There is more to Suite: Estimating partition coefficients for biotic phases using descriptors from the Estimation Program Interface (EPI) Suite^{TM}

By

Syed Anas Rehman

(Reg# 00000204913)

A thesis submitted in partial fulfillment of requirements for the degree of
Master of Science
In
Environmental Science

Institute of Environmental Sciences and Engineering (IESE)

School of Civil and Environmental Engineering (SCEE)

National University of Sciences and technology (NUST)

Islamabad, Pakistan

(2019)

THESIS ACCEPTANCE CERTIFICATE

It is certified that the copy of MS/MPhil thesis entitled by Mr. Syed Anas Rehman Registration No. 00000204913 of **IESE (SCEE)** has been vetted by undersigned, found complete in all aspects as per NUST Statutes/Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members have also been incorporated in the said thesis.

Signature with stamp: ____________________

Name of the Supervisor: Dr. Deedar Nabi

Date: __________________

Signature of HoD with stamp: __________________

Date: __________________

Countersign by

Signature (Dean/Principal): __________________

Date: __________________

CERTIFICATE

It is certified that the contents and form of the thesis entitled

“There is more to Suite: Estimating partition coefficients for biotic phases using descriptors from the Estimation Program Interface (EPI) Suite$^{\mathrm{TM}}$”

Submitted by:

Syed Anas Rehman

have been found satisfactory for the requirement of the degree

Supervisor: ____________________

Dr. Deedar Nabi

Assistant Professor

IESE, SCEE, NUST

Member: _______________________

Dr. Muhammad Arshad

Assistant Professor

IESE, SCEE, NUST

Member: _______________________

Dr. Tahir Mehmood

Assistant Professor

SNS, NUST

DEDICATION

This thesis is dedicated to my affectionate parents.

Acknowledgements

This research work would not have been possible without the guidance of Allah Almighty, the most beneficent and merciful.

I would like to express my utmost gratitude to my Supervisor, Dr. Deedar Nabi for his understanding, wisdom and patience and for pushing me farther than I thought I could go. I would also like to express my gratitude to my GEC members, Dr. Muhammad Arshad and Dr. Tahir Mehmood. Their guidance and expertise generously helped me during my research.

My siblings, friends and research group fellows for helping me in stressful times and not letting me give up.

Lastly, I would like to thank my parents for their unconditional love, support and encouragement.

Table of Contents

List of Figures viii
List of Abbreviations ix
ABSTRACT x
Chapter 1 1
- 1. Introduction 1
  - 1.1. Background 1
  - 1.2. Problem statement 6
  - 1.3. Hypothesis of the study 6
  - 1.4. Objectives of the study 6
  - 1.5. Scope of the study 6
  - 1.6. Significance of the study 7

Chapter 2 8
- 2. Literature Review 8
  - 2.1. Equilibrium partition coefficient 8
  - 2.2. Water and Octanol partition coefficient 9
  - 2.3. Partition coefficient for the Water and Organic Carbon 10
  - 2.4. Biological Phase 11
  - 2.5. Natural Organic Matter 12
  - 2.6. Enthalpy of vaporization 13
  - 2.7. Experimental work 14
  - 2.8. Passive sampling 15
  - 2.9. GCxGC 15
  - 2.10. Abraham Solvation Model 16
  - 2.11. One Parameter-Linear Free Energy Relationship 17
  - 2.12. Estimation method 18
  - 2.13. Knowledge Gap 19

Chapter 3 20
- 3. Methodology 20
  - 3.1. Data acquisition 20
  - 3.2. Statistical analysis 21

Chapter 4 22
- 4. Results and Discussion 22
  - 4.1. Biological phases 22

4.2. NOM Phase 28
4.3. Enthalpy of solvation 30
Chapter 5 33
5. Conclusion and Recommendation 33
5.1. Conclusions 33
5.2. Recommendations 33
References 34

List of Figures

Figure 1.1 - The triad of phases considered in linear free energy relationships. Arrows indicate partitioning equilibria amid various phases 5

Figure 2.1 Schematic of the GC×GC configuration. 1° and 2° denote primary and secondary, respectively 16

Figure 4.1 Experimental versus predicted log K lipid-water 22

Figure 4.2 Biplot of ASM 24

Figure 4.3 Experimental vs predicted log K liposome-water 25

Figure 4.4 Experimental vs predicted log K protein-water 26

Figure 4.5 Experimental vs predicted air-olive oil 27

Figure 4.6 Experimental vs predicted DOC-water 28

Figure 4.7 Experimental vs predicted NOM-air 29

Figure 4.8 Experimental vs predictive log K air-octanol 30

Figure 4.9 Experimental vs predictive log K air-water 31

List of Abbreviations

LRTP	Long Range Transport Potential
UNEP	United Nations Environmental Program
ECHA	European Chemical Agency
DOM	Dissolved Organic Matter
NOM	Natural Organic Matter
LFER	Linear Free Energy Relationship
OP-LFER	One Parameter-Linear Free Energy Relationship
POPs	Persistent Organic Pollutants
PDMS	Polydimethylsiloxane
ASM	Abraham Solvation Model
QSAR	Quantitative Structure-Activity Relationship
PP-LFER	Poly Parameter-Linear Free Energy Relationship
Kow	Octanol-Water Partition Coefficient
Kaw	Air-Water Partition Coefficient
Koc	Organic Carbon-Water Partition Coefficient
PCA	Principle Component Analysis
AIC	Akaike Information Criterion
LOO	Leave One Out

Abstract

In this thesis, two-parameter Linear Free Energy Relationships (2p-LFERs) were developed for diverse set of environmentally-relevant biotic and abiotic phases, and enthalpy of solvation phases. Biological phases evaluated in this study include lipid-water, air-olive oil, liposome-water, and structural protein-water systems. Abiotic phases comprise of dissolved organic carbon (DOC)-water, and natural organic matter (NOM)-air systems. Two systems, air-octanol and air-water were considered for the modeling of enthalpy of phase transfer.

For biological phases, new 2p-LFER models successfully explained the partitioning variability in the datasets with $R^2 = 0.870–0.954$ and root mean square error (RMSE) = 0.435–0.300 log units. For organic carbon systems, my models exhibited R^2 in the range of 0.752–0.958 and RMSE from 0.505 to 0.343. My model explained the variability in enthalpy of solvation data with $R^2 = 0.896$ to 0.733 and RMSE from 9.84 to 7.29. All models developed in this study were tested for various criteria of internal and external validity. These tests show the models are statically robust.

The new models provide the mechanistic insight about how the attributes of hydrophobicity and volatility −octanol-water partition coefficient (Kow) and air-water partition coefficient (Kaw)–control the transfer of contaminants from one phase to another.

My models have an edge over the existing estimation approaches because the previous models are either parameter intensive, or parameters are not readily available for many compounds or in case of one-parameter LFERs, the estimation approach is less accurate. Also, my model is computationally cheap and can easily be integrated in EPI Suite^{TM}, which modules to estimate several environmental properties. Lastly, my model also provides mechanistic understanding about different parameters that control the transfer of pollutants from one phase to another.

Chapter 1

1. Introduction

1.1. Background

Organic chemicals equilibrium partitioning properties stimulate their transportation and circulation among various phases of the environment such as air, soil, water and other natural organic matter(Schwarzenbach *et al.*,, 2002; F. Wania & Mackay, 1999). In order to evaluate their behavior in the environment and the effects of hydrophobic pollutants on the ecology, fate models typically work with equilibrium partitioning properties. They work as input parameters, which are defined as

$$P_{xy,i} = \left\{\frac{C_{x,i}}{C_{y,i}}\right\}_{\text{equilibrium}}$$

Where:

$P_{xy,i}$ = Partition coefficient amid *x* and *y* phases.

$C_{x,i}$, $C_{y,i}$ = Concentrations of contaminant *i* in the two phases, at partitioning equilibrium.

Unfortunately, such partitioning data often have not been experimentally measured and at times they may have significant uncertainties in their results(Schenker *et al.*,, 2005). For example, fairly less data regarding partitioning property, exist for new and recently known pollutants (Daughton, 2004; Howard & Muir, 2010). Moreover, complex fusions may be found in the environment samples, which may contain hundreds or thousands of different organic chemicals, acting as contaminants (Booth *et al.*,, 2006; Korytár *et al.*,, 2005). Many of these contaminants remain uncharacterized. Therefore, mostly it remains uncertain as how to assess the bio-accumulation potential of these chemicals and to parameterize transport models describing inter-phase transfers for these fusions. By addressing these problems, the methods will aid the associated people in their work, intended to assess partitioning properties for the organic pollutants that are detected in those fusions.

Modern societies gain many benefits from numerous synthetic organic chemicals. However, this chemical revolution on one hand is the cause of human domination but on the other hand it leads to the destruction of earth's ecosystems (Vitousek *et al.*,, 1997). Many of these synthetic chemicals are hazardous to both humans and the environment, even in concentrations as low as 1 part per billion (Vitousek *et al.*,, 1997) and are therefore denoted as micro-pollutants. The common properties of many of these chemicals are:

- Persist in the environment
- Toxic to a lot of species
- Bio-accumulate in organisms
- Due to their long-range transport potential (LRTP), they can end up in pure ecosystems around the globe.

Hence, there is a clear need to monitor chemicals that have these properties. International and national regulatory agencies such as the United Nation Environmental Programme (UNEP), Environment Canada, the European Chemicals Agency (ECHA) and the United States Environmental Protection Agency (US EPA) use these attributes for prioritizing these chemicals for monitoring and risk assessment. However, different agencies use different cut-off values for PBT and LRTP (Gawor & Wania, 2013).

The challenge of chemical monitoring and risk assessment is critical, because there are millions of different chemicals accumulated in the environment. Even with an inventory of more than 65 million already registered chemicals and a registration rate of one new chemical per second, there are still many more chemicals to assess(Hendriks, 2013). Human beings and 8 million other species face different levels of risk from these chemicals. There is a dire need of an integrated approach in order to deal with this massive challenge.

Furthermore, many of these chemicals exist in the form of complex mixtures, which makes risk assessment and monitoring even more challenging. These mixtures often include chemical families such as mono-aromatic hydrocarbons, (Scarlett *et al.*,, 2009) dibenzofurans (PHDFs), polychlorinated naphthalene (PCNs), (Walters *et al.*,, 2011) polychlorinated biphenyls (PCBs), (Walters *et al.*,, 2011) polybrominated diphenyl ethers (PBDEs) (Hites, 2004), toxaphene(Xia *et al.*,, 2012), polyhalogenated dibenzo-*p*-dioxins (PHDDs) (van den Berg *et al.*,, 2013), polychlorinated *n*-alkanes (PCAs) (Castells *et al.*,, 2008), and polycyclic aromatic hydrocarbons (PAHs)(Manzano *et al.*,, 2012). Amalgam of these mixtures further contain thousands of compounds, and their diverse physicochemical properties regulate the fate, transport and behavior of these complex mixtures in the ecosystem(van Leeuwen & de Boer, 2008). Once they accumulate in the environment, these families due to LRTP, these mixtures can even be found in unspoiled environments such the arctic regions(Muir & de Wit, 2010). Due to their transport these mixtures may be partly captured at high altitudes by mountain ranges, such as the Himalayas(Wang *et al.*,, 2006) and the Alps (Schmid *et al.*,, 2007), and bio-accumulate in these ecosystems (Schmid *et al.*,, 2007; Shunthirasingham *et al.*,, 2013)as a result of a process called mountain cold-trapping(Frank Wania & Westgate, 2008).

