 200 | 169 |
| % correct total | 89.5 | 93.2 | 83.5 | 82.5 | 89.9 |
| % correct, fast degradable | $\frac{181}{186} = 97.3^{a}$ | $\frac{181}{186} = 97.3^{a}$ | $\frac{101}{108} = 93.5^{b}$ | $\frac{101}{119} = 84.9^{c}$ | $\frac{75}{82} = 91.5^{d}$ |
| % correct, slow degradable | $\frac{83}{109} = 76.1$ | $\frac{94}{109} = 86.2$ | $\frac{66}{92} = 71.7$ | $\frac{64}{81} = 79.0$ | $\frac{77}{87} = 88.5$ |

[a] When the predicted probability is greater than 0.5, the compound is considered to be degrading fast.

[b] A compound displays fast ultimate degradation if the calculated rating (numerical value from the model) is found greater than 2.5.

[c] A compound displays fast primary degradation if the calculated rating (numerical value from the model) is found greater than or equal to 3.5.

[d] When the predicted probability is greater than 0.5, the compound is considered to be degrading fast.

**Table 4.2:** Results of the evaluation of Biowin 5 and Biowin 6

| Set | Parameter | **Biowin 5** Linear MITI Model* | **Biowin 6** Non-Linear Model* |
|---|---|---|---|
| Training Set | Total Correct | 485 | 488 |
| | Compounds in the set | 589 | 589 |
| | % correct total | 82.3 | 82.9 |
| | % Readily Degradable | 79.1 (201/254) | 80.3 (204/254) |
| | % Not Readily Degradable | 84.8 (284/335) | 84.8 (284/335) |
| Validation Set | Total Correct | 240 | 238 |
| | Compounds in the set | 295 | 295 |
| | % Correct Total | 81.3 | 80.7 |
| | % Readily Degradable | 80.2 (105/131) | 103/131 (78.6) |
| | % Not Readily Degradable | 82.3 (135/164) | 135/164 (82.3) |

*During the critical evaluation of the biodegradability, 'readily biodegradable' was assigned a numeric value of 1 and 'not readily biodegradable' was assigned a numeric value of 0. Thus, the probability of degradation ranges from 0 to 1.

The linear MITI biodegradation model is selected for predicting the probability of biodegradation. It belongs to the set of six officially approved protocols as ready biodegradability test guidelines of the *Organization for Economic Cooperation and Development* (OECD). The chemical group probability values that are applied in the MITI biodegradability method were developed from a dataset of 884 chemicals. This set comprised of 499 chemicals that were critically evaluated as 'not readily biodegradable' and 385 chemicals that were critically evaluated as 'readily biodegradable'.

For the derivation of the chemical group values, the dataset of 884 chemical compounds was divided into two separate sets: the training dataset (containing 589 compounds) and the validation dataset (containing 295 compounds). The derivation of the chemical group values was done by using the training dataset. The biodegradability estimates of the chemical compounds from this model might be less accurate if their molecular weight lies outside the molecular weight range of the training set compounds (U.S. Environmental Protection Agency 2009). The molecular weights of the training compounds in the original model ranged from 30.02g/mol to 959.2g/mol. For those compounds which have a chemical group(s) or some structural features which are not represented well in the training set, the predictions are solely based on their molecular weight. In the present study, the maximum and the minimum molecular weights of the compounds were found to be 320.73g/mol and 40.07g/mol, respectively. Hence, it is expected that reasonable estimations of the biodegradability would be obtained from BIOWIN models.

## 4.4. Compilation and preparation of data

In all, 295 compounds were selected in the data pool. The sections 4.4.1, 4.4.2, 4.4.3, 4.4.4 and 4.4.5 provide more detailed information on how their normal boiling temperature, vapor

pressure over a range of temperatures, enthalpy of vaporization, half-life in the troposphere and the biodegradability, respectively, was estimated/ compiled.

### 4.4.1. Compilation of normal boiling temperatures

Horvath (Horvath 2001) reported the normal boiling temperatures (°C) of a number of halogenated organic compounds. The reported normal boiling temperatures have been measured experimentally over a period of three decades. The listed organic compounds either belong to the general classes of halogenated aliphatic ethers, halogenated aliphatic hydrocarbons and halogenated cyclic hydrocarbons or they belong to other closely related compounds.

In this publication, the normal boiling temperatures of all the 295 compounds in the data pool could be found. However, some of the compounds were listed two or more times, with slightly different normal boiling temperatures. In these cases, due to the absence of any way to validate the listed data and because of the closeness of the listed normal boiling temperatures, a normal boiling temperature is randomly picked among the multiple published values and used further in the analysis.

### 4.4.2. Vapor pressures over a temperature range

The vapor pressures over a range of temperatures have been found out for the entire set of 295 compounds. These vapor pressure values were either experimentally measured values, as listed in the NIST web thermo tables (Kazakov, *et al.* 2009) or were estimated values from the vapor pressure module of the SPARC online calculator (Hilal, Karickhoff and Carreira 2003). In the total set of 295 compounds, experimentally measured values could only be found for 44 compounds. The number of different temperatures at which the experimentally measured vapor pressures were available ranged between 16 and 70. For the rest of the 251 compounds, the vapor pressures were estimated using SPARC online calculator. The estimation was performed over a temperature range, at 30 different equally spaced temperature values. The limits of this

temperature range were set at 225 $^0$C and -80 $^0$C, respectively. Figure 4.2 shows the process for using the information derived from vapor pressure values to formulate an objective for optimization.

Hilal *et. al.* (Hilal, Karickhoff and Carreira 2003) tested the performance of the vapor pressure model of the SPARC physical properties calculator on 447 non-polar organic compounds and 300 polar compounds. The *root mean squared* (RMS) deviation error between the observed values and the calculated values and correlation coefficient ($R^2$) values were found to be 0.096 and 0.999, respectively. The calculation of the process parameters is primarily based on equation 4.6.

$$\Delta G_{Process} = \Delta G_{Interaction} + \Delta G_{Other} \tag{4.6}$$

In the estimation of vapor pressure, $\Delta G_{interaction}$ describes the difference in the intermolecular interactions between the molecules in the gaseous and the liquid phase. Under the following assumptions, the intermolecular interactions in the liquid phase are expressed by equation 4.7.

a) The intermolecular forces are additive.

b) Interactions in the gaseous phase can be ignored.

$$\begin{aligned}\Delta G_{Interaction} = {} & \Delta G_{Dispersion\ interaction} + \Delta G_{Induction\ interaction} \\ & + \Delta G_{Dipole-dipole\ interaction} \\ & + \Delta G_{Hydrogen-bonding\ interaction}\end{aligned} \tag{4.7}$$

In equation 4.7, each of the four individual interaction energies is expressed in terms of a finite set of molecular level descriptors such as density based volume, molecular polarizability, molecular dipole and hydrogen bonding parameters which constitute the total interaction. All of these descriptors are estimated from the molecular structure of the compound. According to the model, the vapor pressure P of a pure solute can be expressed as a function of $\Delta G_{Interaction}$ using equation 4.8.

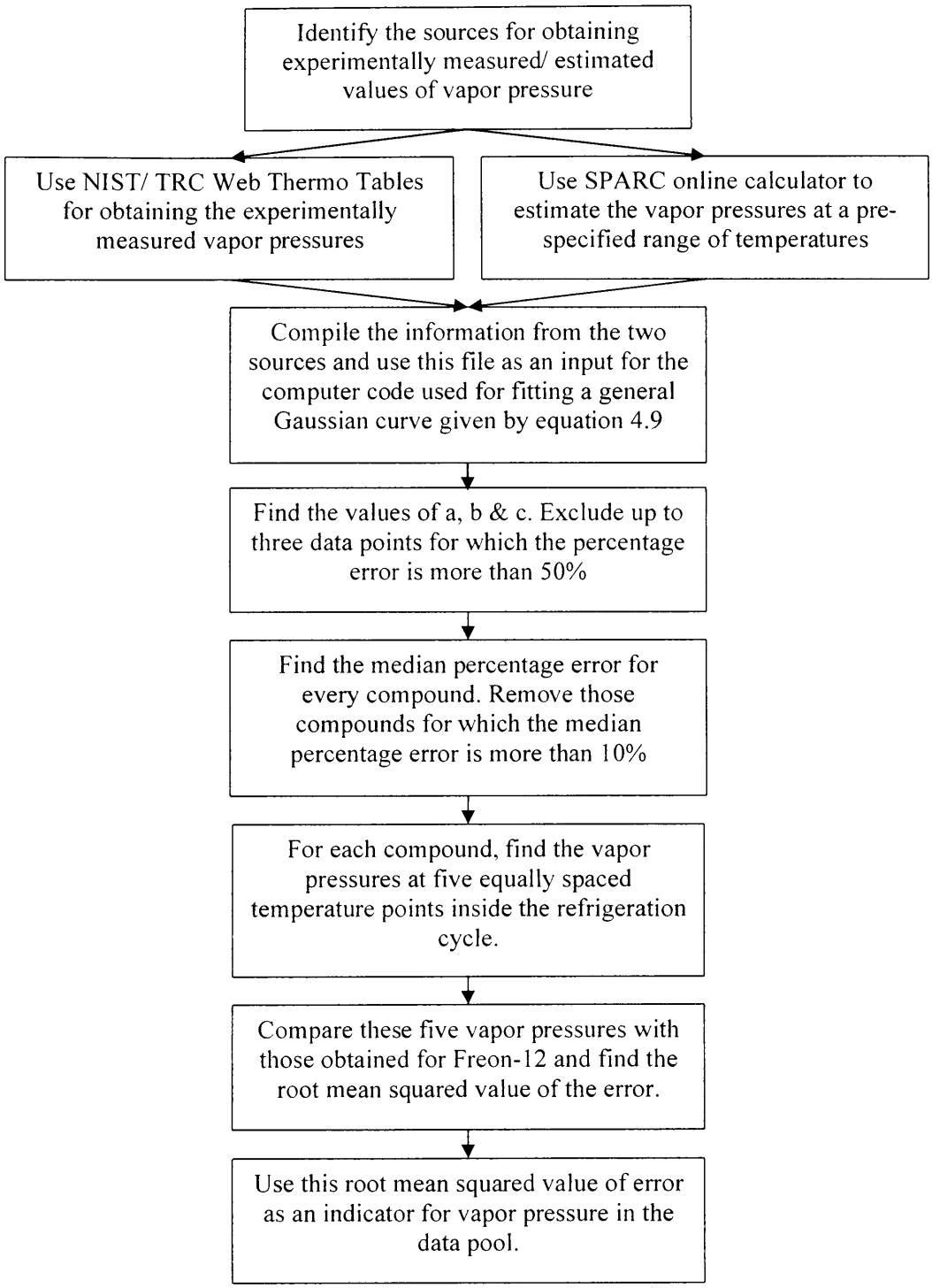



**Figure 4.2:** A schematic diagram showing the process flow for acquiring vapor pressure data and setting up an objective based on the vapor pressure values

$$logP = -\frac{\Delta G_{Interaction}}{2.303RT} + logT + C \tag{4.8}$$

Here, the term (logT + C) describes the change in the entropy contribution during the change in the volume, when the fluid goes from the liquid phase to the gaseous phase. Hilal *et al.* (Hilal, Karickhoff and Carreira 2003) provided further details on the calculations of the energy changes involved in the interaction mechanisms used in SPARC.

Using a computer code written in MATLAB, the experimentally measured and the estimated vapor pressures were then fitted to a general Gaussian curve given by equation 4.9.

$$Vapor\ Pressure\ (T_{sc}) = a \cdot e^{-\left(\frac{T_{sc}-b}{c}\right)^2} \tag{4.9}$$

Here, a, b and c are constants that are uniquely determined for each compound.

A scaled temperature $T_{sc}$ is used for fitting. In general, different scaled temperatures were used for different compounds. The scaled temperatures for each compound were calculated using the equation 4.10.

$$T_{sc} = \frac{T - T_{min}}{T_{max} - T_{min}} \tag{4.10}$$

Here, T is any temperature in the temperature range, $T_{min}$ is the minimum temperature at which an experimentally measured/ estimated value of vapor pressure is known and $T_{max}$ is the maximum temperature at which an experimentally measured/ estimated value of vapor pressure is known.

For each compound the percentage error between fitted vapor pressure and the known value of the vapor pressure was calculated at each data point. Those data points were identified for which this percentage error exceeded 50%. These data points were excluded and the data was refitted using a similar Gaussian curve. The new values of the coefficients a, b and c, were then used to replace the older values of the coefficients for the particular compound. The equation of the Gaussian curve was used to interpolate the vapor pressure values at all the temperatures at which the vapor pressure values were known. The median percentage difference was then

calculated between the known vapor pressure values and the interpolated vapor pressure values. This procedure was repeated for all the compounds in the data set.

Compounds for which the median percentage error was found to be greater or equal to 10% were removed from any further analysis. During this analysis, four compounds were found to have median percentage error greater than 10%. Further analysis was, thus, performed on the remaining 291 compounds. The refrigeration cycle used by Churi and Achenie (Churi and Achenie 1996) was used to set the limits on the temperature, expected in a refrigeration cycle. The following conditions were used for the refrigeration cycle.

a) Evaporating temperature, $T_L$ = -1.1 °C

b) Condensing temperature, $T_H$ = 43.3 °C

c) Mean temperature, $T_{mean}$ = 21.1 °C

d) Superheat temperature $T_{superheat} = T_L$ (under the assumption of saturated conditions)

Using the values of coefficients (a, b and c) and the equation 4.10, the vapor pressure values were calculated at five equally spaced temperatures between $T_L$ and $T_H$. The temperatures at which the vapor pressures were calculated were thus set to be equal to -1.1 °C, 10.0 °C, 21.1 °C, 32.2 °C and 43.3 °C. Freon-12 is one of the compounds selected in the pool of 295 compounds. The values of the vapor pressures were also calculated for Freon-12. Root mean squared values of the difference of the vapor pressures of the particular compound and Freon-12 were calculated. This value was tabulated and then used as an objective in the optimization process. This vapor pressure indicator (equation 4.11) needs to be minimized if the new refrigerant is expected to have similar vapor pressure-temperature curves as that of Freon-12.

$$Vapor\ pressure\ indicator = \sqrt{\frac{\sum_{i=1}^{n}(Estimated\ vapor\ pressure - Observed\ vapor\ pressure)_i}{n}} \quad (4.11)$$

### 4.4.3. Enthalpy of vaporization at 298.15 K and 1 atm pressure

Experimentally measured values of the enthalpies of vaporization (at 298.15 K and 1 atm) of all the 295 compounds used in this study, could not be found in the available literature. Therefore, the enthalpies of vaporization included in the pool of data comprise of 54 experimentally measured values and 241 estimated values. All the experimentally measured values used in the analysis have been reported by Chickos *et al.* (Chickos and Acree Jr. 2003) in a compendium of vaporization enthalpies published within the period 1910-2002.

In the absence of experimentally measured values of enthalpies of vaporization, estimations were performed. As discussed by Chickos *et al.*, Jr. (Chickos and Acree Jr. 2003), vaporization enthalpies can be calculated using the Antoine constants in two different ways. Any known vapor pressure-temperature data can be represented in the form of Antoine equation given by equation 4.12 (Poling, Prausnitz and O'Connell 2004).

$$log_{10}P = A - \frac{B}{(C+T)} \tag{4.12}$$

Here, P represents the vapor pressure, T represents the temperature in $^{o}$C and A, B and C are the Antoine constants. This equation can then be used to estimate the vapor pressures over a temperature range, the mean temperature of which is 298.15K. However, Antoine equation should not be used outside this specified temperature range as this kind of extrapolation may lead to inaccurate results.

This vapor pressure-temperature data can then be formulated in terms of the integrated Clausius-Clapeyron equation, given by equation 4.13.

$$lnP = -\frac{\Delta H_v(T_m)}{R}\left(\frac{1}{T}\right) + C_1 \tag{4.13}$$

Here, $\Delta H_v(T_m)$ represents the enthalpy of vaporization at the mean temperature and $C_1$ is a constant of integration. $\Delta H_v(T_m)$ may then be calculated from a linear regression analysis of lnP

and 1/T. As suggested by Chickos *et al.* (Chickos and Acree Jr. 2003), the $\Delta H_v(T_m)$ can also be calculated by using an alternative approach, utilizing the Antoine constants in equation 4.14.

$$\Delta H_v(T_m) = 2.303RB\left(\frac{T_m}{T_m + C}\right)^2 \tag{4.14}$$

However, none of the previously discussed methods were employed in the estimation of vaporization enthalpy. This was done primarily due to the following two reasons.

a) The variation of enthalpy of vaporization with temperature is not reflected very accurately by the use of Antoine equation.

b) Generally, for a given compound, the temperature values at which Antoine equation can be used satisfactorily are not large. Usually, these small temperature values correspond to vapor pressure values between 0.01 bar to 2 bar. Since, the anticipated vapor pressures inside the refrigeration cycle are expected to be larger; the use of Antoine equation would not have provided accurate estimation of the vapor pressures.

The enthalpy of vaporization values were instead estimated using the enthalpy of vaporization model (Hilal, Karickhoff and Carreira 2003) and implemented in SPARC online calculator V 4.5. In this model, temperature dependence of physical properties has been accounted for by the use of temperature dependant free energies, molecular orientation requirements for dipole-dipole coupling and hydrogen bonding interactions. Hilal *et al.* (Hilal, Karickhoff and Carreira 2003) tested the model's performance at 298.15 K and 1 atm on 422 polar organic compounds and 841 non-polar organic compounds. The RMS deviation error between the measured and the estimated values was found to be 0.41 kcal/ mol. A relatively high correlation coefficient ($R^2$) value of 0.990 suggested that SPARC predicts the enthalpy of vaporization with relatively good accuracy.

### 4.4.4. Half-life in troposphere

As discussed in section 4.3.4., hydroxyl radical concentration in the troposphere can be used to estimate the half-life of an organic compound in the troposphere under the following assumptions:

a) The chemical, with which OH radical is reacting, is present in the troposphere at a very low concentration.

b) A steady state concentration of OH radical is maintained by its generation in the presence of sunlight.

Under the above assumptions the reaction can be considered a pseudo first order reaction and the half-life in the troposphere can be calculated by using equation 4.5.

SMILES representations of all the 295 compounds were imported in the AOPWIN module of the EPI Suite program (U.S. Environmental Protection Agency 2008). Computations were performed for the estimation of the overall hydroxyl radical rate constant. Out of the 295 estimated rate constants, 269 rate constants were found to be non zero. For 53 out of those 269 compounds, AOPWIN also lists the experimentally measured hydroxyl radical rate constants. Wherever applicable, the estimated rate constants were replaced by the experimentally measured rate constants and were used in further calculation of the half-life of the compound in the troposphere, using equation 4.5.

Figure 4.3 reiterates the fact that along with the estimated (or experimentally measured) hydroxyl (OH) radical rate constant of a compound, a suitable OH radical concentration in an appropriate time frame is also needed to estimate the half-life of a given organic compound in the troposphere. Before any computations were made, it was assumed that in general the compounds react more slowly and they remain present in the troposphere for more than a few days. Hence, 24 hours averaged hydroxyl radical concentration at 298 K might be more appropriate. The global average OH radical concentration of $9.2 \times 10^5$ mol/ $cm^3$ was used for the estimation of the half

lives of the compounds in the troposphere. These half lives were tabulated and used as an additional objective during the optimization process.

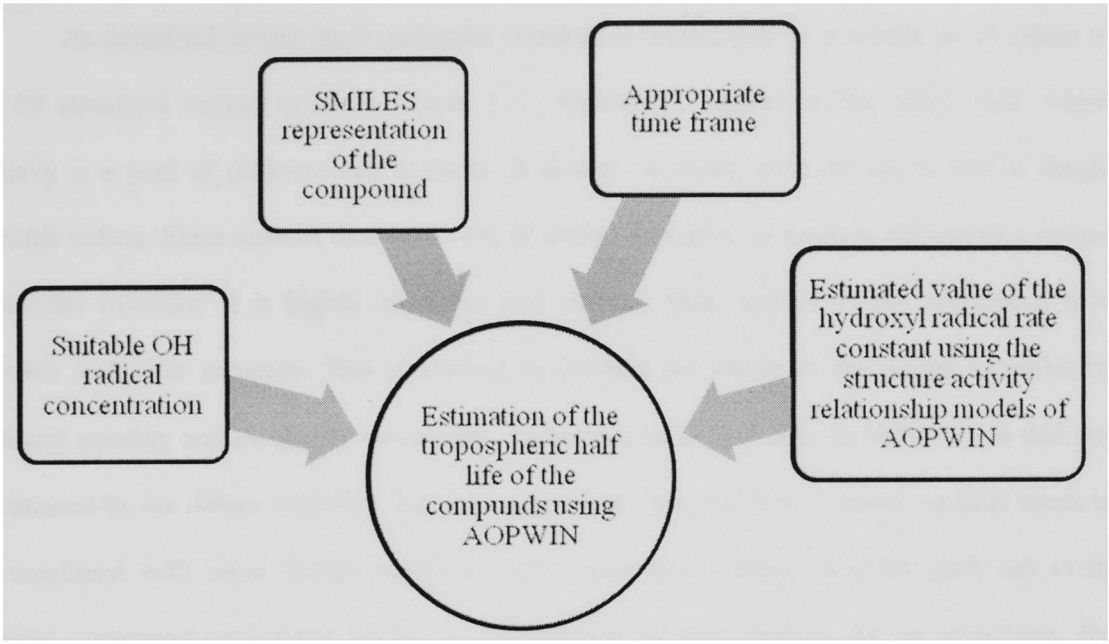



**Figure 4.3:** Process flow chart for the estimation of the half life of an organic compound in the troposphere using AOPWIN

### 4.4.5. Biodegradability

SMILES representations of all the 295 compounds were imported to the BIOWIN module of the EPI Suite program (U.S. Environmental Protection Agency 2009). The numerical rating and biodegradation probability was calculated for each of the 295 compounds using the ultimate biodegradation model (BIOWIN 3) and MITI linear model (BIOWIN 5), respectively. It has been previously reported (Pavan and Worth 2006) that most non-readily biodegradable compounds have their MITI non-linear model prediction value less than 0.5 and the ultimate biodegradation time frame prediction less than or equal to 2.2. For each compound, the results from both of these models were added and a single objective was formulated for the optimization.

