APPLICATION OF TWO RECEPTOR MODELS FOR THE INVESTIGATION OF SITES CONTAMINATED WITH POLYCHLORINATED BIPHENYLS: POSITIVE MATRIX FACTORIZATION AND CHEMICAL MASS BALANCE

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

FİLİZ DEMİRCİOĞLU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ENVIRONMENTAL ENGINEERING

JUNE 2010

APPROVAL OF THE THESIS:

APPLICATION OF TWO RECEPTOR MODELS FOR THE INVESTIGATION OF SITES CONTAMINATED WITH POLYCHLORINATED BIPHENYLS: POSITIVE MATRIX FACTORIZATION AND CHEMICAL MASS

submitted by **FİLİZ DEMİRCİOĞLU** in partial fulfillment of the requirements for the degree of **Master of Science in Environmental Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen ____________________
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Göksel N. Demirer ____________________
Head of Department, **Environmental Engineering**

Assoc Prof Dr. İpek İmamoğlu ____________________
Supervisor, **Environmental Engineering Dept., METU**

Examining Committee Members:

Prof. Dr. Gürdal Tuncel ____________________
Environmental Engineering Dept., METU

Assoc. Prof. Dr. İpek İmamoğlu ____________________
Environmental Engineering Dept., METU

Prof. Dr. Gülen Güllü ____________________
Environmental Engineering Dept., Hacettepe University

Assoc. Prof. Dr. Ayşegül Aksoy ____________________
Environmental Engineering Dept., METU

Dr. Barış Kaymak ____________________
Environmental Engineering Dept., METU

Date: ____________________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : FİLİZ DEMİRCİOĞLU

Signature :

ABSTRACT

APPLICATION OF TWO RECEPTOR MODELS FOR THE INVESTIGATION OF SITES CONTAMINATED WITH POLYCHLORINATED BIPHENYLS: POSITIVE MATRIX FACTORIZATION AND CHEMICAL MASS BALANCE

Demircioğlu, Filiz
M.S., Department of Environmental Engineering
Supervisor : Assoc. Prof Dr. İpek İmamoğlu

June 2010, 148 pages

This study examines the application of two receptor models, namely Positive Matrix Factorization (PMF) and Chemical Mass Balance (CMB), on the investigation of sites contaminated with PCBs. Both models are typically used for apportionment of pollution sources in atmospheric pollution studies, however have gained popularity in the last decade on the investigation of PCBs in soil/sediments. The aim of the study is four-fold; (i) to identify the status of PCB pollution in Lake Eymir area via sampling and analysis of PCBs in collected soil/sediment samples, (ii) to modify the CMB model software in terms of efficiency and user-friendliness (iii) to apply the CMB model to Lake Eymir area PCB data for apportionment of the sources as well as to gather preliminary information regarding degradation of PCBs by considering the history of pollution in the area (iv) to explore the use of PMF for both source apportionment and investigation of fate of PCBs in the environment via use of Monte-Carlo simulated artificial data sets.

Total PCB concentrations (Aroclor based) were found to be in the range of below detection limit to 76.3 ng/g dw with a median of. 1.7 ng/g dw for samples collected from the channel between Lake Mogan and Lake Eymir. Application of the CMB model yield contribution of highly chlorinated PCB mixtures (Aroclor 1254 and Aroclor 1260; typically used in transformers) as sources. The modified CMB model software provided user more efficient and user friendly working environment. Two uncertainty equations, developed and existing in literature, were found to be effective for better resolution of sources by the PMF model.

Keywords: Polychlorinated Biphenyls, Sediment, Positive Matrix Factorization, Chemical Mass Balance

ÖZ

POLİKLORLU BİFENİLLERLE KİRLENMİŞ SAHALARIN ARAŞTIRILMASINDA İKİ RESEPTÖR MODELİNİN KULLANILMASI: POZİTİF MATRİS FAKTORİZASYON VE KİMYASAL KÜTLE DENGESİ

Demircioğlu, Filiz

Yüksek Lisans, Çevre Mühendisliği Bölümü

Tez Yöneticisi : Doç. Dr. İpek İmamoğlu

Haziran 2010, 148 sayfa

Bu çalışma iki reseptör modelinin, Pozitif Matris Faktorizasyon (PMF) ve Kimyasal Kütle Dengesi (CMB), PCBlerle kirlenmiş sahaların araştırılmasında uygulamasını içermektedir. Her iki model genellikle atmosferik kirlilik çalışmalarında kirlilik kaynaklarının belirlenmesi ve ayrıştırılması için kullanılmaktadır. Ancak son on yılda toprak/sediman ve suda PCBlerin araştırılmasında da önem kazanmışlardır. Çalışmanın dört ana amacı vardır; (i) Eymir Gölü çevresinde örnekleme çalışması yaparak toplanan toprak/sediman numunelerinde PCB analizi yapılarak PCB kirlilik durumunun tespit edilmesi, (ii) literatürdeki CMB yazılımının verimlilik ve kullanılabilirlik açısından iyileştirilmesi (iii) Eymir gölünün tarihsel PCB kirlilik durumu düşünülerek PCBlerin degredasyonuyla ilgili olarak bilgi toplayıp göldeki PCBlerin kaynaklarının belirlenerek ayrıştırılması için CMB modelinin uygulanması (iv) Monte Carlo ile oluşturulmuş yapay veri setlerinin kullanılmasıyla PCBlerin hem kaynaklarının belirlenmesi hem de doğadaki akıbetinin incelenmesinde PMF modelinin kullanılmasıdır.

Toplam PCB konsantrasyonu (Aroclor bazında) Eymir ile Mogan’ı bağlayan kanaldan alınan numunelerde tespit değerinin altı ve 76.3 ng/g değerleri arasında olup medyanı 1.7 ng/g olarak ölçülmüştür. CMB modeli uygulaması yüksek klorlu Aroclor karışımlarının (genellikle transformatörlerde görünen Aroclor 1254 ve 1260) numune alma bölgesinde kirliliğe neden olduğunu göstermiştir. CMB model yazılımında yapılan değişikliklerle kullanıcıya hızlı ve kolay kullanım imkanı sağlamıştır. Geliştirilen ve literatürde var olan iki belirsizlik denkleminin PMF modelinin sonuçlarını açıklamada daha etkili olduğu görülmüştür.

Anahtar Kelimeler: Poliklorlu Bifeniller, Sediman, Positif Matris Faktorizasyon, Kimyasal Kütle Dengesi

To My Family
and Kıvanç

ACKNOWLEDGEMENTS

I would like to express my thanks to my supervisor Assoc. Prof Dr. İpek İmamoğlu for her wise guidance and her support throughout this study. She spent countless hours meeting with me and always had a positive attitude which encouraged me to overcome difficulties.

I want to express my gratefulness to academic assistant and my colleague Kadir Gedik for his direct technical and academic support. I am very appreciative of the time spent by him discussing technical concepts with me.

I also want to express my thanks to my colleague Nazan Akduman for her technical support in laboratory.

I would like to offer my thanks to academic assistant Güray Doğan for his help and guidance for PMF model.

I would like to express my gratitude to my family for their endless loving support throughout this study.

Last, but not least, thanks to Kıvanç. He spent his all weekends with me in laboratory and library throughout this study. The completion of this thesis would not have been possible without his support and love.

TABLE OF CONTENTS

CHAPTERS

APPENDICES

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

Polychlorinated biphenyls (PCBs) are persistent organic pollutants which were produced due to their chemical and physical stability. Because of these properties, these chemicals had a widespread use in many industries. After these chemicals were discharged into the environment directly or unintentionally, PCBs were understood to be affected by physicochemical and biological degradations in the environment and by this way, they can pass from one media to another (water, air or soil). Another important point for PCBs is that negative health effects on human and animals have been seen due to accumulation of these chemicals in the tissue. Therefore, it is important to determine the contaminated sites, to examine the sources of the pollution by using receptor models, to discuss fate of the pollutant in the contaminated site and so, to decide and develop effective remediation technologies for these sites.

This study examines the application of two receptor models, namely Chemical Mass Balance and Positive Matrix Factorization, on the investigation of sites contaminated with PCBs. Both models are typically used for apportionment of pollution sources in atmospheric pollution studies, however has gained popularity in the last decade on the investigation of PCBs in soil/sediment and water. Hence, the aim of the study is four fold:

1. Identify the status of PCB pollution in the vicinity of Lake Eymir via sampling and analysis of PCBs in collected soil/sediment samples.

2. Modify the CMB model software for use in PCB data sets so that a more efficient and user-friendly version is developed.

3. Apply the well-established CMB model to Lake Eymir area PCB data for apportionment of the sources as well as gather preliminary information as to the degradation of PCBs by taking into account the history of pollution in the area.

4. Explore the use of PMF for both source apportionment and investigation of fate of PCBs in the environment via use of Monte-Carlo simulated artificial data sets.

Chapter 2 gives background information about PCBs, chemical mass balance model (CMB) and positive matrix factorization (PMF). Physicochemical properties of congeners and PCB mixtures are presented for the identification of these organochlorines, giving a review of studies from Turkey and trends in the world. Lastly, two receptor model approaches are discussed by using recent references of previous applications.

Chapter 3 describes field and modeling studies. First, background information on sampling, laboratory studies and the study area is presented. The study area is the vicinity of Lake Eymir and TEDAS transformer repair and maintenance facility. This region is suspected to be polluted due to leakage and discharge of PCB containing oils in transformers by the channels during repair and maintenance of transformers. Next, generation of the artificial data sets by Monte Carlo Simulation are discussed for model validation. Then, the CMB model as modified in this study is explained in detail, giving the input, output and flowchart of the computer program. Lastly, PMF is discussed, including development of input files, use of parameters during simulation of the model and goodness of fit parameters.

In Chapter 4, results and discussions are presented. The analysis and CMB model results for the sampling site are discussed. Another discussion is done for the sensitivity of the PMF model by using the artificial data sets generated in varying complexity by Monte Carlo simulation.

Chapter 5 discusses conclusions of the study. Additionally, the overall conclusions obtained from the CMB model regarding its application on environmental PCB data are presented. Finally, the results on sensitivity of PMF model are emphasized.

CHAPTER 2

LITERATURE REVIEW

2.1 Polychlorinated Biphenyls (PCBs)

The Polychlorinated biphenyls (PCBs) are a class of chlorinated synthetic organic chemicals. As seen in Figure 2-1, it forms from attachment of two phenyl rings and has 12 carbons and 10 atoms which can be either chlorine or hydrogen (C12 H10-nCln). According to chlorine position and number, there are theoretically 209 PCBs each of which is called as “congener”. 2, 2, 6 and 6’ are called as ortho positions, 3, 3’, 5 and 5’ the meta positions and 4 and 4’ the para positions.

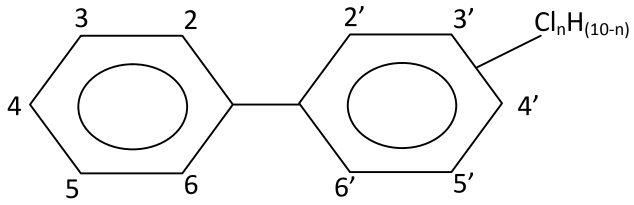


Figure 2-1 General Structure Formula of PCBs

The most accepted naming for congeners is by Ballschmiter and Zell (referred to as B&Z or BZ). The IUPAC nomenclature, the other naming, differs from BZ (Hansen, 1999).). The reason is that "primed" number is considered higher than the corresponding "unprimed" number to IUPAC, however, according to

B&Z, number of primed chlorine positions may never be greater than the number of unprimed positions; e.g.: Ballschmiter specifies the name 2’, 3, 4-trichlorobiphenyl due to more primed position than unprimed one, but in the IUPAC name is 2, 3’, 4’-trichlorobiphenyl (EPA, 2009). IUPAC naming will be used in this study (Appendix A).

PCBs were manufactured between 1930 and 1993 as mixtures, not as individual congeners. These mixtures include between 80 and 110 congeners. Austria, China, Czechoslovakia, France, Germany, Italy, Japan, the Russian Federation (1939-1993) Spain, the United Kingdom and USA (1930-1977) are the countries that produced PCB mixtures (UNEP, 1999). PCBs were produced and sold under specific names. The most common ones were "Aroclor" (USA) series. As looked at the number in the series, chlorine percentage was distinguished. E.g. 1254 Aroclor means that it contains 12 carbons and 54% chlorine by weight (EPA) and this mixture is produced due to good conductivity and nonflammable (ATSDR, 2000). Other common trade names are Kanechlors (Japan) and Clophens (Germany) (Hansen, 1999).

2.1.1 Physicochemical Properties of Congeners

Individual chlorobiphenyls are generally solids. Within mixtures, the physical states are changed depending on weight of chlorinated products: the most lightly chlorinated products are liquid when the heaviest ones are resinous or waxy. The more lightly ones are lost more rapidly during evaporation

PCBs can resist heat due to high boiling point and do not easily break down. In Table 2-1, physico-chemical properties of six congeners of polychlorinated biphenyls are shown. These congeners are given as an example and reference in order to provide general information. As seen in Table 2-1, their octanol water partitioning coefficients (K_{ow}) are high with low aqueous solubility. Additionally, as chlorine substitution increases, $logK_{ow}$ value increases and solubility of congener in water decreases. It indicates that higher chlorinated congeners have affinity for sediment or soil when compared to water. And they are likely to be sorbed to sediment and soil due to their hydrophobicity. Lower solubility while increasing number of chlorine in congener shows decrease in possibility of volatilization. Due to high $logK_{ow}$ value (lipophilicity), they can be bioaccumulated in the tissue by passing through the food chain. Lower chlorinated congeners have higher vapor pressure and higher vaporization tendency.

Table 2-1 Physical and Chemical Properties of Six PCB Congeners

IUPAC No	**Structure**	**Homolog Group**	**Chemical Formula**	**Molar Mass (g/mol)**	**Melting Point °C**	**Vapor Press. (Pa)**	**Aqueous Solub. (g/m^3)**	**$logK_{ow}$**
PCB-7	24	Di-chlorobiphenyl	$C_{12}H_8Cl_2$	223.1	24.4	0.254	1.25	5
PCB-15	4-4’	Di-chlorobiphenyl	$C_{12}H_8Cl_2$	223.1	149	0.0048	0.06	5.3
PCB-29	245	Tri-chlorobiphenyl	$C_{12}H_7Cl_3$	257.5	78	0.013	0.14	5.6
PCB-52	25-2’5’	Tetra-chlorobiphenyl	$C_{12}H_6Cl_4$	292.0	87	0.0049	0.03	6.1
PCB-101	245-2’5’	Penta-chlorobiphenyl	$C_{12}H_5Cl_5$	326.4	76.5	0.001	0.01	6.4
PCB-153	245-2’4’5’	Hexa-chlorobiphenyl	$C_{12}H_4Cl_6$	360.9	103	0.00012	0.001	6.9

Source: (Johnson, Quensen, Chiarenzelli, & Hamilton, 2005)

2.1.2 Physicochemical Properties of Mixtures

Johnson PCBs were produced in hundreds of industrial and commercial applications due to useful properties: fire resistance, low electrical conductivity, high resistance to thermal breakdown, high degree of chemical stability, resistance to many oxidants and other chemicals, low water solubility high tendency to dissolve in fats, hydrocarbons and other organic compounds (UNEP, 1999).

PCBs have persistent property in the environment because of their long half-time. Therefore they can pass one media to another (air, soil or sediment). Physico-chemical properties are presented in Table 2-2. Like in congeners, as the chlorine content increases, water solubility and volatilization decrease.

Table 2-2 Some physical properties of Aroclor mixtures

Aroclor	**State**	**Density @20°C**	**Viscosity@ 99°C**	**Vapor Loss 6 hr@ 100°C**	**Water Solub. µg/L @25°C**
1221	Oil	1.18	30-31	1-1.5%	
1232	Oil	1.27	31-32	1-1.5%	
1016	Oil	1.37			420
1242	Oil	1.38	34-35	0-0.4	240
1248	Oil	1.44	36-37	0-0.3	52
1254	Viscous	1.54	44-48	0-0.2	12
1260	Resin	1.62	72-78	0-0.1	3

Source: (Hansen, 1999)

Henry’s constants of Aroclors 1221, 1016, 1254, 1260 and 1268 are 32.8, 3.4, 273.6 and 719.4 respectively. They increase as chlorine percentage increases. As described by Henry’s constant, congeners with lower chlorine atoms have relatively higher vapor pressure coupled with a volatilization tendency.

PCBs have persistent property in the environment. Due to this property, when chlorination increases, both stability and lipophiliticity will also increase. In other words, the stability which is other property of PCBs causes their slow elimination in the environment (Hansen, 1999).

2.1.3 Fate of PCBs in the Environment

PCBs have three applications in the marketing; (i) closed application e.g. capacitors, transformers, heat exchangers, etc. (ii) semi-closed and (iii) open applications e.g. lubricants, plasticizers, ink solvents, adhesives, carbonless copy papers, etc. PCBs were disposed to the environment with variety of pathways after use. These pathways which with PCBs may enter to the environment may be illegal disposal/discharge of PCB wastes, accidental releases and past open/partially open and uncontrolled uses (Erickson, 1997). Persistence of PCBs in the environment was firstly recognized by Jensen’s research in 1966 (Jensen, 1966). Then, production of PCBs was banned firstly in 1977 in USA, lastly in 1993 in Russia (Breivik et. al., 2002). PCBs were thought to be stable in the environment until 1980s, however, after development of analytical detection techniques, it was understood that PCBs undergo degradation mechanisms in the environment and bioaccumulation in the fat tissues. These mechanisms are physicochemical and biological processes.