In addition to the complexity of the chemicals, the ecosystems in which they are transported and accumulated are also complex. Different environmental phases interact with each other in a complex manner. However, at chemical equilibrium, the distribution of mixtures among different phases is due to their thermodynamic properties. This is defined by equilibrium partitioning theory. There are many conditions in the environment where it is appropriate to consider equilibrium, for example, the distribution of contaminants among the three phases of the sediment-water system: the pore water, dissolved organic matter and amorphous organic carbon portions of the sediment. Even in situations when equilibrium is not attained, knowing the equilibrium partition coefficient is very useful in order to know a chemical's tendency for accumulation in a specific phase (Schwarzenbach *et al.*,, 2002). Equilibrium partition coefficients also tells us the exact direction in which the chemical moves in a complex environment. They also calculate the rate of mass transfer across different interfaces (Schwarzenbach *et al.*,, 2002). However, the main problem is that the experimental partitioning data for majority of these compounds, in environmental phases, does not exist.

In order to overcome the scarcity of experimental property data, scientists assume different estimation strategies, but results usually have different levels of accuracy. Equilibrium partitioning properties are properties which are linked to free energy. In environmental chemistry, it is a common practice to relate and predict one free energy related property to another. These correlations are known as linear free energy relationships (LFERs) (Goss & Schwarzenbach, 2000). Correlations between two partitioning properties are likely if the same kind of interactions manage both properties (Goss & Schwarzenbach, 2003). In general, partition coefficients for three phases — air, water and octanol — are used as substitutions to estimate the partitioning properties for other similar phases (Figure 1). Here, octanol is considered as a substitute phase for natural and biological organic phases such as whole organisms, carbohydrates, amorphous organic carbon, lipids, proteins, and dissolved organic carbon.

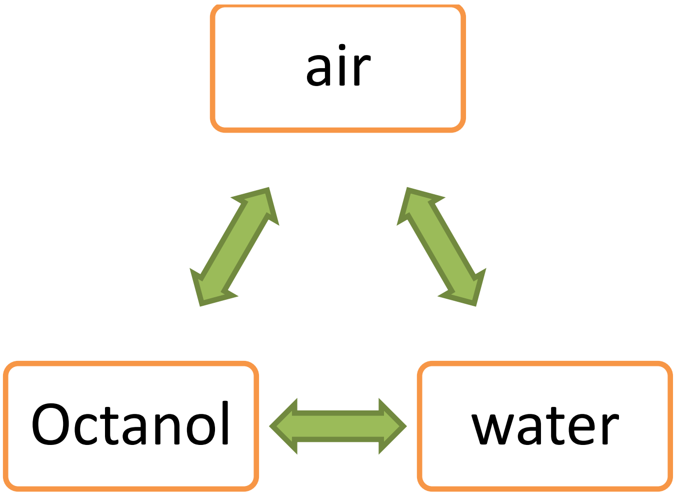

Figure 1.1 - The triad of phases considered in linear free energy relationships. Arrows indicate partitioning equilibria amid various phases

The circulation of complex chemical mixtures among the biotic phases of the environment defines exposure-related concepts such as chemical activity, bio-availability and bio-accessibility and effect-related results like toxicity and bio-accumulation (Reichenberg & Mayer, 2006) under both dynamic (Bayen *et al.*,, 2009) and equilibrium (Schwarzenbach *et al.*,, 2002) conditions. These processes are governed by the complex interaction of partitioning and transport properties and conditions of the environment (EndoBrown *et al.*,, 2013).

In environmental chemistry, the freely dissolved concentration C_{free} of a contaminant is an important factor to evaluate the bio-availability, bio-accumulation, and toxicity (Cui *et al.*,, 2013; Escher & Hermens, 2004; Jahnke *et al.*,, 2008). Dissolved organic matter (DOM) plays the vital role in regulating the freely dissolved concentration of hydrophobic contaminants throughout aquatic environments. DOM reduces the bioavailability and toxicity of hydrophobic chemicals (Haitzer *et al.*,, 1998; Haitzer *et al.*,, 1999). DOM may improve the contaminant's solubility and powers the mobility of contaminants from sediment and soil matrices (ter Laak *et al.*,, 2008).

The aim of this study is to understand different types of inter-molecular interactions, dictating the phase transfer of different media for complex chemicals using simple and low-cost estimation method.

1.2. Problem statement

Partitioning properties are required to do risk assessment. This property data is not readily available and ultimately, we must resort to estimation approaches. Existing estimation approaches are either computationally expensive or requires expensive instrumentation.

1.3. Hypothesis of the study

More than one inter-molecular interaction parameters are required to describe partitioning variability for biotic phases.

1.4. Objectives of the study

- To investigate different types of inter-molecular interactions, dictating the phase transfer for biotic, NOM, and Enthalpy of solvation media.
- To develop, multi-parameter Linear Free Energy Relationship (LFER) for diverse set of biotic and NOM phases.
- To develop, a quasi-LFER model to estimate the enthalpy of phase transfer by exploiting entropy-enthalpy compensation phenomenon of Gibbs Free Energy Change.

1.5. Scope of the study

The research work was divided into three phases:

- In the first phase, two-parameter estimation models were developed for diverse set of biological phases.
- In the second phase, two-parameter estimation models were developed for phases of Natural Organic Matter.

- In the third phase, two-parameter estimation models were developed for enthalpy of solvation media.

1.6. Significance of the study

- This study will create a new estimation method for chemical risk assessment.
- This estimation method can help promote the export and import of chemicals in and from Pakistan required for regulation purpose.
- For risk assessment, a database can be created.
- It will provide a platform to students, researchers, and academia, working in the field of partitioning properties.

Chapter 2

2. Literature Review

2.1. Equilibrium partition coefficient

In a medium when two concentrations are at equilibrium, partition coefficient is basically defined as the ratio of the concentration of a substance present in a media (C_1) to the concentration in other media (C_2). It can be written as, partition coefficient = $(C_1/C_2)_{equil}$. Concentration units can be changed. In case of breathable toxicokinetics, the partition coefficients are air and blood, when we consider gases and the blood and tissue partition coefficient for the rest of the cases. At low concentrations, constant for partitioning between a liquid and gas is known as Henry's law constant. It is written as $H = (C_1/C_2)_{equil}$. Here H denotes vapor and liquid equilibrium. In toxicology, Henry's law is typically applied to the concentration range of interest but deviances happen at higher concentrations. (Schlosser, 2010)

If we take two non-mixable solvents and dissolve an organic compound in them, molecular partitioning takes place. We estimate it by the partition coefficient method, also known as distribution coefficient. Therefore, by finding partition coefficient we basically find the compound's solubility variation in two non-mixable solvents. In most of the cases, water is one of the two solvents and the second are mostly hydrophobic for instance, octan-1-ol. By this method we get to know about the hydrophilic or hydrophobic nature of a substance.

The hydrophobicity of a substance can indicate the relative ease with which that substance will mix in the groundwater and will contaminate it. The above explained method can also be used to already know the possible movement of the substance in the underground water body. In hydrogeology, the water and octanol partition coefficient (K_{ow}) is taken to know and then further model the movement of a hydrophobic organic compounds that gets dissolved in groundwater and is also present in the soil (Speight, 2016).

2.2. Water and Octanol partition coefficient

The frequently used partitioning coefficient in case of liquids is the water and octanol. It is also frequently used in many other media such as atmosphere. As many air pollutants are organic in nature, we can properly estimate and distinguished their dissolving nature. For instance, if a substance can get dissolved easily in organic solvents in a certain medium, under normal conditions of pressure and temperature, then the compound is categorized as lipophilic (soluble in lipids). Alternatively, if it is not soluble in organic solvents then we call it lipophobic (not soluble in lipids).

This estimation about a compound's solubility in lipids and water gives significant information about environmental partitioning; i.e. the octanol–water partition coefficient (K_{ow}). When in equilibrium, it gives the concentration of a compound dissolved in octanol relative to its concentration, dissolved in water. Octanol is an alternate for lipophilic solvents as it cab dissolve in both, water and organic compounds so it is regarded as amphibilic. In this case K_{ow} is represented as $[C_7H_{13}CH_2OH]:[H_2O]$, therefore, as the K_{ow} value increases, it indicates that the compound is more lipophilic (Letcher, 2007).

Table 2.1 shows the values of water solubility, their densities and K_{ow} values of a few prominent chemical compounds.

Table 2.1 - Solubility, Octanol–Water Partitioning Coefficient, and Density Values for Some

Chemical	**Water Solubility ($mg\ l^{-1}$)**	**K_{ow}**	**Density ($kg\ m^{-3}$)**
Atrazine	33	724	
Benzene	1780	135	879
Chlorobenzene	472	832	1110
Cyclohexane	60	2754	780
1,1-Dichloroethane	4960	62	1180
1,2-Dichloroethane	8426	30	1240
Ethanol	Completely miscible	0.49	790
Toluene	515	490	870
Vinyl chloride	2790	4	910
Tetrachlorodibenzo-*para*-dioxin (TCDD)	1.9×10^{-4}	6.3×10^{6}	

2.3. Partition coefficient for the Water and Organic Carbon

Partition coefficient for the water and organic carbon is a non-dimensional number which can be explained as the ratio of a compounds' concentration absorbed per unit mass of soil, to its concentration in the aqueous phase. It denotes the distribution coefficient (K_d) normalized to total organic matter. Mostly the Persistent Organic Pollutants (POPs) have a greater K_{oc} which describes the partitioning of POPs in soils with high organic matter contents. Generally it is understood that, the greater the level of chlorination or bromination the K_{oc} value is more, and consequently the chemical can easily and strongly bind to the organic matter(O’Sullivan & Megson, 2014).

2.4. Biological Phase

Aquatic and terrestrial environments along with the organisms themselves house many complex mixtures of non-polar contaminants. These contaminants include many chemical families such as polychlorinated biphenyls (PCBs), monoaromatic hydrocarbons, polybrominated diphenyl ethers (PBDEs), polyhalogenated dibenzo-*p*-dioxins (PHDDs), dibenzofurans (PHDFs), polychlorinated naphthalenes (PCNs), polycyclic aromatic hydrocarbons (PAHs), toxaphene and polychlorinated *n*-alkanes (PCAs). These chemical families are persistent in the environment and consequently they bio-accumulate. They are toxic in nature and form potent complex mixtures (Eljarrat & Barceló, 2003; Schmitz *et al.*,, 1995).