# 5. SPECIFICATION OF CONSTRAINTS AND THEIR SIGNIFICANCE

## 5.1. Introduction

As described earlier, each molecular structure is represented by a unique set of values of the 28 structural design variables (Table 1.1). Algorithms applied on the initial data, which actually is a pool of different sets of these 28 design variables, generate newer sets of design variable values. Since each of these new sets of design variables, in essence, represents a unique molecular structure, it is highly important that each of these sets of values correspond to a feasible molecular structure. This is ensured by binding the values of the design variables by different equality and inequality constraints. These constraints not only limit the values that can be attained by the design variables; they also place restrictions on how a design variable needs to be combined with other design variables. These equations remain valid for each and every organic compound considered during the construction of meta-models. At the same time, the satisfaction of these equations or constraints remains necessary for any new molecule. The following sections list some of the different constraints which are important from the point of view of obtaining a feasible molecular structure. The sections also explain the necessity and importance of some of the many such constraints.

## 5.2. Limitations on the individual values of design variables

All the 28 design variables are strictly integers as they represent the number of different bonds. It should be noted at this point, that resonance effects in the molecular structure are not considered. The design variables $X_1$, $X_2$, $X_3$ and $X_4$ represent the number of single bonds used by the respective carbon atom to bond itself to the remaining three carbon atoms. In general, no carbon atom can have more than four carbon atoms attached to it *via* single bonds and no carbon atom can have less than zero number of carbon atoms attached to it *via* single bonds. However, since the total number of carbon atoms in the refrigerant molecule, considered in this study, was

restricted to four, a carbon atom may be attached to a maximum of three other carbon atoms *via* carbon-carbon single bonds. Mathematically, this fact can be represented by the following equation.

$$0 \leq X_i \leq 3 \; for \;\; i = 1,2,3,4 \tag{5.1}$$

The design variables $X_5$, $X_6$, $X_7$ and $X_8$ represent the number of carbon atoms attached to each of the four carbon atoms *via* carbon-carbon double bonds. Only a maximum of two carbon atoms may get attached to a given carbon atom and there cannot be fewer than zero carbon atoms that are attached to any carbon atom *via* carbon-carbon double bonds. Hence, the following relationship remains satisfied.

$$0 \leq X_j \leq 2 \; for \;\; j = 5,6,7,8 \tag{5.2}$$

Similarly, the design variables $X_9$, $X_{10}$, $X_{11}$ and $X_{12}$ can only take values between 0 and 4. The reason for this is the fact that in order for the octet rule to be valid, there cannot be more than four hydrogen atoms attached to a single carbon atom and there cannot be fewer than zero hydrogen atoms attached to a carbon atom. Mathematically, this can be represented by the following relationship.

$$0 \leq X_k \leq 4 \; for \;\; k = 9,10,11,12 \tag{5.3}$$

Using a similar argument, it can be concluded that there cannot be more than four chlorine/ fluorine/ bromine atoms attached to a single carbon atom *via* single bonds. Hence, the following relation holds true.

$$0 \leq X_l \leq 4 \; for \;\; 13 \leq l \leq 24 \tag{5.4}$$

The same argument also remains true in the case of oxygen atoms bonded to a carbon atom *via* single bonds. Thus, the number of oxygen atoms attached to a carbon atom *via* a single bond may vary between zero and four. However, the initial pool of 295 data had molecules which either lacked an oxygen atom or contained just a single oxygen atom. Because of this reason, no

carbon atom was found, in the original dataset, to be associated with more than one carbon-oxygen single bond. Hence, equation 5.5 needs to be satisfied.

$$0 \leq X_m \leq 1 \; for \;\; m = 25, 26, 27, 28 \tag{5.5}$$

In cases where the vapor pressure is also represented in one of the objectives, three additional design variables ($X_{29}$, $X_{30}$ and $X_{31}$) are needed. In general, these three design variables are continuous real variables, unlike the other 28 design variables, which strictly attain non-zero values. $X_{29}$, $X_{30}$ and $X_{31}$ represent the coefficients a, b and c, respectively, of the Gaussian curve, represented by equation 4.9.

The least value of coefficient 'a' among all the 295 initial compounds was found to be 0.6726. Since this value was found to be close to zero, the lower bound of $X_{29}$ was set to 0. The upper limit of the design variable $X_{29}$ was set according to the following relation.

$$X_{29(higher)} = a_{max} + 0.1(a_{max} - a_{min}) \tag{5.6}$$

Here, $a_{max}$ represents the maximum value of 'a' among all the 295 initial compounds. This value was found to be 22554.06. The $X_{29(higher)}$ is found to be 24809.399. Hence, $X_{29}$ was limited by the following equation.

$$0 \leq X_{29} \leq 24810 \tag{5.7}$$

Similarly, the higher bound of $X_{30}$ was found out by the following relationship.

$$X_{30(higher)} = b_{max} + 0.10(b_{max} - b_{min}) \tag{5.8}$$

and the lower bound of $X_{30}$ was found out from the following relationship.

$$X_{30(lower)} = b_{min} - 0.10(b_{max} - b_{min}) \tag{5.9}$$

Here, $b_{max}$ and $b_{min}$ are the maximum value of 'b' and the minimum value of 'b', respectively, among all the 295 compounds. $b_{max}$ and $b_{min}$ were found to be 1.2513 and 1.9780, respectively. Thus, $X_{30}$ is limited by the following equation.

$$1.20 \leq X_{30} \leq 2.05 \tag{5.10}$$

Similarly, the higher bound of $X_{31}$ was found out from the following relationship.

$$X_{31(higher)} = c_{max} + 0.10(c_{max} - c_{min}) \quad (5.11)$$

and the lower bound of $X_{31}$ was found out from the following relationship.

$$X_{31(lower)} = c_{min} - 0.10(c_{max} - c_{min}) \quad (5.12)$$

Here, $c_{max}$ and $c_{min}$ were the maximum value of 'c' and the minimum value of 'c', respectively, among all the 295 initial compounds. $c_{max}$ and $c_{min}$ were found to be 0.2852 and 1.7997, respectively. Thus, $X_{31}$ was limited by the following equation.

$$0.17 \leq X_{31} \leq 1.95 \quad (5.13)$$

## 5.3. Equality and inequality constraints involving combinations of multiple design variables

Multiple equality and inequality constraints were formulated to regulate the combinations of the structural design variables. These constraints ensured structural feasibility of the generated sets of the 28 design variables after optimization.

## 5.4. Physical property and other constraints

As discussed earlier, the primary objective of the study is to develop refrigerants which are better than the currently used refrigerants (specifically Freon-12) in as many aspects as possible. Hence, the property constraints depend on the property values of Freon-12.

For limiting the volumetric flow of the refrigerant inside a refrigeration system, the enthalpy of vaporization of the refrigerant should be high. In other words, a larger enthalpy of vaporization reduces the amount of refrigerant required to produce the desired amount of cooling. The enthalpy of vaporization of Freon-12 at 272 K is 18.4 kJ/ mol. Thus, by setting the lower limit of enthalpy of vaporization as 15.0 kJ/ mol (or 3.585 kcal/ mol), it can be ensured that the generated refrigerant will be at least comparable to Freon 12, in terms of the enthalpy of vaporization. A slightly lower value of enthalpy of vaporization was chosen as the lower bound

on the possible values of the enthalpy of vaporization during the design optimization process, because of the following reasons.

a) It ensures that potentially good molecules whose enthalpy of vaporization is slightly lower than that of Freon-12 are not rejected during the enforcement of constraints.

b) The surrogate model (multi-dimensional response surface fit of the initial data pool) which was used to predict the values of the enthalpy of vaporization, based on a set of discrete variables, was expected to have a complex topology. In order to account for errors in the method, the lower value of enthalpy of vaporization was slightly relaxed.

Hence, the following property constraint was enforced.

$$h_v \geq 3.585 \text{ kcal/ mol} \tag{5.14}$$

It has been reported (Sekiya and Misaki 2000), (Calm 2008), that refrigerants which have very short lifetime (atmospheric or tropospheric) tend to decompose near Earth's surface, generally very close to the location of their release into Earth's atmosphere. This results in problems similar to the production of smog or other harmful chemicals. On the other hand, refrigerants with very long lifetimes contribute to the global warming phenomena. Therefore, the desired lifespan of the refrigerants has been a topic of much debate. The desired lifetimes have been reported to be from a few days to a few years depending on the organic compounds used as refrigerants. Under these considerations, this study restricted tropospheric half-life time of the generated molecule to be between 30 days and 365 days. Thus,

$$30\ days \leq troposheric\ half\ life \leq 365\ days \tag{5.15}$$

An organic compound is considered to be readily biodegradable if both of the following conditions are satisfied (Pavan and Worth 2006).

a) The result of the BIOWIN 3 model (Ultimate biodegradation model) is greater than or equal to 2.75.

b) The result of the BIOWIN 5 model (MITI linear model) is greater than or equal to 0.5.

However, slightly lower limits were used in the constraints in this design optimization methodology. It was assumed that the use of slightly lower limits would not generate a molecule which is highly non-biodegradable. Also, the use of slightly lower values reduces the chances of a potentially good molecule been eliminated during the constraint satisfaction process. Hence, the following limits were used.

$$Result_{BIOWIN\,3} \geq 2.25 \tag{5.16}$$

$$Result_{BIOWIN\,5} \geq 0.45 \tag{5.17}$$

# 6. CREATION OF RESPONSE SURFACES AND THEIR ANALYSIS

## 6.1. Introduction

Response surfaces are multi-dimensional equivalents of spline fitting a function given at a number of isolated data points. Since response surface topology is expressed analytically, this means that for any interpolated data point it is extremely easy and fast to read the corresponding interpolated value of the function, off the response surface. For example, enthalpy of vaporization is one such possible function of 28 design variables. Since the initial data set has heats of vaporization for 295 compounds this means that we will have 295 isolated points in a 28-dimensional space. Building a response surface, in this case means fitting a 28-dimensional surface as closely as possible to these original 295 points, while assuring that this hyper-surface has minimal oscillations between these 295 points. For, any set of 28 values of design variables it is very easy to find the corresponding interpolated function value (enthalpy of vaporization) on the fitted response hyper-surface.

Response surfaces generating algorithms (Sigma-Technology Company 2009) were used for fitting the 28-dimensional response surfaces for each of the following design objectives.

a) P1 = Normal boiling temperatures
b) P2 = Enthalpy of vaporization
c) P3 = Tropospheric half-life time
d) P4 = Root mean squared value of the difference in the vapor pressure of the particular organic compound and that of Freon-12
e) P5a = Biodegradability value obtained from BIOWIN 3
f) P5b = Biodegradability obtained from BIOWIN 5

IOSO software (Sigma-Technology Company 2009) offers several methods for automatically generating multi-dimensional response surfaces. In this research, method of

weighted approximations and the method of radial basis functions were used. There are other methods for generating response surfaces available in the open literature. One such method is the wavelet based neural network methodology (Sahoo and Dulikravich 2006). However, this method was not implemented because it has been reported hat radial basis function based polynomial methodology performs better for higher dimensional spaces in terms of accuracy and the required computing time (Colaco, Dulikravich and Sahoo 2007). The weighted approximation method is described in section 6.2. The radial basis functions based methodology for creating response surfaces is briefly explained in section 6.3. In section 6.4, the response surface created by the two methodologies from IOSO are compared and analyzed.

## 6.2. Weighted approximation method used in IOSO software

The value of the function $f(\vec{x})$ whose value is known for the set of $t$ data points (or different designs), is given by the following equation:

$$f_{rs}(\vec{x}) = \frac{\sum_{j=1}^{t} W^j(\vec{x}) \cdot f(\vec{x}^j)}{\sum_{j=1}^{t} W^j(\vec{x})} \tag{6.1}$$

Here, $\vec{x}^j$ represents the j$^{th}$ point (or design) of the plan of the experiment, $f(\vec{x}^j)$ represents the value of the function $f$ at the j$^{th}$ point and $W^j(\vec{x})$ represents the weighting function for the j$^{th}$ point.

$W^j(\vec{x})$ represents a function which takes a value of 1 at the point $\vec{x} = \vec{x}^j$. Its value decreases as the distance between the points $\vec{x}$ and $\vec{x}^j$ increases. Weighting function is given by the following equation:

$$W^j(\vec{x}) = C_1^{\frac{R^j(\vec{x})}{C_2}} \tag{6.2}$$

Here, $C_1$ represents a constant, $C_2$ represents the flexibility parameter. In general, a lower value of the flexibility factor makes the approximation function more flexible. $R^j(\vec{x})$ represents the normalized distance and is given by equation 6.3.

$$R^j(\vec{x}) = \frac{\sum_{i=1}^{n}\left(\overrightarrow{x_i} - \overrightarrow{x_i^j}\right)^2}{max\left(\sum_{i=1}^{n}\left(\overrightarrow{x_i} - \overrightarrow{x_i^j}\right)^2\right)} \tag{6.3}$$

## 6.3. Methods based on radial basis functions used in IOSO software

The radial basis function (RBF) methodology in IOSO software is strongly coupled with the general concept of artificial neural networks. An artificial neural network of RBF type has three layers: the input layer, the hidden layer (or the intermediate layer) and the output layer.

In the algorithm used to build the response surface, the number of neurons in the input layer is equal to the number of the independent variables which affect the output function. The number of neurons in the output layer is determined by the number of parameters which should be approximated simultaneously. Every neuron in the hidden layer (or the output layer) is connected to all the neurons in the input layer (or the hidden layer). The number of neurons in the hidden layer can be varied to suit the particular problem. A density parameter governs the number of neurons in the hidden layer. A value of 0 corresponds to the minimum and the value 1 corresponds to the maximum number of neurons in the hidden layer. A Gauss function (uniquely specified by its center and the radius) is used as the transmission function for neurons in the intermediate layer.

## 6.4. Comparison and analysis of the response surfaces based on the two methods

Table 6.1 shows the results of the analysis of the response surface built using the weighted approximation methods in IOSO software. The mean and the maximum values of the relative errors, for all the six objectives, are found when the flexibility parameter is changed from

0.01 to 0.50. As expected, the errors are the smallest when a flexibility factor of 0.01 is used in the construction of the response surface.

**Table 6.1:** Analysis of the response surfaces built using the weighted approximation method

| Objective | Flexibility factor = 0.01 | | Flexibility factor = 0.02 | | Flexibility factor = 0.50 | |
|---|---|---|---|---|---|---|
| | Mean relative error | Maximum relative error | Mean relative error | Maximum relative error | Mean relative error | Maximum relative error |
| Normal boiling temperature | 0.00067 | 0.00265 | 0.00961 | 0.03484 | 0.16230 | 0.49315 |
| Enthalpy of vaporization | 0.00074 | 0.00317 | 0.00972 | 0.02973 | 0.15430 | 0.43618 |
| Tropospheric half-life time | 0.00097 | 0.00827 | 0.01196 | 0.12953 | 0.10311 | 0.92814 |
| Indicator for vapor pressure | 0.00038 | 0.00500 | 0.00790 | 0.10946 | 0.06109 | 0.83981 |
| BIOWIN 3 result | 0.00086 | 0.00358 | 0.01057 | 0.03812 | 0.15972 | 0.45921 |
| BIOWIN 5 result | 0.00092 | 0.00536 | 0.01130 | 0.04491 | 0.17591 | 0.48671 |

Table 6.2 shows the results of the analysis of the response surfaces built using the radial basis function methodology combined with the artificial neural networks. The response surfaces were built at two different levels of the number of neurons in the hidden layer. The density of

neurons was varied from an almost minimum value of 0.01 to the highest possible value of 1.00. Again the mean and the maximum relative errors were tabulated and it can be seen from Table 6.2 that as the density of neurons in the intermediate layer is increased the relative errors decrease.

**Table 6.2:** Analysis of the response surface built using the method based on radial basis functions

| Objective | Density parameter = 0.01 | | Density parameter = 1.00 | |
|---|---|---|---|---|
| | Mean relative error | Maximum relative error | Mean relative error | Maximum relative error |
| Normal boiling temperature | 0.18429 | 0.58644 | 0.08700 | 0.33795 |
| Enthalpy of vaporization | 0.17801 | 0.53726 | 0.06710 | 0.27462 |
| Tropospheric half-life time | 0.10786 | 0.95752 | 0.07069 | 0.41602 |
| Indicator for vapor pressure | 0.06224 | 0.85123 | 0.03631 | 0.31702 |
| BIOWIN 3 result | 0.18407 | 0.52428 | 0.12739 | 0.40180 |
| BIOWIN 5 result | 0.20350 | 0.55018 | 0.27891 | 2.75827 |

A comparison of the response surfaces generated by IOSO software using the two methodologies suggests that the response surface generated using the weighted approximation is more accurate as the maximum and the mean relative errors are lower than that of the response surface generated using the radial basis function methodology. In fact, the response surface built using a flexibility factor of 0.01 have been found better than any other response surface built and analyzed in this study. Figures 6.1-6.6 show the approximate relative errors in estimating all the

six objectives for all the 265 designs since, only 265 of the original 295 compounds had all six objectives available.

It is interesting to note that the mean relative error in approximating the function for the indicator of the vapor pressure is lower than mean relative errors in approximating other functions. This observation may be explained using the fact that experimentally measured/ estimated vapor pressure values were fitted to a Gaussian curve before any further analysis was performed. During this fitting of the data, few outliers were removed and were not included in the fit. This resulted in an overall smoothing of the data and because of this reason the response surface could conform itself better.