There are two degradation mechanisms affecting PCB congeners in sediments; anaerobic dechlorination and physicochemical weathering. Anaerobic dechlorination occurs biologically in the result of the destruction of higher chlorinated congeners as producing lower chlorinated ones which are less toxic and more biodegradable PCB congeners. However, physicochemical weathering only affects the distribution of the congeners in the environment without eliminating the contamination (Bzdusek, 2005).

Physicochemical Weathering Processes

PCBs which are semi-volatile compounds, readily undergo redistribution in the environment by a variety of processes, including volatilization, atmospheric transport, and wet and dry deposition. Their fate depends on the degree of chlorine. Generally, PCBs that are the most susceptible to alteration are the less chlorinated congeners and the Aroclors that are the most susceptible to alteration are the lighter Aroclors (Aroclors 1016, 1242, 1248) (Johnson, et. al., 2005). Additionally, PCBs may partition into different media (Gouin et. al., 2000). Lightly chlorinated congeners can transport in the aqueous or vapor phase. Ortho-rich PCB congeners can be volatilized easily when compared to nonortho congeners (Johnson et. al., 2005).

If released to the environment, PCBs adsorb strongly onto soil, with generally increasing tendency as the degree of chlorination increases. Vapor loss of PCBs from soil surfaces appears to be an important fate mechanism with the rate of volatilization decreasing with increasing chlorination (Chiarenzelli et. al., 1997). Although the volatilization rate may be low, the total loss by volatilization over time may be significant because of the persistence and stability of PCBs. If released to water, adsorption to sediment and suspended matter will be an important fate process; PCB concentrations in sediment and suspended matter have been shown to be greater than in the associated water column. Although adsorption can immobilize PCBs (especially the higher chlorinated congeners) for relatively long periods of time, eventual resolution into the water column has been shown to occur. If released to the atmosphere, PCBs will primarily exist in the vapor-phase; the tendency to become associated with the particulate-phase will increase as the degree of chlorination of the PCB increases. Physical removal of PCBs from the atmosphere, which is very important environmentally, is accomplished by wet and dry deposition (EPA, 2009).

Biological Degradation

The persistence of PCBs increases with an increase in the degree of chlorination. Mono-, di- and trichlorinated biphenyls biodegrade relatively rapidly, when compared to tetrachlorinated and higher chlorinated biphenyls. Although biodegradation of higher chlorinated congeners may occur very slowly on an environmental basis, no other degradation mechanisms have been shown to be important in natural water and soil systems; therefore, biodegradation may be the ultimate degradation process in water and soil (Pakdeususuk, 2002). There are two biological processes, aerobic and anaerobic.

Aerobic degradation: When congeners including chlorine less than five, aerobic degradation will be possible, however, reductive processes are valid. Aerobic bacteria which live in oxygenated environments and anaerobic bacteria which live in oxygen free environments are capable of using lower chlorinated PCBs as C-source in aquatic environment. The main degradation product is 4-chlorobenzoic acid. In general, formation of chlorinated benzoic acids is a major degradation pathway for aerobic degradation of PCBs. PCB congeners can be dechlorinated under anaerobic conditions. Reductive dechlorination of PCBs act to remove chlorines from a biphenyl ring and chlorine is replaced with hydrogen atom as seen Figure: 2-2. It is a two electron transfer process in H2 or reduced organic substrates are assumed to be electron donors and water is thought to be a proton source.

Anaerobic degradation: If congeners including chlorine less than five, series of hydroxilations and ring cleavage are possible. Under anaerobic conditions, PCBs are electron acceptors (Pakdeususuk, 2002).

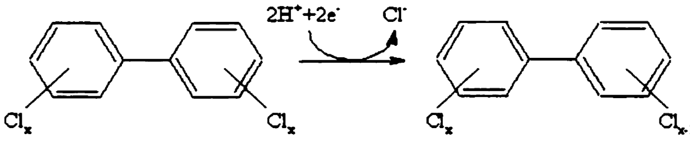


Figure: 2-2 General Scheme of Dechlorination of PCBs

As a result, studies on this fate of PCBs investigate which mechanisms PCBs undergo in the environment. According to congener specific analysis results of contaminated sites, fate of PCBs can be investigated via modeling tools.

2.2 Current Regional/Global Trends in PCB Pollution

The Stockholm Convention is a global treaty accepted to ban the uses and the production of persistent organic pollutants (POPs) due to effects on environment and human health. By the Convention effectiveness, a global status has been developed for environmental background concentrations. The monitoring of global trends in POPs concentrations is the Convention’s first objective to protect human health and environment. Therefore, air monitoring, human exposure and other environmental data in soil, sediment and water are used to evaluate POPs pollution.

As a result of the monitoring studies in soil/sediment and air media, in this part, it is dealt with background information about PCB pollution in various media. Gedik and Imamoglu (2010) give summarized table about these global trends in their study. They give background PCB levels in various media, with an emphasis on soil, sediment, air, freshwater and mussel. The soil sampling is in EMEP station in Czech Republic between 1996 and 2005, Central and Southern Europe (CEE) in 2005, background surface soils across the world in 1998, Rural/remote soils across the US in 2003, Great Britain and Lithuania in 2006. The sediment sampling is EMEP in station in Czech Republic between 1996 and 2005, CEE between 2005 and 2006, and Baltic Sea between 1968 and 1997. The air sampling is in Lithuania in 2006, EMEP station in Czech Republic between 1996 and 2005, CEE in 2005. The freshwater sampling is in EMEP station in Czech Republic between 1996 and 2005. The mussel sampling is in Asia-Pacific in 1994, 1997, 1998, 1999, 2001. That study indicates the global trends in the world. The concentrations, sampling frequency, survey year, number of samples measured are given in detail in Gedik and Imamoglu (2010).

2.3 Pollution Monitoring Regions in Turkey

PCBs were never produced in Turkey, however, they were used as PCB oils in closed equipments and for open/partially open applications. Turkey started to use PCBs approximately in 1960s. In Turkey, the legal prohibition and restriction on use of PCB equipments and oils started after Stockholm

Convention. The current regulations are listed in Table 2-3. However, there has not been a regulation for contaminated sites yet.

Table 2-3 An overview of the current national regulatory actions on/mentioning PCBs (Gedik and Imamoglu, 2010)

Year/Official Gazette No.	By-law	Medium/Acceptor	Regulating limit
2005/25755	Control of Hazardous Wastes	Liquid, Waste oil, Equipment	<1 mg/kg
2005/25831	Control Soil Pollution	Soil, Waste Sludge, Compost	<0.5 mg/kg
2007/26739	Control of PCBs and PCTs	Material, Equipment	<5 L for equipment <50 ppm for material
2008/26952	Control of Waste Oils	Waste oil	CategoryI< 10 ppm CategoryII< 50 ppm CategoryIII> 50 ppm
2008/27092	Restrictions on the Production, Marketing and Use of Certain Dangerous Substances and Preparations	Any type of product	Not allowed to be produced or marketed
2009/27277	Control of Air Pollution Originating from Industrial Activities	Stack gas	<0.1 ng/Nm^3

There is no monitoring station in Turkey for PCBs. For this reason, background PCB concentrations don’t exist. However, official import records, official governmental records, and reports prepared by international organizations and scientific literature help PCB contribution to evaluate. By this way, PCB distribution is determined in seas as measuring in water and hot spot regions are Izmit Bay, Ordu and Sinop. Another distribution media is soil and regions are Ankara, Aliaga, Golbasi Ankara and Iskenderun Hatay. Izmir Aliaga is potential area for PCB contribution into air from electric arc furnaces used during steel making. In biological media, Gedik and Imamoglu (2010) states that the monitoring was done for marine organisms in Mediterranean Sea, İstanbul, Mersin, Kahramanmaras, etc; fish species in Mediterranean Sea; mussel in Izmir Bay and mid-black sea; adipose tissue in Ankara; and human milk in Ankara, Kahramanmaras, İstanbul, etc. In those regions, PCBs in measurable level are indicated.

In the scope of this study, the region between Lake Eymir and TEDAS repair and maintenance center was examined for PCB pollution. Eymir Lake and its vicinity are in the boundary of Special Environmental Protection Area, which is a region in Gölbaşı, Ankara. Various studies carried out in the area showed that the area is contaminated with domestic and industrial wastes. In the study area and its vicinity, there are small-scale industries and TEDAŞ Gölbaşı Maintenance and Repair Facility where, among other activities malfunctioning transformers from various power plants in Turkey were maintained. PCBs were not produced in Turkey, but they have been used. Therefore, transformers were brought from countries which use PCBs in the manufacturing process (UNIDO, 2006). Then, when needed, these transformers used in power plants around Turkey were brought to repair this facility of TEDAS to be opened for repair and/or change of oil.

Potential of a PCB related contamination is first pointed out by Yeniova in 1997. Yeniova (1998) indicates high level of PCB pollution in this region due to TEDA*S* maintenance facility. Therefore, TEDAS was chosen as hot spot region as a part of this study.

2.4 Receptor Modeling

Receptor modeling based on mass conservation is used to apportion sources in the environment and to identify source composition to each sample and altered source patterns (Hopke, 1991 and Henry et. al., 1984). There are two types of mass approach models; chemical mass balance-CMB (source known) and multivariate models (source unknown) such as Principal Component Analysis-PCA, and Positive Matrix Factorization-PMF, UNMIX, SAFER, Cluster Analysis, etc. In this study we have used CMB and PMF models.

2.4.1 Chemical Mass Balance, CMB

The chemical mass balance model is a receptor modeling technique that uses the concept of mass balance approach to solve for contributions from different sources to a sample at a receptor site. The equation of the model is expressed by (Rachdawong, 1997) as;

$$F_j = \sum_{i=1}^{n} \phi_{ji} a_i + e_j \qquad (1<j<m) \tag{2-1}$$

where

F_j = Measured concentration of the j^{th} chemical species in the sample

Φ_{ji}= Concentration of the j^{th} chemical species in the i^{th} source

a_i = Source contribution factor of the i^{th} source

e_j = Error associated with the j^{th} chemical species

n= number of sources

m= number of chemical species used in the model

The chemical mass balance model indicates contributions from different sources to a sample at a receptor site. There are two basic assumptions for the model: (i) the amount of a chemical in a sample at the receptor site is equal to sum of the amount of this chemical coming from each source, and (ii) The concentrations of a source coming from sampling location and the source material have the same ratio for each chemical considered and it is expressed by a contribution factor (Li et. al., 2001).

The basic idea of the CMB is that a given measured profile can be reproduced by linear combinations of several sources. Model is based on a χ2-weighted least squares procedure, where the numbers of variables (congeners) exceed the numbers of sources (Aroclors), m>n. The weighted error (equal to χ2) is minimized in a least squares method used in the multi linear regression by which contribution factors (a_i values) are calculated in the model (Su, 1997). The relative contribution Pi from source is calculated for each source according to:

$$P_i = \frac{a_i \sum_{j=1}^{m} \phi_{ji}}{a_1 \sum_{j=1}^{m} \phi_{j1} + a_2 \sum_{j=1}^{m} \phi_{j2} + ... + a_n \sum_{j=1}^{m} \phi_{jn}} \qquad (2\text{-}2)$$

Goodness of fit parameters are χ2 and multiple correlation coefficient R^2 used to assess fit between predicted and measured data. During calculation of χ2, relative errors are assumed by constant coefficient variation in the uncertainty of sample and source profiles. The multiple correlation coefficient R^2 is expressed as Su, (1997):

$$R^2 = \frac{\sum_{j=1}^{m} w_j F_j^{'2}}{\sum_{j=1}^{m} w_j F_j^2} \tag{2-3}$$

where

F_j'=Calculated concentration PCB species = $\sum_{i=1}^{n} \phi_{ji} a_i$

F_j= Measured concentration of PCB species

w_j=Weighting factor= $\frac{1}{\sigma_j^2}$

σ_{ij}=Uncertainty of the measured value of the ^{jth} PCB species

In the CMB used in this study, we have used the effective variance least-squares method discussed by Henry et. al. (1984). In this method, errors are associated with concentration of chemical species in both measured observables and sources. Therefore, the χ2 is computed based on

$$\chi^2 = \sum_{j=1}^{m} \frac{(F_j - F_j^{'})}{(r.e.)_m^2 F_j^2 + \sum_{i=1}^{n} (r.e.)_i^2 (a_i \Phi_{ji})^2} \tag{2-4}$$

where F_j and F_j' are the measured and calculated concentrations, respectively, of j^{th} PCB in the sample. This equation is derived from equation (11) of Henry et. al. (1984) using relative errors $(r.e.)_m$ and $(r.e.)_i$ of the measurement and the source profile, respectively.

$$(r.e.)_m = \frac{\sigma F_j}{F_j} \tag{2-5}$$

$$(r.e.)_i = \frac{\sigma \Phi_{ji}}{\Phi_{ji}} \tag{2-6}$$

where σ is standard error

χ2 should be equal to the number of degrees of freedom (df) in order to obtain a good fit between measured and predicted values (m-n=df) (Su, 1997). For the measurement error, six values are tried (0-0.1-0.3-0.5-0.6-0.9). Estimation of the actual (r.e.)$_m$ corresponding to χ2=df are made by inter- or extrapolation. To understand if fit is satisfactory, the value of measured error corresponding to χ2 is investigated. When the measured relative error is less than 50%, the fit is considered to be satisfactory. The fit is considered acceptable if the relative error is between 50 and 70%, and is inadequate for larger relative errors (Rachdawong, 1997, Su, 1997 and Christensen et. al., 1997).

Uncertainties in source profiles are calculated by the following equation:

$$\delta P_i = P_i \sqrt{(\frac{\delta N}{N})^2 + (\frac{\delta D}{D})^2} \tag{2-7}$$

where Pi=N/D

N= Numerator in equation (2-2)

D= Denominator in equation (2-2)

$$\delta N = \delta a_i \sum_{J=1}^{M} \varphi_{ji} \tag{2-8}$$

$$\delta D = ((\delta a_1 \sum_{j=1}^{m} \varphi_{j1})^2 + (\delta a_2 \sum_{j=1}^{m} \varphi_{j2})^2 + (\delta a_3 \sum_{j=1}^{m} \varphi_{j3})^2 + \ldots + (\delta a_n \sum_{j=1}^{m} \varphi_{jn}) \qquad (2\text{-}9)$$

where, δa_i is calculated as in Henry et. al. (1984)

Then, relative contribution P_i of a source is calculated as:

$$P_i = P_i \pm \delta P_i \qquad (2\text{-}10)$$

Use of CMB in modeling of environmental pollution studies (e.g. PAHs, PCBs)

CMB model was typically used in air pollution studies. Later its use in other fields such as in aquatic environment was demonstrated (Christensen et. al., 1997, Li et. al., 2003, Christensen et. al., 1999, Li et. al., 2001, etc.). The use of the model for application of pollution in aquatic sediment samples was shown by Christensen and his team for aquatic environment and this study is explained in Su et. al. (2000). They used CMB model apportionment of PAH sources in sediments. Other PAH applications are Su (1997), Su et. al. (2000), Li et. al. (2001), Cheng & Hopke (1986), Li et. al. (2003), and Bzdusek & Christensen (2004). Additionally, EPA CMB 8.2 is used for the source estimation of sediment PAHs by Li et. al. (2003).

In the literature, application of the CMB model is present for the investigation of PCB pollution in soil, air or sediments (Imamoglu et. al., 2002, Cetin et. al., 2007, Honda et. al., 2008, etc.). In Turkey, there is only one CMB model application for PCBs. It has been performed by Cetin et. al., 2007 to partition sources of atmospheric PCBs around an industrial area in İzmir and EPA’s CMB 8.2 has been used for this purpose.

The FORTRAN code of the CMB model, as applied by Imamoglu (2001) was converted to excel VBA to be used in the present study.

2.4.2 Positive Matrix Factorization Model (PMF)

The PMF model was introduced by Paatero and Tapper (1994) in 1994 and in the following years, it has found many applications in various areas for source apportionment. It was applied generally in atmospheric studies (Lee et. al. (1999), Hopke (2003), Juvela et. al. (1996), Amota et. al (2009), Du et. al. (2007), etc.) However, in the previous years, it was also used in sediment and soil investigations (Buzdusek et al, 2006a, Buzdusek et. al., 2006b, Du et. al., 2008, etc.). PMF do not require previous knowledge of sources like the CMB model.

In 1997, Paatero (1997) introduced PMF2 which is a two dimensional model. The equation of PMF is expressed as;

$$X_{ij} = G_{ik}F_{kj} + E_{ij} \qquad (2\text{-}11)$$

where

X= Data Matrix

G= Score Matrix (source contribution)

F = Loading Matrix (source profile)

E=Error matrix

In equation (2-11), data matrix, X, is separated to G and F matrixes. E is used to add the error as a result of multiplication of F and G matrices. Besides, i, j and k indicates number of congeners, samples and sources respectively (Paatero & Tapper, 1994). The aim of the PMF model is to minimize Q(E) with the constraint of keeping non-negative element of G and F factors. Hopke (2010).

$$Q(E) = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{E_{ij}}{\sigma_{ij}}\right)^2 \qquad (2\text{-}12)$$

$$E_{ij} = X_{ij} - \sum_{k=1}^{p} G_{ik}F_{kj} \qquad (2\text{-}13)$$

i=1,….,m j=1,….,n

E_{ij} is an estimate of the uncertainty in the ith variable measured in the jth sample. Hopke (2010) states that preparation of uncertainty file is the critical point in the PMF analysis.