For ecological risk assessment of aquatic toxicity, the most commonly used biological endpoint, for aquatic species, is effective lethal concentration of chemicals at different trophic levels(Schwarzenbach *et al.*,, 2002). A compound can be toxic to an organism through different modes of action. For instance, non-covalent interactions of the chemical with the organism can cause narcosis, which is the non-specific toxicity(Schwarzenbach *et al.*,, 2002). The modes of action which cause specific toxicities include mechanisms of electrophilic covalent interactions or receptor mediated functional toxicity (Hoover *et al.*,, 2005). Narcosis disturbs the normal functioning of biological membranes. Narcosis is further sub-divided into non-polar narcosis, amine narcosis, polar narcosis and ester narcosis. This further classification occurs because these families have different partition coefficients for membrane proteins and lipids. Nonpolar narcosis is also termed as baseline toxicity or minimum toxicity. When a chemical is present at above the threshold concentration level, it is called median lethal concentration, and it exerts baseline toxicity. Non-polar baseline toxicity is exerted by more than 60 to 70% of known industrial chemicals whereas, all the other known chemicals exert much greater toxicity than the non-polar baseline toxicity (Casalegno & Sello, 2013). Hence, baseline toxicity of these persistent chemicals is a major effect being observed in different ecosystems due to addition of toxicities (Öberg, 2004).

Many other properties like biodynamics, bioavailability and toxicities of these mixtures are a little difficult to assess, as thousands of these compounds are acting simultaneously in a system (Altenburger *et al.,*, 2013; Donnelly *et al.,*, 2004). Even when a complex mixture composition has been interpreted, it is still not possible to categorize different components according to their bioavailability and bioaccumulation potential along with their toxicity (Redman *et al.,*, 2012) because the partition coefficient, effective lethal concentration data, and kinetic parameters which are needed to investigate these processes are not known for many compounds.

2.5. Natural Organic Matter

In the field of environmental chemistry, a key parameter used to assess bioavailability, bioaccumulation, and toxicity is the freely dissolved concentration (C_{free}) of a contaminant. In the aquatic environments, dissolved organic matter (DOM) plays a vital role in controlling the C_{free} of hydrophobic contaminants. DOM not only reduces the bioavailability and toxicity of hydrophobic chemicals (Haitzer *et al.,*, 1998; Haitzer *et al.,*, 1999) but also enhances the contaminant’s apparent solubility and aids the mobility of contaminants from sediments and soil matrices (ter Laak *et al.,*, 2008). It enables the mass transfer processes and increases the exchange kinetics between the aqueous medium and sorbents such as soils, organisms, sediments as well as the passive samplers (ter Laak *et al.,*, 2008). It also provides a micro-heterogeneous environment to hydrophobic organic compounds(Latch & McNeill, 2006). It can either guard or expose the bound hydrophobic compounds from transient species such as hydrated electrons and single oxygen and subsequently give it a reduced or an enhanced reactivity zone (Burns *et al.*,, 1996; Burns *et al.*,, 1997; Latch & McNeill, 2006). For last few decades, an increasing concentration trends of DOM in natural waters have been observed(Evans *et al.*,, 2005). This increasing trend is being linked to the global warming and decreased acid rain(Evans *et al.*,, 2005; Freeman *et al.*,, 2001). In the future, the environmental significance of DOM may increase and become more noticeable. The sorption coefficients of DOM and C_{free} should be known in order to quantify their role in the fate of chemical contaminants.

Traditional methods used to determine C_{free} and DOM-water partition coefficients of hydrophobic contaminants in different ecosystems have failed to provide consistent results due to very low concentrations, lack of reliable sorption coefficients and binding to the matrices. The DOM phase is dispersed within the aqueous phase and the equilibrium distribution of the contaminant between pure water and DOM is contradiction in terminis (Durjava *et al.*,, 2007). In traditional methods, separating two phases physically without upsetting the equilibrium state is extremely difficult to accomplish (Schwarzenbach *et al.*,, 2002). Furthermore, aqueous solubility of hydrophobic organic compounds are extremely low, thus it is difficult to maintain stable concentrations to carry out the experiments (Fauser & Thomsen, 2002; Yang *et al.*,, 2007).

2.6. Enthalpy of vaporization

Latent heat of vaporization, ΔH_v is also called the enthalpy of vaporization. It is basically the difference between the enthalpy of the saturated vapor and that of the saturated liquid at the same temperature. The information from this method is further utilized in the process calculations. For instance, for the design of a relief systems which includes volatile compounds, in distillation process, ΔH_v is required to estimate the heat loads for the condenser and the reboiler and for designing heat exchanger for condensing liquids, the data of ΔH_v is vital.

An improved Watson equation for the enthalpy of vaporization as a function of temperature is given in the following equation(Coker, 2014).

$\Delta Hv=A(1-TTc)n$

Where

ΔH_v = enthalpy of vaporization (kJ/mol)

A, T_c, and n = regression coefficients for chemical compound

T = temperature (K)

2.7. Experimental work

In regulatory toxicology, for the hazard assessment animal tests are employed to identify the toxicological properties of chemicals. When the chemicals are used in the manufacturing of a product or when they are used for a specific purpose, humans or the environment get exposed to their toxicological impacts. These toxicological impacts are then identified. Presently, safe tests of these potential chemical compounds should be regulated and there must be a system of standard and uniform animal testing in place. Regulations must be made for the safety of the employees during the manufacturing process of different products involving the chemical compounds. Additionally, regulations should be in place for the safety of the environment and the user (Liebsch & Spielmann, 2002).

Laboratory animals, when exposed to the chemical compounds in question, can undergo distress, considerable pain and in severe cases, can prove fatal. In order to meet regulatory requirements, bio-concentration experiments must have specific number of test animals. For example, over 2.6 million test animals are essentially required as per European chemical management program, REACH (Van der Jagt *et al.*,, 2004).

2.8. Passive sampling

Polydimethylsiloxane (PDMS) consists of different polymers and solvents, having different capacities to carry out kinetic process. Material which is made up of a single polymer is usually considered as a favorable option, for chemical monitoring, in the matter of partitioning of the chemical between different interfaces. Exchange rates and diffusion coefficients between water phase and the sampler decides the amount of chemical move toward passive sampler. By following an integrated approach between exposure and time, PDMS can monitor Polychlorinated biphenyls (PCBs) with an extensive range of hydrophobicity and complex organic chemicals(Nabi *et al.,*, 2013).

When simple diffusion model is employed, by taking into account the surface and volume ratio of PDMS, equilibrium times can be easily estimated for hydrophobic compounds. For the monitoring studies using PDMS in field, sampling time and equilibrium constant are considered to study the sorption behavior of chemical.

At molecular level, van der Waals forces and hydrogen bond interactions govern the affinity of a chemical to be absorbed at passive sampler. The attainment of equilibrium state depends mostly upon size and symmetry of the chemical, movement of water, sampling area and physiochemical properties of the chemicals. Aqueous boundary layer that limits the transfer of chemicals between two media, can act as a limiting factor, affecting uptake kinetics(Nabi *et al.,*, 2013).

2.9. GCxGC

Comprehensive GC×GC system is two-dimensional gas chromatography, employing a pair of GC columns. It is an efficient analytical technique which is used to resolve thousands of compounds present in complex mixtures (Hamilton *et al.,*, 2004). It has a primary column connected in series with a secondary column though a modulator (Figure 2). The modulator acts as an inlet for the secondary dimension and it constantly slices the partly resolved primary peak, which is broad, into sub-peaks. These sub-peaks are further separated in the secondary column(Liu & Phillips, 1991). Generally, the secondary column is much shorter in length than the primary column. The basic reason is to retain the separation, which was achieved in the first dimension. Different chemical selectivity is present in the stationary phase, that is, in the secondary column than that of the primary column. This technique helps to attain very sharp peaks, which improve the system's analytical detection limits. This technique requires fast sampling detectors.

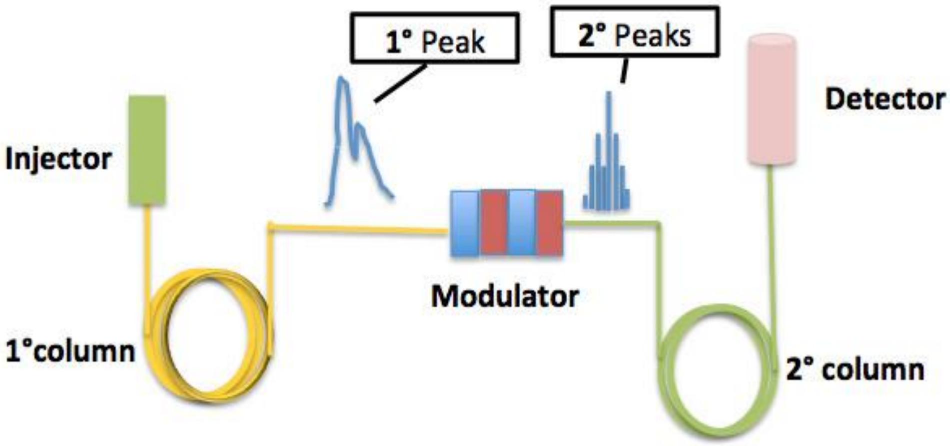

Figure 2.1 Schematic of the GC×GC configuration. 1° and 2° denote primary and secondary, respectively

2.10. Abraham Solvation Model

The Abraham solvation model (ASM) is a multiple parameter equation (eq.2 a-d) that has been described extensively in environmental chemistry for the estimation of transport and partitioning properties(Abraham & Ibrahim, 2006; DiFilippo & Eganhouse, 2010; Endo *et al.*,, 2012; EndoBrown *et al.*,, 2013; EndoDroge *et al.*,, 2011; EndoEscher *et al.*,, 2011; Endo & Goss, 2011; EndoHale *et al.*,, 2011; EndoMewburn *et al.*,, 2013; Geisler *et al.*,, 2012; Hoover *et al.*,, 2005; Kamprad & Goss, 2007; Lohmann, 2011; Poole *et al.*,, 2013; Sprunger *et al.*,, 2007).

$$\log \boldsymbol{P_{xy,i}} = \boldsymbol{c + e\,E_i + s\,S_i + a\,A_i + b\,B_i + v\,V_i} \qquad (\boldsymbol{2a})$$

$$\log \boldsymbol{P_{xy,i}} = \boldsymbol{c + e\,E_i + s\,S_i + a\,A_i + b\,B_i + l\,L_i} \qquad (\boldsymbol{2b})$$

$$\log \boldsymbol{P_{xy,i}} = \boldsymbol{c\ + s\,S_i + a\,A_i + b\,B_i + l\,L_i + v\,V_i} \qquad (\boldsymbol{2c})$$

$$\log \boldsymbol{P_{xy,i}} = \boldsymbol{c\ + s\,S_i + a\,A_i + b\,B_i + l\,L_i} \qquad (\boldsymbol{2d})$$

Where

$P_{xy,I}$ = partitioning property of chemical (dispersed between two phases denote the value of intercept)

e,s,a,b,v = value of coefficient of each Abraham solvation parameters, describing the importance of each parameter.

E_i = excess molar refractivity that is used to check the polarizability of the solute (cm^3/mole)/10

S_i = characteristics of polarity or di-polarity of the solute

A_i = Acidity of the solute

B_i = Basicity of the chemical

V_i = McGowan characteristic of molecular volume of solute in specific unit of (cm^3/mole)/100

L = gas- hexadecane partitioning coefficient of solute at 25°C (Bradley et al., 2015).