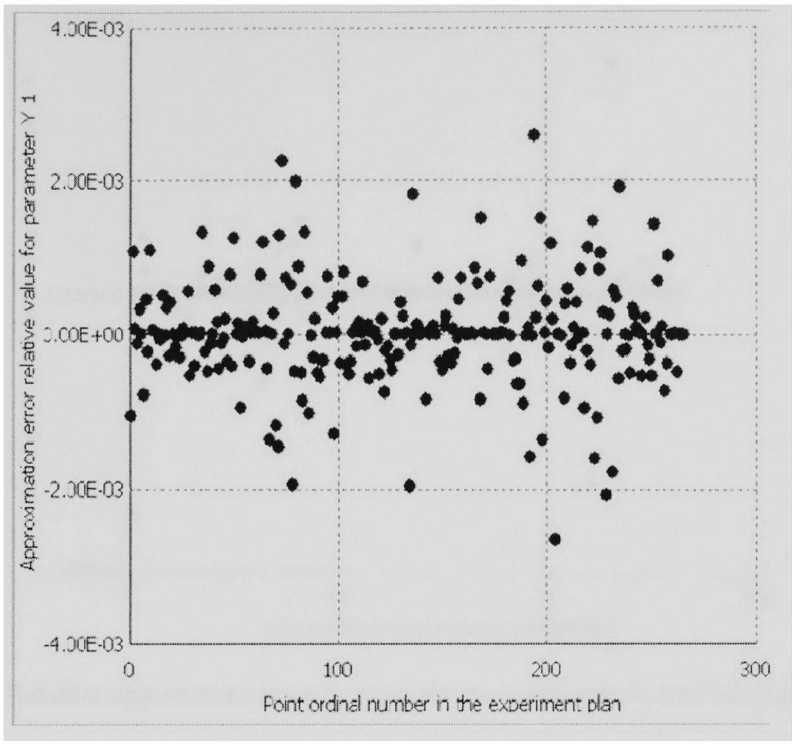



**Figure 6.1:** Relative approximation error value for the normal boiling temperature using the weighted approximation method with a flexibility parameter of 0.01

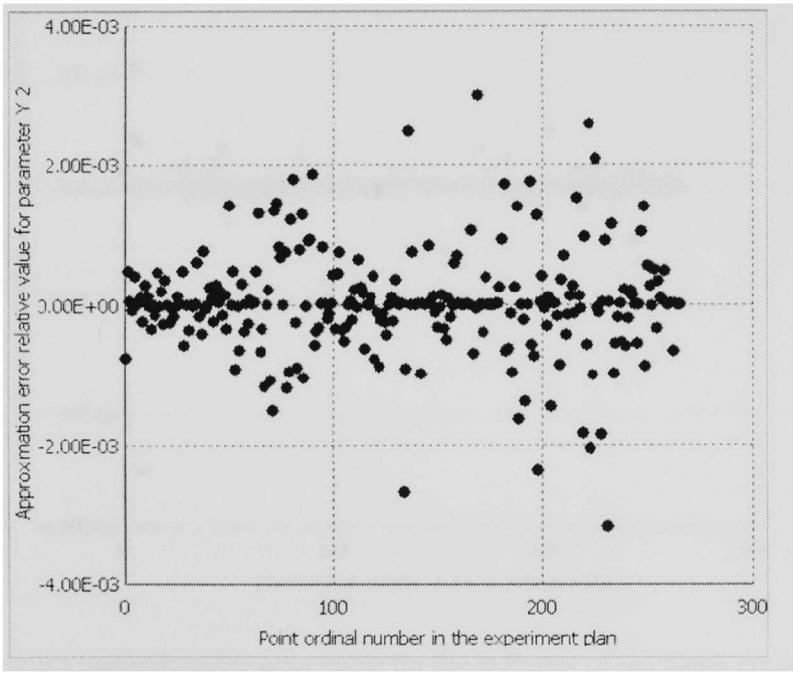



**Figure 6.2:** Relative approximation error value for the enthalpy of vaporization using the weighted approximation method with a flexibility parameter of 0.01

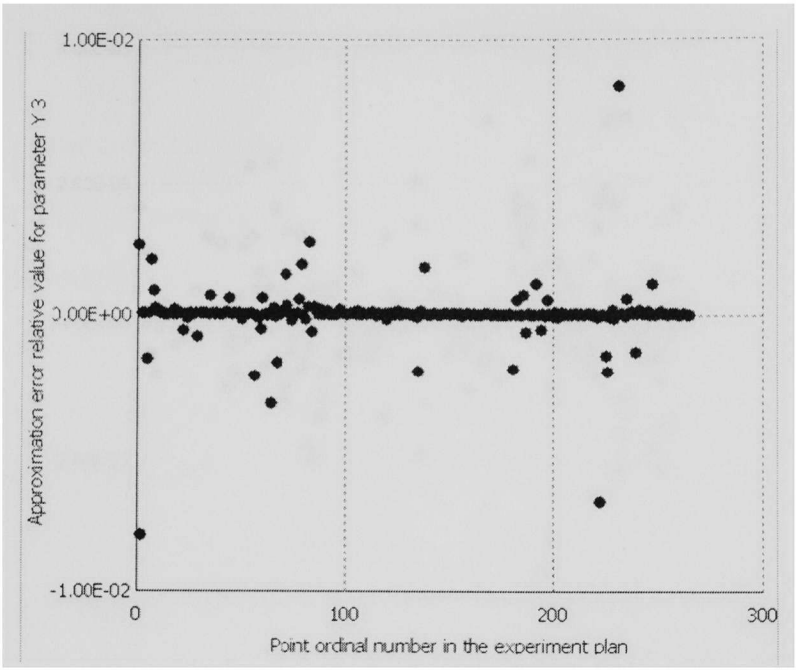



**Figure 6.3:** Relative approximation error value for the tropospheric half life time using the weighted approximation method with a flexibility parameter of 0.01

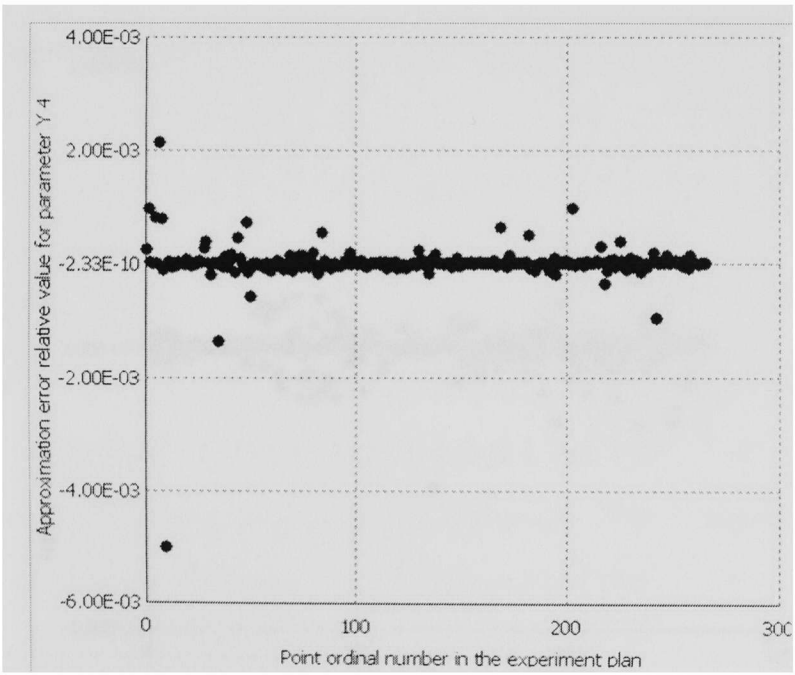



**Figure 6.4:** Relative approximation error value for the indicator of the vapor pressure using the weighted approximation method with a flexibility parameter of 0.01

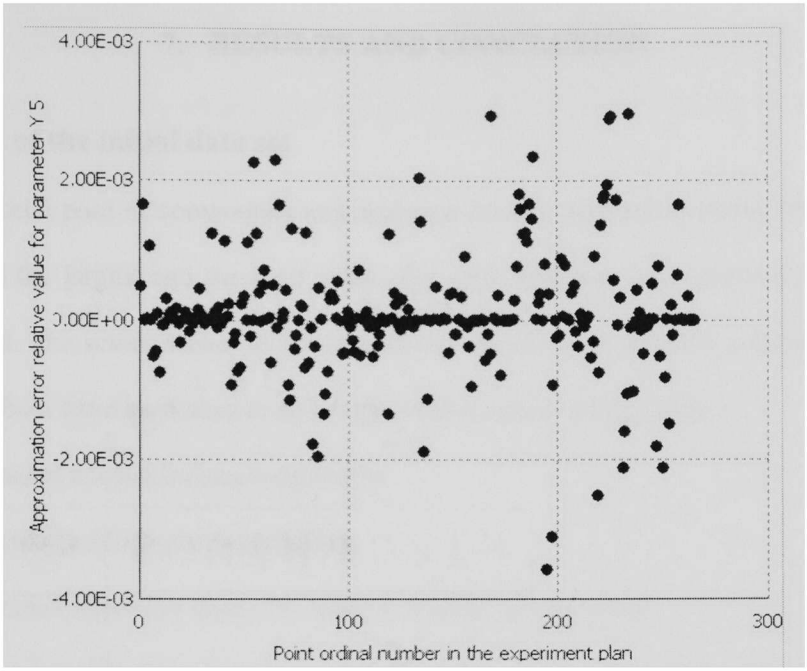



**Figure 6.5**: Relative approximation error value for the results of BIOWIN 3 model, using the weighted approximation method with a flexibility parameter of 0.01

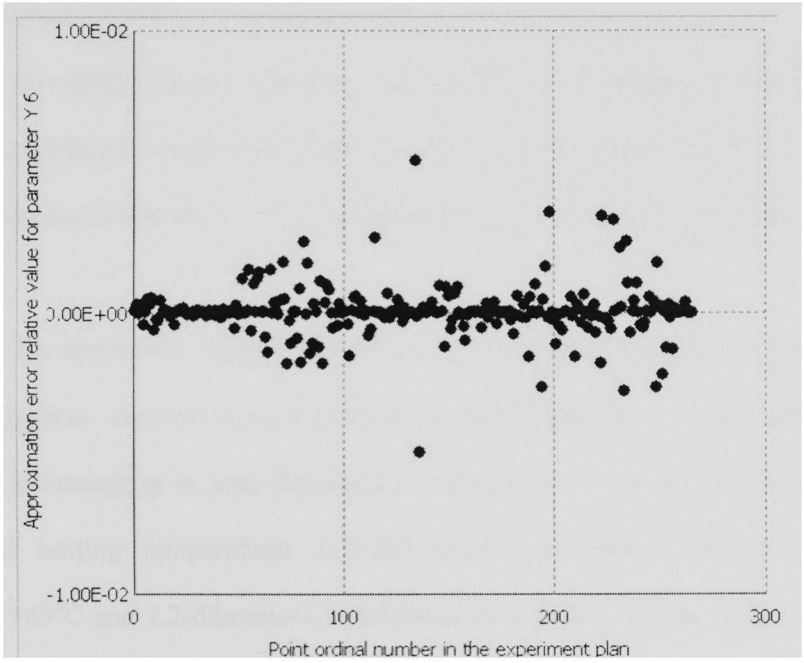



**Figure 6.6:** Relative approximation error value for the results of BIOWIN 5 model, using the weighted approximation method with a flexibility parameter of 0.01

# 7. RESULTS AND CONCLUSION

## 7.1. Analysis of the initial data set

The initial pool of compounds was analyzed for any noticeable trends. Two compounds which showed the largest and the least values for each objective were selected and radar plots were generated. The scaled values of the objective functions were plotted on the radar plots. The following symbols have been used to denote the scaled values of objectives.

a) P1 – scaled normal boiling temperature

b) P2 – scaled enthalpy of vaporization

c) P3 – scaled logarithm (base 10) of tropospheric half-life time

d) P4 – scaled indicator for vapor pressure difference with respect to Freon-12 (equation 4.11)

e) P5a – scaled BIOWIN 3 estimation result

f) P5b – scaled BIOWIN 5 estimation result

The set of compounds (for which the values of all the six objectives are known) numbers 265 out of the original 295 compounds. Table 7.1 lists the maximum and the minimum values of all the six objectives, for the set of 265 compounds, for which the values of all the objectives are known.

Figure 7.1 shows the objective function values of the four organic compounds which have the two highest and two lowest normal boiling temperatures in the entire set of 265 compounds. It is interesting to note that the two compounds which are found to have the two highest normal boiling temperatures (1,2-dibromo-1-chloroethane with a normal boiling temperature of 163°C and 1,2-dibromo-1,1-dichloroethane with a normal boiling temperature of 161.5°C) also have relatively high enthalpies of vaporization. The compounds which are found to have two of the lowest normal boiling temperatures (trifluoromethane with a normal boiling

temperature of -82°C and bromotrifluoromethane with a normal boiling temperature of -59°C) also have relatively small enthalpies of vaporization. These compounds also show relatively higher tropospheric half-life times.

**Table 7.1:** Maximum and minimum values of the six objective functions for the set of 265 compounds (for which the values of all the six objective functions were known)

| Objective | Maximum value | Minimum value |
|---|---|---|
| Normal boiling temperature(°C) | 163 | -82 |
| Enthalpy of vaporization (kcal/mol) | 13.07 | 2.94 |
| Tropospheric half-life time (days) | 89120.37 | 0.16943 |
| Vapor pressure indicator (kPa) | 8859.464 | 0 |
| BIOWIN 3 estimation result | 3.3691 | 1.1031 |
| BIOWIN 5 estimation result | 0.6415 | -0.0569 |

Figure 7.2 shows the objective function values of the four organic compounds which have the two highest and the two lowest enthalpies of vaporization in the entire set of 265 compounds. Figure 7.2 also suggests a possible relationship between the normal boiling temperature of a compound and its enthalpy of vaporization. The compounds which are found to have two of the highest enthalpies of vaporization (1,1,1-trichloro-2-(dichloromethoxy)-2,2-difluoroethane with an enthalpy of vaporization of 13.07 kcal/mol and 2,2-dichloro-1-(dichloromethoxy)-1,1-difluoroethane with an enthalpy of vaporization of 12.54 kcal/mol) also have relatively large boiling temperatures. The compounds which were found to have two of the lowest enthalpies of vaporization (1,1,2-trifluoroethene with an enthalpy of vaporization of 2.94 kcal/mol and 1,1,1-trifluoroethane with an enthalpy of vaporization of 3.297 kcal/mol) also have relatively small normal boiling temperatures. It is interesting to observe from Figure 7.1 and

Figure 7.2 that the compounds which have larger normal boiling temperatures and enthalpies of vaporization, not only have moderately large tropospheric lifetimes, they are also found to be not very easily biodegradable and their vapor pressures *vs.* temperature curve is similar to the vapor pressure *vs.* temperature curve of Freon-12.

Figure 7.3 shows the objective function values of the four organic compounds which have the two highest and the two lowest tropospheric half-life times in the entire set of 265 compounds. Figure 7.3(a) shows that the compounds which were found to have two of the largest tropospheric half-lives (bromotrifluoroethane with an estimated tropospheric half-life of 89120.37 days and 1,2-dibromo-1,1,2,2-tetrafluoroethane with an estimated tropospheric half-life of 68554.13 days) have similar vapor pressure *vs.* temperature curves to that of Freon-12. These compounds also do not biodegrade very easily.

This observation supports the fact that organic compounds containing higher number of halogen atoms or perhalogenated organic compounds generally have higher atmospheric life-times. On the other hand compounds represented in Figure 7.3 (b) are not only found to have two of the lowest tropospheric half-life times ((ethenyloxy) ethane with a tropospheric half-life of 0.16943 days and ethoxyethene with a tropospheric half-life of 0.2208 days), they are also found to have relatively higher probability to undergo biodegradation. As expected compounds with unsaturations are usually unstable and are found to be the ones which have the least tropospheric half- lives.

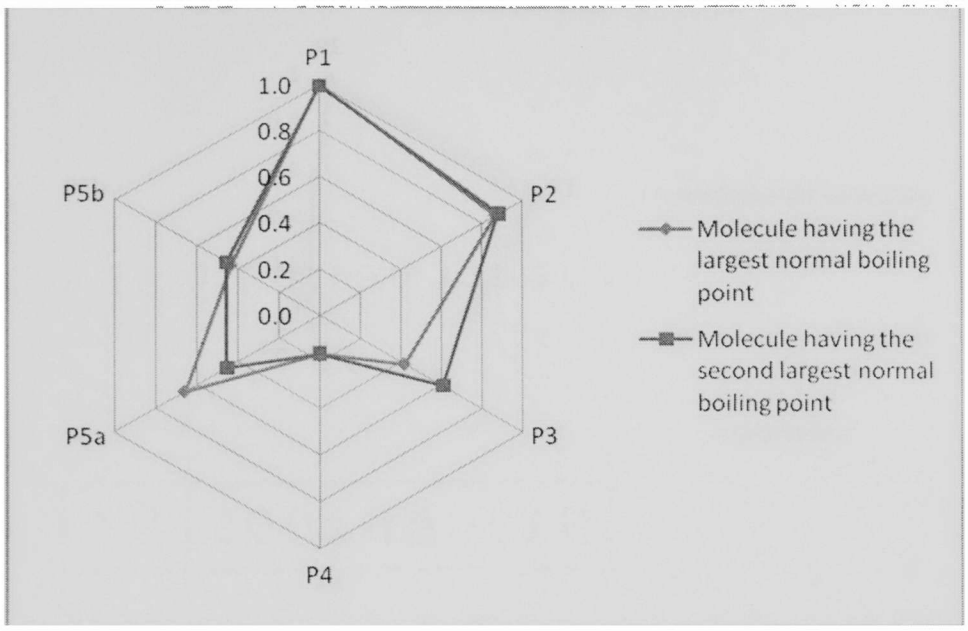
Figure 7.4 (b) shows that among all the considered 265 compounds, 1,1-difluoroethane is the closest to Freon-12 in terms of the vapor pressure variation with temperature. Moreover, both 1,1-difluoroethane and Freon-12 have higher biodegradation probability.

Figure 7.5 (a) shows that butane (BIOWIN 3 biodegradation probability prediction is 3.3691) and propa-1,2-diene (BIOWIN 3 biodegradation probability prediction is 3.1106) have two of the highest BIOWIN 3 prediction values. It should be noted that none of the two

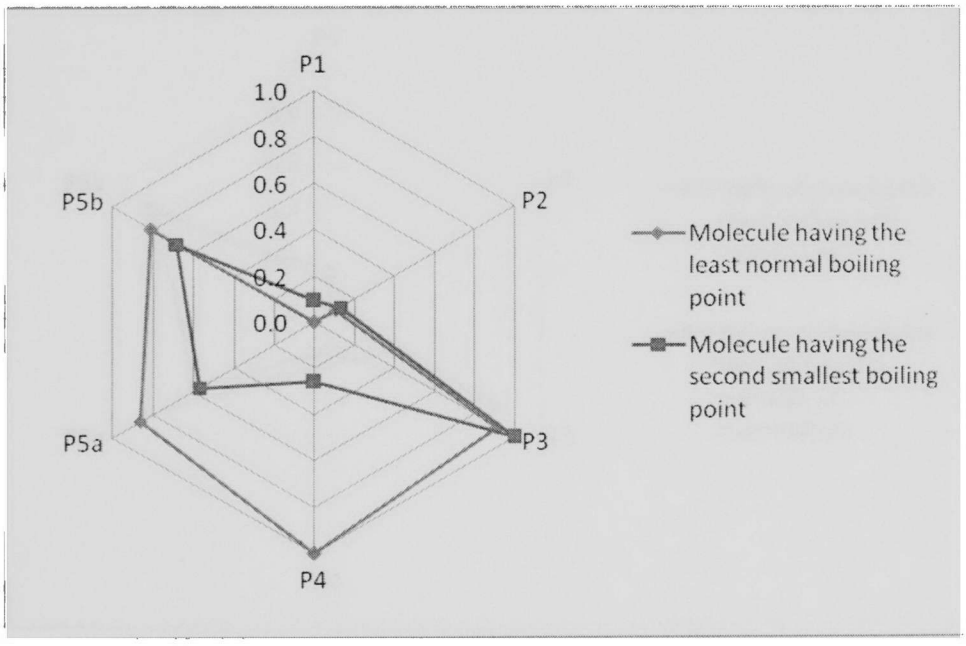
compounds have any halogen atoms in their molecular structures. Both these compounds have relatively smaller normal boiling temperatures, smaller enthalpies of vaporization, lower tropospheric half-live times and variation of vapor pressures with temperature similar to that of Freon-12. The compounds having two of the lowest BIOWIN 3 estimation results have a large number of halogen atoms in their molecular structures (1,1-dichloro-2-(1-chloro-2,2,2-trifluoroethoxy)-1,2,2-trifluoroethane has a BIOWIN 3 prediction value of 1.1031 and 1,1,1-trichloro-2-(dichloromethoxy)-2,2-difluoroethane has a BIOWIN 3 prediction value of 0.0402). Figure 7.5 (b) shows that these compounds also have relatively large tropospheric half-lives. It should be noted that they also have a large number of halogen atoms in their molecular structures.

Figure 7.6 (a) shows the objective values of the two compounds that are found to have largest BIOWIN 5 estimation results (difluoromethane has the BIOWIN 5 prediction result of 0.6415 and butane has the BIOWIN 5 prediction result of 0.6389). Both these compounds also have relatively larger values of the BIOWIN 3 estimation results. Figure 7.6 (b) shows the objectives values of the compounds which have two of the smallest values of the BIOWIN 5 estimation result (1-chloro-1-(1-chloro-2,2-difluoroethoxy)-2,2-difluoroethane has the BIOWIN 5 prediction value of -0.0569 and 1-chloro-1-(dichloromethoxy)-2,2,3,3-tetrafluoropropane has the BIOWIN 5 prediction value of -0.04). As seen from their IUPAC nomenclature, both of these compounds have multiple halogen atoms in their molecular structures.

Such an analysis of the pool of 265 compounds reiterates the fact that different families of organic compounds are suitable for optimizing different objectives. A single compound may have optimal values for a few of the objectives, but for some other objectives it may have undesired values. Therefore, multi-objective optimization needs to be used as a tool in order to find the molecular structures that are considered suitable with respect to a number of simultaneous, often conflicting objectives subject to a number of constraints discussed in Chapter 5.