The advantage of the PMF model is individual weighting of data by uncertainties. It enables objective function to reduce emphasis on data points having high uncertainties. Therefore, the below detection limit and missing values do not have much effect on factor formation (Polissar et. al., 2001).

Generally, sources of the samples taken from many locations are not known. In this case, statistical models like FA, PVA, UNMIX and PMF provide advantages to find the sources of pollution. PCBs cause widespread pollution as transporting to long range and accumulating in sediments. The studies indicate that they undergo biological and physicochemical degradations although they are known to be persistent in the environment. Chiarenzelli et. al. (1997) states in his study that lower chlorinated PCB congeners having relatively higher volatility can partition into atmosphere from sediment/soil by physicochemical weathering and they can be transported long ranges. Another possibility is the biological degradation within the depth sediment. Bedard and May (1996) point out partial dechlorination of Aroclor 1260 in Wonds Pond sediments. In this scope, the studies in recent years indicate that receptor models help to identify the fate and degradation of PCBs in the environment. One model is PMF application for PCBs; however, few studies are present on source explanations of PCB contamination in soil or sediments by PMF since PMF was used for atmospheric pollution studies until the 2000s.

Bzdusek and Christensen (2006a) compared a new variant of PMF model with factor analysis (FA) and other existing methods of PMF using an artificial data set (20 congeners by 40 samples) and a real data set (32 congeners by 106 sediment samples of the Sheboygan River). New variant of PMF model was developed with and without penalty terms. Their study showed that this model provided more realistic source profiles than FA. When PMF new variant was compared with PMF2 model with penalty terms (Paatero, 1997) in the literature, PMF2 was difficult to reproduce nonzero values, however, the source profiles were still distinguishable.

Du and Rodenburg (2007) partitioned sources of gas-phase PCBs in atmosphere of Camden, NJ, USA by PMF since it caused atmospheric depositions of PCBs to the tidal Delaware River. The tidal Delaware River was also examined for PCBs in river sediments (32 and 58 PCB congeners measured in between 32 and 106 samples) by PMF model to identify covarying congener patterns that are associated with sources (Du et. al., 2008).

CHAPTER 3

MATERIALS AND METHODS

3.1 Field Studies

3.1.1 Sampling Site Description

In On 22.10.1990, the Council of Ministers issued the decision to declare the marshy areas of Gölbaşı as a Special Protected Zone (SEPZ) (decision # 90/117). The area located 20 km south of Ankara is declared as SEPZ with respect to 9th clause of the Environmental Protection Law; due to their ecologic and recreational significance. In the study area and its vicinity, there are small-scale industries and TEDAŞ Gölbaşı Maintenance and Repair Facility where, among other activities malfunctioning transformers from various power plants in Turkey were maintained (Figure 3-1). As mentioned in the second chapter, PCBs were not produced in Turkey, but has been used. Therefore, transformers were brought from countries which use PCBs in the manufacturing process (UNIDO, 2006). Then, when needed, these transformers used in power plants around Turkey were brought to repair this facility of TEDAS to be opened for repair and/or change of oil. Potential of a PCB related contamination in the area was first pointed out by Yeniova (1998).

In the National Implementation Plan (NIP) prepared at 2006 by the Ministry of Environment and Forestry, officials from EÜAŞ Environment – New and Renewable Energy Sources Department stated that until 1995 damaged transformers were opened for repair and/or dielectric fluid replacement but after that time the workers were notified for not opening PCB containing transformers. Officials also stated that before 1995, malfunctioning transformers from various places of Turkey were collected in one place; Gölbaşı Maintenance and Repair facility and only opened in this place. It is presumed that before this date PCB containing transformers were opened here and grease in the transformers are discharged without any control. In 1997, Yeniova (1998) took 15 samples from inside the facility and observed Aroclor 1260 mixture with a dilution value of 0,53 – 464 µg/g (average 105±150 µg/g, median 50 µg/g) in all of these samples. 3 of these samples were taken from the exact area where transformer oil was renewed and in these samples a PCB concentration of 383±78 µg/g is found. Other 12 samples were taken from the vicinity which confirmed gradually decreasing concentration values. Turkish regulations applied today, classifies soils with a PCB contamination higher than 1 µg/g as polluted soils.

Above mentioned facility is located at the Gölbaşı, adjacent to the Lake Eymir territory of METU and to the channel which connects Mogan and Eymir lakes. This channel carries excess water of Lake Mogan to Lake Eymir and it is known that increasing and decreasing amounts of water is flowed through this channel over the years. Yeniova (2008) under the scope of the same study has collected 20 samples from the channel during the summer period when water levels are decreased. Although concentrations in the sediment samples are lower she could observe PCB as Aroclor 1260 mixtures (average 26.6±58.5 ng/g, median 3.9 ng/g). Results are the proof of waste entry from the facility to the channel. This also indicates that any PCB containing waste could be entered to the Lake Eymir through this channel during 20 years period.

Starting from the findings of Yeniova (2008), Akduman (2009) took soil and sediment samples from the territory on July 2007. Studies shows TEDAŞ facility originated PCB penetration to Lake Eymir. Historical information suggests cease of PCB entry into the lake approximately 10 yrs ago. Sampling of soil and sediment should take this into account, such that, for example sampling from sediment surface and within sediment depth may be required. Besides, measurement of high PCB concentrations in the soil samples taken from inside the TEDAŞ facility suggests that these chemicals could still emerge due to old transformers. Additionally observation of 84 ppb and 45 ppb PCB concentrations from the surface sediments taken from the lake indicates the transportation of these chemicals to the Lake Eymir in the recent years. Usage of these chemicals is prohibited with the regulations on 1996 but usage in the closed systems like transformers is still legal till 2025. In this time period opening the transformers containing these chemicals without taking proper caution or replacing the oils in them can result in inference of PCBs to the environment.

Starting from these findings, this region has been chosen as the potential area polluted by PCBs.

3.1.2 Sampling Study

Sampling was mainly concentrated on the channel between Lake Mogan and Lake Eymir, which is suspected to carry PCBs into the lake.

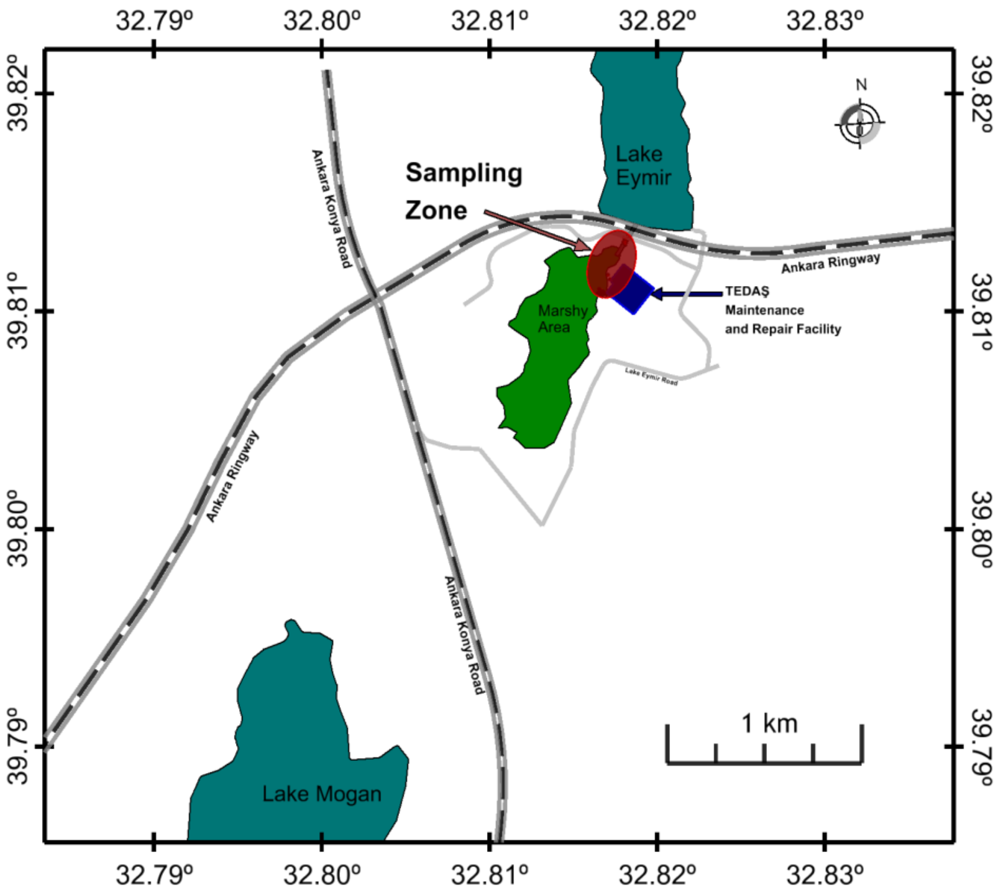


Figure 3-1 Location of the Sampling Site near Lake Eymir

A total of 29 soil/sediment samples were taken from the marshy area shown in Figure 3-2 in two sampling campaigns (02.07.2008 and 24.09.2008). The samples were collected from two regions: first is the region where there is a channel connecting to Eymir and second is in and the vicinity of a channel near boundary of TEDAS. GPS Coordinates are indicated in Table 3-1 for each sampling location. Since the sampling points are very close to each other, the uncertainty of the GPS used may be of concern. For this purpose, the coordinates of the sampling locations were checked using Google Earth.

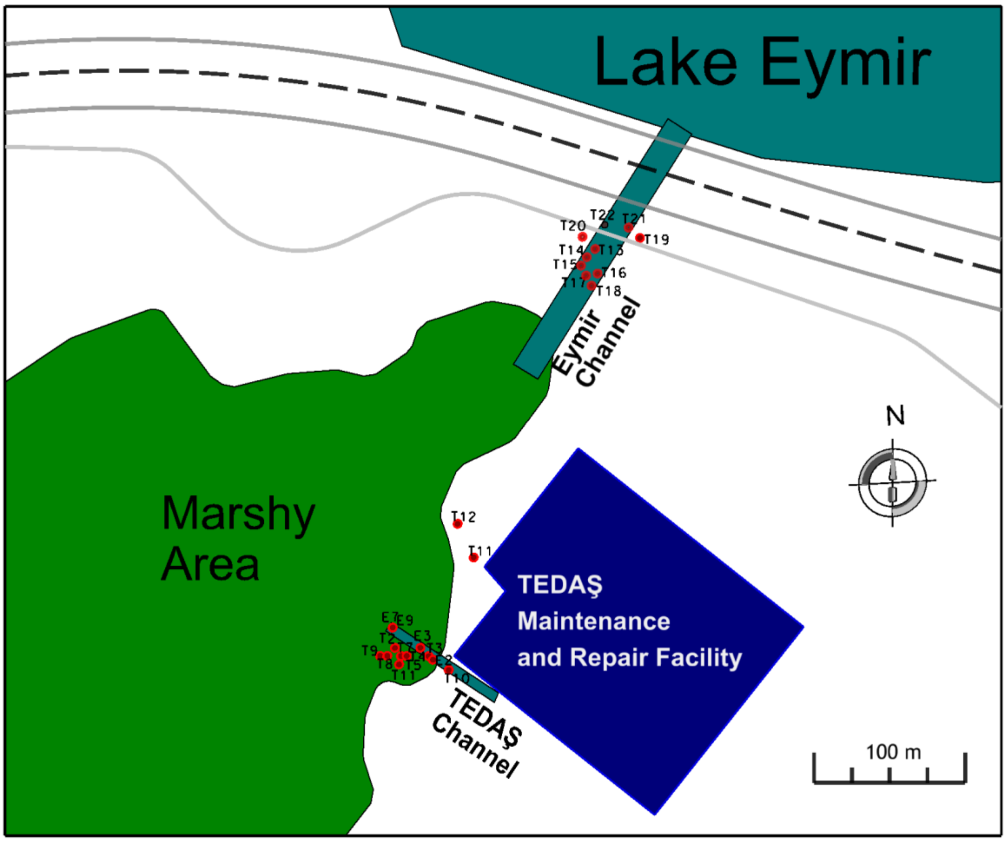


Figure 3-2 Sampling Points Vicinity of Lake Eymir and TEDAS

Sampling was focused on the channel between TEDAS and Lake Eymir. The sampling zone can remain under water seasonally, so the samples are in wet or dry as understood from moisture content of them in Table 3-1. In July, 21 samples were collected from various outlets of the TEDAS discharge channel and dried Eymir sediments in the vicinity. In September, 8 samples were collected from the mouth of the channel known to contain PCBs. In September, samples were collected in cores to be able to understand the effect of the depth to concentrations of PCBs. Detail information is given about samples and results in section 4.1.

Table 3-1 Coordinates, Moisture Content and Organic Matter Content of Samples Taken from the Vicinity of the Channel between Lake Eymir and Lake Mogan

Sample No	**Moisture Content (%)**	**Organic Content (%)**	**Latitude**	**Longitude**
T1	-	-	39°48'38.32"N	32°48'57.15"E
T2	-	-	39°48'38.67"N	32°48'57.03"E
T3	65.7	16.4	39°48'38.50"N	32°48'57.97"E
T4	45.5	7.5	39°48'38.77"N	32°48'57.36"E
T5	43.4	7.7	39°48'38.77"N	32°48'57.36"E
T7	47.2	3.0	39°48'39.14"N	32°48'57.20"E
T8	50.6	11.2	39°48'39.20"N	32°48'56.82"E
T9	23.2	4.4	39°48'39.66"N	32°48'56.62"E
T10	71.6	16.0	39°48'38.20"N	32°48'58.53"E
T11	27.9	5.0	39°48'40.62"N	32°48'59.22"E
T12	21.7	4.3	39°48'41.34"N	32°48'58.77"E
T13	35.7	9.6	39°48'47.26"N	32°49'2.58"E
T14	21.1	4.5	39°48'47.08"N	32°49'2.35"E
T15	35.4	6.8	39°48'46.90"N	32°49'2.19"E
T16	24.8	4.3	39°48'46.73"N	32°49'2.65"E
T17	32.0	5.9	39°48'46.69"N	32°49'2.33"E
T18	28.0	5.6	39°48'46.47"N	32°49'2.48"E
T19	27.4	5.0	39°48'47.50"N	32°49'3.83"E
T20	22.0	3.3	39°48'47.52"N	32°49'2.23"E
T21	24.6	8.1	39°48'47.72"N	32°49'3.51"E
T22	55.9	6.8	39°48'47.78"N	32°49'2.85"E
E1	26.7	4.5	39°48'38.42"N	32°48'58.09"E
E2	26.9	4.9	39°48'38.42"N	32°48'58.09"E
E3	40.3	12.4	39°48'38.68"N	32°48'57.74"E
E4	28.0	3.6	39°48'38.68"N	32°48'57.74"E
E5	31.6	6.8	39°48'38.68"N	32°48'57.74"E
E6				
E7	32.0	7.7	39°48'39.11"N	32°48'56.97"E
E9	36.3	6.8	39°48'39.11"N	32°48'56.97"E

Sample numbers starting with T indicate those taken in July, and E indicates samples taken in September 2008. E6 sample is taken from the reef root sample

3.2 Laboratory Studies

Soil dry sediment sampling (10 cm top) was done using an auger and/or a shovel. In the September sampling, a PVC tube was inserted into the ground and sectioned into three and two slices, to obtain depth-wise information. The samples which were put into amber colored glass jars with Teflon lined lids were kept at coolers in the field and at -20 ^{o}C in the laboratory until prepared for the analysis. Nearly 30 percent of samples were separated for moisture and loss or ignition. Nearly 5 g of each sample (taken from 30 percent of samples) was used to determine moisture content as dried at 105 ^{o}C in oven for one day and volatile organic matter content as dried at 550 ^{o}C in oven for 4 hours. Moisture content and organic content were calculated as shown below:

$$\text{Moisture Content, }\% = \frac{g\ \ of\ \ dry\ sample}{g\ \ of\ \ sample} \times 100 \tag{3-1}$$

$$\text{Organic Content, }\% = \frac{g\ \ of\ \ ignited\ sample}{g\ \ of\ \ dry\ sample} \times 100 \tag{3-2}$$

Dry sample: The sample which is kept at 105 ^{o}C in oven for one day.

Ignited sample: The sample which is kept at 550 ^{o}C in oven for 4 hours.

For soil/sediment samples collected in July, the range of moisture content is between 21-71 % and the range of organic matter content is between 3-16.4 %. For soil/sediment samples collected in September, the range of moisture content is between 27-40 % and the range of organic matter content is between 4-12 %.

EPA methods 3540C (Soxhlet extraction), 3630C (Silica gel cleanup), 3660B (Sulfur cleanup), 3665A (Sulfiric acid cleanup) were carried out for PCB extractions, cleanup and analysis. Detailed information about the methods is given in Table 3-2. The procedure is briefly explained below. At the laboratory, the soil sediment samples were kept refrigerated at -20 ºC. Then, samples were freeze dried in Christ Alpha 1-4 model freeze drier at pressure around 0.006 mbar with a -45ºC condenser temperature for 16 hours. PCBs were extracted from soil sediment samples using soxhlet apparatus using 300 ml hexane and acetone solvent mixtures (1:1, v/v) for 17 hours. The solution was passed from anhydrous sodium sulfate column in order to clean from water interferences. After volume reduction by Kuderna-Danish (K–D) concentrator, extract was passed from a silica gel chromatographic column (silica gel activated for 16 hours at 130ºC and deactivated to 4.5% with deionised water) in order to clean from other chemicals. To remove elemental sulfur, sulfuric acid rinsed copper was added into the soxhlet flask. The clean extracts were concentrated again to volume between 2- 6 ml.