2.11. One Parameter-Linear Free Energy Relationship

In the environmental evaluation, double logarithmic correlations for the unidentified partition constant and an identified partition constant of the compounds, gives the value for the equilibrium partitioning. For instance, the equilibrium partition between organic compound and air and alternatively organic compound and water is associated with the octanol and air or octanol and water partition constant, respectively(Goss & Schwarzenbach, 2000). The limitation is that these relationships can only calculate the variations for the class of single substances. These relationships provide no answer to understand the difference between substance classes as well as the difference between different natural organic phases(Goss & Schwarzenbach, 2000). These limitations are because:

- The comprehensive variations of the compound are not explained by only one parameter as the partitioning is because of various exchanges, not dependent on each other.
- In different phases, the specific properties are not denoted in the variable parameter. They depend on the slope and intercept of the double logarithmic correlation.

2.12. Estimation method

Estimation methods which are being extensively used are Group methods, LFER, Ab initio and Fragmental. These methods help to screen chemicals at initial stage. Fragmental method has an extensive application for developing a QSAR model. This in turn is used to forecast and estimate site specific physiochemical profiling, biological and physical properties of chemicals, active skeleton and functional group of the chemicals (Japertas *et al.,*, 2003).

Group estimation method use solvent activity in polymer. In this method, input of first order group gets assessed at initial stage and at advance stage, 2nd order group gets assessed and valued. Isomers identification and estimation along with the accurate application has been used to ascertain the properties (Constantinou & Gani, 1994). Poly parameter linear free energy relation model (PP- LFER) has greater efficiency for estimation of the partitioning coefficient in several technical systems. The model has been deemed applicable as it describes the linear relationship between numerous environmental properties. LFERs model equation also describes strong relationship between the descriptors (Endo & Goss, 2014; Endo *et al.,*, 2009). Ab initio is a quantum mechanical modeling method, which is used to find the mechanical structure of molecule at an optimized level. It is based on simulation of the system to explore various local methods. Structural measurements of the chemical as well as its individuality from other chemicals can be assessed in this method. Ab initio method scatter data with the help of different modelling methods like bead method and envelop function(Nedyalkova *et al.*,, 2019; Singh & Kollman, 1986).

2.13. Knowledge Gap

- It is not fully known which free energy related properties are suitable to develop a low-cost simple estimation approach.
- It is not known to what extent the addition of K_{aw} would improve the performance of OP-LFERs.

Chapter 3

3. Methodology

Details about materials and methods are described in this chapter which was used for the investigation purpose of the study. Analysis on available data was performed by authentic computational software on computer.

3.1. Data acquisition

Experimental values of biological phases involving diverse set of chemicals were taken from literature (Abraham & Ibrahim, 2006; Endo *et al.*,, 2012; EndoEscher *et al.*,, 2011; Geisler *et al.*,, 2012). To avoid over-representation, multiple values reported for a single chemical were averaged using arithmetic mean. All the inorganic values were omitted from the data sets. The estimated values of K_{ow}, K_{aw}, and k_{oc} were acquired from EPI Suite ^{TM} 4.1 – KOWWIN v1.68, HenryWin v3.20, KOCWIN v2.00 (US-EPA, 2018). The diversity of the chemical set used in the study can be gauged in term of the wide ranges spanned by K_{ow} (8 order of magnitude), K_{aw} (7 order of magnitude), and K_{oc} (6 order of magnitude).

Experimental values of Natural Organic Matter phases involving diverse set of chemicals were taken from literature (Neale *et al.*,, 2012; Niederer *et al.*,, 2007). To avoid over-representation, multiple values reported for a single chemical were averaged using arithmetic mean. All the inorganic values were omitted from the data sets. The estimated values of K_{ow}, K_{aw}, and k_{oc} were taken from EPI Suite ^{TM} 4.1 – KOWWIN v1.68, HenryWin v3.20, KOCWIN v2.00 (US-EPA, 2018). The diversity of the chemical set used in the study can be gauged in term of the wide ranges spanned by K_{ow} (8 order of magnitude), K_{aw} (7 order of magnitude), and K_{oc} (5 order of magnitude).

Experimental values of Enthalpy of solvation phase involving diverse set of chemicals were taken from literature (Mintz *et al.*,, 2006). To avoid over-representation, multiple values reported for a single chemical were averaged using arithmetic mean. All the inorganic values were omitted from the data sets. The estimated values of K_{ow}, K_{aw}, and k_{oc} were obtained from EPI Suite ^{TM} 4.1 – KOWWIN v1.68, HenryWin v3.20, KOCWIN v2.00 (US-EPA, 2018). The diversity of the chemical set used in the study can be gauged in term of the wide ranges spanned by K_{ow} (7 order of magnitude), K_{aw} (8 order of magnitude), and K_{oc} (4 order of magnitude).

3.2. Statistical analysis

Statistical analyses such as multiple linear regression, Principle Component Analysis (PCA) and cross-validation were performed using R statistical environment (version - 3.5.3) (R(3.5.3), n.d.) and XLSTAT (XLSTAT, 2013). Contribution of a variable in the model was considered statistically significant if the computed t-value of the variable coefficient is less than or equal to the critical t-values reported at the significance level (p-value <0.05) for a given degree of freedom. The Akaike Information Criterion was employed for the selection of ideal variables count in the model. AIC penalizes the model upon adding new variables that do not impart sufficient information to the model. Hence, a model with minimum AIC value was selected. Analysis of correlation was also performed to check any overlapping information bring by different descriptors.

After selection of variables, regression diagnostic (Studentidized Residuals, Hat Values and Cook's Distance) were analysed to identify influential values in each model or approach proposed in this study. Standard errors of the fitting coefficients in each model were computed using bootstrapping algorithm. Different models were developed for each data set with various combination of descriptors. Following cross-validation tests were performed to assess the predictive capability of models: K-Fold, repeated K- Fold (r =10), Leave-One-Out (LOO), and bootstrapping (n=1000) for internal validation of each model and for external validation, the dataset was randomly split into training and test sets. Principle Component Analysis (PCA) was performed on model descriptors to investigate the dimensionality in all the data sets.

Chapter 4

4. Results and Discussion

4.1. Biological phases

Several partitioning phases of biological systems were analyzed namely, lipid-water, liposome-water, protein-water, and air-olive oil partition coefficients. The approach used to explore the variability in different biotic phases, and development of predictive models are discussed in the following section.

4.1.1. Lipid-water

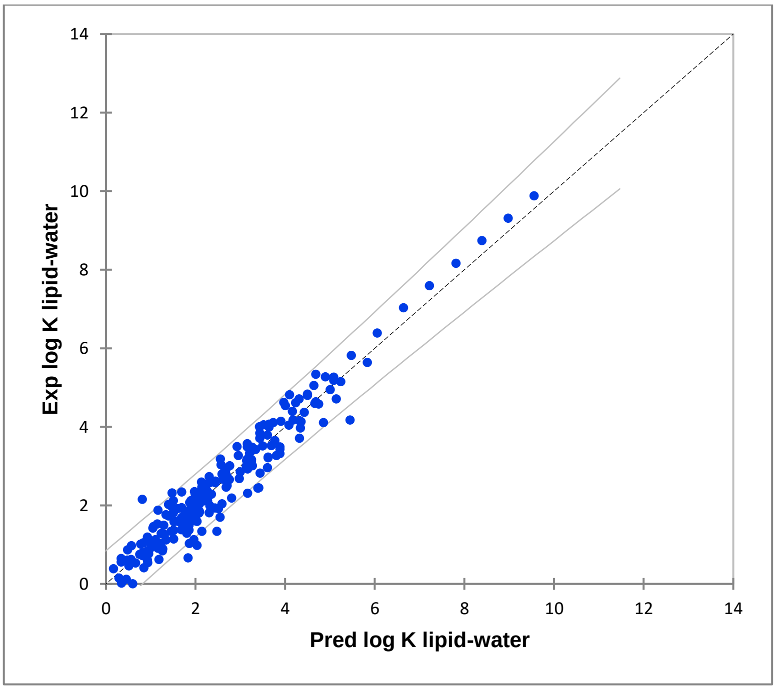

Figure 4.1 Experimental versus predicted log K lipid-water

A multiple linear regression of log K lipid-water against three descriptors namely, Kow, Kaw, and Koc was performed the variables retained were Kow partition coefficient of water and octanol, and kaw partition coefficient of water and air. These descriptors were acquired from freely accessible software, EPI Suite and can be justified scientifically. Kow is exerting dominant effect on lipid-water partition coefficient which can be rationalized easily because historically octanol is considered as a good surrogate solvent for lipid phase, however, there are intermolecular interaction parameters which are coded differently in octanol than in lipid phase. Specifically, the hydrogen bonding interaction parameters are not coded sufficiently by Kow. The air water partition coefficient captures the hydrogen bonding interaction parameters which are not sufficiently capture by octanol-water.

$$\log K_{lipid\text{-}water} = -0.27(\pm 0.06) + 1.16(\pm 0.02)\log K_{ow} + 0.\ 0.10(\pm 0.01)\log K_{aw} \qquad \textbf{Eq. 4-1}$$

n= 247, R^2 = 0.954, Adj. R2 = 0.953, RMSE = 0.404, PRESS RMSE = 0.409, Q2 = 0.953

With experimental values (n=247) Eq. 4-1 explains the 95% of variability in lipid-water data with RMSE of 0.404 log unit. The model is internally valid as indicated by the closeness of RMSE and PRESS RMSE values, and of R^2, Adj.R^2, Q^2 values. This is further supported by other cross-validation tests.

4.1.1.1. Comparison with the ASM

The Abraham Solvation model was used in the literature from where the experimental values were acquired. The predictive performance of the estimation models developed in this study was compared to the ASM using the common Abraham Solvation descriptors resulting in eq. 4, which returned an RMSE value of 0.218 log unit and R2 = 0.987. In comparison, ASM has performed just slightly better than the model explained in the study but as discussed in the literature review, Abraham solvation is a parameter intensive model and is available for less than 8000 chemicals.

$$\log K_{lipid\text{-}water} = 0.67-0.38E-1.53S-1.98A-4.29B+1.43V+0.74L \qquad \textbf{Eq. 4-2}$$

n= 247, R^2= 0.987, Adj.R^2= 0.986, Q^2= 0.986, RMSE= 0.218, PRESS RMSE= 0.227

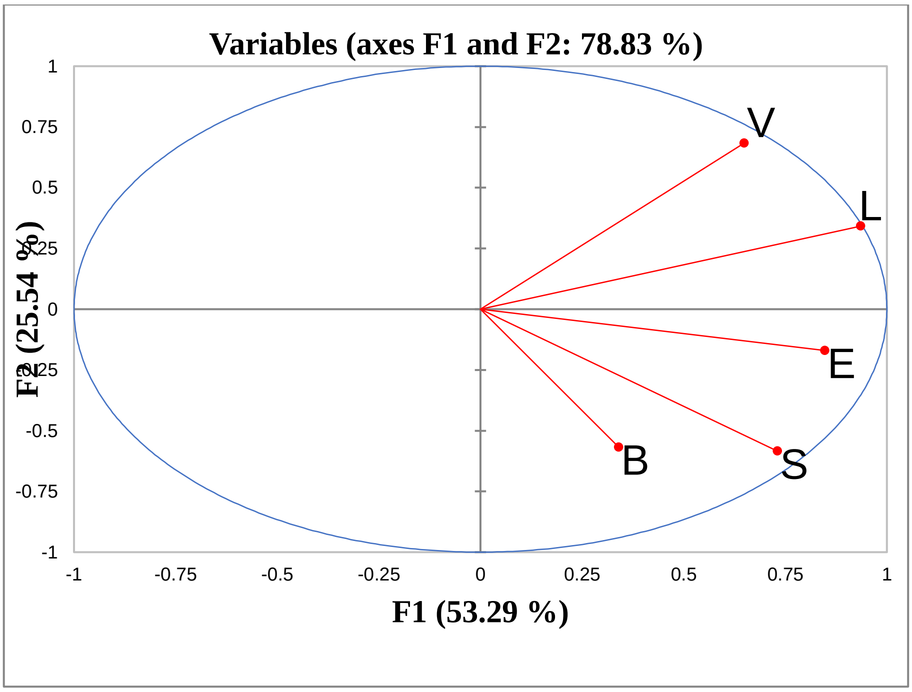

Figure 4.2 Biplot of ASM

Figure.4.2 explains the justification for two parameter model, which is further reinforced by the dimensionality analysis of Abraham Solvation parameters, which shows that only two dimensions are enough to explain more than 78% of the information encoded in the Abraham solvation descriptors for organic chemicals.