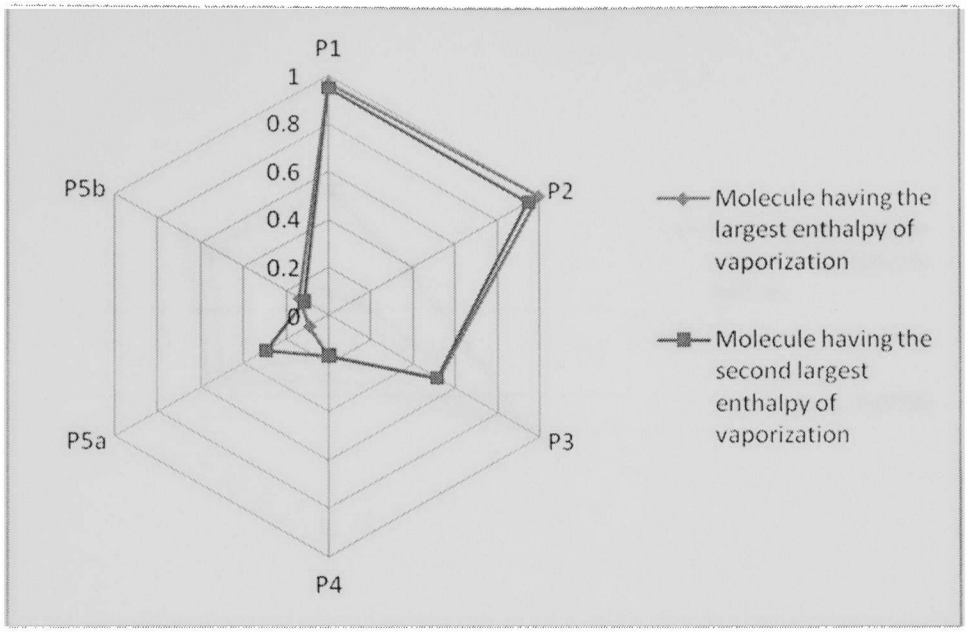



(a)

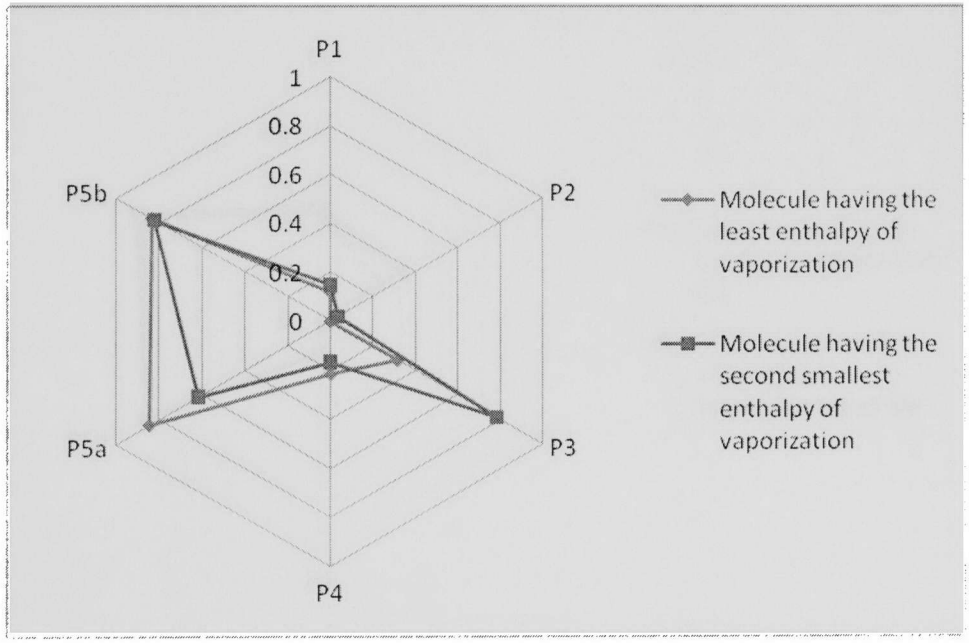



(b)

**Figure 7.1:** Representation of the performance of compounds which have (a) two of the largest normal boiling temperatures, and (b) two of the lowest normal boiling temperatures

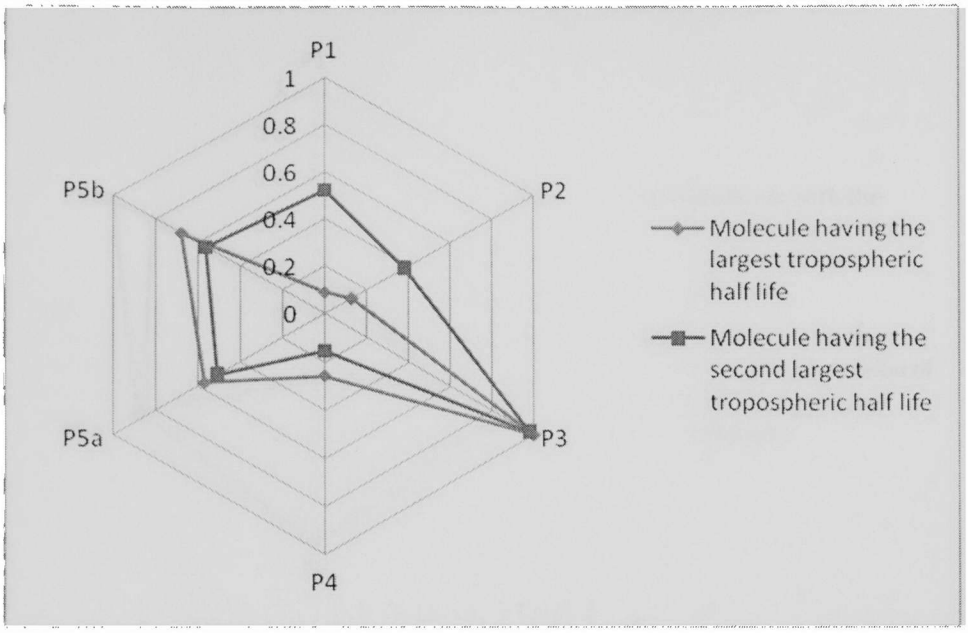



(a)

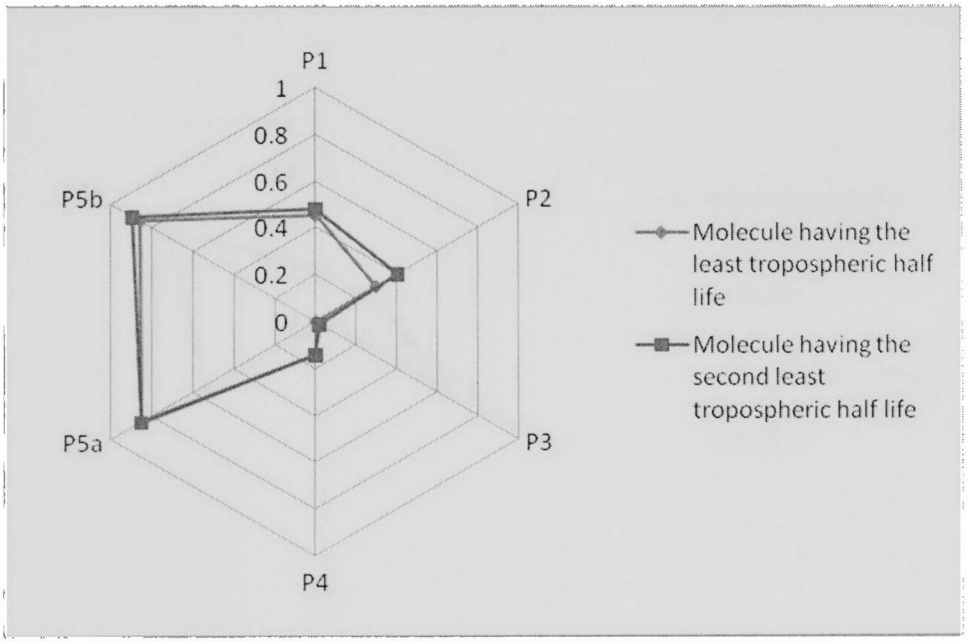



(b)

**Figure 7.2:** Representation of the performance of compounds which have (a) two of the largest enthalpies of vaporization, and (b) two of the lowest enthalpies of vaporization

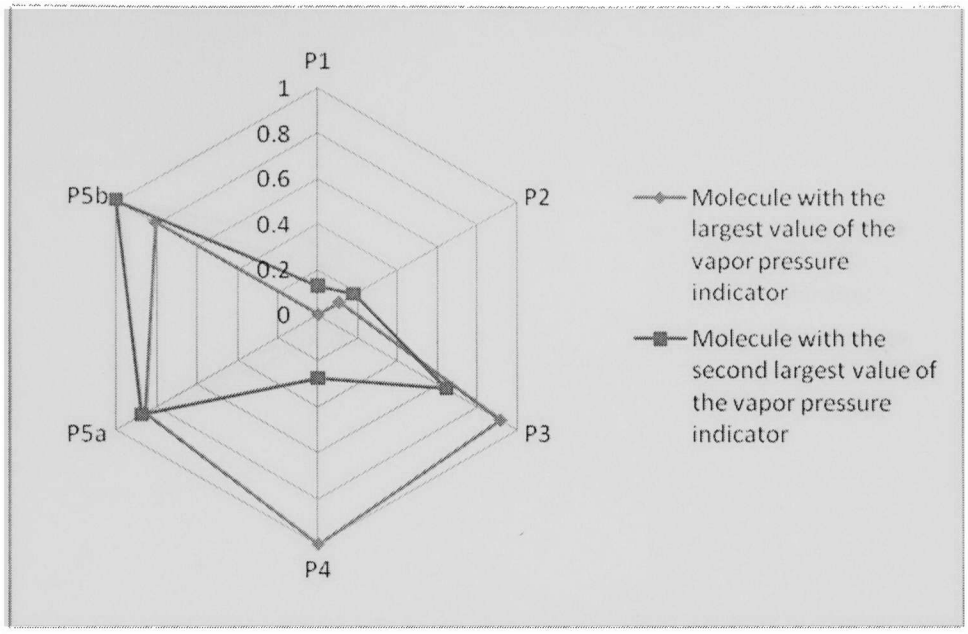



(a)

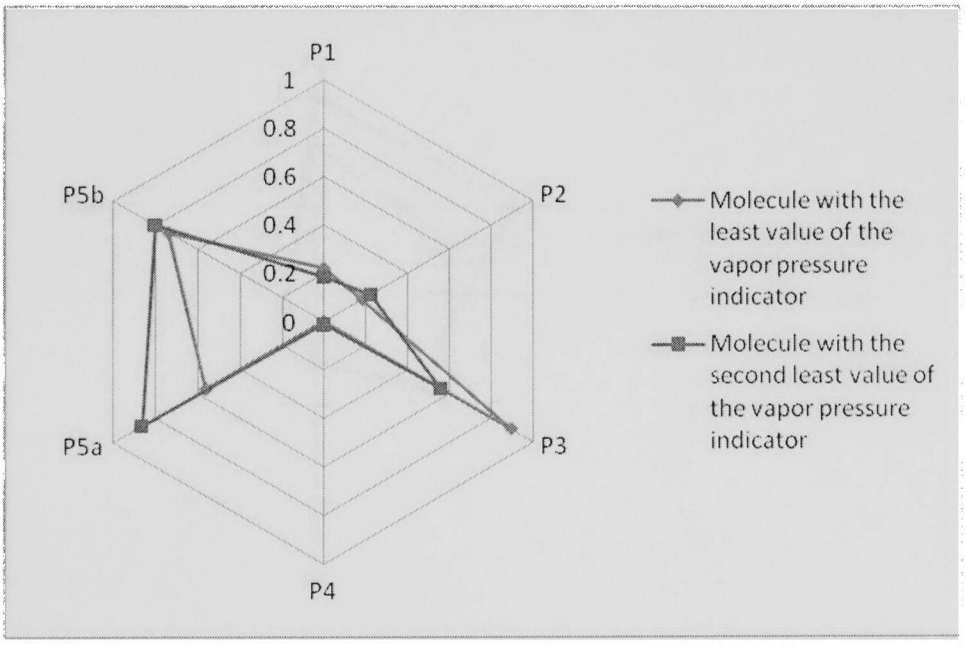



(b)

**Figure 7.3:** Representation of the performance of compounds which have (a) two of the largest tropospheric half-life times, and (b) two of the smallest tropospheric half-life times

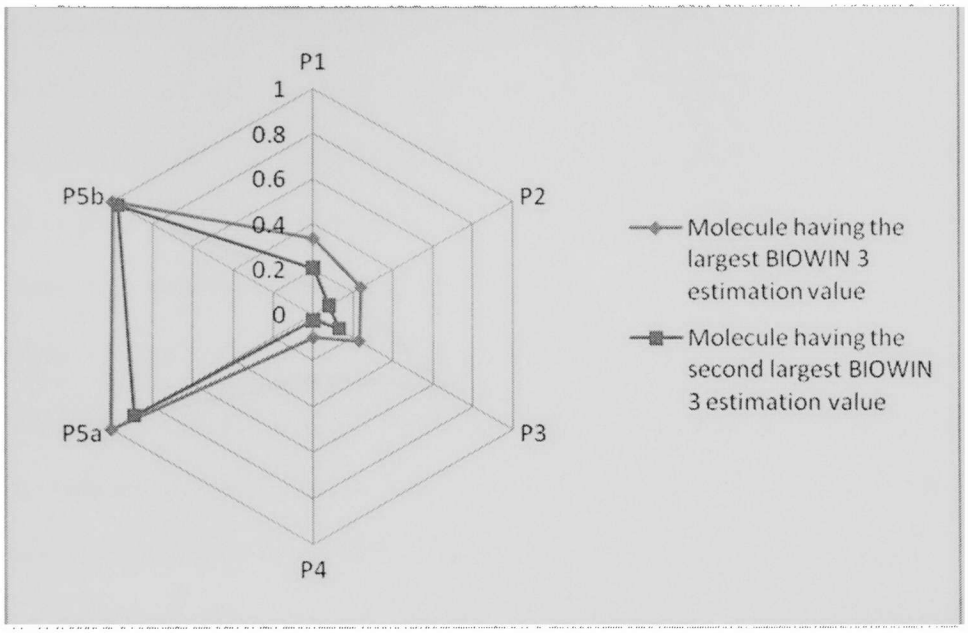



(a)

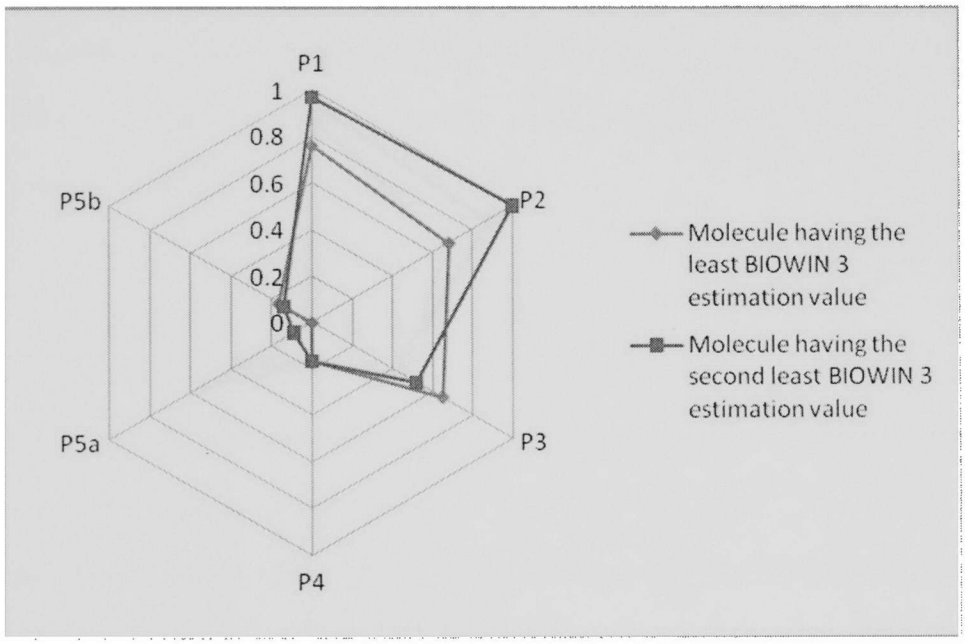



(b)

**Figure 7.4:** Representation of the performance of compounds which have (a) two of the largest values of the indicator of the vapor pressure, and (b) two of the smallest values of the indicator of the vapor pressure

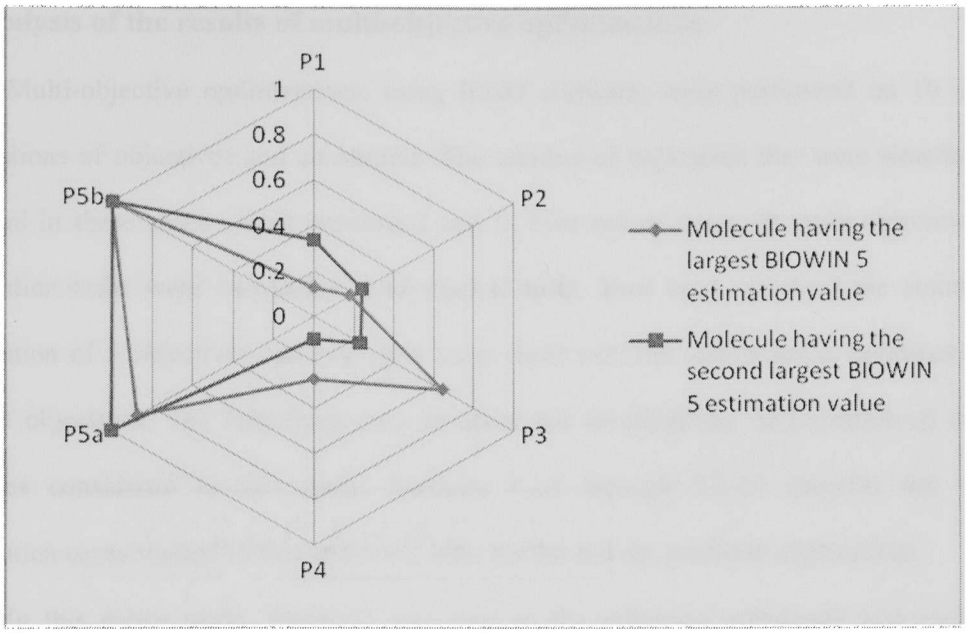



(a)

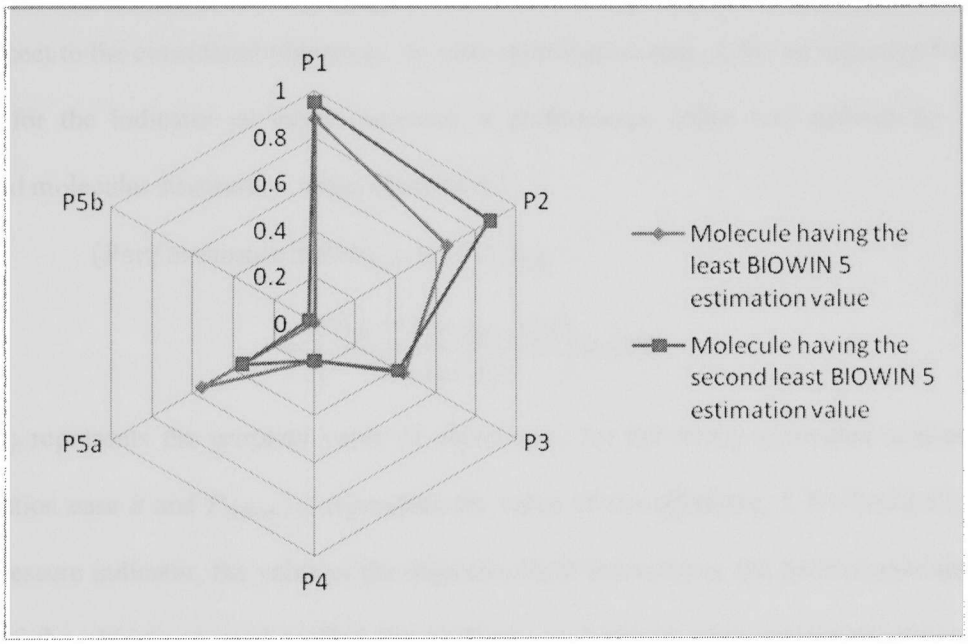



(b)

**Figure 7.5:** Representation of the performance of compounds which have (a) two of the largest BIOWIN 3 estimation values, and (b) two of the smallest BIOWIN 3 estimation values




(a)




(b)

**Figure 7.6:** Representation of the performance of compounds which have (a) two of the largest BIOWIN 5 estimation values, and (b) two of the smallest BIOWIN 5 estimation values

## 7.2. Analysis of the results of multi-objective optimizations

Multi-objective optimizations, using IOSO software, were performed on 10 different combinations of objectives and constraints. The number of objectives that were simultaneously optimized in these cases varied between 2 and 5. Five out of these 10 multi-objective design optimization tasks were 2-objective optimization tasks. Two tasks involved the simultaneous optimization of 3 objectives and two other cases dealt with the simultaneous optimization of 4 different objectives. The remaining case involved the simultaneous optimization of all the 5 objectives considered in this study. Sections 7.2.1 through 7.2.10 describe the different optimization cases studied in this endeavor, their results and the probable implications.

In this entire study, Freon-12 was used as the reference refrigerant and each of the optimization tasks was aimed at finding a molecule that showed improvement over Freon-12, with respect to the considered objectives. In each optimization case, $k$, for an objective function, $i$, (except for the indicator of vapor pressure), a performance index was defined for a newly generated molecular structure, $j$, using equation 7.1.

$$(Performance\ index)_{ijk} = (P.I.)_{ijk} = \left(\frac{P_{ijk} - P_{(Freon-12)i}}{P_{(Freon-12)i}}\right) \times 100 \tag{7.1}$$

Here, $P_{ijk}$ represents the property value of objective $i$, for the newly generated molecule, $j$, in optimization case $k$ and $P_{(Freon\text{-}12)i}$ represents the value of the objective, $i$, for Freon-12. For the vapor pressure indicator, the value of the objective itself is treated as the performance index. The values of the objectives used to find the performance indexes were estimated using different property models. For each of the newly generated molecular structures, the normal boiling temperatures, the enthalpies of vaporization and the vapor pressure indicators were estimated using the SPARC online calculator. Similarly, their tropospheric half-life times and the biodegradabilities were predicted using AOPWIN and BIOWIN, respectively. These estimated

property values were not only used to calculate the performance indexes; they were also used for an overall verification and validation of the results obtained from the IOSO software. Performance indexes were used in comparing the performance of each of the newly generated molecular structures with respect to that of Freon-12.

Scaled values of the performance indexes were then calculated for each of the objectives, $i$, for each of the newly generated molecular structures, $j$, in each of the optimization cases denoted by $k$. This *scaled performance index* (S.P.I.) was calculated using the following equation.

$$(Scaled\ Performance\ Index)_{ijk} = (S.P.I.)_{ijk} = \frac{P_{ijk} - P_{(Freon-12)i}}{(P_{high})_{ik} - (P_{low})_{ik}} \tag{7.2}$$

Here, $(P_{low})_{ik}$ represents the lowest value of the property for objective $i$ and in the case $k$ and $(P_{high})_{ik}$ represents the highest property value for all the generated molecules for the objective $i$, in the case $k$. The scaled performance index was used to gauge the performances of each of the generated molecules with respect to each other and Freon-12. Any generated molecule was considered to be highly preferable if it possessed a relatively higher S.P.I. for an objective that was supposed to be maximized. Similarly, any newly generated molecule was considered desirable if it possessed relatively lower S.P.I. for an objective that was supposed to be minimized.