Table 3-2 Method used for PCB extraction, cleanup and detection purposes

US EPA Method No	***Target:***
3540C Soxhlet Extraction (Revision 3 December 1996)	extraction of nonvolatile and semivolatile organic compounds from soil and sediment
3630C Silica Gel Cleanup (Revision 3 December 1996)	separation of analytes from interfering compounds of a different chemical polarity
3660B Sulfur Cleanup (Revision 2 December 1996)	elimination of sulfur interference using copper powder
3665A Sulfuric Acid Cleanup (Revision 1 December 1996)	destruction of most organic chemicals in sample extracts prior to analysis for PCBs
8082A PCBs by Gas Chromatography (Revision 1 November 2000)	determination of PCBs as Aroclors or individual congeners in extracts from soil and sediment matrices using GC/ECD or GC/MS

Laboratory blanks, laboratory control samples, analysis of certified reference material were performed as a part of Quality Assurance / Quality Control (QA/QC) procedures to check whether the recoveries are in the acceptable range of 70-130 %. Stability of instrumental analyses was also checked at regular intervals using calibration standards. In Laboratory control samples (LCS) consisting of A1016/A1260 and PCB 209

External and internal standard calibration methods are performed for quantification of PCBs as Aroclors and individual congeners. Aroclor specific analysis was carried out using a Varian CP-3800 series Gas Chromatograph (GC) coupled with an Electron Capture Detector (ECD). Chromatographic separation was performed on a WCOT fused silica CP-Sil 8 CB Varian capillary column (30m length x 0.32mm internal diameter, 0.25 µm film thickness) with high purity helium and nitrogen as the carrier and make-up gases, respectively. GC oven temperature program was started at 100^{o}C (held 2 minutes), ramped to 160^{o}C at a rate of 8^{o}C/min, then to 250^{o}C at a rate of 3^{o}C/min (held 10 minutes), further ramped at 20^{o}C/min to 290^{o}C with a final hold for 5 minutes. Injector and detector temperature were set at 250^{o}C and 350^{o}C, respectively. 1 µm of extract was injected under split mode. US EPA method 8082A was followed during the analysis of samples.

Aroclor based PCB analysis was performed to initially identify whether PCBs are present in the samples, by identification of typical congener patterns of PCB mixtures, e.g. Aroclor 1016. The samples were then analyzed for individual PCB congeners for congener-specific quantification and further modeling studies.

In addition to Aroclor based PCB measurement, seven individual PCB congeners were analyzed. EU 86/278/CEE Directive and various regulations require measurement of 7 PCB congeners (IUPAC No. 28, 52, 101, 118, 138, 153 and 180). For these seven congeners, congener specific analysis was done. The analysis details are given below:

Congener specific analysis was carried out using an Agilent 6890N Gas Chromatograph (GC)/Agilent 5973 Mass Detector (MS). Calibration type is internal standard calibration. Chromatographic separation was performed on HP-5 MS capillary GC column (30m x0.25mm x 0.25 µm) with helium (constant flow and velocity are 1 ml/min and rate 36 cm/s, respectively) as carrier gas and methane as reagent gas. GC oven temperature program was started at 70 ^{o}C (held 2 minutes), increased to 150^{o}C at a rate of 25 ^{o}C/min, then to 200 ^{o}C at a rate of 3^{o}C/min, further increased at a rate of 8 ^{o}C/min to 280 ^{o}C with a final hold for 10 minutes. Injector and detector temperature were set at 250^{o}C and 350^{o}C, respectively. Total injection time is 42 min. 1 µm of extract was injected under split mode.

3.3 Modeling Studies

3.3.1 Monte Carlo Simulation

The sensitivity and verification of the PMF model was done by using artificial data sets. These data sets were generated by Monte Carlo simulation. Briefly, Monte Carlo is for those data sets are generated by using source profiles obtained from the literature and considering deviation in source profiles with normal distribution for source apportionments. The equation of it is expressed below equation (3-3). The assumptions accepted for the CMB model which is explained in section 2.4.1 are also valid for Monte Carlo Simulation.

$$D_{ijk} = \sum_{j=1}^{n} \{A_{ij} + C_j A_{ij}[\sqrt{2} erf^{-1}(2R_{jk} - 1)]\} \qquad (3\text{-}3)$$

D_{ijk}= the generated concentration of compound i in sample k
A_{ij}= the starting concentration of compound i in sample k
C_{ij}= the coefficient of variation (CV) for each source
R_{jk}= a uniformly distributed random number between 0 and 1
erf^{-1}= inverse Gaussian error function

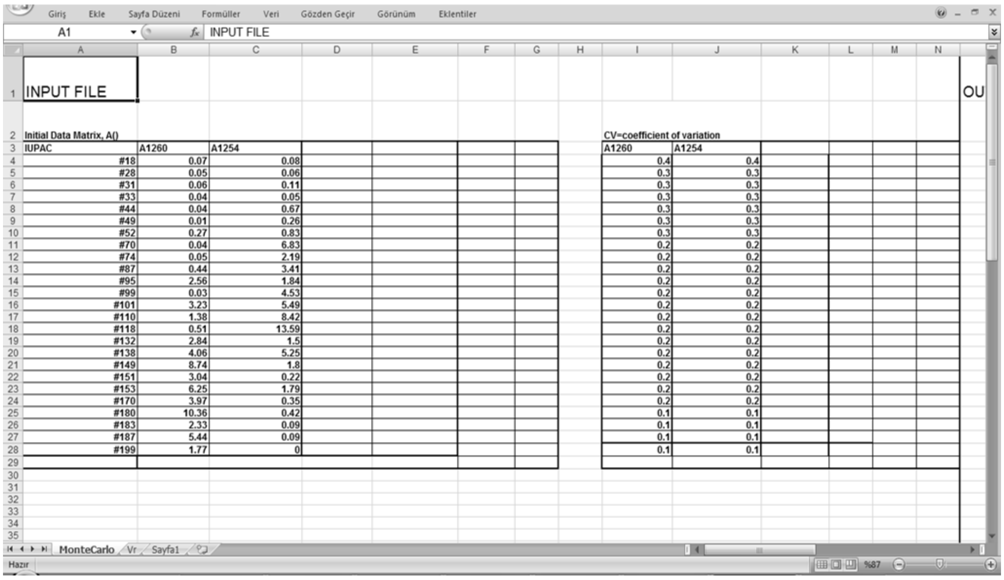
According to the equation, all PCB congeners change together since all are affected from the same random number, R_{jk}. Different random number is used for different source. Additionally, random numbers vary from sample to sample.

Coefficient of variation, C is provided for each PCB congener and molecular weight and physico-chemical properties of congeners are considered. Higher coefficient of variation is used for PCB congener having lower molecular weight since they have a tendency to volatilize or solubilize, etc. (Bzdusek, 2005).

In this study, Monte Carlo simulated data sets were used for both the CMB model and the PMF model validation during investigation of the uncertainty matrix. More than one artificial data set can be generated via Monte Carlo Simulation.

3.3.2 Artificial Data Set Using Known Source Profiles

The artificial data sets are created by using equation (3-3). The program codes given in Rachdawong (1997) were converted to Microsoft Excel 2003 Visual Basic Application (VBA) and the program codes can be seen in Appendix B. Input and output of the programs are displayed in one excel sheet as shown in Figure 3-3 and Figure 3-4, respectively.



INPUT FILE

Initial Data Matrix, A()

IUPAC	A1260	A1254
#18	0.07	0.08
#28	0.05	0.06
#31	0.06	0.11
#33	0.04	0.05
#44	0.04	0.67
#49	0.01	0.26
#52	0.27	0.83
#70	0.04	6.83
#74	0.05	2.19
#87	0.44	3.41
#95	2.56	1.84
#99	0.03	4.53
#101	3.23	5.49
#110	1.38	8.42
#118	0.51	13.59
#132	2.84	1.5
#138	4.06	5.25
#149	8.74	1.8
#151	3.04	0.22
#153	6.25	1.79
#170	3.97	0.35
#180	10.36	0.42
#183	2.33	0.09
#187	5.44	0.09
#199	1.77	0

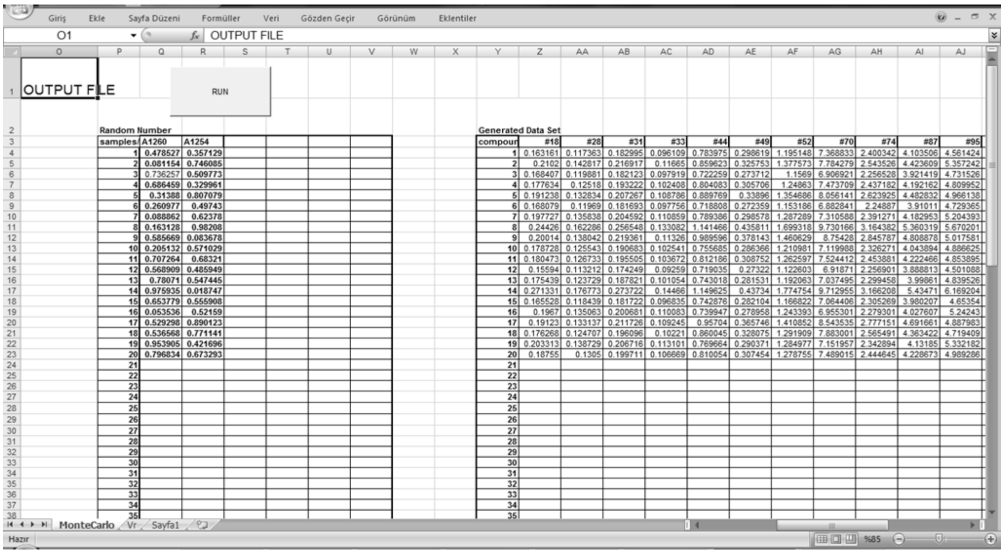
CV=coefficient of variation

A1260	A1254
0.4	0.4
0.3	0.3
0.3	0.3
0.3	0.3
0.3	0.3
0.3	0.3
0.3	0.3
0.2	0.2
0.2	0.2
0.2	0.2
0.2	0.2
0.2	0.2
0.2	0.2
0.2	0.2
0.2	0.2
0.2	0.2
0.2	0.2
0.2	0.2
0.2	0.2
0.2	0.2
0.2	0.2
0.1	0.1
0.1	0.1
0.1	0.1
0.1	0.1

Figure 3-3 Input File of Monte Carlo; Known Source Profile and CV Values

First, degraded A1260 profiles to create some data sets and known literature source profiles are multiplied together to form a data matrix or some are multiplied together separately then, their matrices are combined to form unique data matrix. Next, coefficients of variation in Figure 3-3 are varied to each congener regarding the fate of PCBs in the sediment. Lastly, two dimensional data matrix is obtained as the program adds partitions of sources.

$$D_{ik} = \sum_{j=1}^{n} D_{ijk} \tag{3-4}$$



OUTPUT FILE

Random Number

samples/	A1260	A1254
1	0.478527	0.357129
2	0.081154	0.746085
3	0.736257	0.509773
4	0.686459	0.329961
5	0.31388	0.807079
6	0.260977	0.49743
7	0.088862	0.62378
8	0.163128	0.98208
9	0.585669	0.083678
10	0.205132	0.571029
11	0.707264	0.68321
12	0.568909	0.485949
13	0.78071	0.547445
14	0.975935	0.018747
15	0.653779	0.555908
16	0.053536	0.52159
17	0.529298	0.890123
18	0.536568	0.771141
19	0.953905	0.421696
20	0.796834	0.673293
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		
31		
32		
33		
34		
35		

Generated Data Set

compour	#18	#28	#31	#33	#44	#49	#52	#70	#74	#87	#95
1	0.163161	0.117363	0.182995	0.096109	0.783975	0.298619	1.195148	7.368833	2.400342	4.103506	4.561424
2	0.2102	0.142817	0.216917	0.11665	0.859623	0.325753	1.377573	7.784279	2.543526	4.423609	5.357242
3	0.168407	0.119881	0.182123	0.097919	0.722259	0.273712	1.1569	6.906921	2.256528	3.921419	4.731526
4	0.177634	0.12518	0.193222	0.102408	0.804083	0.305706	1.24863	7.473709	2.437182	4.192162	4.809952
5	0.191238	0.132834	0.207267	0.108786	0.889769	0.33896	1.354686	8.056141	2.623925	4.482832	4.966138
6	0.168079	0.11969	0.181693	0.097756	0.718808	0.272359	1.153186	6.882841	2.24887	3.91011	4.729365
7	0.197727	0.135838	0.204592	0.110859	0.789386	0.298578	1.287289	7.310588	2.391271	4.182953	5.204393
8	0.24426	0.162286	0.256548	0.133082	1.141466	0.435811	1.699318	9.730166	3.164382	5.360319	5.670201
9	0.20014	0.138042	0.219361	0.11326	0.989596	0.378143	1.460629	8.75428	2.845787	4.808878	5.017581
10	0.178728	0.125543	0.190683	0.102541	0.755685	0.286366	1.210981	7.119988	2.326271	4.043894	4.886625
11	0.180473	0.126733	0.195505	0.103672	0.812186	0.308752	1.262597	7.524412	2.453881	4.222466	4.853895
12	0.15594	0.113212	0.174249	0.09259	0.719035	0.27322	1.122603	6.91871	2.256901	3.888813	4.501088
13	0.175439	0.123729	0.187821	0.101054	0.743018	0.281531	1.192063	7.037495	2.299458	3.99861	4.839526
14	0.271331	0.176773	0.273722	0.14466	1.149625	0.43734	1.774754	9.712955	3.166208	5.43471	6.169204
15	0.165528	0.118439	0.181722	0.096835	0.742876	0.282104	1.166822	7.064406	2.305269	3.980207	4.65354
16	0.1967	0.135063	0.200681	0.110083	0.739947	0.278958	1.243393	6.955301	2.279301	4.027607	5.24243
17	0.19123	0.133137	0.211726	0.109245	0.95704	0.365746	1.410852	8.543535	2.777151	4.691661	4.887983
18	0.176268	0.124707	0.196096	0.10221	0.860045	0.328075	1.291909	7.883001	2.565491	4.363422	4.719409
19	0.203313	0.138729	0.206716	0.113101	0.769664	0.290371	1.284977	7.151957	2.342894	4.13185	5.332182
20	0.18755	0.1305	0.199711	0.106669	0.810054	0.307454	1.278755	7.489015	2.444645	4.228673	4.989286
21											
22											
23											
24											
25											
26											
27											
28											
29											
30											
31											
32											
33											
34											
35											

Figure 3-4 Output File of Monte Carlo; Generated Data Set (sampleXcongener) and Random number used to create Data Set

The program is initialized by “RUN” button in output sheet. It calls “INPUT FILE” graphical user interface (GUI) (Figure 3-5). Then, numbers of congeners and sources determined in the input sheet are asked in the monitor. The number of samples to be used in the calculations is also asked in the GUI. Lastly, the program computes the COD values by “RUN” button in the GUI.

Source # : 2
Congener # : 25
Sample # : 20

Figure 3-5 Monte Carlo GUI

3.3.3 CMB Modeling

The model used by Rachdawong (1997) was written in FORTRAN code. This was converted to visual basic application of Microsoft Excel in order to develop a more user friendly and efficient version of the model. To run it;

1. An output file including number of congeners, sources, source compositions and environmental profiles was prepared in text ".txt" format.

2. CMB model tested the three sources, as given in input and computed the best predicted profile according to equation (2-1).

Properties of the modified CMB model software:

Elimination of not detected congeners in each sample: When CMB model in Rachdawong (1997) is used, value of a congener which cannot be measured is removed by hand for each sample run since the model cannot simulate for "0" value of measured data in a sample. It is difficult to remove different congeners which have "0" values, in all samples. By this program, zero congeners in each sample are eliminated automatically with one run, so these congeners in source profiles are also eliminated. This becomes especially important when sampling is performed at a site having a heterogeneous character in terms of the concentration observed in various samples. This situation, in the case of PCBs, may lead to a number of congeners are not detected which should be removed from each environmental sample profile individually. In short, the modified CMB model software enables source profiles to be prepared according to each sample, individually.

All samples can be determined with one run: The modified CMB software enables execution of multiple samples with one run.

Trial of combination two and three sources: The model checks that possible combination from all sources in combination of two and three source profiles in order to obtain the best predicted profiles according to goodness of fit parameters.

Elimination of a source if all congeners in this source are "0": The model runs for Aroclor and Clophen source inputs. Since the model automatically eliminates not detected values in environmental profiles during this procedure, it may happen that a source profile may have "0" or not detected values for the rest of the congeners. If sum of all congeners coming from a source becomes zero, developed CMB model will not run and will disregard this source.

Graphics and sorting data: After program calculates all trials of 2 and 3 combinations of sources for a sample, the results are sorted from the smallest measured relative errors to the highest which coincides with largest R^2 values, so it enables user to see the best fit (the smallest RESA value) easily since it is sorted from the smallest to the highest.