4.1.2. Liposome-water

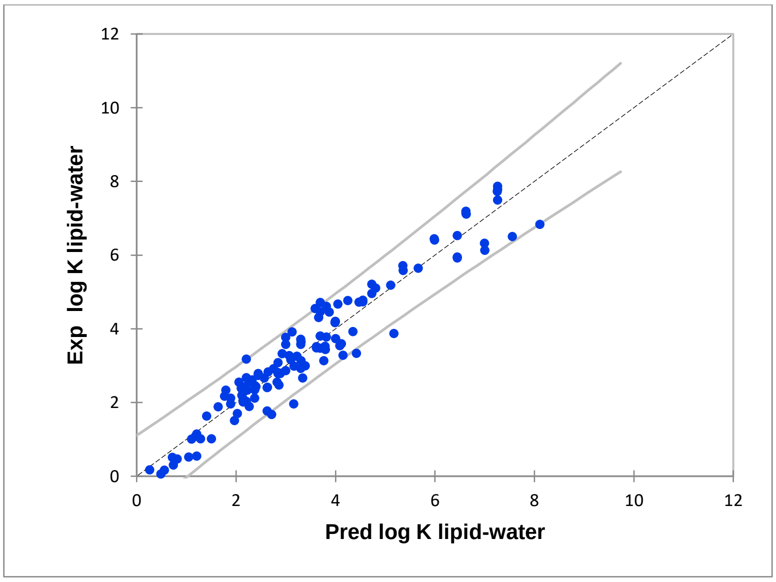

Figure 4.3 Experimental vs predicted log K liposome-water

A multiple linear regression of log K liposome-water against three descriptors namely, Kow, Kaw, and Koc was performed the variables retained were Kow partition coefficient of octanol-water, and koc partition coefficient of organic carbon-water. These descriptors were retained can be justified scientifically by the fact that Kow is exerting dominant effect on liposome-water partition coefficient which can be rationalized easily because historically octanol is considered as a good surrogate solvent for lipid phase, however, there are intermolecular interaction parameters which are coded differently in octanol than in liposome phase. Koc is the second descriptor that is retained in this model which also explains the intermolecular interactions but the residual intermolecular interactions that weren't explained by Kow.

$$\log K_{\text{liposome-water}} = -0.10(\pm 0.09)+0.72(\pm 0.05)\log K_{ow}+0.39(\pm 0.06)\log_{Koc} \qquad \textbf{Eq. 4-3}$$

n= 131, R^2 = 0.938, Adj. R2 = 0.937, RMSE = 0.456, PRESS RMSE = 0.490, Q2=0.934

With experimental values (n=131) Eq.4-3 explains 93% of variability in liposome-water data with RMSE of 0.456 log unit. The model is internally valid as indicated by the closeness of

RMSE and PRESS RMSE values, and of R^2, Adj.R^2, Q^2 values. This is further supported by other cross-validation tests.

4.1.3. Protein-water

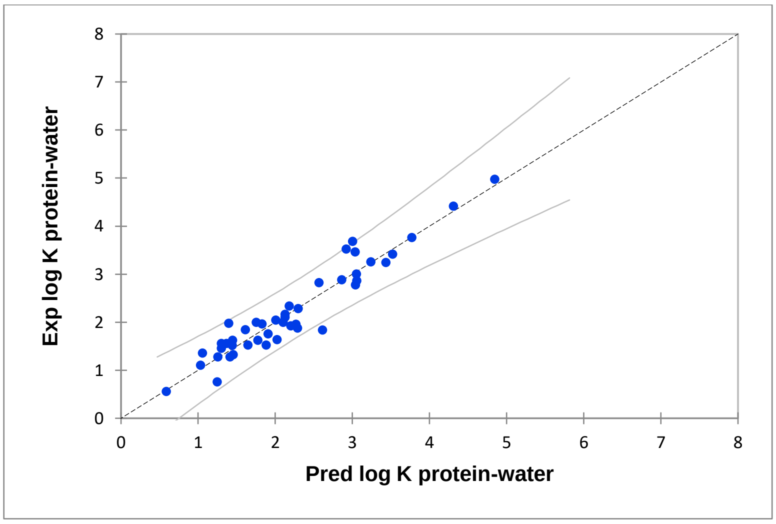

Figure 4.4 Experimental vs predicted log K protein-water

A multiple linear regression of log K Protein-water against three descriptors namely, Kow, Kaw, and Koc was performed the variables retained were Kow partition coefficient of octanol-water, and koc partition coefficient of organic carbon-water. These descriptors were retained can be justified scientifically by the fact that Kow is exerting dominant effect on protein-water partition coefficient which can be rationalized easily because octanol is considered as a good surrogate solvent for lipid phase, however, there are intermolecular interaction parameters which are coded differently in octanol than in protein phase. Koc is the second descriptor that is retained in this model which also explains the intermolecular interactions but the residual intermolecular interactions that weren't explained by Kow.

$$\log K_{protein\text{-}water} = -0.88(\pm 0.15)+0.59(\pm 0.06)\log K_{ow}+0.36(\pm 0.07)\log K_{oc} \qquad \textbf{Eq. 4-4}$$

n= 45, R^2 = 0.911, Adj. R^2 = 0.907, RMSE = 0.292, PRESS RMSE = 0.312, Q2=0.899

With experimental values (n=45) Eq.4-4 explains 91% of variability in protein-water data with RMSE of 0.292 log unit. The model is internally valid as indicated by the closeness of RMSE and PRESS RMSE values, and of R^2, Adj.R^2, Q^2 values. This is further supported by other cross-validation tests.

4.1.4. Air-olive oil

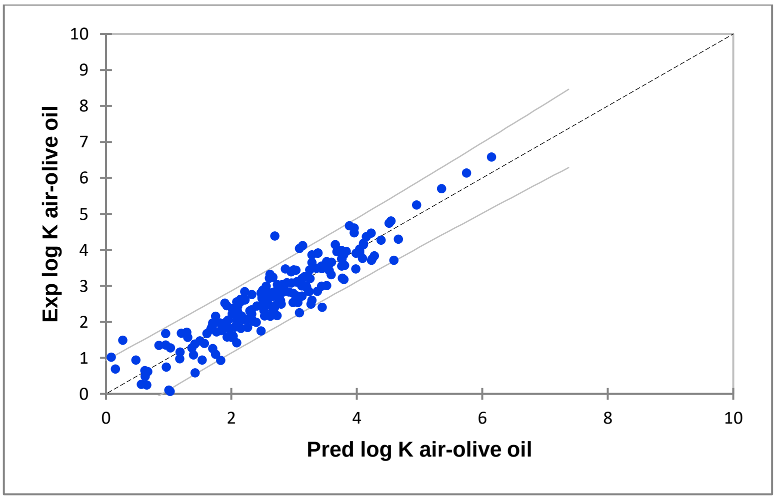

Figure 4.5 Experimental vs predicted air-olive oil

A multiple linear regression of log K air-olive oil against three descriptors namely, Kow, Kaw, and Koc was performed the variables retained were Kow partition coefficient of octanol-water, and kaw partition coefficient of air-water. These descriptors were acquired from freely accessible software, EPI Suite and can be justified scientifically. Kow is exerting dominant effect on air phase which can be explained by the fact that here intermolecular interaction parameters are explained by Kow. The hydrogen bonding interaction parameters are not coded sufficiently by Kow therefore, Kaw is used to captures the hydrogen bonding interaction parameter in olive oil phase.

$$log\ K_{air\text{-}olive\ oil} = 0.16(\pm0.07)+0.97(\pm0.02)log\ K_{ow}-0.65(\pm0.02)log\ K_{aw} \qquad \textbf{Eq. 4-5}$$

n= 200, R^2 = 0.870, Adj. R2 = 0.869, RMSE = 0.435, PRESS RMSE = 0.443, Q2=0.865

With experimental values (n=200) Eq.6 explains 87% of variability in air-olive oil data with RMSE of 0.435 log unit. The model is internally valid as indicated by the closeness of RMSE and PRESS RMSE values, and of R^2, Adj.R^2, Q^2 values. This is further supported by other cross-validation tests.

4.2. NOM Phase

Several partitioning phases of Natural Organic Matter systems were analyzed namely, DOC-water and NOM-air partition coefficients. The approach used to explore the variability in different NOM phases, and development of predictive models are discussed in the following section.

4.2.1. DOC-water

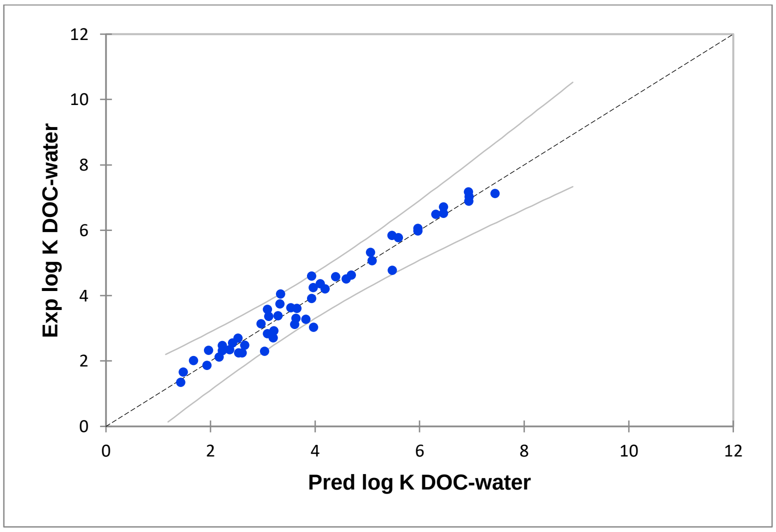

Figure 4.6 Experimental vs predicted DOC-water

A multiple linear regression of log K DOC-water against three descriptors namely, Kow, Kaw, and Koc was performed the variables retained were Kow partition coefficient of octanol-water, and koc partition coefficient of organic carbon-water. These descriptors were acquired from freely accessible software, EPI Suite and can be justified scientifically. Conventionally octanol is taken as a surrogate phase for Dissolved Organic Carbon, but DOC is quite heterogenous than octanol solvent. So, the introduction of a closer relative of DOC into the model equation would improve the outcome a lot. The appropriate descriptor which is available in EPI Suite is log Koc, which is estimated through MCI (Molecular Connectivity Index).

$$\log K_{DOC\text{-}water} = -0.37(\pm0.14)+0.58(\pm0.05)\log K_{ow}+0.54(\pm0.08)\log K_{oc} \qquad \textbf{Eq. 4-6}$$

n= 52, R^2 = 0.958, Adj. R2 = 0.956, RMSE = 0.343, PRESS RMSE = 0.358, Q2=0.954

With experimental values (n=52) Eq.4-6 explains 95% of variability in DOC-water data with RMSE of 0.343 log unit. The model is internally valid as indicated by the closeness of RMSE and PRESS RMSE values, and of R^2, Adj.R^2, Q^2 values. This is further supported by other cross-validation tests.