The importance of each of the objectives depends on the particular practical situation for which an optimized refrigerant is needed. Therefore, this importance can be the topic of a lengthy subjective discussion. In this study, it was assumed that each of the objectives is equally important and for the sole purpose of ranking the generated molecules, a single entity was created. This entity was called the average scaled performance index and it was used to rank the candidate solutions in the order of their importance. The average scaled performance index of any newly generated molecular, $j$, was then calculated using the following equation.

$$(Average\ S.P.I.)_j = \frac{\sum_i^n (c_i \cdot (S.P.I.)_{ijk})}{n} \quad (7.3)$$

Here, $c_i$ represents user specified weighting coefficient for the particular objective, $i$. In this study, its value is specified to be 1 if the objective is maximized and -1 if the objective is minimized. $n$ represents the total number of simultaneous objectives involved in the optimization process. The generated molecules were ranked according to the average value of the scaled performance indexes of all the objectives.

### 7.2.1. Case-I: The simultaneous maximization of the normal boiling temperature and the maximization of the enthalpy of vaporization

This design optimization task was aimed at finding molecular structures that showed the largest possible increase in the normal boiling temperatures and the enthalpies of vaporization, over that of Freon-12. Constraints used in this two-objective constrained optimization problem were specified using equations 5.1 through 5.14. Other multiple equality and inequality constraints involving combinations of design variables were also enforced. Information related to all the 295 compounds in the initial pool of data was used in the creation of the response surfaces.

Figure 7.7 shows significant differences in the normal boiling temperatures obtained by IOSO software and those generated by SPARC online calculator. The points are labeled 1 through 6 for easy identification. 6 unique chemically possible molecular structures were obtained as a result of the optimization. Table 7.2 shows the results obtained for task 1 and the comparison of the properties (IOSO software estimations) of the six generated molecular structures with those obtained from the SPARC online calculator. The differences in the SPARC estimated normal boiling temperatures and those predicted by IOSO can also be attributed to the relatively larger inaccuracy of the normal boiling temperature model used in the SPARC online calculator.

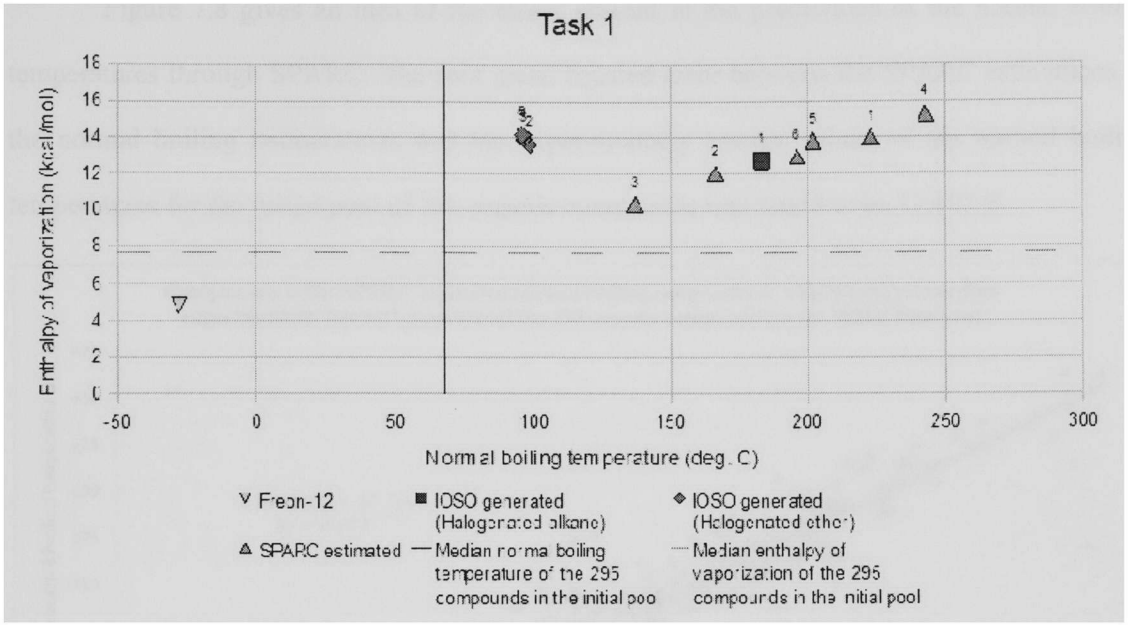



**Figure 7.7:** Properties of the molecular structures obtained for case-1 using IOSO software and their comparison with those generated through SPARC online calculator

**Table 7.2:** Comparison of the IOSO generated and SPARC predicted properties of all the six generated molecular structures in case-1

| Results for Pareto-optimized molecules | IOSO predicted normal boiling temperature | | SPARC predicted normal boiling temperature | | IOSO predicted enthalpy of vaporization | | SPARC predicted enthalpy of vaporization | |
|---|---|---|---|---|---|---|---|---|
| | K | % | K | % | kcal/mol | % | kcal/mol | % |
| Molecule 1 | 456.58 | 186.24 | 496.15 | 202.39 | 12.73 | 266.24 | 13.97 | 292.38 |
| Molecule 2 | 371.78 | 151.65 | 439.85 | 179.42 | 13.62 | 284.95 | 12.06 | 252.41 |
| Molecule 3 | 369.59 | 150.76 | 410.45 | 167.43 | 13.91 | 291.02 | 10.37 | 217.04 |
| Molecule 4 | 369.50 | 150.72 | 515.85 | 210.42 | 14.14 | 295.73 | 15.25 | 319.17 |
| Molecule 5 | 369.26 | 150.62 | 475.15 | 193.82 | 14.17 | 296.51 | 13.78 | 288.41 |
| Molecule 6 | 369.15 | 150.58 | 469.45 | 191.50 | 14.19 | 296.86 | 12.96 | 271.24 |

Figure 7.8 gives an idea of the errors present in the predictions of the normal boiling temperatures through SPARC. The root mean squared error between the SPARC estimations of the normal boiling temperatures and the experimentally known values of the normal boiling temperatures for the initial pool of 295 organic compounds was found to be 17.605 K.

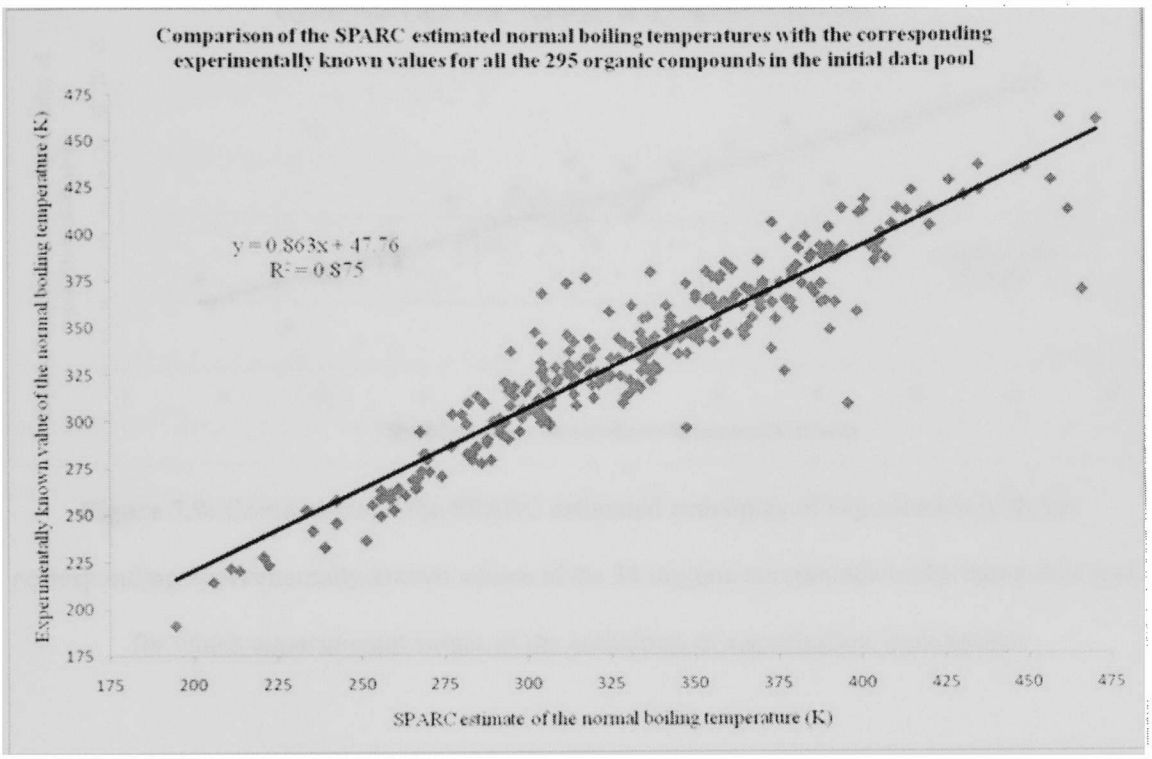



**Figure 7.8:** Comparison of the SPARC estimated normal boiling temperatures with the corresponding experimentally known values for all the 295 organic compounds in the original data pool

The root mean squared error between the IOSO generated normal boiling temperature and the SPARC estimated normal boiling temperature for all the 6 generated molecules was found to be 91.79 °C. Similarly, the root mean squared error between the IOSO generated enthalpies of vaporization for all the 6 generated molecules was found to be 1.80 kcal/ mol. Again, the large differences between the SPARC estimated enthalpies of vaporization and the IOSO predictions of the enthalpies of vaporization can also be partially attributed to the errors in

the estimations of the enthalpies of vaporization using SPARC online calculator. Figure 7.9 gives an idea of the errors present in the estimation of the enthalpies of vaporization of 54 compounds from the initial pool of data. The causes of errors will be discussed in detail in Section 7.3.

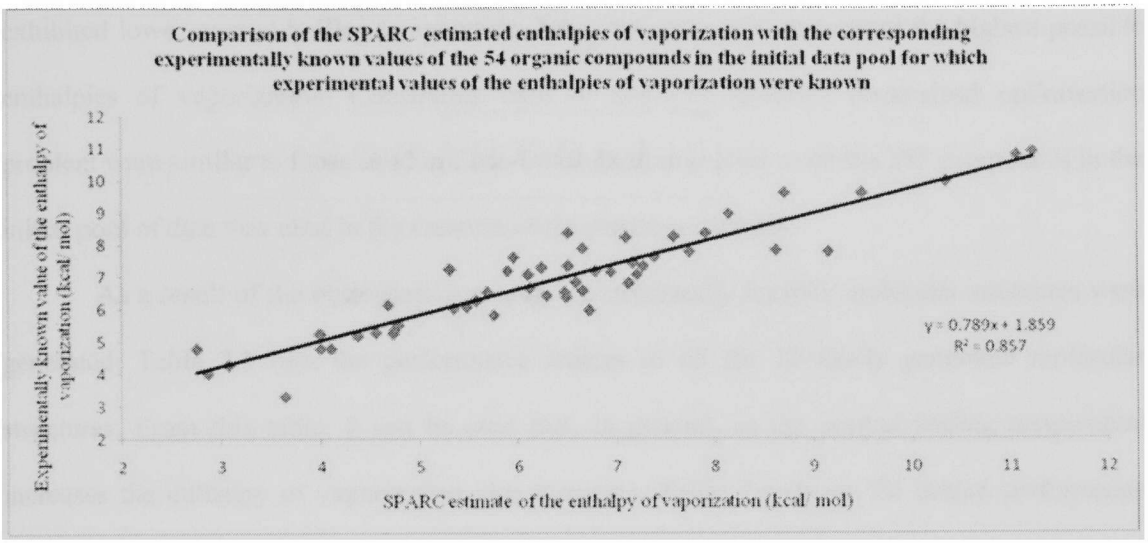



**Figure 7.9:** Comparison of the SPARC estimated enthalpies of vaporization with the corresponding experimentally known values of the 54 organic compounds in the initial data pool for which experimental values of the enthalpies of vaporization were known

On an average, the 6 generated molecules were found to have their normal boiling temperatures 139.16 °C more than that of Freon-12, if these values are calculated using the normal boiling temperatures generated by IOSO software. However, average increase in the normal boiling temperatures was found to be 222.67 °C, if SPARC estimations of the normal boiling temperatures were considered.

Similarly, when IOSO generated enthalpies of vaporization were used, it was found that the average of the enthalpies of vaporization of the 6 generated molecules surpassed the enthalpy of vaporization of Freon-12 by 9.02 kcal/ mol. However, when SPARC generated values were used the average difference between the enthalpy of vaporization of Freon-12 and those of the 6 newly generated molecules, was found to be 8.29 kcal/ mol.

### 7.2.2. Case-II: The simultaneous minimization of the normal boiling temperature and the maximization of the enthalpy of vaporization

This two-objective optimization task was aimed at generating molecular structures which exhibited lower normal boiling temperature, but at the same time possessed the highest possible enthalpies of vaporization. Constraints used in this two-objective constrained optimization problem were similar to those used in Case-I. Information related to all the 295 compounds in the initial pool of data was used in the creation of the response surfaces.

As a result of the optimization process 12 structurally feasible molecular structures were generated. Table 7.3 lists the performance indices of all the 12 newly generated molecular structures. From this table, it can be seen that, in general, as the normal boiling temperature increases the enthalpy of vaporization also increases. Table 7.4 shows the scaled performance indices of all the generated molecules and their ranks based on the value of their average S.P.I.

**Table 7.3:** Performance indices for all the newly generated molecules in case-2

| Results for Pareto-optimized molecules | Performance index (%) for | |
|---|---|---|
| | Normal boiling temperature | Enthalpy of vaporization |
| Molecule 1 | -36.83 | -58.37 |
| Molecule 2 | -19.84 | -39.96 |
| Molecule 3 | 17.01 | 16.53 |
| Molecule 4 | 33.62 | 45.61 |
| Molecule 5 | 34.11 | 71.76 |
| Molecule 6 | 34.55 | 71.13 |
| Molecule 7 | 61.55 | 98.54 |
| Molecule 8 | 58.94 | 126.36 |
| Molecule 9 | 64.98 | 136.61 |
| Molecule 10 | 68.94 | 141.00 |
| Molecule 11 | 83.7 | 187.66 |
| Molecule 12 | 91.5 | 171.13 |

**Table 7.4:** Scaled performance indices of the molecules generated in case-2 and their ranks

| Results for Pareto-optimized molecules | Scaled performance index (S.P.I.) for | | Average S.P.I. | Rank |
|---|---|---|---|---|
| | Normal boiling temperature | Enthalpy of vaporization | | |
| Molecule 1 | -0.29 | -0.24 | 0.02 | 3 |
| Molecule 2 | -0.15 | -0.16 | 0 | 8 |
| Molecule 3 | 0.13 | 0.07 | -0.03 | 11 |
| Molecule 4 | 0.26 | 0.19 | -0.04 | 12 |
| Molecule 5 | 0.27 | 0.29 | 0.01 | 6 |
| Molecule 6 | 0.27 | 0.29 | 0.01 | 6 |
| Molecule 7 | 0.48 | 0.4 | -0.04 | 12 |
| Molecule 8 | 0.46 | 0.51 | 0.03 | 2 |
| Molecule 9 | 0.51 | 0.56 | 0.02 | 3 |
| Molecule 10 | 0.54 | 0.57 | 0.02 | 3 |
| Molecule 11 | 0.65 | 0.76 | 0.06 | 1 |
| Molecule 12 | 0.71 | 0.7 | -0.01 | 10 |
| Freon-12 | 0 | 0 | 0 | 8 |

Based solely on the average S.P.I., the molecule 11 was ranked 1. This molecule showed an increase of 83.7% and 187.66% in the normal boiling temperature and the enthalpy of vaporization, respectively over that of Freon-12. For the newly generated molecules, the value of the root mean squared error between the SPARC estimations of the normal boiling temperatures and IOSO predicted normal boiling temperatures was found to be 56.99 K. Similarly, the rot mean squared error between the SPARC estimated enthalpies of vaporization and IOSO predicted enthalpies of vaporization for the newly developed molecules was found to be 1.21 kcal/ mol. Figure 7.10 graphically shows the performance of each of the newly generated molecular structures on a normal boiling temperature-enthalpy of vaporization plane. The molecules are numbered from 1 to 12 for easy identification.

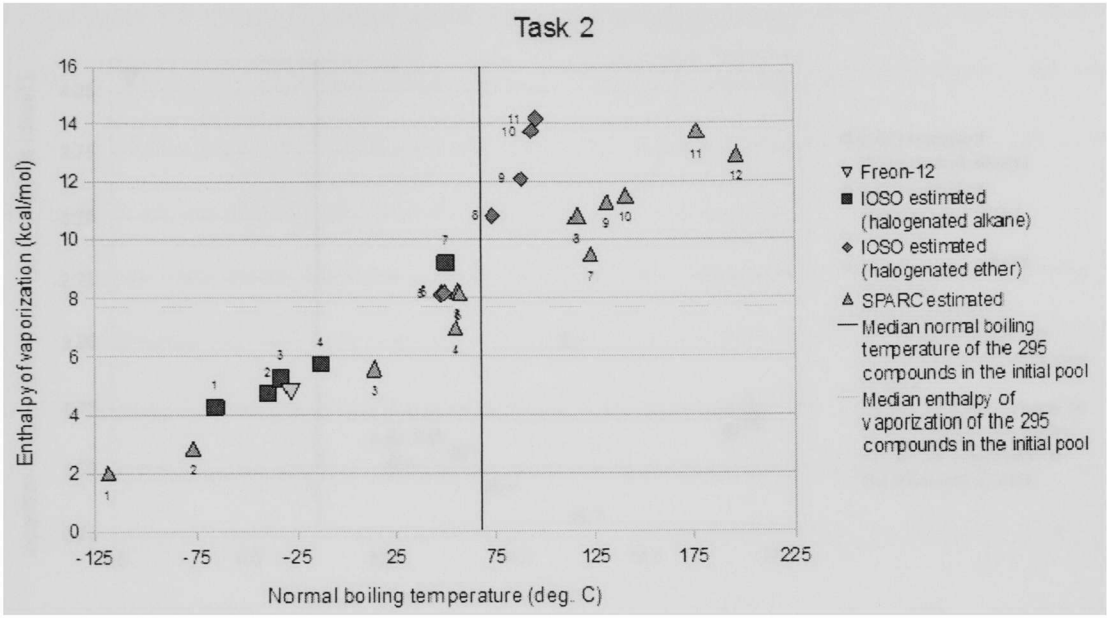



**Figure 7.10:** Performance of the newly generated molecular structures in case-2

7.2.3. Case-III: The simultaneous maximization of the enthalpy of vaporization and the minimization of the tropospheric half-life time

The aim of this two-objective optimization task was to find a molecular structure which showed the maximum possible enthalpy of vaporization, but had the least possible tropospheric half-life time. Information regarding 269 compounds was used during the optimization process. The constraints involved in this task were given by equations 5.1 through 5.15.

As a result of the optimization, IOSO software produced 5 chemically possible molecular structures. The distribution of these molecular structures on the enthalpy of vaporization-tropospheric half-life plane is shown in Figure 7.11. The optimization process generated 5 different chemically feasible molecular structures. The performance indexes for all the 5 newly generated molecular structures, are shown in Table 7.5. In the calculation of the performance indices, the enthalpy of vaporization and the tropospheric half-life time values of the newly

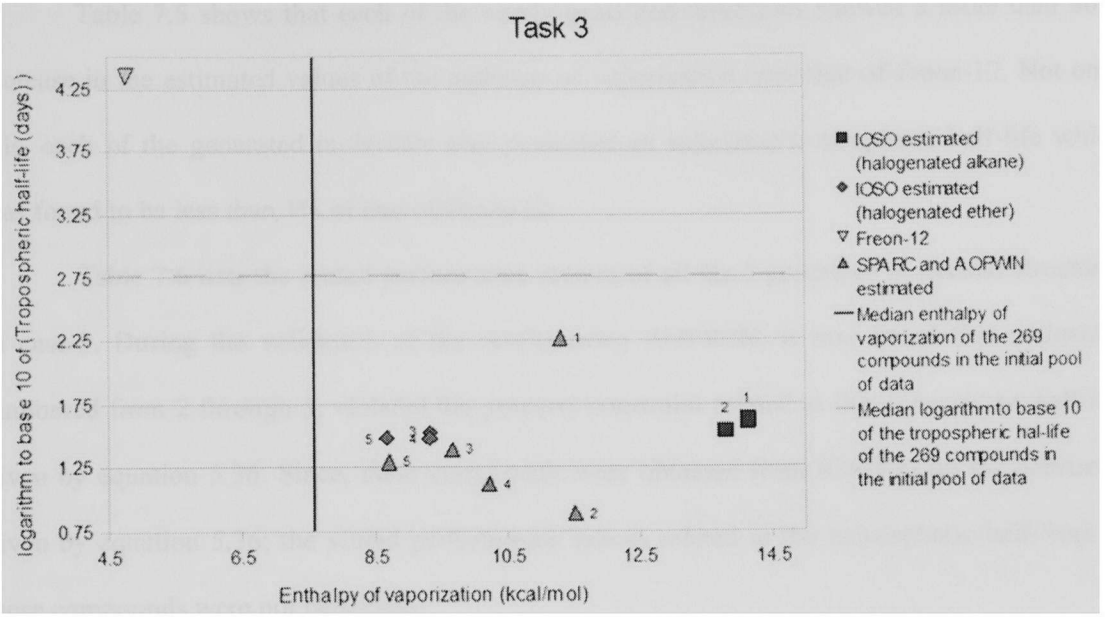



**Figure 7.11:** Distribution of the newly generated molecular structures on the enthalpy of vaporization-tropospheric half-life plane

generated molecules were estimated using SPARC online calculator and AOPWIN program, respectively.