Variance inflation factor (VIF): An important difficulty in least square models is that the model cannot separate the sources which have same or very similar compound compositions. This is called collinearity. In other words, collinearity means there is a strong correlation among some or all of the independent variables in a regression model. This will be a problem in apportionment of sources during modeling using linear combination like CMB. Additionally, the collinearity may cause negative source contributions in the CMB model. VIF is an indicator of multicollinearity, as used by Cheng & Hopke (1986). Therefore, VIF is incorporated into the model to indicate potential for multicollinearity that affects the estimation of source contribution (Cheng & Hopke, 1986). It is used before performing CMB and it runs with CMB visual basic program. The smaller the VIF, the smaller the risk of multicollinearity among sources. The VIF equation is expressed by

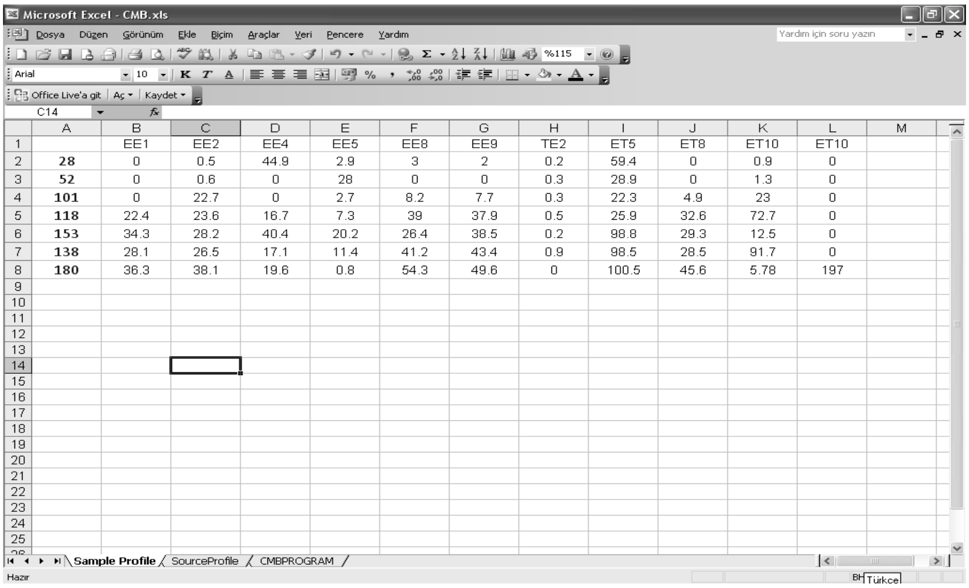
$$VIF = \frac{1}{(1 - R^2)} \tag{3-5}$$

where R^2 is the multiple correlation coefficient of a profile with the rest of the the profiles to be used in a run of the CMB model.

The set of improvements made on the CMB model does not change the inner workings of the model, in the sense that the same equation is applied with the same set of goodness of fit parameters. Yet, the model can now test thousands of source combinations automatically, with an indicator for multicollinearity. This leads to a greater chance to find the best predicted profile with the smallest effort, since all possible combinations of potential sources are tested and sorted automatically by the model.

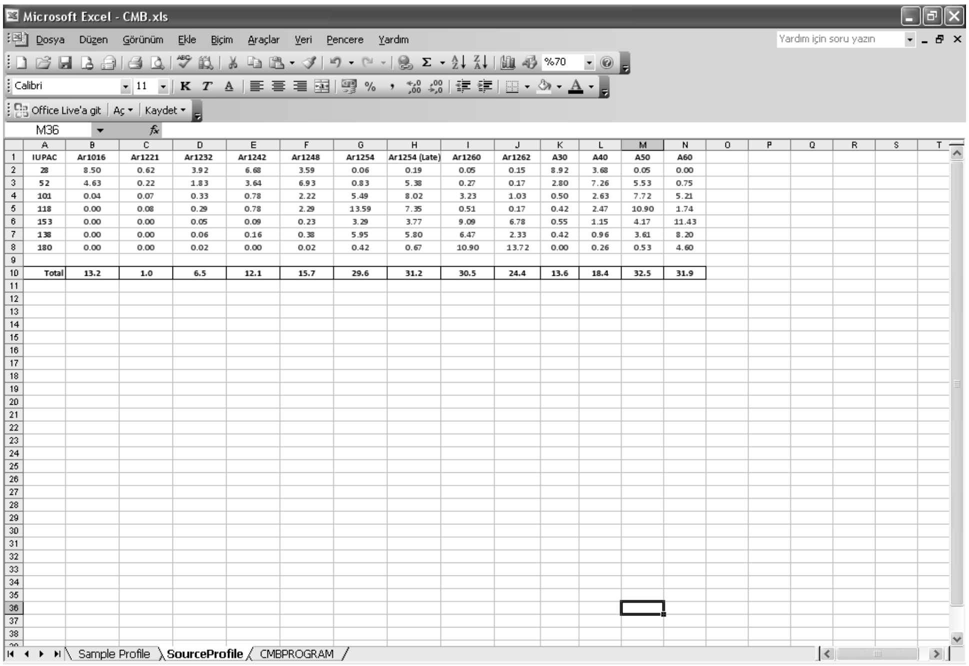
Input Files: There are two input files. One is for measured data which is called as “Sample Profile” (Figure 3-6). The concentrations of individual congeners for all samples are written into the sheet. The first column indicates the congener name and first row shows sample name. Additionally, the number of samples to be used in the calculations is also asked in the GUI. Therefore, the program computes for samples as specified in the GUI.

The other input file is the “Source Profile” sheet (Figure 3-7). The full list of potential sources and their compositions are written in this sheet. The program then uses this big set to pick sources in combinations of two and three to explain the environmental concentrations. Source composition information for PCBs is well established. For example, Frame et al. (1996) and Ishikawa et al. (2007) contain congener specific detailed information on composition of various PCB sources. Names of congeners should be written in the same order in both input files. For example, congener 52 which is in third row and is second congener in “Sample Profile” sheet, cannot be measured in sample E1. Therefore, the program eliminates it. In “Source Profile” sheet, the program also eliminates second congener in all sources during modeling. In the program, number of the sources is not asked to user in GUI since the program runs for all sources which are present in the sheet.



	EE1	EE2	EE4	EE5	EE8	EE9	TE2	ET5	ET8	ET10	ET10
28	0	0.5	44.9	2.9	3	2	0.2	59.4	0	0.9	0
52	0	0.6	0	28	0	0	0.3	28.9	0	1.3	0
101	0	22.7	0	2.7	8.2	7.7	0.3	22.3	4.9	23	0
118	22.4	23.6	16.7	7.3	39	37.9	0.5	25.9	32.6	72.7	0
153	34.3	28.2	40.4	20.2	26.4	38.5	0.2	98.8	29.3	12.5	0
138	28.1	26.5	17.1	11.4	41.2	43.4	0.9	98.5	28.5	91.7	0
180	36.3	38.1	19.6	0.8	54.3	49.6	0	100.5	45.6	5.78	197

Figure 3-6 Input File: Sample Profile Sheet



IUPAC	Ar1016	Ar1221	Ar1232	Ar1242	Ar1248	Ar1254	Ar1254 (Late)	Ar1260	Ar1262	A30	A40	A50	A60
28	8.50	0.62	3.92	6.68	3.59	0.06	0.19	0.05	0.15	8.92	3.68	0.05	0.00
52	4.63	0.22	1.83	3.64	6.93	0.83	5.38	0.27	0.17	2.80	7.26	5.53	0.75
101	0.04	0.07	0.33	0.78	2.22	5.49	8.02	3.23	1.03	0.50	2.63	7.72	5.21
118	0.00	0.08	0.29	0.78	2.29	13.59	7.35	0.51	0.17	0.42	2.47	10.90	1.74
153	0.00	0.00	0.05	0.09	0.23	3.29	3.77	9.09	6.78	0.55	1.15	4.17	11.43
138	0.00	0.00	0.06	0.16	0.38	5.95	5.80	6.47	2.33	0.42	0.96	3.61	8.20
180	0.00	0.00	0.02	0.00	0.02	0.42	0.67	10.90	13.72	0.00	0.26	0.53	4.60
Total	13.2	1.0	6.5	12.1	15.7	29.6	31.2	30.5	24.4	13.6	18.4	32.5	31.9

Figure 3-7 Input File: Source Profile Sheet

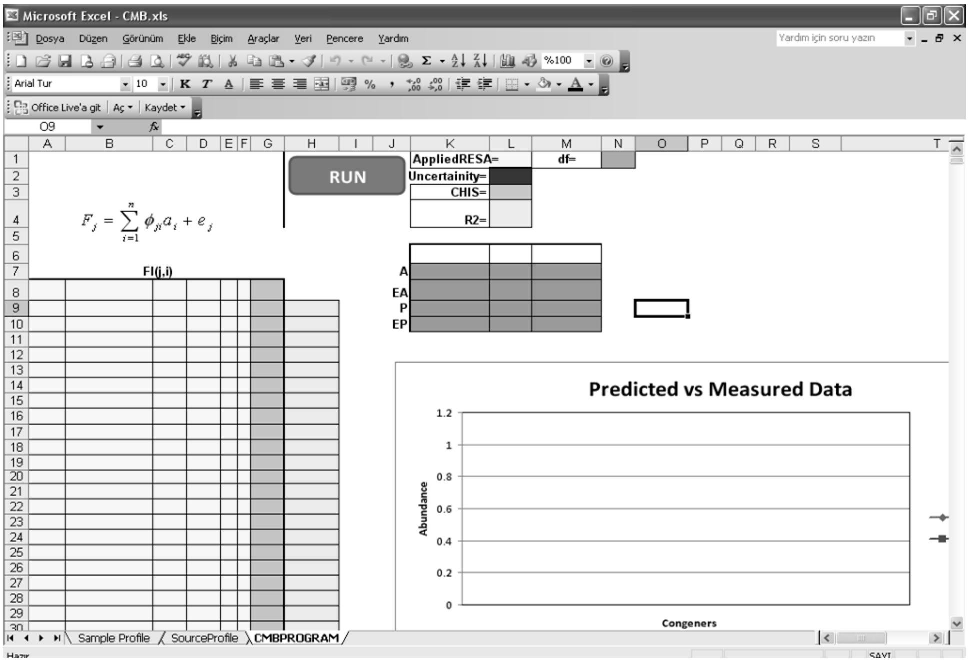
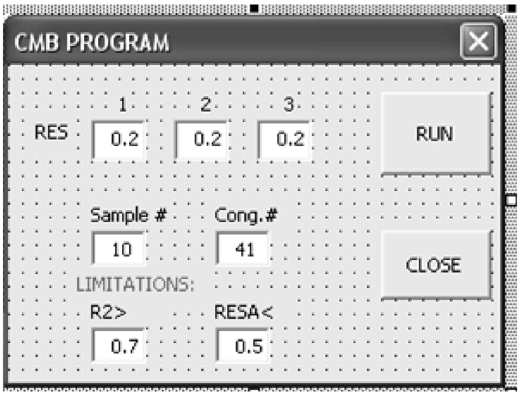


Figure 3-8 Simulation Sheet

Program is executed with “RUN” button in “CMBPROGRAM” sheet (Figure 3-8). This displays a user interface. Displayed interface (Figure 3-9) asks for the last set of information as input into the program. In the CMB model GUI, user enters the source profile relative error (RES) values for maximum three sources, numbers of samples and numbers of congeners. “RUN” button triggers the “initialize” procedure and the program executes. If “Close” button is selected, program will not run and GUI will be closed. There are two conditions specified in the GUI to write the results of source combinations to output sheets: first one is lower limit of multiple correlation coefficients, R2 and second one is upper limit of relative error of samples, RESA.



RES 1: 0.2
RES 2: 0.2
RES 3: 0.2
Sample #: 10
Cong.#: 41
LIMITATIONS:
R2>: 0.7
RESA<: 0.5

Figure 3-9 CMB Model GUI

Output Files: The program provides a separate output worksheet for each sample. First column indicating sample name in Sample Profile file also describes sheet of each sample output after computation of the program. In each sheet, there are two parts; in first part, trial numbers, source names combined, df, RESA, source contributions (P), source contributions errors (EP), source contribution factor (A), source contribution factor error (EA) and the variance inflation factor (VIF) values are tabulated. In the second part, congener names which are non zero for measured data, measured values and predicted values are written. After sorting out trials according to ascending RESA values, five graphs are drawn to plot measured vs. predicted profiles belonging to the first five predicted data. All calculations and procedures are written in visual basic (Figure 3-10).

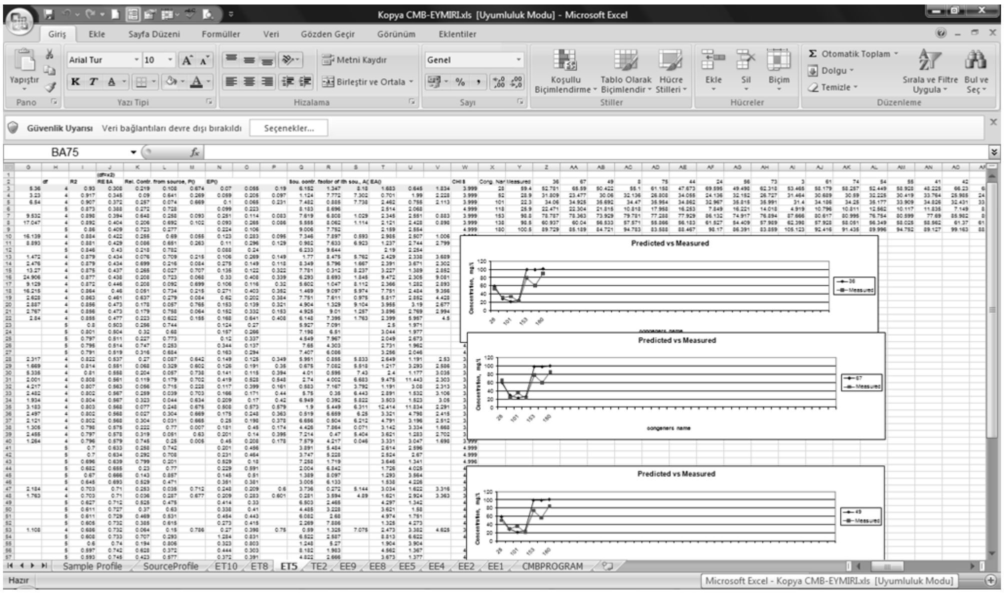


Figure 3-10 An Output sheet for a sample

Main flow chart of the program subs and subroutines is given in Figure 3-11. Each one is explained below:

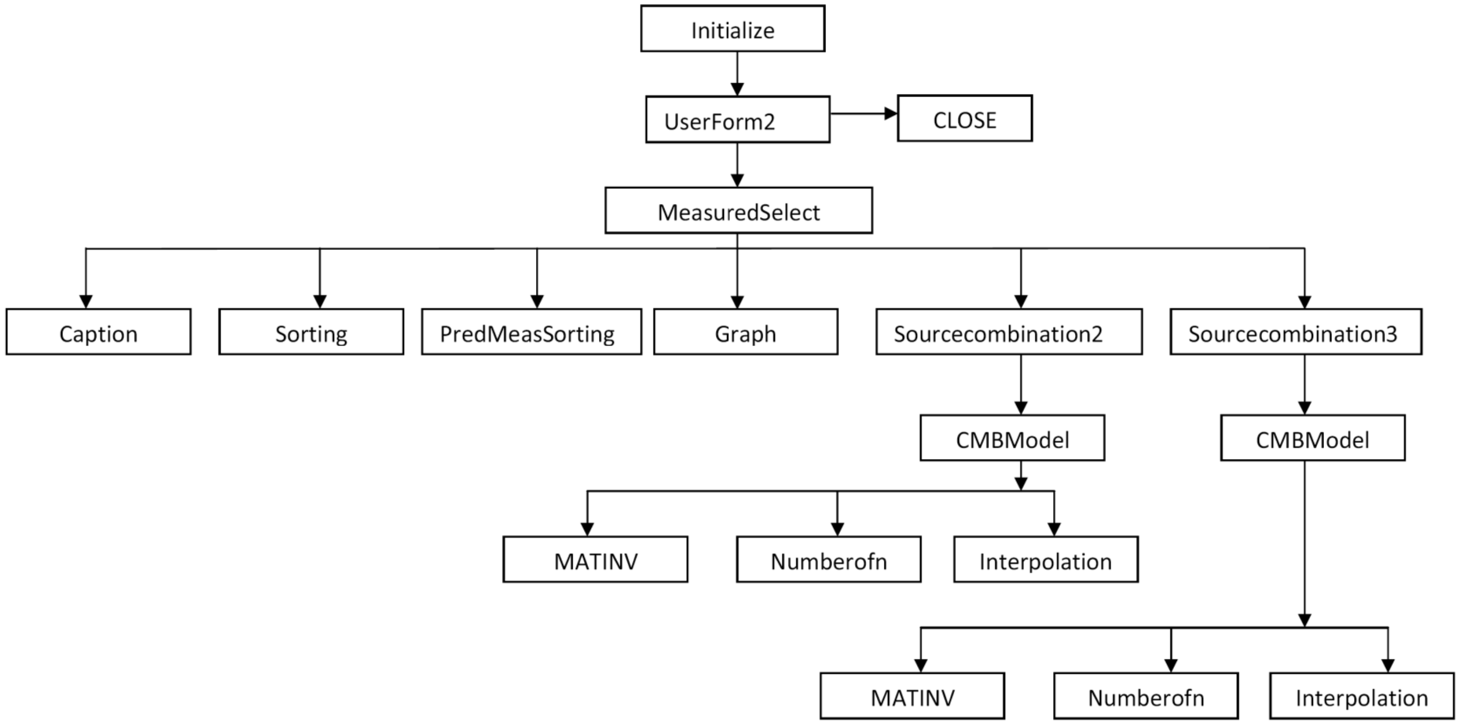


Figure 3-11 Main Flowchart of Program Subs and Subroutines

Initialize: Calls MeasuredSelect procedure and program starts to run.