4.2.2. NOM-air

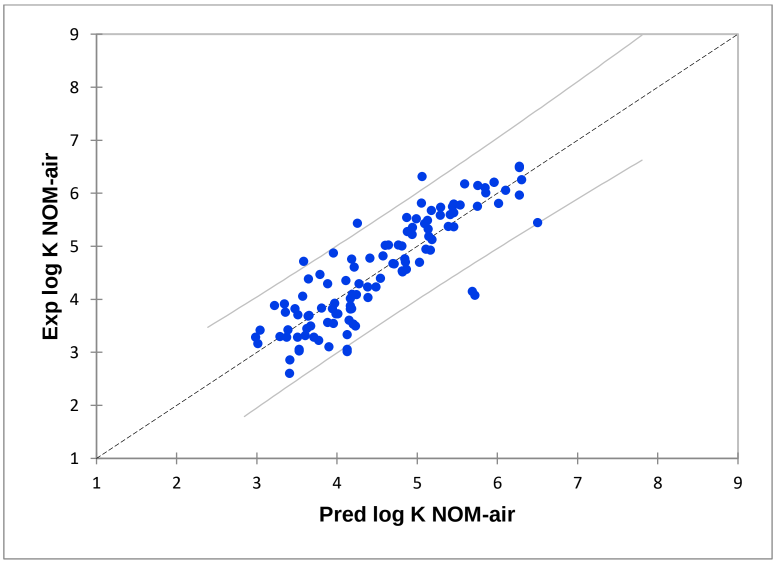

Figure 4.7 Experimental vs predicted NOM-air

A multiple linear regression of log K NOM-air against three descriptors namely, Kow, Kaw, and Koc was performed the variables retained were Kow partition coefficient of octanol-water, and kaw partition coefficient of air-water. These descriptors were acquired from freely accessible software, EPI Suite and their scientific justification is that Kow is exerting relatively less dominant effect on NOM phase hence explaining the intermolecular interactions. Whereas, the hydrogen bonding interaction parameters are coded by Kaw which is exerting dominant effect on the air phase.

$$\log K_{Nom\text{-}air} = -0.16(\pm0.26)+0.82(\pm0.05)\log K_{ow}-0.84(\pm0.04)\log K_{aw} \qquad \textbf{Eq. 4-7}$$

n= 113, R^2 = 0.752, Adj. R2 = 0.748, RMSE = 0.505, PRESS RMSE = 0.521, Q2=0.735

With experimental values (n=113) Eq.4-7 explains 75% of variability in DOC-water data with RMSE of 0.505 log unit. The model is internally valid as indicated by the closeness of RMSE and PRESS RMSE values, and of R^2, Adj.R^2, Q^2 values. This is further supported by other cross-validation tests

4.3. Enthalpy of solvation

Two very important systems of enthalpy of solvation were analyzed namely, air-octanol and air-water partition coefficients. The approach used to explore the variability in different enthalpy of solvation phases, and development of predictive models are discussed in the following section.

4.3.1. Air-octanol

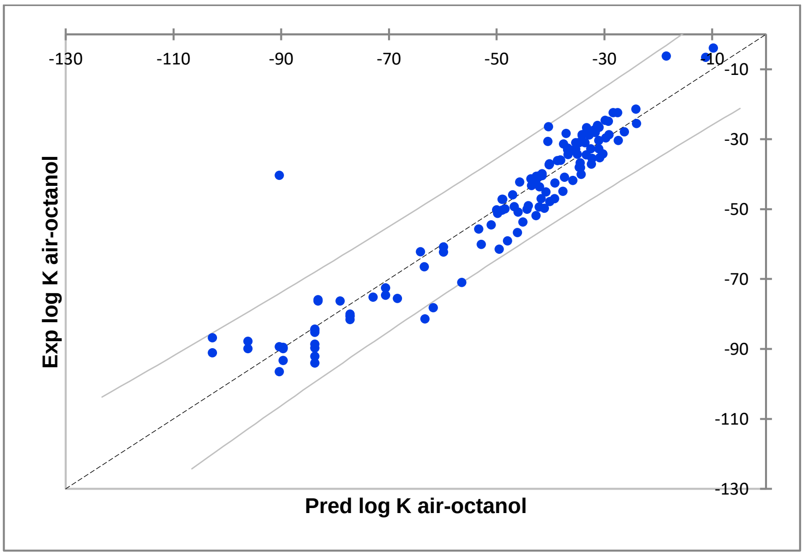

Figure 4.8 Experimental vs predictive log K air-octanol

A multiple linear regression of log K air-octanol against three descriptors namely, Kow, Kaw, and Koc was performed. These descriptors were acquired from freely accessible software, EPI Suite and their scientific justification is that, even though enthalpy is not a free energy related property. Exploiting enthalpy to entropy compensation mechanism, we have developed this 2-parameter model based on octanol-water and air-water partition coefficient. The compounds which are very hydrophobic would be more soluble in the octanol phase and would require higher energy of enthalpy to move to air phase, on the other hand the compounds which have higher air-water partition coefficient would not need high enthalpy of phase transfer to be released into the air phase.

$$\Delta H = -4.53(\pm1.54)-8.75(\pm0.28)\log K_{ow}+7.01(\pm0.39)\log K_{aw} \qquad \textbf{Eq. 4-8}$$

n= 131, R^2 = 0.896, Adj. R2 = 0.894, RMSE = 7.292, PRESS RMSE = 7.552, Q2=0.889

With experimental values (n=131) Eq.4.8 explains 89% of variability in air-octanol data with RMSE of 7.292 log unit. The model is internally valid as indicated by the closeness of RMSE and PRESS RMSE values, and of R^2, Adj.R^2, Q^2 values. This is further supported by other cross-validation tests.

4.3.2. Air-water

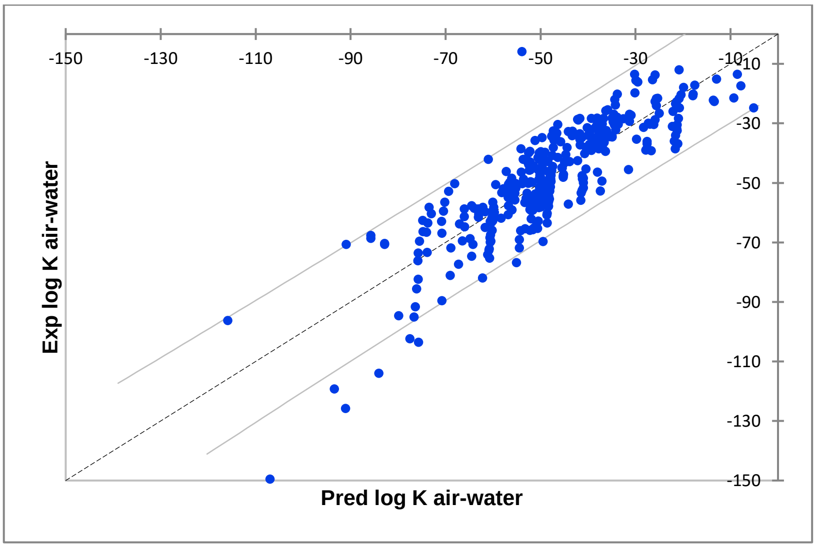

Figure 4.9 Experimental vs predictive log K air-water

A multiple linear regression of log K air-water against three descriptors namely, Kow, Kaw, and Koc was performed. These descriptors were acquired from freely accessible software, EPI Suite and their scientific justification is that, even though enthalpy is not a free energy related property. Exploiting enthalpy to entropy compensation mechanism, we have developed this 3-

parameter model based on octanol-water, air-water, and organic carbon-water coefficient partition. The compounds which are very hydrophilic would be more soluble in the water phase and would require higher energy of enthalpy to move to air phase, on the other hand the hydrophobic compounds which have relatively higher organic carbon-water partition coefficient would not need high enthalpy of phase transfer to be released into the air phase. The most dominant descriptor is the air-water partition coefficient itself explaining the hydrogen bonding in the system.

$\Delta H = -30.007(\pm 1.45)-5.63(\pm 0.88)\log K_{ow}+5.62(\pm 1.19\log K_{oc}+8.34(\pm 0.30)\log K_{aw}$ **Eq. 4-9**

n= 349, R^2 = 0.733, Adj. R2 = 0.730, RMSE = 9.848, PRESS RMSE = 10.029, Q2=0.723

With experimental values (n=343) Eq.4-9 explains 73% of variability in air-water data with RMSE of 9.848 log unit. The model is internally valid as indicated by the closeness of RMSE and PRESS RMSE values, and of R^2, Adj.R^2, Q^2 values. This is further supported by other cross-validation tests.

Chapter 5

5. Conclusion and Recommendation

5.1. Conclusions

Our models were successfully able to describe the factors governing chemical dynamics for different biological, natural organic matter, and enthalpy of vaporization phases.

Two-parameter models were developed for a diverse set of organic chemicals. Our two-parameter models were as good as five-parameter, Abraham Solvation Model. Explaining variance almost equally as good as the previous models.

One of the best things about these proposed LFER method is that they are easily accessible than different multi-parameter estimation methods.

Like every model, our models have limitations too; these proposed models are not applicable for ionizable, inorganic, metallic and organo-metallic compounds. These models require machine readable SMILE codes, which are available only for well-defined chemical structures.

5.2. Recommendations

Following are some of the recommendations from my study;

1. These models can be integrated with EPI suite as separate modules.
2. Several other properties, which follow the same principle of solvation, should be trained parsimoniously using the skeleton developed in this study.