**Table 7.5:** Performance indexes of the newly generated molecular structures in case-III

| Molecular Structure ID | Performance Index (%) for the | |
|---|---|---|
| | Enthalpy of vaporization | Tropospheric half-life |
| 1 | 136.50 | -99.16 |
| 2 | 141.11 | -99.97 |
| 3 | 101.97 | -99.89 |
| 4 | 113.69 | -99.94 |
| 5 | 81.88 | -99.91 |

Table 7.5 shows that each of the newly generated molecules showed a more than 80% increase in the estimated values of the enthalpy of vaporization over that of Freon-12. Not only this, each of the generated molecules also possessed an estimated tropospheric half-life which was found to be less than 1% of that of Freon-12.

Table 7.6 lists the scaled performance indices of all the 5 generated molecular structures in case-3. During the validation of the results using AOPWIN, it was found that molecules numbered from 2 through 5, violated the property constraint related to the tropospheric half-life given by equation 5.36. Since, these compounds were obtained from IOSO using the constraint given by equation 5.36; the scaled performance indices related to the tropospheric half-lives of these compounds were not penalized.

**Table 7.6:** Scaled performance indices of the generated compounds and their ranks for case-3

| | | **Molecule 1** | **Molecule 2** | **Molecule 3** | **Molecule 4** | **Molecule 5** |
|---|---|---|---|---|---|---|
| **S.P.I. for** | Enthalpy of vaporization | 0.9674 | 1.0000 | 0.7226 | 0.8057 | 0.5802 |
| | Tropospheric half-life time | -0.9920 | -1.0000 | -0.9992 | -0.9998 | -0.9995 |
| Average S.P.I. | | 0.9797 | 1.0000 | 0.8609 | 0.9027 | 0.7899 |
| Rank | | 2 | 1 | 4 | 3 | 5 |

These compounds were then ranked according to the values of their average scaled performance indices. Thus, molecule 2 was ranked 1. It should be noted that this molecule showed a 141.11% increase over the estimated enthalpy of vaporization of Freon-12. At the same time, its tropospheric half-life time was approximately equal to 0.03% of that of Freon-12. For the newly generated molecular structures, the root mean squared error between the SPARC estimated values of the enthalpies of vaporization and those predicted by IOSO was found to be 1.681 kcal/ mol. Similarly, for the newly generated molecules, the root mean squared error between the

estimations of the tropospheric half-life using AOPWIN and those predicted using IOSO software was found to be 66.77 days.

### 7.2.4. Case-IV: The simultaneous maximization of the enthalpy of vaporization and the minimization of the vapor pressure indicator

This optimization task was aimed at generating a molecular structure that possessed the highest possible enthalpy of vaporization, but at the same time had vapor pressure variation over a temperature range as similar as possible to that of Freon-12. The optimization process used information about 291 different organic compounds for which the enthalpies of vaporization and the variation of vapor pressure over a temperature range was known (or estimated). This constrained 2-objective design optimization problem involved all the constraints used in case-II. Three additional constraints on the values of a, b and c were also used. The optimization resulted in the generation of 6 new chemically possible molecular structures.

Table 7.7 lists the performance indices for all the 6 newly generated molecules in case-4. It shows that each of the generated molecules, except molecule 6, shows improvement in the enthalpy of vaporization over that of Freon-12.

**Table 7.7:** Performance indices for all the 6 generated molecules in case-4

| Results for Pareto-optimized molecules | Performance index (%) for Enthalpy of vaporization | Vapor pressure indicator |
|---|---|---|
| Molecule 1 | 216.032 | 577.381 |
| Molecule 2 | 162.453 | 577.002 |
| Molecule 3 | 100.712 | 574.773 |
| Molecule 4 | 59.481 | 558.614 |
| Molecule 5 | 33.738 | 507.819 |
| Molecule 6 | -16.492 | 101.835 |

Table 7.8 shows the scaled performance indices for the 6 generated molecular structures. SPARC predicted values were used in the calculation of the scaled performance indices. Despite of the fact that 5 out of the 6 generated molecules had higher enthalpies of vaporization than Freon-12, the average scaled performance indices of the molecules suggested that Freon-12 was better than any of the generated molecules. However, molecule 1, which was ranked just below Freon-12, showed 216.032% higher estimated enthalpy of vaporization than that of Freon-12.

**Table 7.8:** Scaled performance indices for the 6 generated molecules and their ranks in case-4

| Molecular structure ID | Scaled performance index (S.P.I.) for | | Average S.P.I. | Rank |
|---|---|---|---|---|
| | Enthalpy of vaporization | Vapor pressure indicator | | |
| Molecule 1 | 0.929 | 1.000 | -0.035 | 2 |
| Molecule 2 | 0.699 | 0.999 | -0.150 | 4 |
| Molecule 3 | 0.433 | 0.995 | -0.281 | 5 |
| Molecule 4 | 0.256 | 0.967 | -0.356 | 6 |
| Molecule 5 | 0.145 | 0.880 | -0.367 | 7 |
| Molecule 6 | -0.071 | 0.176 | -0.124 | 3 |
| Freon-12 | 0.000 | 0.000 | 0.000 | 1 |

The root mean squared error between the SPARC predicted enthalpies of vaporization of all the newly generated molecular structures and the corresponding predictions using the IOSO software was found to be 1.63 kcal/ mol. Similarly, the root mean squared error between the estimations of vapor pressure indictor using SPARC online calculator and the corresponding predictions using the IOSO software was found to be 1163.909 kPa.

### 7.2.5. Case-V: The simultaneous maximization of the normal boiling temperature, the maximization of the enthalpy of vaporization and the minimization of the tropospheric half-life time

The primary aim of this optimization task was to find molecules that possessed the largest possible normal boiling temperatures and the largest possible values of the enthalpy of vaporization, but which had relatively low tropospheric half-life. The least value of the tropospheric life-time was specified to be 30 days. Data related to 269 compounds was used in the creation of the response surfaces. The specified constraints were similar to those specified in the case-III.

13 new chemically possible molecular structures were generated as a result of the optimization process using IOSO software. IOSO also predicted the values of their normal boiling temperatures, enthalpies of vaporization and the tropospheric half-life times. For the validation of results, the normal boiling temperatures and the enthalpies of vaporizations of the 13 newly generated molecular structures were then estimated using the SPARC online calculator. AOPWIN was used to validate the tropospheric half-life times of the generated molecules. Table 7.9 shows that considerable difference between the IOSO estimations and those obtained from other prediction models were found. A relatively large difference in the predictions of the normal boiling temperatures can also be partially attributed to the relatively higher inaccuracy of the normal boiling temperature prediction model used in the SPARC online calculator. However, it should be noted that the molecule structures generated from IOSO, show significant improvements with respect to Freon-12, with respect to all the three objectives. This observation was confirmed by the performance indices listed in Table 7.10.

Table 7.11 lists the scaled performance indices and the ranks of the 13 newly generated molecular structures in this case.

**Table 7.9:** Validation of the results obtained in task 5 and their comparison with Freon-12

| For the 13 newly generated molecular structures | Normal boiling temperature (°C) | Enthalpy of vaporization (kcal/mol) | Tropospheric half-life time (days) |
|---|---|---|---|
| RMS error between the IOSO generated values and those predicted by other models | 45.81* | 2.73* | 102.63# |
| Average change in the property values, as compared to those of Freon-12 | +141.94* | +4.83* | -22200.13# |

*For the purpose of the validation of the results, this property was estimated using SPARC online calculator

#For the purpose of the validation of the results, this property was estimated using AOPWIN module

**Table 7.10:** Performance indices for all the generated molecules in case-5

| Results for Pareto-optimized molecules | Performance Index (%) for | | |
|---|---|---|---|
| | Normal boiling temperature | Enthalpy of vaporization | Tropospheric half-life |
| Molecule 1 | 102.386 | 192.382 | -97.862 |
| Molecule 2 | 87.701 | 171.452 | -99.794 |
| Molecule 3 | 56.659 | 95.061 | -99.818 |
| Molecule 4 | 66.979 | 116.409 | -99.837 |
| Molecule 5 | 27.367 | 37.505 | -99.811 |
| Molecule 6 | 31.809 | 44.621 | -99.454 |
| Molecule 7 | 45.854 | 72.039 | -99.975 |
| Molecule 8 | 73.343 | 130.013 | -99.806 |
| Molecule 9 | 59.147 | 101.967 | -99.889 |
| Molecule 10 | 62.288 | 113.688 | -99.941 |
| Molecule 11 | 30.948 | 52.574 | -99.515 |
| Molecule 12 | 49.733 | 81.875 | -99.911 |
| Molecule 13 | 58.495 | 104.270 | -99.721 |
| Freon-12 | 0.000 | 0.000 | 0.000 |

It should be noted that the molecule 1 was ranked 1. It not only showed vast improvements in terms of the normal boiling temperature and enthalpy of vaporization over Freon-12. Its estimated tropospheric half-life time was only 2.138% of that of Freon-12.

**Table 7.11:** Scaled performance indices for the 13 newly generated molecules and their ranks in case-5

| Molecular structure ID | Scaled performance index (S.P.I.) for | | | Average S.P.I. | Rank |
|---|---|---|---|---|---|
| | Normal boiling temperature | Enthalpy of vaporization | Tropospheric half-life | | |
| Molecule 1 | 1.000 | 1.000 | -0.979 | 0.993 | 1 |
| Molecule 2 | 0.857 | 0.891 | -0.998 | 0.915 | 2 |
| Molecule 3 | 0.553 | 0.494 | -0.998 | 0.682 | 8 |
| Molecule 4 | 0.654 | 0.605 | -0.999 | 0.753 | 4 |
| Molecule 5 | 0.267 | 0.195 | -0.998 | 0.487 | 13 |
| Molecule 6 | 0.311 | 0.232 | -0.995 | 0.512 | 12 |
| Molecule 7 | 0.448 | 0.374 | -1.000 | 0.607 | 10 |
| Molecule 8 | 0.716 | 0.676 | -0.998 | 0.797 | 3 |
| Molecule 9 | 0.578 | 0.530 | -0.999 | 0.702 | 7 |
| Molecule 10 | 0.608 | 0.591 | -1.000 | 0.733 | 5 |
| Molecule 11 | 0.302 | 0.273 | -0.995 | 0.524 | 11 |
| Molecule 12 | 0.486 | 0.426 | -0.999 | 0.637 | 9 |
| Molecule 13 | 0.571 | 0.542 | -0.997 | 0.704 | 6 |
| Freon-12 | 0.000 | 0.000 | 0.000 | 0.000 | 14 |

7.2.6. Case-VI: The simultaneous maximization of the normal boiling temperature, the maximization of the enthalpy of vaporization and the minimization of the indicator of vapor pressure

The aim of this 3-objective optimization task was to find a molecular structure which possessed the highest possible normal boiling temperature and the highest possible enthalpy of vaporization, but at the same time had similar variations in the vapor pressure over a temperature range, to that of Freon-12. Data related to 291 compounds was used in the optimization process. Constraints used in this optimization task ensured structural feasibility of the newly generated molecular structures, set the lower bound of the enthalpy of vaporization and enforced limits on the coefficients a, b and c.

The optimization process resulted in the generation of 6 chemically feasible molecular structures. Table 7.12 shows the performance indices of all the 6 newly generated molecular structures in this task. As seen from Table 7.13, molecule 1 is ranked 1 among the generated molecular structures. It shows an increase of 110.749% in the normal boiling temperature over that of Freon-12. It also has a 213.102% higher enthalpy of vaporization as compared to that of Freon-12. However, among all the 6 newly generated molecular structures, this molecule is found to be the one having the most different vapor pressure variations over a range of temperatures, than Freon-12.

The root mean squared errors were calculated for all the 6 newly generated molecular structures between the SPARC estimated values of the normal boiling temperature, enthalpy of vaporization and vapor pressure indicator, and the corresponding predicted values using IOSO software. The results are listed below.

a) RMS error between the SPARC estimated normal boiling temperatures and the corresponding predictions using IOSO software: 46.62 K

b) RMS error between the SPARC estimated enthalpies of vaporization and the corresponding predictions using IOSO software: 1.94 kcal/ mol

c) RMS error between the estimated vapor pressure indicator using SPARC and the corresponding predictions using IOSO software: 688.16 kPa

**Table 7.12:** Performance indices of the 6 newly generated molecular structures in case-6

| Results for Pareto-optimized molecules | Performance Index (%) for | | Vapor pressure indicator |
|---|---|---|---|
| | Normal boiling temperature | Enthalpy of vaporization | |
| Molecule 1 | 110.749 | 213.102 | 577.383 |
| Molecule 2 | 103.732 | 216.032 | 577.381 |
| Molecule 3 | 63.512 | 100.712 | 574.773 |
| Molecule 4 | 43.373 | 59.481 | 558.614 |
| Molecule 5 | 27.196 | 33.738 | 507.819 |
| Molecule 6 | 2.509 | -16.492 | 101.835 |

**Table 7.13:** Scaled performance indices of all the newly generated molecular structures in case-6

| Molecular structure ID | Scaled performance index (S.P.I.) for | | | Average S.P.I. | Rank |
|---|---|---|---|---|---|
| | Normal boiling temperature | Enthalpy of vaporization | Vapor pressure indicator | | |
| Molecule 1 | 1.000 | 0.916 | 1.000 | 0.305 | 1 |
| Molecule 2 | 0.937 | 0.929 | 1.000 | 0.289 | 2 |
| Molecule 3 | 0.573 | 0.433 | 0.995 | 0.004 | 3 |
| Molecule 4 | 0.392 | 0.256 | 0.967 | -0.107 | 6 |
| Molecule 5 | 0.246 | 0.145 | 0.880 | -0.163 | 7 |
| Molecule 6 | 0.023 | -0.071 | 0.176 | -0.075 | 5 |
| Freon-12 | 0.000 | 0.000 | 0.000 | 0.000 | 4 |

### 7.2.7. Case-VII: The simultaneous minimization of the indicator of the vapor pressure and the maximization of the sum of the BIOWIN 3 and the BIOWIN 5 values

This optimization task was aimed at generating molecules similar to Freon-12 with respect to the variation of vapor pressure, over a temperature range. However, the generated molecules were required to be as highly biodegradable as possible. Apart from the constraints used in case-6, two additional constraints were also enforced. These constraints were used to set the lower bounds on the BIOWIN 3 and BIOWIN 5 results. Data related to 265 compounds from the initial pool was used during the optimization process.

The optimization process resulted in the generation of only 3 chemically feasible molecular structures. The performance indices for all those 3 newly generated molecules are shown in Table 7.14. The performance indices and the ranks of each of these 3 molecular structures are listed in Table 7.15. It should be noted that the molecule ranked 1 (molecule 2) showed an increase of 41.946% and 39.803% in the BIOWIN 3 and BIOWIN 5 values, respectively over the corresponding values of Freon-12. As compared to the other generated molecular structures, its vapor pressure variation with temperature was not found very much different than that of Freon-12.

**Table 7.14:** Performance indices of all the 3 newly generated molecules in case-7

| Results for Pareto-optimized molecules | Performance Index (%) for | | Vapor pressure indicator |
|---|---|---|---|
| | BIOWIN 3 result | BIOWIN 5 result | |
| Molecule 1 | 28.637 | 9.431 | 101.835 |
| Molecule 2 | 41.946 | 39.803 | 319.630 |
| Molecule 3 | 31.443 | 28.337 | 558.714 |
| Freon-12 | 0.000 | 0.000 | 0.000 |

**Table 7.15:** Scaled performance indices of all the newly generated molecular structures in case-7

| Molecular structure ID | Scaled performance index (S.P.I.) for | | | Average S.P.I. | Rank |
|---|---|---|---|---|---|
| | BIOWIN 3 result | BIOWIN 5 result | Vapor pressure indicator | | |
| Molecule 1 | 0.683 | 0.237 | 0.182 | 0.246 | 2 |
| Molecule 2 | 1.000 | 1.000 | 0.572 | 0.476 | 1 |
| Molecule 3 | 0.750 | 0.712 | 1.000 | 0.154 | 3 |
| Freon-12 | 0.000 | 0.000 | 0.000 | 0.000 | 4 |

For all the 3 newly generated molecular structures, the RMS errors between the IOSO predicted values of BIOWIN 3, BIOWIN 5 and vapor pressure indicator and their corresponding estimations using BIOWIN and SPARC were calculated. They are presented below.

a) RMS error between SPARC estimated values of indicator of vapor pressure and the corresponding values predicted by IOSO software: 355.51 kPa

b) RMS error between BIOWIN estimated values of BIOWIN 3 model and the corresponding values predicted by IOSO software: 0.14

c) RMS error between BIOWIN estimated values of BIOWIN 5 model and the corresponding values predicted by IOSO software: 0.03

### 7.2.8. Case-VIII: The simultaneous maximization of the normal boiling temperature, the maximization of the enthalpy of vaporization, the minimization of the tropospheric half-life time and the minimization of the vapor pressure indicator

The aim of this 4-objective optimization task was to find molecular structures which possessed highest possible normal boiling temperatures and enthalpies of vaporization, had relatively low tropospheric half-life time and had vapor pressure variations similar to that of

Freon-12. Properties of 265 compounds in the initial data set were used in the optimization process. Constraints used in this optimization process not only ensured structural feasibility of the generated molecular structures, they also limited the values of the coefficients a, b and c. Constraints also limited the tropospheric half-life of any newly generated compound.

12 new structurally feasible molecules were generated as the result of the optimization. The performance indices are listed in Table 7.16 and the scaled performance indices are listed in Table 7.17. On the sole basis of the average scaled performance index, molecule 2 was ranked 1. It showed an increase of 103.732% in terms of the normal boiling temperature and 216.032% in terms of the enthalpy of vaporization over the corresponding values of Freon-12. Its tropospheric half-life was also estimated to be equal to 0.699% of that of Freon-12. However, the value of the indicator of vapor pressure for this particular molecule is relatively higher as compared to the other newly generated molecular structures.

**Table 7.16:** Performance indices of all the 12 newly generated molecular structures in case-8

| Results for Pareto-optimized molecules | Performance index (%) for | | | Vapor pressure indicator |
|---|---|---|---|---|
| | Normal boiling temperature | Enthalpy of vaporization | Tropospheric half-life | |
| Molecule 1 | 63.512 | 100.712 | -95.829 | 574.773 |
| Molecule 2 | 103.732 | 216.032 | -99.301 | 577.381 |
| Molecule 3 | 56.659 | 95.061 | -99.818 | 573.067 |
| Molecule 4 | 29.684 | 36.040 | -99.879 | 520.539 |
| Molecule 5 | 27.196 | 33.738 | -99.500 | 507.819 |
| Molecule 6 | 49.733 | 81.875 | -99.911 | 568.925 |
| Molecule 7 | 29.684 | 36.040 | -99.879 | 520.539 |
| Molecule 8 | 27.196 | 33.738 | -99.500 | 507.819 |
| Molecule 9 | -27.983 | -54.793 | -99.856 | 2758.556 |
| Molecule 10 | 5.495 | -1.214 | -99.832 | 140.539 |
| Molecule 11 | 0.090 | -19.841 | -99.771 | 187.675 |
| Molecule 12 | 2.509 | -16.492 | -98.876 | 101.835 |

**Table 7.17:** Scaled performance indices for all the newly generated molecular structures in case-8

| Results for Pareto-optimized molecules | Scaled performance index (S.P.I.) for | | | | Average S.P.I. | Rank |
|---|---|---|---|---|---|---|
| | Normal boiling temperature | Enthalpy of vaporization | Tropospheric half-life | Vapor pressure indicator | | |
| Molecule 1 | 0.482 | 0.372 | -0.959 | 0.208 | 0.401 | 2 |
| Molecule 2 | 0.788 | 0.798 | -0.994 | 0.209 | 0.592 | 1 |
| Molecule 3 | 0.430 | 0.351 | -0.999 | 0.208 | 0.393 | 3 |
| Molecule 4 | 0.225 | 0.133 | -1.000 | 0.189 | 0.292 | 5 |
| Molecule 5 | 0.206 | 0.125 | -0.996 | 0.184 | 0.286 | 7 |
| Molecule 6 | 0.378 | 0.302 | -1.000 | 0.206 | 0.368 | 4 |
| Molecule 7 | 0.225 | 0.133 | -1.000 | 0.189 | 0.292 | 6 |
| Molecule 8 | 0.206 | 0.125 | -0.996 | 0.184 | 0.286 | 7 |
| Molecule 9 | -0.212 | -0.202 | -0.999 | 1.000 | -0.104 | 13 |
| Molecule 10 | 0.042 | -0.004 | -0.999 | 0.051 | 0.246 | 9 |
| Molecule 11 | 0.001 | -0.073 | -0.999 | 0.068 | 0.214 | 11 |
| Molecule 12 | 0.019 | -0.061 | -0.990 | 0.037 | 0.228 | 10 |
| Freon-12 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 12 |

The RMS errors between the IOSO software predicted values and the corresponding property predictions using other models were also computed. The results are listed below.

a) RMS error between SPARC estimated values of the normal boiling temperatures and the corresponding values predicted by IOSO software: 56.849 K

b) RMS error between SPARC estimated values of the enthalpies of vaporization and the corresponding values predicted by IOSO software: 2.27 kcal/ mol

c) RMS error between AOPWIN estimated values of the tropospheric half-lives and the corresponding values predicted by IOSO software: 249.69 days

d) RMS error between SPARC estimated values of indicator of vapor pressure and the corresponding values predicted by IOSO software: 805.70 kPa

7.2.9. Case-IX: The simultaneous maximization of the normal boiling temperatures, the maximization of the enthalpy of vaporization, the minimization of the indicator of vapor pressure and maximization of the sum of the BIOWIN 3and the BIOWIN 5 values

This 4-objective design optimization was aimed at obtaining structurally feasible organic compounds that showed highest possible normal boiling temperatures, enthalpies of vaporization and biodegradabilities, but at the same time displayed similar vapor pressure variations with temperature as that of Freon-12. Constraints used in this optimization task were similar to the constraints used in case-8. However, unlike case-8, no constraints were used to restrict the tropospheric half life of the newly generated compounds. Data related to 265 organic compounds from the initial data pool was used in the optimization process.