UserForm2: Interface that allows user to arrange and enter needed variables.

MeasuredSelect: Main procedure of the program. All functions and procedures are called in this subroutine. Also selection of non-zero congener concentrations of measured data is performed here.

Caption: This sub is for captions of variables written in generated sheets/outputs.

Sorting: The results of the runs except for predicted and measured data in output sheet are sorted in terms of ascending RESA values.

PredMeasSorting: All predicted and measured data are arranged to ascending RESA values.

Graph: After sorting, graphs for measured versus predicted data are drawn for the first five runs with the smallest RESA values.

Sourcecombination2: CMB model calculates all possible two source combinations.

Sourcecombination3: CMB model calculates all possible three source combinations.

CMBModel: The CMB model is called in Sourcecombination2 and Sourcecombination3. This subroutine tries to partition relevant sources to measured data to relevant sources as using effective variance least square method.

MATINV: It is a function used in the CMB model to take inverse of matrix.

Numberofn: It is a function to make resize the number of sources for different number of sources due to use of 2 and 3 source numbers.

Interpolation: Function to make interpolation between RESA and χ^2 values in order to find df= χ^2.

3.3.4 PMF Modeling

There are a lot of versions of the Positive Matrix Factorization (PMF model). These are versions of PMF2 and PMF3 prepared and developed by Paatero and Tapper (1994), and EPA PMF1.1 and EPA PMF3 prepared by USEPA. PMF3 prepared by Paatero is a three-dimensional program and is mostly used for atmospheric studies which have much data dependent on time. PCBs are studied with soil and sediment samples and these environmental samples do not show great changes in a short time. Therefore, there is no time dimension in the measurements. This is the reason for using the two dimensional model.

The PMF model has two different bases, Windows and DOS. Although PMF model based on Windows is more user-friendly, types and format of input and output files create problem during the execution of the program. Furthermore, recent studies using the PMF model also use the DOS version such as Amato (2009), Du et. al. (2007), etc. Therefore, PMF2.v.4.2 based on DOS was preferred in this study.

There are two input files prepared in Microsoft Excel; the data matrix for measured samples and the standard deviation matrix for uncertainties assigned to each measurement. The first matrix consists only of measured data.

The methods which were carried out during preparation of the uncertainty matrix were;

- Use of constant values in uncertainty matrix for all data ,
- Use of equations created by Polissar et. al. (1998) and Hopke (2010),
- Use of combinations of constant values for some data with equations created by Polissar et.al. (1998) and Hopke (2010),
- Use of the equations developed in this study.

3.3.4.1 Formation of Model Input Files

There are two input files; one is for data matrix including measured data and second is for uncertainty matrix including standard deviation of measured data. Data below detection limit and missing values can also be included in data set.

Hopke (2010) states that preparation of standard deviation file is the critical point. Therefore, while using weighting which is an advantage of the PMF model, standard deviation matrix can be controlled. By this way, higher uncertainty can be assigned for small values in data matrix in order not to allow them to affect the objective function too much or effects of missing values on factor profiles can be prevented by giving high uncertainties for very high data (Paatero & Tapper, 1994). Ozturk (2009) summarized several approaches of formation of input files in Table 3-3.

Table 3-3 Examples of input data formulas for PMF in the literature (Ozturk, 2009).

	Measured Data Set File		**Uncertainty File of Measure Data Set**	
References	**BDL Data**	**Missing Data**	**BDLData σ**	**Missing Data σ**
Poirot et. al., (2001), Polissar et. al., (2001)	DLx0.5	Median Concentration	1.5x DL	4x Median Concentration
Anttila et. al. (1995)		Mean Concentration	DL	DL
Song et. al., (2001)	DLx0.5		DL	DL
Begum et. al., (2005), Xie et. al., (2006)	DLx0.5	Geometric Mean	5/6x DL	4x Geometric Mean
Brown et. al., (2007)	DLx0.5	Mean Concentration	1.5x DL	4x Median Concentration
Chung et. al., (2005)	DLx0.5	Geometric Mean	DL	4x Geometric Mean
Chueinta et. al., (2000)	DLx0.5	Mean Concentration	DLx0.5	

DL: Detection limit, BDL: Below Detection Limit

In this study, effects of standard deviation formula prepared by Polissar et al. (1998) were inspected. Its general formula of data and standard deviation of data is given below for measured, BDL and missing data, respectively

$x_{ij} = v_{ij}$ For determined (measured) values $\sigma_{ij} = u_{ij} + d_{ij}/3$

$x_{ij} = d_{ij}/2$ For below limit of detection values $\sigma_{ij} = \bar{d}_{ij}/2 + d_{ij}/3$

$x_{ij} = \bar{v}_{ij}$ For missing values $\sigma_{ij} = 4\bar{v}_{ij}$

3.3.4.2 Output Files of the Model

Output files are F FACOR_txt containing F Loading profiles, G FACTOR_txt containing G scores, and MISC_txt containing Explained Variance of F and Scaled Residual matrix.

3.3.4.3 Use and Arrangement of Parameters in the Program

- Selection of the Error Model (EM): EM is selected as -14 since it is recommended as suitable for environmental data (Paatero, 2002) and this value has been used in other studies (Du, et. al., 2008, Juvela et. al., 1996 and Bzdusek, 2005).
- SEED: The program uses different starting points to find global minima with the Seed function. Aim of the program is to minimize the objective function. Therefore, Seed is used to find global minima by using different starting points. As Seed is changed, program is executed with different pseudorandom values (Willis, 2000). Du and Rodenburg (2007) state that different seeds should be tried in order to obtain global result and not local ones. Du et. al. (2007) suggested that seed values should be changed between 1 and 10 in order to identify global minimum.
- *FPEAK:* After number of factors is determined, the true Fpeak rotation should be decided for the solution. Hopke (2010) states that non zero Fpeak values provide more realistic results on real data sets. Moreover, he explains that generally the highest fpeak is observed before an important increase is observed in the Q value. There is no determined

range defined for the Fpeak value. In the literature, everybody uses different ranges and values. In the determination of PAH and PCB sources, Bzdusek (2005) has used fpeak values in the range of -2 and +2, and even until -4. Amota (2009) states that fpeak rotations are changed between ranges of 0.8 and -0.8, step by 0.1 in PMF model for atmospheric PM pollution.

- *Goodness of Fit Parameters:* Finally, the results obtained from the model are explained by using goodness of fit parameters for the fit of the model results with measured data.

3.3.4.4 Goodness of Fit Parameters

Objective Function: Qtheo should be approximately equal to Q robust. Qtheo is also called as degrees of freedom, df and it is calculated as

df=mxn-px(m+n) or

df=mxn (Bzdusek, 2005 and Paatero & Tapper, 1994).

where, n= number of samples, m_ number of congeners and p= number of sources.

Paatero and Tapper (1994) specify that Qtheo can also be expressed by being equal to total number of data points. The degrees of freedom used in this study is calculated as df= nxm —px(n+m). Although it is stated that Q robust is approximately equal to df (Qtheo.) for the good fit model, Ozturk (2009) indicates that Q robust can change 50 % from calculated Q value.

Multiple Correlation Coefficient, R^2: total congeners in each sample are drawn for predicted data found by the model vs. measured data. The better the fit, the closer R^2 gets to “1”.

Coefficient of Determination, COD: This value was used to evaluate the performance of other statistical models like FA and PVA (Bzdusek and Christensen, 2004, Imamoglu, 2001 and Johnson et. al., 2000). It provides information on the goodness of fit on a congener specific analysis. COD values of each variable are expected to be 1 or close to 1 to find perfect fit between variables modeled and measured data sets. COD equation is;

$$R_j^2 \square \frac{s(x)_j^2 - d_j^{*\square 2}}{S(x)_j^2} \qquad 0 < R_j^2 < 1 \tag{3-6}$$

where

R_j^2 = the coefficient of determination for congener j

$s(x)_j^2$ =variance of the jth variable in original data

d_j^{**} = standard deviation of d_{ji} over all samples

In this study, the COD program used by Imamoglu (2001) was recorded with the visual basic language (Appendix C).

Explained Variance-F for Each Source (EV(F)): It is one of outputs of the program and dimensionless. It indicates the importance of each factor in explaining one row or column of the observed matrix. The values of EV are between 0.0 and 1.0.

The percentage contribution of each source and residual is calculated as

$$\%\ for\ each\ i = \frac{100 \times \sum_{j=1}^{m} EV(F)_{ji}}{m} \qquad (3\text{-}7)$$

i=1…...n+1

j=1…..m

m= number of congeners

n=number of factors

$(n+1)^{st}$= residual factor

3.3.4.5 Other Important Parameters

Scaled Residuals: Another output is the scaled residual matrix. Plots of the scaled residuals for all species can be used as an indicator to determine the number of factors. It is desirable to have symmetric distributions and to have all the residuals within ±3 standard deviations. If there is asymmetry or a larger spread in the residuals, then the number of factors should be reexamined (Willis, 2000).

Factor Loading Matrix, F: Each column of F matrix obtained from model reflects a source profile. Therefore, these factor profiles can be compared with PCB source compositions in the literature to be able to understand whether they resemble each other or not.

G Scores, G: G matrix defines the effects of each source on samples. Each column of G matrix vs. samples is drawn to understand correction of factor numbers as regarding to contribution and distribution of factors to samples. It is discussed together with the results of compositions in F Loading matrix.

3.3.4.6 Artificial Data Sets Used in PMF Modeling

The environmental PCB data set obtained in the first part of this study is not sufficiently large to be used in the PMF model for the identification of dominant congener patterns. Hence, in order to investigate the use of the PMF model, artificial data sets were created as a first stage of the PMF modeling effort. Additionally, artificial data sets are important to be able to discuss on outputs of the PMF model by the help of the known inputs while understanding this complex PMF model in first.

The aim for testing of PMF model with artificial data sets was to understand the effects of values in the uncertainty matrix on the results and to develop some suggestions before using on environmental PCB data set.

Artificial data sets in varying complexity were generated to examine effects of different uncertainties for the model, so some results are deduced. These results are discussed in Section 4.3.2.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Status of PCB Pollution in Lake Eymir

Yeniova's study in 1997 indicated major PCB contamination within the grounds of the TEDAS facility near Lake Eymir, with decreasing concentration in the area, finding 313-586 ng/g in dry weight total PCBs in soils a few meters, away from the transformer repair facility. Akduman (2009) also took sandy samples from concrete floor of the facility, finding concentrations between 5700-10000 ng/g in dry weight. These studies confirm the presence of a PCB contamination source in the vicinity of Lake Eymir. In order to investigate in more detail the extent of contamination within the boundaries of the specially protected zone, outside the boundaries of the TEDAS facility, a total of 20 samples were collected in two sampling campaigns from the area between Lake Mogan and Lake Eymir adjacent to the outer wall of the TEDAS facility. Moisture Content and organic content of samples range between 21-71% and 3-16.4% respectively. The results were evaluated in terms of total PCBs and as profiles of seven individual PCB congeners. Source apportionment for the congener specific data was performed with CMB model, which is dealt with in Section 4.2.

Determination and Assessment of Pollution as Total PCBs: the total PCB concentrations of samples collected in the first sampling campaign are tabulated in Table 4-1. PCB concentrations change between 76.3 ng/g and 0.1

ng/g with an average of 16.7 ng/g in dry weight. As can be seen from Figure 2-2, samples collected in the first campaign can further be divided into two: those taken closest to the discharge channel of TEDAS (samples No: 1-10), and those taken from the channel leading to Lake Eymir. High Aroclor 1260 is measured in the samples (No: 3, 4, 5, 7, 8, 10) especially taken from near transformer repair and maintenance center of TEDAŞ. Total PCB concentration of these samples changes between 41-76.3 ng/g in dry weight. In the other sample points, PCB is detected in measurable level (0.1-7.2 ng/g). When compared with other samples, PCB concentrations are found as low (No: 13, 22) in the samples taken from inside of and close to the channels which are expected to be active for transportation of PCBs. The reason for low PCB concentrations in the channel may be due to soil filling done during bridge construction around channel region. PCB concentrations of samples in vicinity of Lake Eymir in July are given in Figure 4-1 and Figure 4-2.

Table 4-1 Total PCB Concentration in Samples taken first campaign

	$\Sigma_{Ar1016+1260}$ **PCB (ng/g dry weight)**
Minimum-Maximum	0.1 – 76.3
Mean ± Standard deviation	17.5 ± 26
Median	1.7
Geometric Mean	2.0

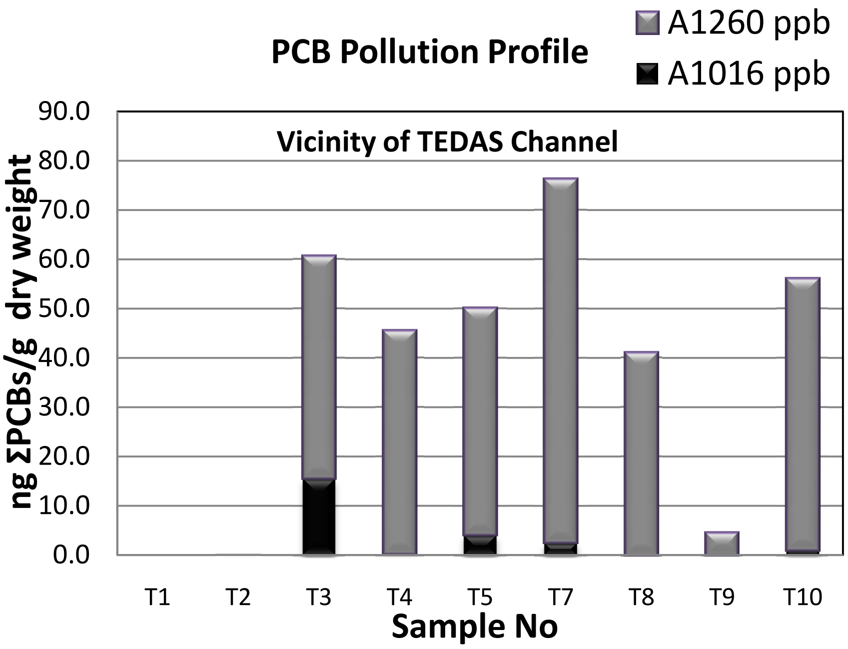


Figure 4-1 Total PCB Concentration in Samples Taken in First Sampling Campaign in the Vicinity of TEDAS Channel

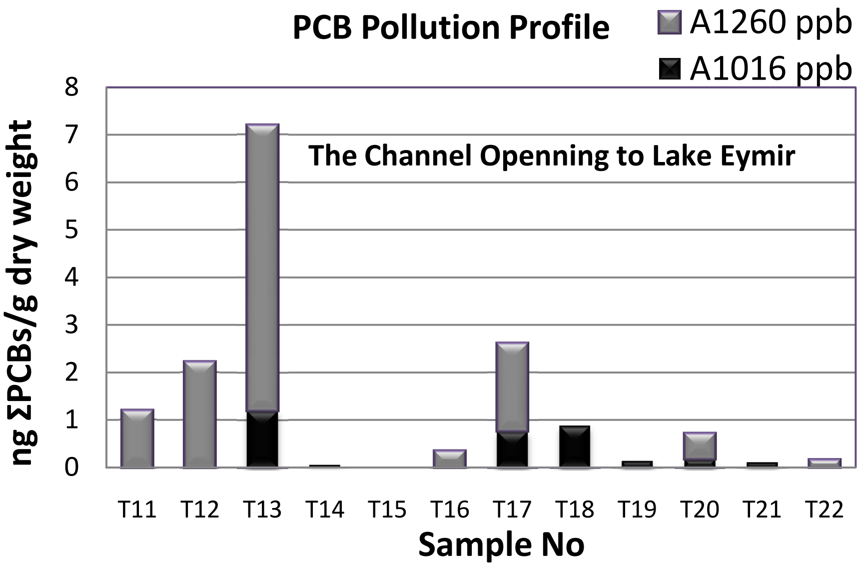


Figure 4-2 Total PCB Concentration in Samples Taken in First Sampling Campaign from the Channel Opening to Lake Eymir

After the first sampling campaign, a second one was conducted to gather further information on regions where high PCB concentrations were measured in July. The main aim was to understand the effect of depth on PCB concentration. The relationship between depth and PCB concentrations is displayed in Figure 4-3.

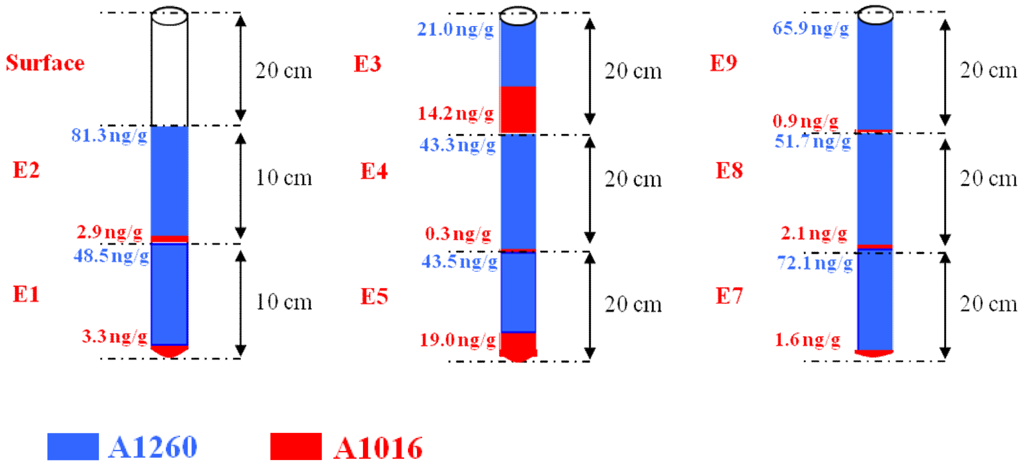


Figure 4-3 The relationship between depth and PCB concentrations in Second Campaign

Total PCB concentrations for three samples are given in Table 4-2. No consistent increase or decrease with depth was observed. The highest concentration in this set was observed for the surface samples, E2. Even though this was the case, for the other two locations, the highest concentration was observed at around 40-60 cm of depth. Overall, a much more homogeneous concentration of PCBs were observed for these samples, which is a significant finding, as can be seen from the low standard deviations in Table 4-2.