References

Abraham, M. H., & Ibrahim, A. (2006). Gas to Olive Oil Partition Coefficients: A Linear Free Energy Analysis. *Journal of Chemical Information and Modeling, 46*(4), 1735-1741. doi:10.1021/ci060047p

Altenburger, R., Backhaus, T., Boedeker, W., Faust, M., & Scholze, M. (2013). Simplifying complexity: Mixture toxicity assessment in the last 20 years. *Environmental Toxicology and Chemistry, 32*(8), 1685-1687. doi:10.1002/etc.2294

Bayen, S., ter Laak, T. L., Buffle, J., & Hermens, J. L. M. (2009). Dynamic Exposure of Organisms and Passive Samplers to Hydrophobic Chemicals. *Environmental Science & Technology, 43*(7), 2206-2215. doi:10.1021/es8029895

Booth, A. M., Sutton, P. A., Lewis, C. A., Lewis, A. C., Scarlett, A., Chau, W., . . . Rowland, S. J. (2006). Unresolved Complex Mixtures of Aromatic Hydrocarbons: Thousands of Overlooked Persistent, Bioaccumulative, and Toxic Contaminants in Mussels. *Environmental Science & Technology, 41*(2), 457-464. doi:10.1021/es0615829

Burns, S. E., Hassett, J. P., & Rossi, M. V. (1996). Binding Effects on Humic-Mediated Photoreaction: Intrahumic Dechlorination of Mirex in Water. *Environmental Science & Technology, 30*(10), 2934-2941. doi:10.1021/es950906i

Burns, S. E., Hassett, J. P., & Rossi, M. V. (1997). Mechanistic Implications of the Intrahumic Dechlorination of Mirex. *Environmental Science & Technology, 31*(5), 1365-1371. doi:10.1021/es960581l

Casalegno, M., & Sello, G. (2013). Determination of toxicant mode of action by augmented top priority fragment class. *Journal of chemical information and modeling, 53*(5), 1113-1126.

Castells, P., Parera, J., Santos, F. J., & Galceran, M. T. (2008). Occurrence of polychlorinated naphthalenes, polychlorinated biphenyls and short-chain chlorinated paraffins in marine sediments from Barcelona (Spain). *Chemosphere, 70*(9), 1552-1562. doi:http://dx.doi.org/10.1016/j.chemosphere.2007.08.034

Coker, A. K. (2014). *Ludwig's applied process design for chemical and petrochemical plants*: gulf professional publishing.

Constantinou, L., & Gani, R. (1994). New group contribution method for estimating properties of pure compounds. *AIChE Journal*, *40*(10), 1697-1710.

Cui, X., Mayer, P., & Gan, J. (2013). Methods to assess bioavailability of hydrophobic organic contaminants: Principles, operations, and limitations. *Environmental Pollution, 172*(0), 223-234. doi:http://dx.doi.org/10.1016/j.envpol.2012.09.013

Daughton, C. G. (2004). Non-regulated water contaminants: emerging research. *Environmental Impact Assessment Review, 24*(7-8), 711-732. doi:10.1016/j.eiar.2004.06.003

DiFilippo, E. L., & Eganhouse, R. P. (2010). Assessment of PDMS-Water Partition Coefficients: Implications for Passive Environmental Sampling of Hydrophobic Organic Compounds. *Environmental Science & Technology, 44*(18), 6917-6925. doi:10.1021/es101103x

Donnelly, K. C., Lingenfelter, R., Cizmas, L., Falahatpisheh, M. H., Qian, Y., Tang, Y., . . . Mumtaz, M. M. (2004). Toxicity assessment of complex mixtures remains a goal. *Environmental Toxicology and Pharmacology, 18*(2), 135-141. doi:http://dx.doi.org/10.1016/j.etap.2004.03.013

Durjava, M. K., ter Laak, T. L., Hermens, J. L. M., & Struijs, J. (2007). Distribution of PAHs and PCBs to dissolved organic matter: High distribution coefficients with consequences for environmental fate modeling. *Chemosphere, 67*(5), 990-997. doi:http://dx.doi.org/10.1016/j.chemosphere.2006.10.059

Eljarrat, E., & Barceló, D. (2003). Priority lists for persistent organic pollutants and emerging contaminants based on their relative toxic potency in environmental samples. *TRAC Trends in Analytical Chemistry, 22*(10), 655-665. doi:http://dx.doi.org/10.1016/S0165-9936(03)01001-X

Endo, S., Bauerfeind, J., & Goss, K.-U. (2012). Partitioning of Neutral Organic Compounds to Structural Proteins. *Environmental Science & Technology*, *46*(22), 12697-12703. doi:10.1021/es303379y

Endo, S., Brown, T. N., & Goss, K.-U. (2013). General Model for Estimating Partition Coefficients to Organisms and Their Tissues Using the Biological Compositions and Polyparameter Linear Free Energy Relationships. *Environmental Science & Technology*. doi:10.1021/es401772m

Endo, S., Droge, S. T. J., & Goss, K.-U. (2011). Polyparameter Linear Free Energy Models for Polyacrylate Fiber−Water Partition Coefficients to Evaluate the Efficiency of Solid-Phase Microextraction. *Analytical Chemistry, 83*(4), 1394-1400. doi:10.1021/ac102868e

Endo, S., Escher, B. I., & Goss, K.-U. (2011). Capacities of Membrane Lipids to Accumulate Neutral Organic Chemicals. *Environmental Science & Technology, 45*(14), 5912-5921. doi:10.1021/es200855w

Endo, S., & Goss, K.-U. (2011). Serum Albumin Binding of Structurally Diverse Neutral Organic Compounds: Data and Models. *Chemical Research in Toxicology, 24*(12), 2293-2301. doi:10.1021/tx200431b

Endo, S., & Goss, K.-U. (2014). Applications of polyparameter linear free energy relationships in environmental chemistry. *Environmental science & technology, 48*(21), 12477-12491.

Endo, S., Grathwohl, P., Haderlein, S. B., & Schmidt, T. C. (2009). LFERs for soil organic carbon− water distribution coefficients (K OC) at environmentally relevant sorbate concentrations. *Environmental science & technology, 43*(9), 3094-3100.

Endo, S., Hale, S. E., Goss, K.-U., & Arp, H. P. H. (2011). Equilibrium Partition Coefficients of Diverse Polar and Nonpolar Organic Compounds to Polyoxymethylene (POM) Passive Sampling Devices. *Environmental Science & Technology, 45*(23), 10124-10132. doi:10.1021/es202894k

Endo, S., Mewburn, B., & Escher, B. I. (2013). Liposome and protein–water partitioning of polybrominated diphenyl ethers (PBDEs). *Chemosphere, 90*(2), 505-511. doi:http://dx.doi.org/10.1016/j.chemosphere.2012.07.069

Escher, B. I., & Hermens, J. L. M. (2004). Peer Reviewed: Internal Exposure: Linking Bioavailability to Effects. *Environmental Science & Technology, 38*(23), 455A-462A. doi:10.1021/es0406740

Evans, C. D., Monteith, D. T., & Cooper, D. M. (2005). Long-term increases in surface water dissolved organic carbon: Observations, possible causes and environmental impacts. *Environmental Pollution, 137*(1), 55-71. doi:http://dx.doi.org/10.1016/j.envpol.2004.12.031

Fauser, P., & Thomsen, M. (2002). Sensitivity analysis of calculated exposure concentrations and dissipation of DEHP in a topsoil compartment: The influence of the third phase effect and Dissolved Organic Matter (DOM). *Science of the Total Environment, 296*(1—3), 89-103. doi:http://dx.doi.org/10.1016/S0048-9697(02)00053-0

Freeman, C., Evans, C., Monteith, D., Reynolds, B., & Fenner, N. (2001). Export of organic carbon from peat soils. *Nature, 412*(6849), 785-785.

Gawor, A., & Wania, F. (2013). Using quantitative structural property relationships, chemical fate models, and the chemical partitioning space to investigate the potential for long range transport and bioaccumulation of complex halogenated chemical mixtures. *Environmental Science: Processes & Impacts, 15*(9), 1671-1684. doi:10.1039/c3em00098b

Geisler, A., Endo, S., & Goss, K.-U. (2012). Partitioning of Organic Chemicals to Storage Lipids: Elucidating the Dependence on Fatty Acid Composition and Temperature. *Environmental Science & Technology, 46*(17), 9519-9524. doi:10.1021/es301921w

Goss, K.-U., & Schwarzenbach, R. P. (2000). Linear Free Energy Relationships Used To Evaluate Equilibrium Partitioning of Organic Compounds. *Environmental Science & Technology, 35*(1), 1-9. doi:10.1021/es000996d

Goss, K.-U., & Schwarzenbach, R. P. (2003). Rules of Thumb for Assessing Equilibrium Partitioning of Organic Compounds: Successes and Pitfalls. *Journal of Chemical Education, 80*(4), 450. doi:10.1021/ed080p450

Haitzer, M., Höss, S., Traunspurger, W., & Steinberg, C. (1998). Effects of dissolved organic matter (DOM) on the bioconcentration of organic chemicals in aquatic organisms — a review. *Chemosphere, 37*(7), 1335-1362. doi:http://dx.doi.org/10.1016/S0045-6535(98)00117-9

Haitzer, M., Höss, S., Traunspurger, W., & Steinberg, C. (1999). Relationship between concentration of dissolved organic matter (DOM) and the effect of DOM on the bioconcentration of benzo[a]pyrene. *Aquatic Toxicology, 45*(2—3), 147-158. doi:http://dx.doi.org/10.1016/S0166-445X(98)00097-6

Hamilton, J., Webb, P., Lewis, A., Hopkins, J., Smith, S., & Davy, P. (2004). Partially oxidised organic components in urban aerosol using GCXGC-TOF/MS. *Atmospheric Chemistry and Physics, 4*(5), 1279-1290.

Hendriks, A. J. (2013). How To Deal with 100,000+ Substances, Sites, and Species: Overarching Principles in Environmental Risk Assessment. *Environmental Science & Technology, 47*(8), 3546-3547. doi:10.1021/es400849q

Hites, R. A. (2004). Polybrominated Diphenyl Ethers in the Environment and in People: A Meta-Analysis of Concentrations. *Environmental Science & Technology, 38*(4), 945-956. doi:10.1021/es035082g

Hoover, K. R., Acree, W. E., & Abraham, M. H. (2005). Chemical Toxicity Correlations for Several Fish Species Based on the Abraham Solvation Parameter Model. *Chemical Research in Toxicology, 18*(9), 1497-1505. doi:10.1021/tx050164z

Howard, P. H., & Muir, D. C. G. (2010). Identifying New Persistent and Bioaccumulative Organics Among Chemicals in Commerce. *Environmental Science & Technology, 44*(7), 2277-2285. doi:10.1021/es903383a

Jahnke, A., McLachlan, M. S., & Mayer, P. (2008). Equilibrium sampling: Partitioning of organochlorine compounds from lipids into polydimethylsiloxane. *Chemosphere, 73*(10), 1575-1581. doi:http://dx.doi.org/10.1016/j.chemosphere.2008.09.017

Japertas, P., Didziapetris, R., & Petrauskas, A. (2003). Fragmental methods in the analysis of biological activities of diverse compound sets. *Mini reviews in medicinal chemistry, 3*(8), 797-808.

Kamprad, I., & Goss, K.-U. (2007). Systematic Investigation of the Sorption Properties of Polyurethane Foams for Organic Vapors. *Analytical Chemistry, 79*(11), 4222-4227. doi:10.1021/ac070265x

Korytár, P., Leonards, P. E. G., de Boer, J., & Brinkman, U. A. T. (2005). Group separation of organohalogenated compounds by means of comprehensive two-dimensional gas chromatography. *Journal of Chromatography A, 1086*(1-2), 29-44. doi:http://dx.doi.org/10.1016/j.chroma.2005.05.087

Latch, D. E., & McNeill, K. (2006). Microheterogeneity of Singlet Oxygen Distributions in Irradiated Humic Acid Solutions. *Science, 311*(5768), 1743-1747. doi:10.1126/science.1121636

Letcher, T. M. (2007). *Thermodynamics, solubility and environmental issues*: Elsevier.