Table 7.18 shows that most of the newly generated molecular structures showed improvements in the objective values when compared with Freon-12. Table 7.19 lists the scaled performance indices of each of the molecules and their ranks which were solely based on the value of the average scaled performance index.

**Table 7.18:** Performance indices of all the newly generated molecular structures in case-9

| Results for Pareto-optimized molecules | Performance index (%) for | | | | Vapor pressure indicator |
|---|---|---|---|---|---|
| | Normal boiling temperature | Enthalpy of vaporization | BIOWIN 3 result | BIOWIN 5 result | |
| Molecule 1 | 29.684 | 36.040 | 25.861 | 16.958 | 520.539 |
| Molecule 2 | 19.572 | 24.529 | 25.477 | 1.028 | 437.319 |
| Molecule 3 | 13.017 | 14.483 | 2.191 | 4.201 | 321.632 |
| Molecule 4 | -27.983 | -54.793 | 31.987 | 36.433 | 2758.556 |
| Molecule 5 | 0.090 | -19.841 | 30.314 | 39.256 | 187.675 |
| Molecule 6 | 27.196 | 33.738 | 24.188 | -12.845 | 507.819 |
| Molecule 7 | 37.683 | 52.156 | 18.871 | -4.289 | 549.226 |
| Molecule 8 | 2.509 | -16.492 | 28.637 | 9.431 | 101.835 |
| Freon-12 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

**Table 7.19:** Scaled performance indices and the ranks of the generated molecules in case-9

| Results for Pareto-optimized molecules | Scaled performance index (S.P.I.) for | | | | | Average S.P.I. | Rank |
|---|---|---|---|---|---|---|---|
| | Normal boiling temperature | Enthalpy of vaporization | BIOWIN 3 result | BIOWIN 5 result | Vapor pressure indicator | | |
| Molecule 1 | 0.452 | 0.337 | 0.808 | 0.325 | 0.189 | 0.35 | 1 |
| Molecule 2 | 0.298 | 0.229 | 0.796 | 0.020 | 0.159 | 0.24 | 4 |
| Molecule 3 | 0.198 | 0.135 | 0.068 | 0.081 | 0.117 | 0.07 | 7 |
| Molecule 4 | -0.426 | -0.512 | 1.000 | 0.699 | 1.000 | -0.05 | 9 |
| Molecule 5 | 0.001 | -0.186 | 0.948 | 0.753 | 0.068 | 0.29 | 2 |
| Molecule 6 | 0.414 | 0.315 | 0.756 | -0.247 | 0.184 | 0.21 | 5 |
| Molecule 7 | 0.574 | 0.488 | 0.590 | -0.082 | 0.199 | 0.27 | 3 |
| Molecule 8 | 0.038 | -0.154 | 0.895 | 0.181 | 0.037 | 0.18 | 6 |
| Freon-12 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.00 | 8 |

Molecule 1, which was ranked 1 among all the generated molecules, showed a higher normal boiling temperature, enthalpy of vaporization, BIOWIN 3 value and the BIOWIN 5 value than the corresponding values of Freon-12. The value of the vapor pressure indicator was neither found to be the minimum among all the newly generated molecules nor was it was the highest. As done in previously discussed optimization cases, the RMS errors between the predictions using the IOSO software and the estimations based on the property prediction models were calculated for all the 8 generated molecular structures. These RMS errors are listed below.

a) RMS error between SPARC estimated values of the normal boiling temperatures and the corresponding values predicted by IOSO software: 56.31 K

b) RMS error between SPARC estimated values of the enthalpies of vaporization and the corresponding values predicted by IOSO software: 2.04 kcal/ mol

c) RMS error between SPARC estimated values of indicator of vapor pressure and the corresponding values predicted by IOSO software: 859.97 kPa

d) RMS error between BIOWIN estimated values of BIOWIN 3 result and the corresponding values predicted by IOSO software: 0.24

e) RMS error between BIOWIN estimated values of BIOWIN 5 result and the corresponding values predicted by IOSO software: 0.098

7.2.10. Case-X: The simultaneous maximization of the normal boiling temperature, the maximization of the enthalpy of vaporization, the minimization of the tropospheric half-life time, the minimization of the indicator of vapor pressure and the maximization of the sum of the BIOWIN 3 and the BIOWIN 5 values

The aim of this five-objective design optimization task was to find a molecular structure which not only possesses higher normal boiling temperature, enthalpy of vaporization, BIOWIN 3 value and the BIOWIN 5 value, but also shows lower tropospheric life-time and has a similar vapor pressure-temperature profile as that of Freon-12. Figure 7.12 shows the location of these new molecules in the objective function space.

Table 7.20 shows the performance index of all the 7 newly generated molecular structures. It can be seen from this table that all the generated molecules perform better than Freon-12 with respect to the tropospheric half-life time and the biodegradability values. However, not all the molecules are found to have higher normal boiling temperature and a higher enthalpy of vaporization than Freon-12.

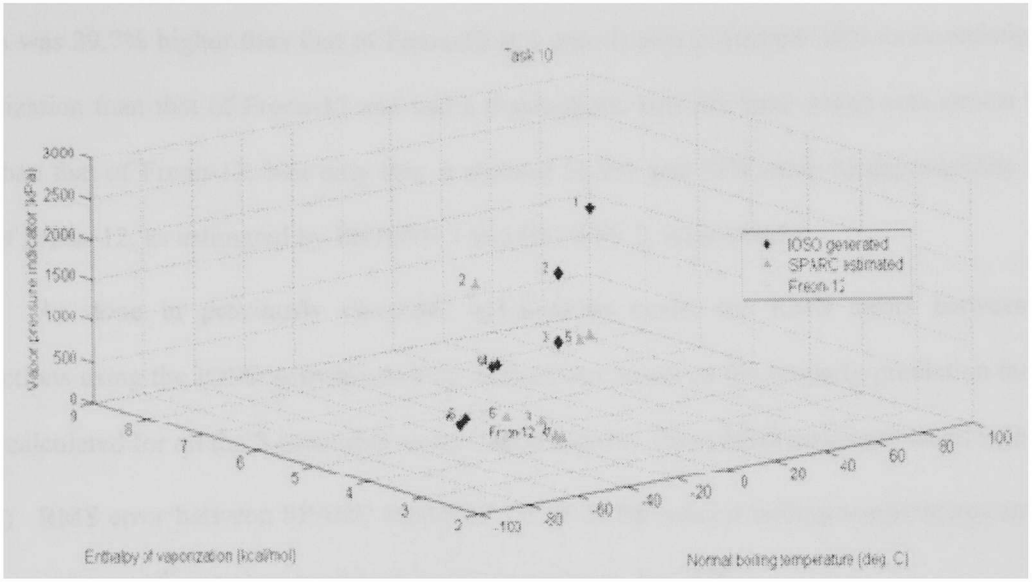
Table 7.21 shows the scaled performance indexes of all the 7 generated molecular structures with respect to each of the objectives. This table gives an indication of the performance of each of the molecules as compared to the other generated molecules and also ranks them according to their average performance indices.

**Table 7.20:** Performance indexes for all the 7 molecular structures generated in case-10

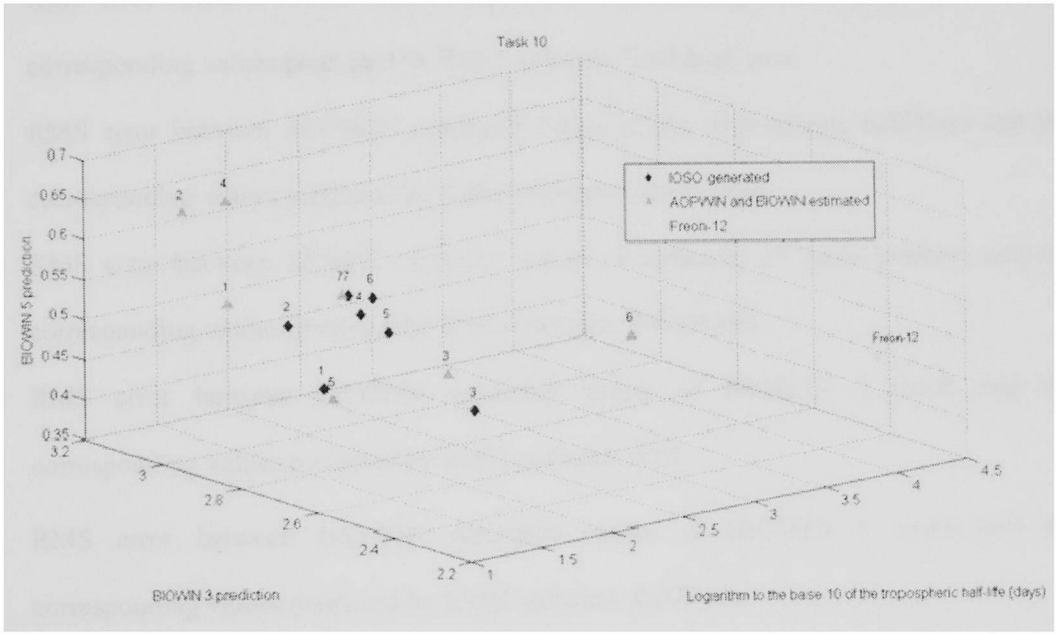
| Results for Pareto-optimized molecules | Performance index (%) for | | | | | Vapor pressure indicator |
|---|---|---|---|---|---|---|
| | Normal boiling temperature | Enthalpy of vaporization | Tropospheric half-life | BIOWIN 3 result | BIOWIN 5 result | |
| Molecule 1 | 29.7 | 36.0 | -99.9 | 25.9 | 17.0 | 520.539 |
| Molecule 2 | -28.0 | -54.8 | -99.9 | 32.0 | 36.4 | 2758.556 |
| Molecule 3 | 5.50 | -1.2 | -99.8 | 4.00 | 14.3 | 140.539 |
| Molecule 4 | 1.00 | -19.8 | -99.8 | 30.3 | 39.3 | 187.675 |
| Molecule 5 | 27.2 | 33.7 | -99.5 | 24.2 | -12.8 | 507.819 |
| Molecule 6 | -9.90 | -23.8 | -95.9 | 5.40 | 12.6 | 600.693 |
| Molecule 7 | 2.50 | -16.5 | -98.9 | 28.6 | 9.40 | 101.835 |

**Table 7.21:** Calculation of the scaled performance indices of the generated molecular structures and determination of their ranking in case-10

| | | Molecule structure ID | | | | | | | Freon-12 |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
| Scaled performance index (S.P.I.) for | Normal boiling temperature | 0.515 | -0.485 | 0.095 | 0.002 | 0.472 | -0.172 | 0.044 | 0.000 |
| | Enthalpy of vaporization | 0.397 | -0.603 | -0.013 | -0.218 | 0.371 | -0.262 | -0.182 | 0.000 |
| | Tropospheric half-life | -1.000 | -1.000 | -1.000 | -0.999 | -0.996 | -0.961 | -0.990 | 0.000 |
| | Vapor pressure indicator | 0.189 | 1.000 | 0.051 | 0.068 | 0.184 | 0.218 | 0.037 | 0.000 |
| | BIOWIN 3 result | 0.808 | 1.000 | 0.126 | 0.948 | 0.756 | 0.167 | 0.895 | 0.000 |
| | BIOWIN 5 result | 0.325 | 0.699 | 0.274 | 0.753 | -0.247 | 0.242 | 0.181 | 0.000 |
| Average S.P.I. | | 0.476 | 0.102 | 0.239 | 0.403 | 0.361 | 0.120 | 0.315 | 0.000 |
| Rank | | 1 | 7 | 5 | 2 | 3 | 6 | 4 | 8 |




(a)




(b)

**Figure 7.12:** Distribution of the newly generated molecules in the (a) thermodynamic property space (b) environmental property space

On the basis of the average scaled performance index values the best three molecular structures were found to be molecule 1, molecule 4 and molecule 5. Molecule 1 boiled at a temperature

which was 29.7% higher than that of Freon-12 at 1 atm. It also possessed 36% more enthalpy of vaporization than that of Freon-12 and had a tropospheric half-life time which was almost 99% less than that of Freon-12. Not only this, it showed 25.9% and 17% more biodegradability than that of Freon-12, as estimated by BIOWIN 3 and BIOWIN 5, respectively.

As done in previously discussed optimization cases, the RMS errors between the predictions using the IOSO software and the estimations based on the property prediction models were calculated for all the 8 generated molecular structures. These RMS errors are listed below.

a) RMS error between SPARC estimated values of the normal boiling temperatures and the corresponding values predicted by IOSO software: 61.54 K
b) RMS error between SPARC estimated values of the enthalpies of vaporization and the corresponding values predicted by IOSO software: 2.40 kcal/ mol
c) RMS error between AOPWIN estimated values of the tropospheric half-lives and the corresponding values predicted by IOSO software: 226.45 days
d) RMS error between SPARC estimated values of indicator of vapor pressure and the corresponding values predicted by IOSO software: 746.94 kPa
e) RMS error between BIOWIN estimated values of BIOWIN 3 result and the corresponding values predicted by IOSO software: 0.24
f) RMS error between BIOWIN estimated values of BIOWIN 5 result and the corresponding values predicted by IOSO software: 0.079

### 7.3. Causes of errors

There are many potential sources of errors which affect the accuracy and validity of the final results. Many of the errors are related to the quality of the initial data itself, while some others are attributed to the accuracy of response surfaces. Section 7.3.1 discusses the main sources of errors in the initial pool of data and their overall effects on the results, while Section

7.3.2 explains the influence of the inaccuracy of the meta-models (response surfaces) on the newly designed molecules.

### 7.3.1. Quality of the data used in this study

The values of the properties (normal boiling temperature, vapor pressure, enthalpy of vaporization, half-life time in the troposphere and biodegradability) of the 295 compounds were either estimated using a suitable prediction model or compiled from the experimentally measured values as published in peer reviewed journals.

The estimations were made using those models which were found to possess relatively higher accuracies for the classes of compounds that have been considered in this study. However, even the best of the prediction models produce results which are not 100% accurate.

The biodegradability was estimated using the models which are based on the group contribution methodology. It is highly likely that not all the chemical groups of significance with respect to the set of 295 compounds were included and represented adequately in the training set of the estimation model. Even if they were represented sufficiently, it is very hard to prepare a training set which provides equal representation to all the unique structural components. Moreover, biodegradability of a compound also depends on the temperature of the surroundings in which the micro-organisms exist. Temperature variations are known to have direct and indirect influence on such biological processes. In most cases, the transformations that micro organisms perform cannot be scaled with temperature by the use of equations such as the Arrhenius equation (equation 7.4).

$$k = A \cdot e^{-E_a/RT} \tag{7.4}$$

Here, k and T represent the rate constant of the chemical reaction and the absolute temperature, respectively. $E_a$ and A represent the activation energy and the pre-exponential factor, respectively. R represents the gas constant.

Errors in estimating the half-life of compounds in troposphere also creep up because of the similar reasons. Discrepancies in experimentally measured values of the tropospheric half-life may exist because of the human error, mechanical errors in the devices used for measurements and due to any differences in the techniques used for measurement. Moreover, during the estimation of the half-life of an organic compound in the troposphere it was assumed that the hydroxyl radical concentration is constant. However, there are variations in the hydroxyl radical concentration depending on the time of the day and the month of the year (season).

The SPARC model used for the estimation of vapor pressure in this study also induces additional errors in the data set. The root mean squared error value of the vapor pressure model used in SPARC has already been mentioned as 0.096, earlier in Section 4.4.2. Figure 7.14 and Figure 7.14 are shown as examples of compounds for which a difference between the vapor pressures estimated by SPARC and those listed in the NIST experimental database (Kazakov, *et al.* 2009), was found. It should be noted that the difference in the vapor pressures is more significant only at higher temperatures. In general, these vapor pressure differences are found to be relatively small inside the temperature range considered as the operating temperature range of the refrigeration system.

Errors were also observed in the estimation of the enthalpies of vaporization when they were estimated using the SPARC online calculator. The root mean squared value of the difference between the experimentally observed values and the estimated values of the enthalpy of vaporization of the organic compounds used in the development of SPARC has been reported to be 0.41 (Hilal, Karickhoff and Carreira 2003).

The regression line as computed by Hilal *et al.* (Hilal, Karickhoff and Carreira 2003) is given by equation 7.5. The relative error between the estimated values and the experimentally observed values of the $\Delta H_v$ is given by equation 7.6.

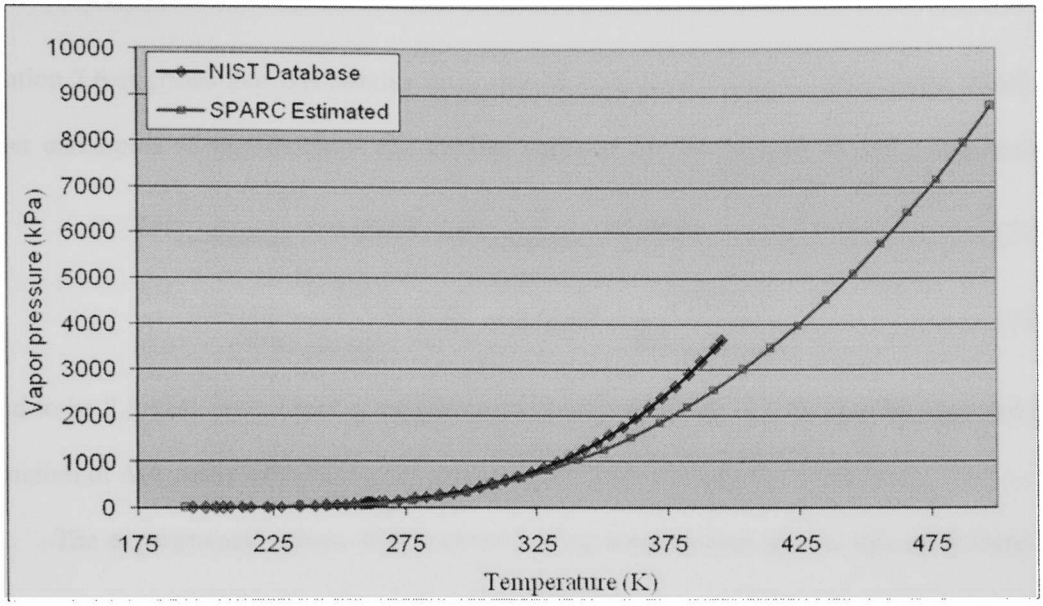



**Figure 7.14:** Comparison of the SPARC estimated vapor pressure against NIST experimental database listed vapor pressure for 2-chloro-1,1,1,2-tetrafluoroethane

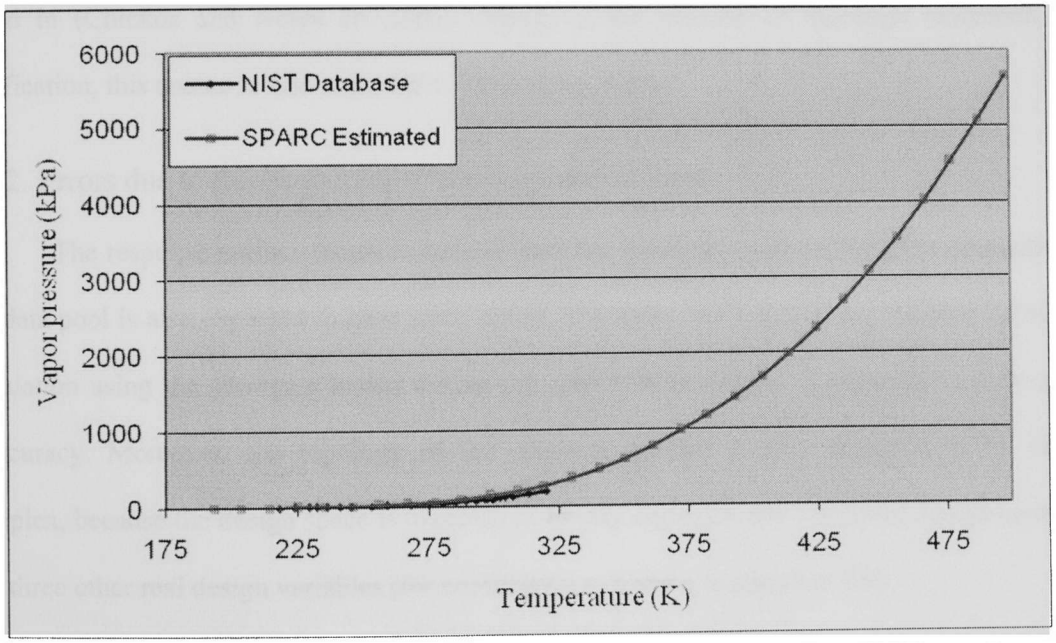



**Figure 7.14:** Comparison of the SPARC estimated vapor pressure against NIST experimental database listed vapor pressure for 2-bromo-1,1,1-trifluoroethane

Equation 7.6 suggests that this relative error should decrease for organic compounds which have higher enthalpies of vaporization. The median value of $\Delta H_v$ for the set of 265 compounds was

$$\Delta H_{v_{estimated}} = 1.0563 \cdot \Delta H_{v_{observed}} - 0.7061 \quad (7.5)$$

$$\frac{\Delta H_{v_{estimated}} - \Delta H_{v_{observed}}}{\Delta H_{v_{observed}}} = 0.0563 - \frac{0.7061}{\Delta H_{v_{observed}}} \quad (7.6)$$

found to be 7.5 kcal/ mol. Thus, a median relative error of about -3.79% may be expected in the estimation of $\Delta H_v$ using SPARC.