Table 4-2 Total PCB Concentration in Samples taken in Second Campaign

	$\Sigma_{Ar1016+1260}$ PCB (ng/g dry weight)
Minimum-Maximum	35.2-84.2
Mean ± Standard deviation	58.9 ± 16.1
Median	58.1
Geometric Mean	57

As seen in Figure 4-4, Aroclor 1260, the PCB mixture including the higher chlorine content was found in the samples. This is consistent with the use of such mixtures especially in electrical equipments such as transformers. Relatively much smaller amount of low chlorinated congeners was observed. This means these congeners were not discharged to environment. Besides, Chiarenzelli et al. (1997) states that relatively low chlorinated congeners evaporate when sediments polluted by PCBs expose to wetting-drying processes continuously, like the sediments of this region close to Lake Eymir.

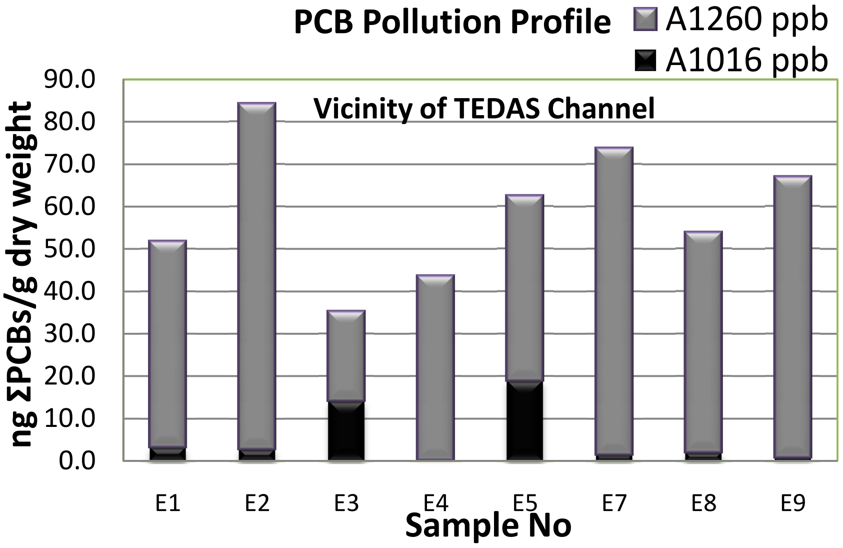


Figure 4-4 Total PCB Concentration in Samples taken in Second Campaign

Concentrations obtained in this study are lower than results of PCB pollution specified in Yeniova's (1998) study which is done in the vicinity of Lake Eymir and TEDAŞ facility. According to Yeniova's (1998) study in 1997, PCB concentration in TEDAŞ facility was measured up to 464000 ng/g. However, PCB concentrations in samples taken from the vicinity of channels opening to Lake Eymir are not as high as in TEDAS. There are 10 years between two studies. Decrease in PCB concentration may be explained with rehabilitation of the field after Yeniova's (1998) study in 1997. Also no disposal of high PCB concentration is done over the years as like before. On the other hand, today PCB pollution may still have an impact on the region since PCBs are observed in surface sediments and soils. In order to discuss PCB pollution sources in more detail, congener specific analysis are given below. As PCB congener specific analysis results are used, sources of PCB pollution in the region are tried to be identified by the use of a Chemical Mass Balance Model which is discussed in Section 4.2.

Determination and Assessment of Pollution as Specific PCB Congeners: Analysis results of samples of Lake Eymir are presented in Table 4-3 for seven congeners (IUPAC NAMES: #28, #52, #101, #118, $138, #153 and #180). These congeners are listed in the “Working Document on Sludge” that complements the European Directive 86/278/CEE. Congener specific analysis could only be quantifiable for sample numbers 2, 5, 7, 8 and 10 from the first campaign and 1, 2, 4, 5, 6, 8 and 9 from the second campaign. There are two reasons for selecting these samples: (1) only the samples including high PCB concentrations in both two campaigns have been selected, (2) also the samples including adequate amount have been selected for congener specific measurement. Congener profiles are shown in Figure 4-5.

Table 4-3 Concentrations of Specific PCB Congeners in Samples

	$\Sigma_{Ar1016+1260}$ PCB (ng/g dry weight)
Minimum-Maximum	ND*-5.4
Mean ± Standard deviation	13.7 ± 3.8
Median	13.9
Geometric Mean	13.2

*ND: Below Detection Limit (<0.2 ng/g in dry weight)

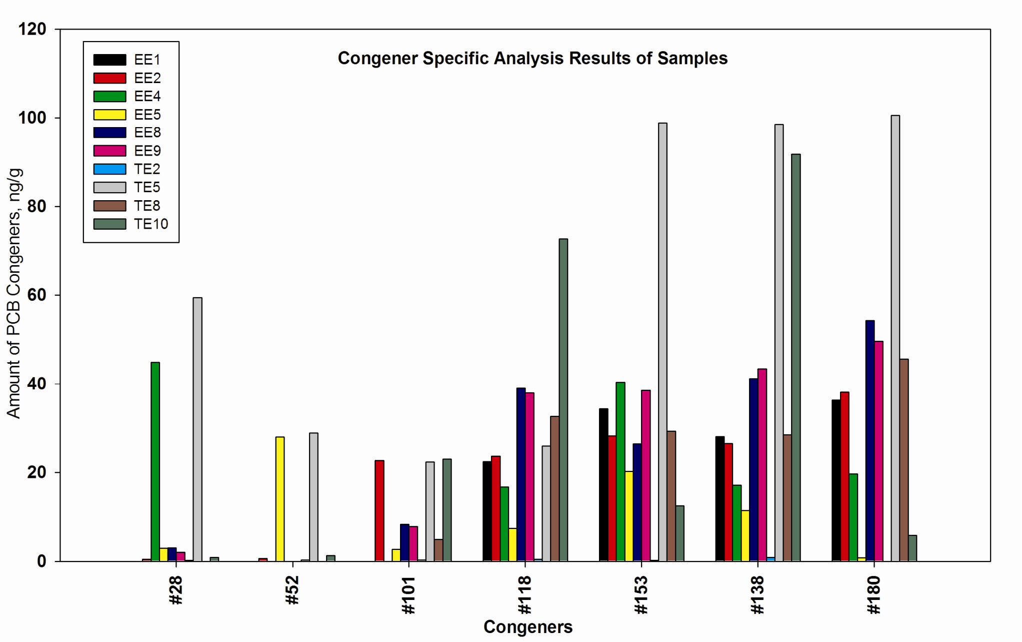


Figure 4-5 Congener Profiles for 10 Samples of Vicinity of Eymir Lake Reproduced Using 7 Congeners

The average congener distribution for the samples specified above is given in Figure 4-6 with standard deviation of means. PCBs were always released to the environment as mixtures, so potential mixtures used in Turkey and sold in high amounts in the world are Aroclor (USA and England) and Clophen (Germany). So, the average congener profile is compared with Aroclor and Clophen mixtures reproduced using the same seven congeners. The profiles of these mixtures are given in Figure 4-7 and Figure 4-8. Average congener profile resembles the highly chlorinated Aroclor mixture, A1260.

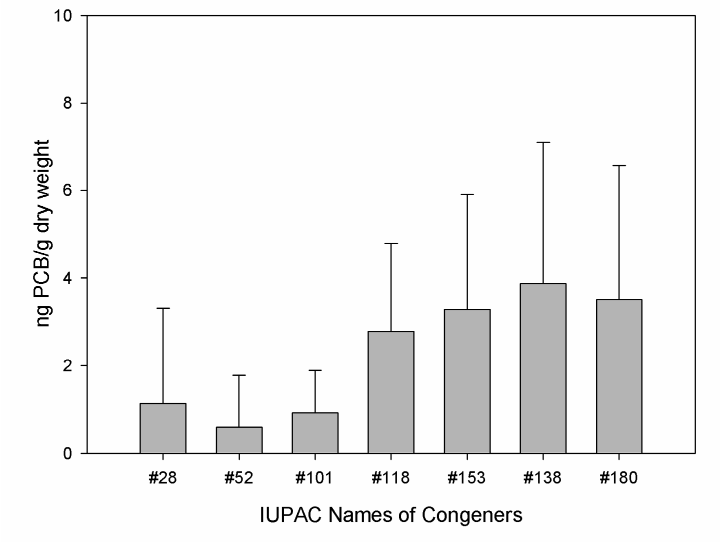


Figure 4-6 Average Congener Profiles of Vicinity of Eymir Lake Samples

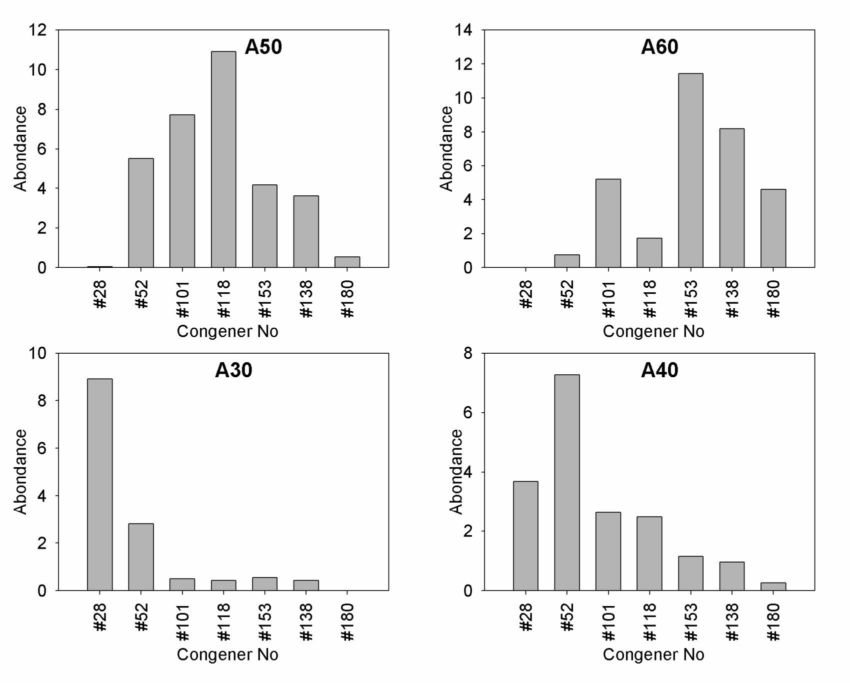


Figure 4-7 Clophen Profiles Reproduced Using 7 Congeners (Ishikawa et al., 2007)

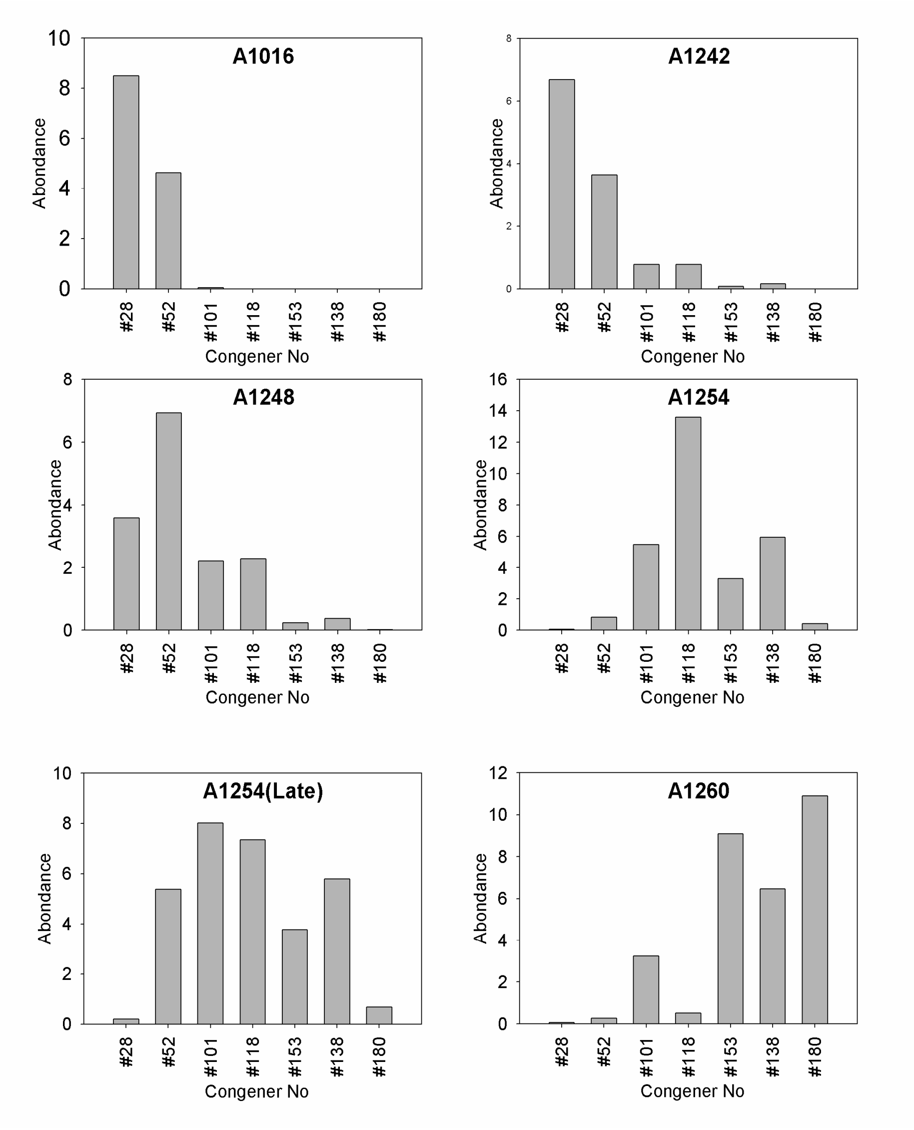


Figure 4-8 Aroclor Profiles Reproduced Using 7 Congeners (Frame et al., 1996)

The PCB concentration values found from the world are Ashtabula River 4.500 -0.300 ppm, for Fox River 44.000 – 0.046 ppm and at the Lake Michigan, Green Bay gulf between 0.300 -0.019 ppm (Imamoglu I. , 2001). This shows the PCB concentrations measured in the Lake Eymir should not be underestimated. Also it is possible to make comparison with the other studies carried out in Turkey. Firstly the findings of this study are consistent with Yeniova’s (1998) study. When the PCB concentrations are compared with findings, samples in the vicinity of TEDAS facility is found to be relatively polluted zones. This indicates the need of more extensive studies and rehabilitation suggestions.

Apart from sediment samples taken, a reef roof sample, E6, was taken in the second campaign. Its total PCB concentration was found as 600 ppb (ng/g). It indicates that the reef roofs can probably absorb PCB contamination. More samples are needed to confirm that event.

4.2 CMB Applications

The Chemical Mass Balance Model is used for the vicinity of Lake Eymir samples to find sources and their contributions. Aroclor (A1242, A1248, A1254, A1254 (late production), A1260 and A1016) and Clophen (A30, A40, A50 and A60) mixtures were evaluated in CMB model. CMB results applied to sediment samples are given in Table 4-4 with goodness of fit parameters. As specified in Table 4-4, source profiles and relative contributions of sources. PCB profiles in samples of vicinity of Lake Eymir cannot be explained by CMB model. In other words, the most of PCB profiles cannot be explained by linear combinations of Aroclor and Clophen profiles. One of the reasons for this is the nondetection of a number of PCB congeners. The model was run with 20% uncertainty for source profiles. VIF values calculated for the corresponding sample-source pairs were in the range of 1.06-6.99 with a median of 1.60, indicating insignificant collinearity among source profiles. The average relative error corresponding to χ^2=df is 0.51 for all, with a range of 0.28 to 0.74. Although these are acceptable, they also show that some congeners cannot be predicted satisfactorily with the case of these PCB sources.

Table 4-4 CMB Model Results of Lake Eymir Samples

Sample	Apportionment Results (%)					Model Statistics		
	Ar1016	Ar1254	A1260	ClA30	ClA40	χ^{2b}	$R^{2\,a}$	R.E^{b}
E1			57.0±24.8	43.0±19.8		2	94.9	0.28
E2			99.5±40.3	0.5±0.5		5	76.9	0.55
E4			40.3±14.4		59.7±19.4	3	92.8	0.31
E5 [d]					79.4±51.9	5	63.3	0.71
E8	2.7±2.0		97.3±43.4			4	75.7	0.58
E9	1.7±1.3		98.3±45.0			4	75.1	0.59
T2		58.3±23.6			41.7±17.0	4	88.3	0.39
T5	21.9±7.0	10.8±5.5	67.4±19.0			4	93.0	0.31
T8		1.4±48.5	98.6 ± 118.1			3	65.3	0.74
T10		82.7±45.5	17.3±14.2			5	68.9	0.64
Avg.	**8.77±3.11**	**38.3±20.18**	**71.96±34.08**	**21.75±13.65**	**60.27±19.49**		**79.42**	**0.51**

[a] R^2: multiple correlation coefficient, df: degrees of freedom (no. of congener – no. of sources), χ^2: Objective Function and R.E.: Relative Error. [b] χ^2=df. [c] Variance Inflation Factor. [d]: includes two sources; one (Cl60) is not shown, percentage of Cl60 is given as 20.6 ± 20.1.