Liebsch, M., & Spielmann, H. (2002). Currently available in vitro methods used in the regulatory toxicology. *Toxicology letters, 127*(1-3), 127-134.

Liu, Z., & Phillips, J. B. (1991). Comprehensive Two-Dimensional Gas Chromatography using an On-Column Thermal Modulator Interface. *Journal of Chromatographic Science, 29*(6), 227-231. doi:10.1093/chromsci/29.6.227

Lohmann, R. (2011). Critical Review of Low-Density Polyethylene's Partitioning and Diffusion Coefficients for Trace Organic Contaminants and Implications for Its Use As a Passive Sampler. *Environmental Science & Technology, 46*(2), 606-618. doi:10.1021/es202702y

Manzano, C., Hoh, E., & Simonich, S. L. M. (2012). Improved Separation of Complex Polycyclic Aromatic Hydrocarbon Mixtures Using Novel Column Combinations in GC × GC/ToF-MS. *Environmental Science & Technology, 46*(14), 7677-7684. doi:10.1021/es301790h

Mintz, C., Clark, M., Acree, W. E., & Abraham, M. H. (2006). Enthalpy of Solvation Correlations for Gaseous Solutes Dissolved in Water and in 1-Octanol Based on the Abraham Model. *Journal of Chemical Information and Modeling, 47*(1), 115-121. doi:10.1021/ci600402n

Muir, D. C. G., & de Wit, C. A. (2010). Trends of legacy and new persistent organic pollutants in the circumpolar arctic: Overview, conclusions, and recommendations. *Science of the Total Environment, 408*(15), 3044-3051. doi:http://dx.doi.org/10.1016/j.scitotenv.2009.11.032

Nabi, D., Gros, J., Dimitriou-Christidis, P., & Arey, J. S. (2013). *"Mapping environmental partitioning properties of nonpolar complex mixtures using Comprehensive Two Dimensional Gas Chromatography" Unpublished Manuscript*. Unpublished Manuscript. Environmental Chemistry Modeling Laboratory. Environmental Chemistry Modeling Laboratory, Swiss Federal Institute of Technology at Lausanne (EPFL), Lausanne, Switzerland. Lausanne.

Neale, P. A., Escher, B. I., Goss, K.-U., & Endo, S. (2012). Evaluating dissolved organic carbon–water partitioning using polyparameter linear free energy relationships: Implications for the fate of disinfection by-products. *Water Research, 46*(11), 3637-3645. doi:http://dx.doi.org/10.1016/j.watres.2012.04.005

Nedyalkova, M., Madurga, S., Tobiszewski, M., & Simeonov, V. (2019). Calculating the Partition Coefficients of Organic Solvents in Octanol/Water and Octanol/Air. *Journal of chemical information and modeling*.

Niederer, C., Schwarzenbach, R. P., & Goss, K.-U. (2007). Elucidating Differences in the Sorption Properties of 10 Humic and Fulvic Acids for Polar and Nonpolar Organic Chemicals. *Environmental Science & Technology, 41*(19), 6711-6717. doi:10.1021/es0709932

O'Sullivan, G., & Megson, D. (2014). Brief overview: discovery, regulation, properties, and fate of POPs. In *Environmental forensics for persistent organic pollutants* (pp. 1-20): Elsevier.

Öberg, T. (2004). A QSAR for baseline toxicity: validation, domain of application, and prediction. *Chemical research in toxicology, 17*(12), 1630-1637.

Poole, C. F., Ariyasena, T. C., & Lenca, N. (2013). Estimation of the environmental properties of compounds from chromatographic measurements and the solvation parameter model. *Journal of Chromatography A*(0), http://dx.doi.org/10.1016/j.chroma.2013.1005.1045. doi:http://dx.doi.org/10.1016/j.chroma.2013.05.045

Redman, A. D., Parkerton, T. F., McGrath, J. A., & Di Toro, D. M. (2012). PETROTOX: An aquatic toxicity model for petroleum substances. *Environmental Toxicology and Chemistry, 31*(11), 2498-2506. doi:10.1002/etc.1982

Reichenberg, F., & Mayer, P. (2006). Two complementary sides of bioavailability: Accessibility and chemical activity of organic contaminants in sediments and soils. *Environmental Toxicology and Chemistry, 25*(5), 1239-1245. doi:10.1897/05-458r.1

Scarlett, A., Dissanayake, A., Rowland, S. J., & Galloway, T. S. (2009). Behavioral, physiological, and cellular responses following trophic transfer of toxic monoaromatic hydrocarbons. *Environmental Toxicology and Chemistry, 28*(2), 381-387. doi:10.1897/08-298.1

Schenker, U., MacLeod, M., Scheringer, M., & Hungerbuhler, K. (2005). Improving Data Quality for Environmental Fate Models: A Least-Squares Adjustment Procedure for Harmonizing Physicochemical Properties of Organic Compounds. *Environmental Science & Technology, 39*(21), 8434-8441. doi:10.1021/es0502526

Schmid, P., Kohler, M., Gujer, E., Zennegg, M., & Lanfranchi, M. (2007). Persistent organic pollutants, brominated flame retardants and synthetic musks in fish from remote alpine lakes in Switzerland. *Chemosphere, 67*(9), S16-S21. doi:http://dx.doi.org/10.1016/j.chemosphere.2006.05.080

Schmitz, H.-J., Hagenmaier, A., Hagenmaier, H.-P., Bock, K. W., & Schrenk, D. (1995). Potency of mixtures of polychlorinated biphenyls as inducers of dioxin receptor-regulated CYP1A activity in rat hepatocytes and H4IIE cells. *Toxicology, 99*(1–2), 47-54. doi:http://dx.doi.org/10.1016/0300-483X(94)02994-6

Schwarzenbach, R. P., Gschwend, P. M., & Imboden, D. M. (2002). *Environmental Organic Chemistry* (2nd ed.).

Shunthirasingham, C., Wania, F., MacLeod, M., Lei, Y. D., Quinn, C. L., Zhang, X., . . . Alaee, M. (2013). Mountain Cold-Trapping Increases Transfer of Persistent Organic Pollutants from Atmosphere to Cows,Äô Milk. *Environmental Science & Technology, 47*(16), 9175-9181. doi:10.1021/es400851d

Singh, U. C., & Kollman, P. A. (1986). A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH3Cl+ Cl− exchange reaction and gas phase protonation of polyethers. *Journal of Computational Chemistry, 7*(6), 718-730.

Speight, J. G. (2016). *Environmental organic chemistry for engineers*: Butterworth-Heinemann.

Sprunger, L., Proctor, A., Acree Jr, W. E., & Abraham, M. H. (2007). Characterization of the sorption of gaseous and organic solutes onto polydimethyl siloxane solid-phase microextraction surfaces using the Abraham model. *Journal of Chromatography A, 1175*(2), 162-173. doi:http://dx.doi.org/10.1016/j.chroma.2007.10.058

ter Laak, T. L., Busser, F. J. M., & Hermens, J. L. M. (2008). Poly(dimethylsiloxane) as Passive Sampler Material for Hydrophobic Chemicals: Effect of Chemical Properties and Sampler Characteristics on Partitioning and Equilibration Times. *Analytical Chemistry, 80*(10), 3859-3866. doi:10.1021/ac800258j

van den Berg, M., Denison, M. S., Birnbaum, L. S., DeVito, M. J., Fiedler, H., Falandysz, J., . . . Peterson, R. E. (2013). Polybrominated Dibenzo-p-Dioxins, Dibenzofurans, and Biphenyls: Inclusion in the Toxicity Equivalency Factor Concept for Dioxin-Like Compounds. *Toxicological Sciences, 133*(2), 197-208. doi:10.1093/toxsci/kft070

Van der Jagt, K., Munn, S., Tørsløv, J., & de Bruijn, J. (2004). *Alternative approaches can reduce the use of test animals under REACH. Addendum to the report "Assessment of additional testing needs under REACH. Effects of (Q)SARs, risk based testing and voluntary industry initiatives". Report EUR 21405. European Commission, Joint Research Centre, Ispra, Italy*. Retrieved from

van Leeuwen, S. P. J., & de Boer, J. (2008). Advances in the gas chromatographic determination of persistent organic pollutants in the aquatic environment. *Journal of Chromatography A, 1186*(1-2), 161-182. doi:http://dx.doi.org/10.1016/j.chroma.2008.01.044

Vitousek, P. M., Mooney, H. A., Lubchenco, J., & Melillo, J. M. (1997). Human Domination of Earth's Ecosystems. *Science, 277*(5325), 494-499. doi:10.1126/science.277.5325.494

Walters, D. M., Mills, M. A., Cade, B. S., & Burkard, L. P. (2011). Trophic Magnification of PCBs and Its Relationship to the Octanol−Water Partition Coefficient. *Environmental Science & Technology, 45*(9), 3917-3924. doi:10.1021/es103158s

Wang, X.-p., Yao, T.-d., Cong, Z.-y., Yan, X.-l., Kang, S.-c., & Zhang, Y. (2006). Gradient distribution of persistent organic contaminants along northern slope of central-Himalayas, China. *Science of the Total Environment, 372*(1), 193-202. doi:http://dx.doi.org/10.1016/j.scitotenv.2006.09.008

Wania, F., & Mackay, D. (1999). The evolution of mass balance models of persistent organic pollutant fate in the environment. *Environmental Pollution, 100*(1-3), 223-240. doi:http://dx.doi.org/10.1016/S0269-7491(99)00093-7

Wania, F., & Westgate, J. N. (2008). On the Mechanism of Mountain Cold-Trapping of Organic Chemicals. *Environmental Science & Technology, 42*(24), 9092-9098. doi:10.1021/es8013198

Xia, X., Hopke, P. K., Crimmins, B. S., Pagano, J. J., Milligan, M. S., & Holsen, T. M. (2012). Toxaphene trends in the Great Lakes fish. *Journal of Great Lakes Research, 38*(1), 31-38. doi:http://dx.doi.org/10.1016/j.jglr.2011.11.001

XLSTAT. (2013) (Version 2013.1.01). Paris: Addinsoft. Retrieved from http://www.xlstat.com/en/

Yang, Z.-Y., Zhao, Y.-Y., Tao, F.-M., Ran, Y., Mai, B.-X., & Zeng, E. Y. (2007). Physical origin for the nonlinear sorption of very hydrophobic organic chemicals in a membrane-like polymer film. *Chemosphere, 69*(10), 1518-1524. doi:http://dx.doi.org/10.1016/j.chemosphere.2007.05.080

Supporting information of my thesis comprises of all tables, plots, figures and other relevant detailed description of analytical tools.

Supporting information of my thesis can be accessed by the following link:

https://pern-my.sharepoint.com/:w:/g/personal/syed_0189005_talmeez_pk/EU_QS9ch2ARDkVKeVmf4IucB8GIvfsYjl8tIhD4p4WqGaw?e=UkbOAf