The experimental values of the normal boiling temperatures used in this study were listed in a peer reviewed publication (Horvath 2001). However, the normal boiling temperatures listed in this publication were extracted from other original documents without any corrections or validations. In the absence of any verification, this source is bound to have some errors. Those values of the enthalpy of vaporization which were not estimated using SPARC calculator were listed in (Chickos and Acree Jr. 2003). Again, in the absence of thorough validation and verification, this source is also expected to have some errors.

### 7.3.2. Errors due to the inaccuracy of the response surfaces

The response surface which is created from the information about the 295 compounds in the data pool is also expected to have some errors. Therefore, the result of any objective function evaluation using the surrogate model during the optimization process is expected to have some inaccuracy. Moreover, the topology of the response surface is also expected to be highly complex, because the design space is made up of twenty eight discrete structural design variables and three other real design variables (for coefficients a, b and c in equation 4.9).

## 7.4. Conclusion

Multi-objective design optimizations were successfully performed using an initial pool of data, constraints which also ensure the structural feasibility of any generated molecule and meta models. It was found that the methodology successfully generated several candidate solutions which were generally found better than Freon-12 in one or more objectives. Errors between the property predictions using IOSO software and those obtained through property prediction models used in this study were also calculated. The inaccuracies of the property prediction models are also partially responsible for these relatively larger errors, apart from the different sources of inaccuracies present in this methodology.

# 8. SCOPE FOR FUTURE WORK

Many properties and constraints which are not included in this study might be considered important for designing a new molecule. For example, a constraint on the Gibb's free energy of formation, at a particular temperature, can be used to ensure the thermodynamic feasibility of a compound. For a molecule to be feasible at a particular temperature, the Gibb's free energy of formation, $G_f$, at that temperature should be less than zero. Mathematically, for obtaining a stable compound

$$G_f(T) < 0 \tag{8.1}$$

Although, this constraint ensures the feasibility of the compound, it may eliminate those compounds which are not stable at the particular temperature, but have very slow dissociation rates. In other words, some of these compounds may exist without considerable dissociation, long enough for them to be useful as commercial refrigerants. Thus, the study of kinetics may be involved in further studies.

It is expected that even with a slight increase in the number of elements that could constitute the molecule of a new refrigerant, the number of ways in which these elements could be combined to form the molecule would increase very rapidly. This combinatorial explosion may drastically increase the computational time to solve the problem. It has been reported in the past (Harper, *et al.* 1999) that a multilevel generate-and-test approach may be used to address this issue of combinatorial explosion and to make the whole optimization process more robust, flexible and computer efficient. With an increase in the size of the molecules, the present methodology would need to tackle the problem of combinatorial explosion. A multi-level approach used with the methodology discussed may prove to be a better overall approach.

In the absence of reliable experimentally measured ODP values of a large number of compounds, expressions for estimating the ODP by utilizing the number of abstractable hydrogen atoms, number of chlorine/bromine/iodine atoms along with their positions in the molecule need

to be used. Similar expressions have been reported previously (Duvedi and Achenie 1996). However, the application of these expressions has been limited because the expressions were developed for a small family of organic compounds with a number of restrictions on the type of atoms that could constitute the organic molecule.

Freezing temperature and the liquid/ vapor phase change specific heat of a particular refrigerant are of much importance in the evaluation of the performance of the refrigeration cycle. Due to the lack of available experimental/ estimated data for a large number of promising organic compounds, these properties were not considered in this study. It would be highly desirable to search a credible and accurate model which could estimate the liquid specific heat of compounds and integrate the estimated values in the form of another objective.

The developed methodology can handle only those organic compounds which have up to four carbon atoms. In order to handle compounds containing a larger number of carbon atoms, the methodology would need to be modified. Apart from the number of bonds of a particular kind associated with each carbon atom, the way in which carbon atoms themselves are arranged in the molecule would also need to be addressed. In addition to this, a number of constraints would need to be modified, since some constraints might become irrelevant and some others might need to be formulated from scratch.

With an inclusion of a number of additional elements such as nitrogen, boron, phosphorous, *etc.*, the complexity of the design problem should increase significantly. In many such situations, the octet rule would not remain valid. Odd electron species (such as nitrogen oxide NO, nitrogen dioxide $NO_2$, *etc.*) and compounds constituted by elements having incomplete octets (such as boron trifluoride $BF_3$) and/ or expanded octets (such as phosphorous pentachloride $PCl_5$) would need to be considered. Although some of these compounds may be toxic, flammable and unsuitable for use as a refrigerant due to other reasons, an ideal compound may belong to the category of organic compounds in which one or more atoms violate the octet rule.

It is expected that in the immediate future, new refrigerants will be required to have low GWP. Atmospheric lifetimes of organic compounds have been directly/ indirectly linked to their GWP. Incorporation of the GWP as a separate objective should be given a high priority in any of the future work related to the design optimization of new refrigerants.

Many of the currently used refrigerants are also mixtures of two or more organic compounds. In many cases, it has been found that mixtures when used as refrigerants perform better than any of their constituting components. An attempt to find the optimal mixture with two or more constituents would require a considerable change in the methodology used in this study. Consequently, such attempts may result in the generation of better refrigerants that may be obtained by the methodology suggested in this study.

# REFERENCES


[1] Achenie, L. E. K., R. Gani, and V. Venkatasubramanian. *Computer Aided Molecular Design: Theory and Practice.* Amsterdam: Elsevier, 2003.

[2] Bahm, K, and M. A. K. Khalil. "A new model of tropospheric hydroxyl radical concentrations." *Chemosphere* (Elsevier) 54 (2004): 143-166.

[3] Basarova, P., and V. Svoboda. "Prediction of the enthalpy of vaporization by the group contribution method." *Fluid Phase Equilibria* (Elsevier), no. 105 (1995): 27-47.

[4] Bhattacharjee, S. "Geometric volume and new refrigerants - II. Haloethanes." *Computers and Chemical Engineering* 19, no. 1 (1995): 51-58.

[5] Boethling, R. S., P. H. Howard, W. Meylan, J. Stitleler, J. Beauman, and N. Tirado. "Group contribution method for predicting probability and rate of aerobic biodegradation." *Environmental Science & Technology* 28 (1994): 459-465.

[6] Brown, J. S. "HFOs: New, low global warming potential refrigerants." *ASHRAE Journal*, August 2009: 22-29.

[7] Calm, J. M. "The next generation of refrigerants - Historical review, considerations, and outlook." *International Journal of Refrigeration* (Elsevier) 31 (2008): 1123-1133.

[8] Calm, J. M., and D. A. Didion. "Trade-offs in refrigerant selection: Past, present and future." *International Journal of Refrigeration* 21, no. 4 (1998): 308-321.

[9] Carr, S., D. E. Heard, and M. A. Blitz. "Comment on "Atmospheric hydroxyl radical production from electronically excited NO2 and H20"." *Science* 324 (2009): 336b.

[10] Chickos, J. S., and W. E. Acree Jr. "Enthalpies of vaporization of organic and organometallic compounds, 1880-2002." *Journal of Physical and Chemical Reference Data* 32, no. 2 (2003): 519-878.

[11] Cholakov, G. S., W. A. Wakeham, and R. P. Stateva. "Estimation of normal boiling points of hydrocarbons from descriptors of molecular structure." *Fluid Phase Equilibria* (Elsevier), no. 163 (1999): 21-42.

[12] Churi, N., and L. E. K. Achenie. "Novel mathematical programming model for computer aided molecular design." *Industrial and Engineering Chemistry Research* (ACS Publications) 35 (1996): 3788-3794.

[13] Colaco, M. J., G. S. Dulikravich, and D. Sahoo. "A comparison of two methods for fitting high dimensional response surfaces." *Inverse Problems, Design and Optimization Symposium.* Miami, 2007.

[14] Duvedi, A. P., and L. E. K. Achenie. "Designing environmentally safe refrigerants using mathematical programming." *Chemical Engineering Science* 51, no. 15 (1996): 3727-3739.

[15] Duvedi, A., and L. E. K. Achenie. "On the design of environmentally benign regrigerant mixtures: a mathematical programming approach." *Computers & Chemical Engineering* 21, no. 8 (1997): 915-923.

[16] Egorov, I. N., G. V. Kretinin, I. A. Leshchenko, and S. V. Kuptzov. "Multi-objective approach for robust design optimization problems." *Inverse Problems in Science and Engineering* 15, no. 1 (2007): 47-59.

[17] Egorov-Yegorov, I. N., G. V. Kretinin, I. A. Leshchenko, and S. V. Kuptcov. "Multi-objective robust optimization using IOSO technology Part 1: Main features." *Fifth Conference on Evolutionary Methods for Design, Optimization and Control Applications to Industrial and Societal Problems.* 2003. 120-121.

[18] Fermeglia, M., M. Ferrone, and S. Pricl. "Development of an all atom force field from ab initio calculations for alternative refrigerants." *Fluid Phase Equilibria* (Elsevier) 210 (2003): 105-116.

[19] Gani, R. "Case Studies in Chemical Product Design - Use of CAMD techniques." Chap. 14 in *Chemical product design: Toward a perspective through case studies*, by R Gani, K Dam-Johansen Ka M. Ng. Elsevier B.V., 2007.

[20] Granryd, E. "Hydrocarbons as refrigerants - an overview." *International Journal of Refrigeration* 24 (2001): 15-24.

[21] Halimic, E., D. Ross, B. Agnew, A. Anderson, and I. Potts. "A comparison of the operating performance of alternative refigerants." *Applied Thermal Engineering* 23 (2003): 1441-1451.

[22] Harper, P. M., R. Gani, P. Kolar, and T. Ishikawa. "Computer aided molecular design with combined molecular modeling and group contribution." *Fluid Phase Equilibria* 158-160 (1999): 337-347.

[23] Hilal, S. H., S. W. Karickhoff, and L. A. Carreira. "Prediction of the vapor pressure, boiling point, heat of vaporization, and diffusion coefficient of organic compounds." *QSAR & Combinatorial Science* 22 (2003): 565-574.

[24] Horvath, A. L. "Boiling points of halogenated organic compounds." *Chemosphere*, no. 44 (2001): 897-905.

[25] Joback, K. G., and G. Stephanopoulos. "Designing molecules possessing desired physical property values." *Foundations of Computer Aided Process Design.* 1989.

[26] Kazakov, A. F., C. D. Muzny, R. D. Chirico, V. Diky, R. A. Stevenson, and M. Frenkel. "NIST/ TRC Web Thermo Tables (WTT)." *NIST Standard Reference Subscription Database 3.* Professional. Edited by R. D. Chirico, V. V. Diky, D. G. Friend, A. Laesecke, E. W. Lemmon, J. W. Magee, C. D. Muzny, R. A. Perkins, R. A. Stevenson M. Frenkel. Thermodynamics Research Center (TRC), Physical and Chemical Properties Division. Boulder, CO, 2009.

[27] Khetib, Y., A. H. Meniai, and A. Lallemand. "Computer-aided design of CFC and HCFC substitutes using group contribution methods." *Desalination* 239 (2009): 82-91.

[28] Kopko, W. L. "Beyond CFCs: Extending the search for new refrigerants." *International Journal of Refrigeration* 13, no. 2 (March 1990): 79-85.

[29] Li, H., H. Higashi, and K. Tamura. "Estimation of boiling and melting points of light, heavy and complex hydrocarbons by means of a modified group vector space method." *Fluid Phase Equilibria* (Elsevier), no. 239 (2006): 213-222.

[30] Li, P., Y. H. Liang, P. S. Ma, and C. Zhu. "Estimations of enthalpies of vaporization of pure compounds at different temperatures by a corresponding-states group contribution method." *Fluid Phase Equilibria* (Elsevier), no. 137 (1997): 63-74.

[31] Li, S., J. Matthews, and A. Sinha. "Atmospheric hydroxyl radical production from electronically excited NO2 and H20." *Science* 319 (2008): 1657-1660.

[32] Marrero, J., and R. Gani. "Group contribution based estimation of pure component properties." *Fluid Phase Equilibria*, 2001: 183-208.

[33] Martin, T. M., and D. M. Young. "Prediction of the acute toxicity (96-h LC50) of organic compounds to the fathead minnow (Pimephales promelas) using group contribution method." *Chemical Research in Toxicology* 14 (2001): 1378-1385.

[34] Michalewicz, Z. "Evolutionary algorithms in engineering optimization." In *Evolutionary Algorithms and Intelligent Tools in Engineering Optimization*, by W. Annicchiarico, J. Periaux, M. Cerrolaza and G. Winter, 35-37. Barcelona: WIT Press, 2005.

[35] Mohanraj, M., S. Jayaraj, and C. Muraleedharan. "Environment friendly alternatives to halogenated refrigerants - A review." *International Journal of Greenhouse Gas Control* 3 (2009): 108-119.

[36] Moran, M. J., and H. N. Shapiro. *Fundamentals of engineering thermodynamics.* 4th Edition. John Wiley & Sons, 2000.

[37] Nannoolal, Y., J. Rarey, D. Ramjugernath, and W. Cordes. "Estimation of pure component properties Part 1. Estimation of the normal boiling point of non electrolyte organic compounds via group contributions and group interactions." *Fluid Phase Equilibria* (Elsevier), no. 226 (2004): 45-63.

[38] Pavan, M., and A. P. Worth. "Review of QSAR models for ready biodegradation." Toxicology and Chemical Substances Unit, European Chemicals Bureau, Institute for Health and Consumer Protection , Ispra (VA), Italy, 2006.

[39] Pfeiffenberger, U. "Comparison of refrigerants." *International Journal of Refrigeration*, 1982: 74-78.

[40] Poling, B. E., J. M. Prausnitz, and J. P. O'Connell. *The properties of gases and liquids.* 5th Edition . The McGraw-Hill Companies, 2004.

[41] Sablijic, A., and W. Peijnenburg. "Modeling lifetime and degradability of organic compounds in air, soil and water systems." *Pure and Applied Chemistry* (IUPAC) 73, no. 8 (2001): 1331-1348.

[42] Sahinidis, N. V., and M. Tawarmalani. "Applications of global optimization to process and molecular design." *Computers and Chemical Engineering* 24 (2000): 2157-2169.

[43] Sahoo, D., and G. S. Dulikravich. "Evolutionary wavelet neural network for large scalefunction estimation in optimization." *11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference.* Portsmouth, 2006.

[44] Saleh, B., and M. Wendland. "Screening of pure fluids as alternative refrigerants." *International journal of refrigeration* 29 (2006): 260-269.

[45] Sastri, S. R.S., S. Mohanty, and K. K. Rao. "A new method for predicting critical volumes of organic compounds." *Fluid Phase Equilibria* (Elsevier) 129 (1997): 49-59.

[46] Sekiya, A., and S. Misaki. "The potential of hydofluorothers to replace CFCs, HCFCs and PFCs." *Journal of Fluorine Chemistry*, no. 101 (2000): 215-221.

[47] Sigma Technology. "IOSO NM version 1.0 Multiobjective non-linear optimization software." *Reference Manual.* Moscow.

[48] Sigma-Technology Company. "Approx-Version 1." Moscow, August 2009.

[49] U.S. Environmental Protection Agency. *Atmospheric oxidation program for microsoft windows (AOPWIN) v 1.92a.* June 2008.

[50] U.S. Environmental Protection Agency. *BIOWIN v 4.10.* March 2009.

[51] *SMILES Tutorial.* February 2009. http://www.epa.gov/med/Prods_Pubs/smiles.htm (accessed December 2009).

[52] Walters, A. E., P. B. Myrdal, and S. H. Yalkowsky. "A method for estimating the boiling points of organic compounds from their melting points." *Chemosphere* 31, no. 4 (1995): 3001-3008.

[53] Wang, Q., P. Ma, C. Wang, and S. Xia. "Position group contribution method for predicting the normal boiling point of organic compounds." *Chinese Journal of Chemical Engineering* 17, no. 2 (2009): 254-258.

[54] Zitzler, E. "Evolutionary algorithms for multiobjective optimization." In *Evolutionary Methods for Design, Optimization and Control*, by K. C. Giannakoglou, D. T. Tsahalis, J. Periaux, K. D. Papailiou and T. Fogarty, 19-20. Barcelona: International Center for Numerical Methods in Engineering (CIMNE), 2002.

# APPENDICES

## APPENDIX A – SMILES Notation

SMILES (Simplified molecular input line entry system) is a way of representing the structure of a chemical compound, which is compatible with many of the computer codes (U.S. Environmental Protection Agency 2009). SMILES requires five basic rules of Chemistry to be observed. They are the following:

1) Atoms and bonds

SMILES notation supports all the elements of the periodic table. Any atom is represented by its respective atomic symbol. If this symbol contains two letters, the first letter would be upper case and the second letter should be in lower case. Letters in upper case are used to represent non-aromatic atoms. Lower case atoms represent aromatic atoms. A SMILES string is terminated by a blank. Bonds are represented as:

- \- single bond
- = double bond
- # triple bond
- * aromatic bond
- . disconnected structures

2) Simple Chains

Simple chain structures are represented by using a combination of atomic and bond symbols. Hydrogen bonds may or may not be explicitly specified. However, a presence of a Hydrogen bond in the SMILES string requires the explicit identification of all the Hydrogen atoms in the molecule.

3) Branches

A branch is specified by placing the SMILES string corresponding to the branch inside parenthesis. The string is placed directly after the symbol of the atom to which the branch is attached and the left parenthesis is followed by the appropriate bond symbol.

4) Rings

Ring structures are identified in SMILES notation by using numerals. The numerals are used to identify the opening and closing ring atom. In case the chemical compound contains multiple rings, individual rings are assigned different numbers. The bond symbol before the ring closure number identifies the type of bond.

5) Charged Atoms

Charged atoms, if any, are identified by placing the charge inside brackets and using this directly after the symbol of the atom which is charged.

APPENDIX B – Zonally and monthly averaged hydroxyl radical concentrations ($10^5$ mol/cm$^3$)

| Latitude | December | March | June | September | Annual average | Hemisphere average |
|---|---|---|---|---|---|---|
| **85°N** | - | - | 6.3 | 0.1 | 1.6 | |
| **75°N** | - | 0.1 | 8.1 | 0.4 | 2.1 | |
| **65°N** | - | 0.4 | 10.3 | 1.2 | 3.0 | |
| **55°N** | 0.0 | 1.3 | 10.9 | 2.9 | 3.8 | |
| **45°N** | 0.5 | 3.4 | 14.9 | 6.1 | 6.2 | 9.8 NH |
| **35°N** | 2.3 | 6.9 | 18.7 | 10.6 | 9.6 | |
| **25°N** | 5.8 | 11.6 | 17.9 | 13.3 | 12.1 | |
| **15°N** | 8.8 | 13.6 | 16.7 | 14.9 | 13.5 | |
| **5°N** | 13.2 | 17.0 | 16.0 | 18.1 | 16.0 | |
| **5°S** | 16.9 | 15.8 | 12.6 | 19.7 | 16.3 | |
| **15°S** | 17.4 | 12.4 | 8.3 | 15.6 | 13.4 | |
| **25°S** | 15.5 | 9.7 | 4.7 | 11.2 | 10.3 | |
| **35°S** | 11.0 | 6.6 | 2.3 | 6.4 | 6.6 | |
| **45°S** | 7.7 | 4.1 | 0.8 | 3.8 | 4.1 | 8.5 SH |
| **55°S** | 5.1 | 2.1 | 0.1 | 1.9 | 2.3 | |
| **65°S** | 3.5 | 0.8 | - | 0.6 | 1.2 | |
| **75°S** | 4.0 | 0.2 | - | 0.1 | 1.1 | |
| **85°S** | 3.1 | 0.0 | - | - | 0.8 | |
| | | | | | | **Global Average** |
| **Average** | 8.4 | 8.3 | 10.0 | 9.9 | | 9.2 |