As seen in Table 4-4, Aroclor profiles are typically determined as source and A1260 mixture comes across as the most common source identified for the sediments. Source contribution of A1260 including the highest chlorine content is between 17.3 % and 99.5 %. In Addition to that, A1254 mixture which contains lower chlorine content when compared to A1260 also contributes as to some extent to a number of samples (1.4 %-82.7 %). Clophen A40 has come across in two samples in high content (41.7 %- 79.4 %).

When the apportionment results in Table 4-4 are viewed, it can be seen that samples taken from the same locations in different depths are generally consistent with each other. Such that, E8 and E9 have almost the same apportionment results, E1 and E2 share A1260 as the dominant source while Clophen A40 is the dominant so for both E4 and E5.

Sample T8 has 0.74 sample relative error value which is higher than 0.7 is not in the confident interval specified in Rachdawong (1997). The literature sources selected do not explain T8 sample profiles satisfactorily.

CMB model reveals that the main sources for samples taken from the vicinity of Lake Eymir are Aroclor or Clophen mixtures including highly chlorinated PCBs. These mixtures have been typically used in the transformers which are used for electricity production and transmission. Although these are closed systems, it is known that they can enter into the nature as a result of accident or inappropriate/ illegal storage for long periods of time. The result of this study refers to PCB pollution in the region due to transformers around the TEDAŞ repair and maintenance facility. Aroclor mixtures are mostly determined as sources in this study. It is also compatible with the import information for PCBs (Gedik & İmamoğlu, 2010). Although A1016 is not used in transformers as oils, it is observed in 3 samples (E8, E9 and T5) with a high degree of uncertainty.

Regardless of the time and locations of sampling, A1260 comes up as a major contributor of PCBs to the sediments of the region. An example of a congener distribution calculated by the CMB model is illustrated in Figure 4-10 for sample T5. Congener 138 is underestimated in the sample. It decreases the fitting between measured and predicted profiles. Other congeners fit in the sample well.

The measured data vs. predicted PCB data for all samples are given in Figure 4-9. These plots indicate some over and under estimated congeners exist in the predicted sample. Generally, model underestimates values when compared with measured results. As specified in study of Chiarenzelli et al. (1997), significant volatile loss in low chlorinated congeners (#28, #52 and #101) is seen in soil/sediments to the ambient air. It causes increase in the relative error as tabulated in Table 4-4. The CMB results of PCB profiles congeners are given in Appendix D.

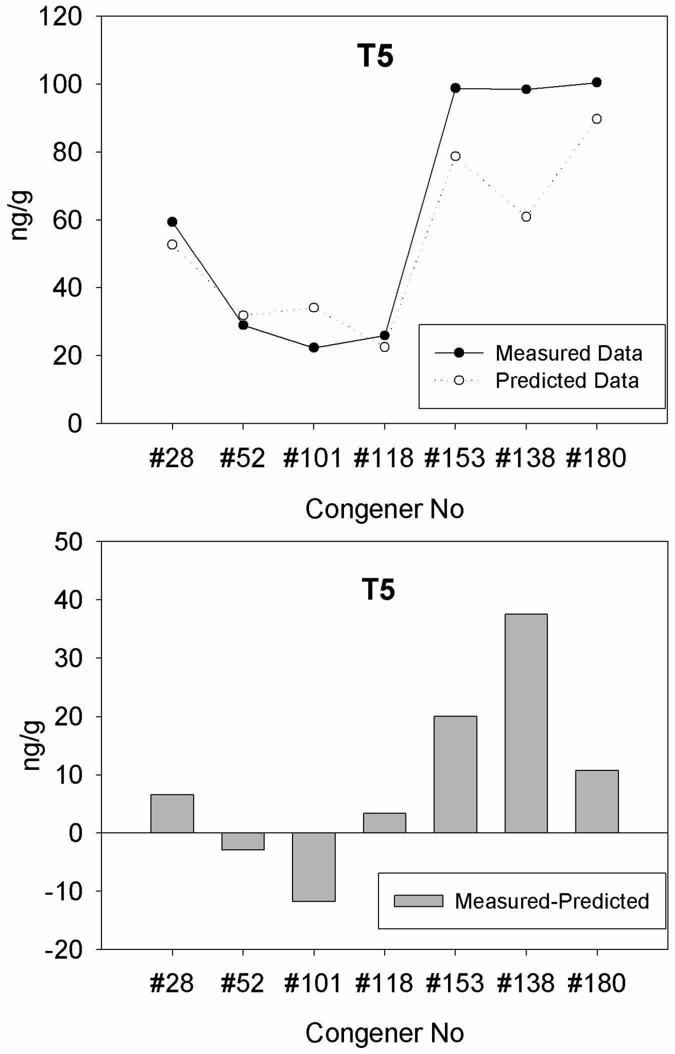


Figure 4-9 A: The measured and predicted congener profile of a sample B: The difference congener profile of predicted and measured profiles

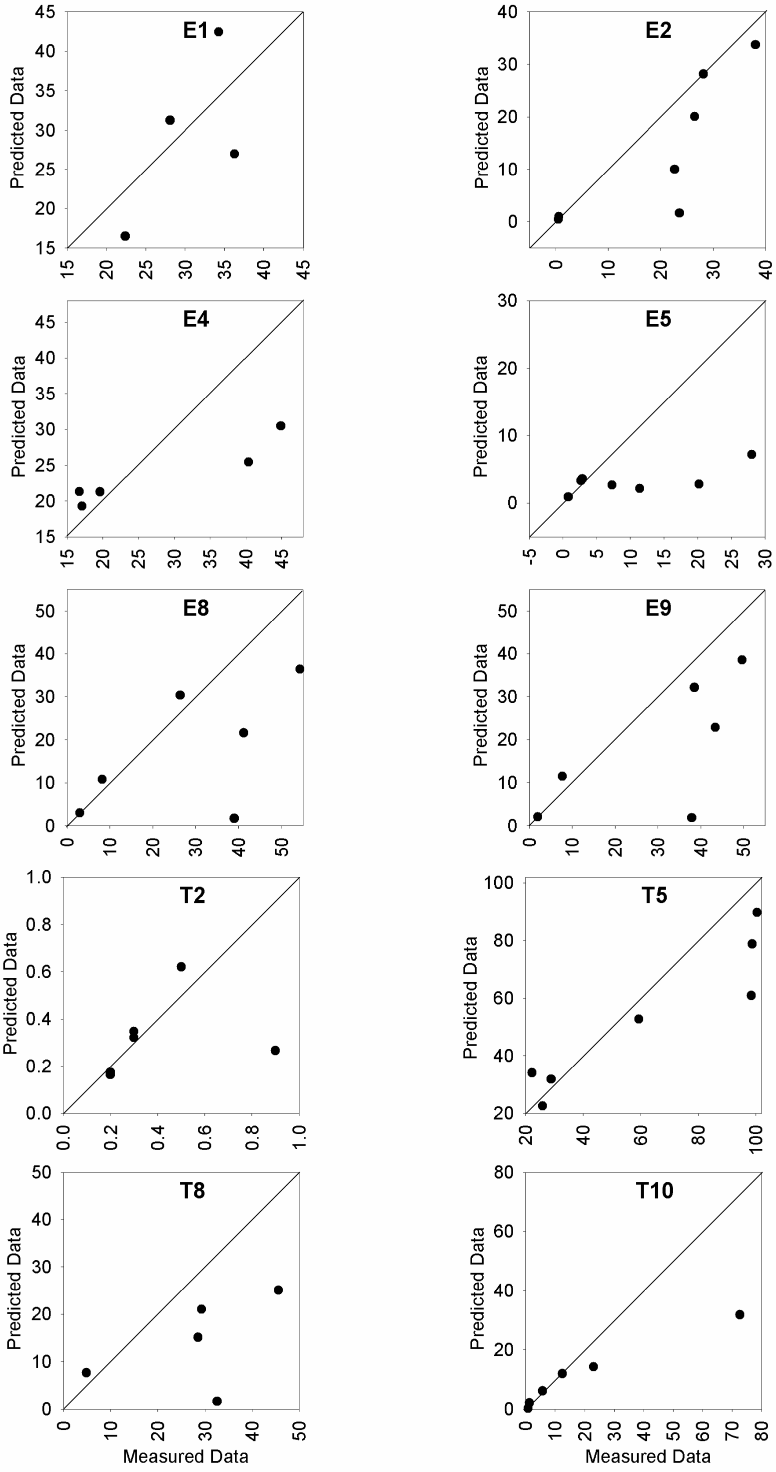


Figure 4-10 Measured vs. Predicted PCB Pollution for Vicinity of Lake Eymir

4.3 PMF Applications

4.3.1 Creation of Artificial Data Sets and Error estimates: Varying Complexity

As stated by Hopke (2010), the PMF model is a complex model. Therefore, experience is needed to understand the model behaviour. So, artificial data sets which have known source profiles were used to be able to compare outputs of the model with known input data. A detail discussion about the reasons of using artificial data sets in the PMF modelling is also presented in 3.3.4.6.

4.3.1.1 Simple Data Set (PD0)

PD0 data set has 5 PCB congeners and 2 sources from Frame (1996). In this data set, Monte Carlo simulation was not used. Data set was formed as putting percentage of sources from Frame (1996) to data set directly (Figure 4-11). The data set, average and standard deviation of PD0 are given in Appendix E.

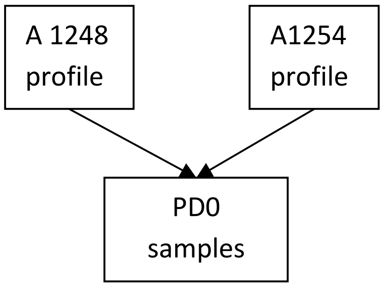


Figure 4-11 The Sources Used in PD0 Data Set

The congeners selected were; #70, #99, #101, #110 and 118, and sources were; Aroclor 1248 (shortly A1248) and Aroclor 1254 (shortly A1254). There are total 10 samples where congener percentages in A1248 and A1254 were put into directly for first 5 samples and last 5 samples, respectively (Figure 4-12). When this data set was formed, these points were considered;

- The congeners selected should not be below detection limit or “0”. By this way, effects of values of each uncertainty can be observed for measured data alone.
- Percentages of same congeners selected in two sources should not differ much so that the separation power of the model can be tested.
-

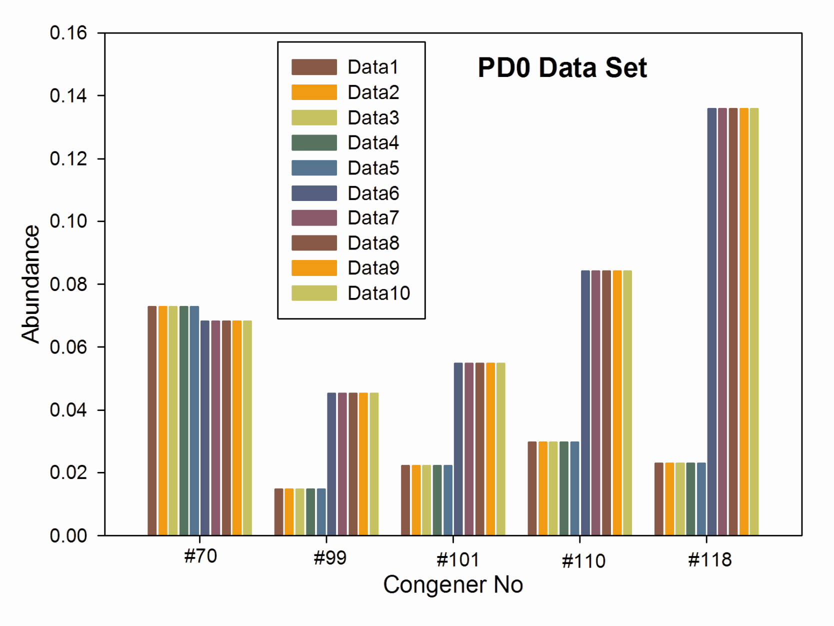


Figure 4-12 PCB profile of PD0 Data Set

4.3.1.2 Complicated Data Set - I (PD1)

Secondly, PD1 which is more complex than PD0 was developed for verification of PMF. Monte Carlo Simulation was executed to create 70 samples by using 4 sources and 26 congeners. Two sources, A1016 and A1260, used for this data set are taken from Frame (1996). A1260 and A1016 have very little same congeners. Moreover, two degradation profiles are used to observe degradation and physico-chemical weathering which PCBs can undergo in the environment. The detailed information about degradations is given in section 2.1.3. These two profiles are; (i) an anaerobically degraded Aroclor 1260 profile, as obtained from the field study by Bedard et. al. (1996), namely 9WP and (ii) a volatilized A1260 profile, as obtained from the experimental study by Chiarenzelli et. al. (1997). In, the sources are simulated by Monte Carlo to generate PD1 data set (Figure 4-13). The PCB distribution of PD1 data set is given in Figure 4-14.

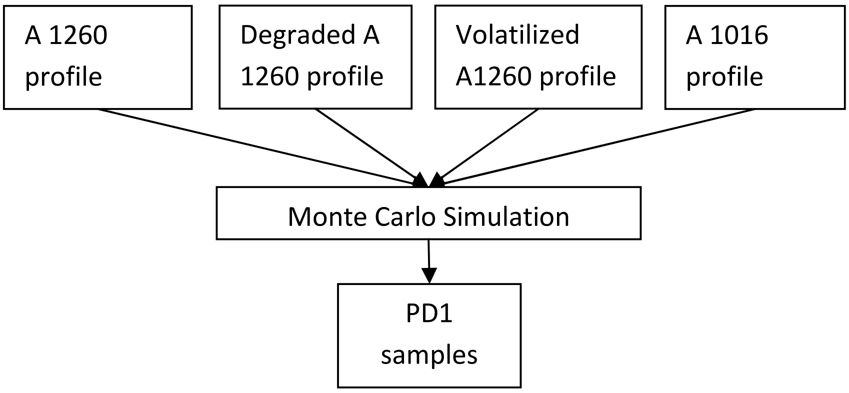


Figure 4-13 The Sources Used in PD1 Data Set

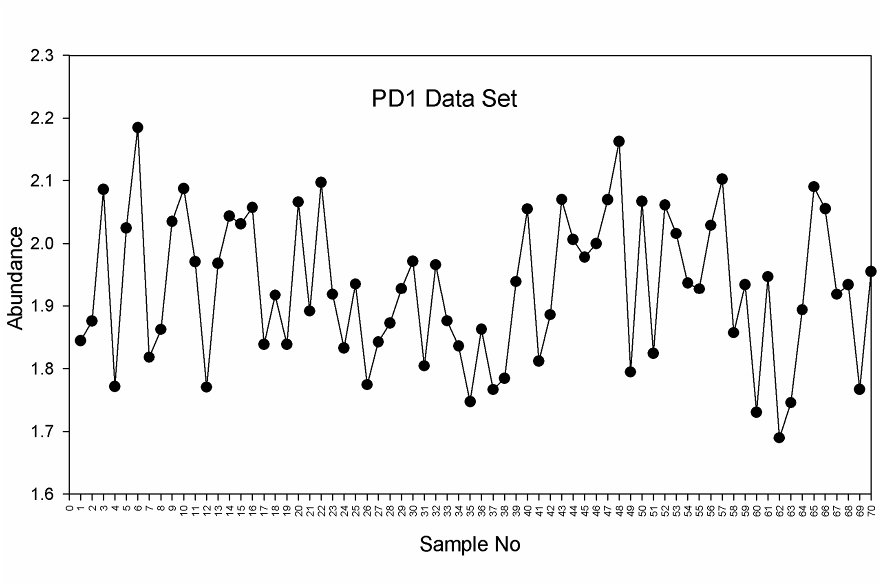


Figure 4-14 PCB Distribution of PD1 Data Set

In Monte Carlo Simulation, different coefficients of variation are used for each congener to simulate the data set. These coefficients are displayed in Table 4-5. While giving these coefficients, there were two important considerations: phyisico-chemical properties of PCBs and the results of field and laboratory studies in terms of fate of PCBs in the environment. Here, a higher value indicates potential for greater variability, while smaller values indicate persistency and smaller degree of effect of environmental fate on the congener. The same CV Values are used for same congeners in different sources. Monte Carlo is executed once for 70 samples, 4 sources and 26 congeners. Sources are normalized since they are prepared according to different units in literature. It is interpreted in the following section.

Table 4-5 Coefficient of Variation Values Used for Monte Carlo Simulation of PD1

Source Congener	**A1016, Vol1260, 9WP, A1260**
#18	0.4
#25	0.3
#26	0.3
#28	0.3
#31	0.3
#32	0.3
#44	0.3
#52	0.3
#60	0.2
#64	0.2
#66	0.2
#70	0.2
#87	0.2
#99	0.2
#101	0.2
#110	0.2
#118	0.2
#132	0.2
#138	0.2
#141	0.2
#149	0.2
#153	0.2
#170	0.2
#174	0.2
#180	0.1
#187	0.1

4.3.1.3 Complicated Data Set - II (PD2)

The third artificial data set which is more complex than others is to verify PMF model. Monte Carlo Simulation was executed for 40 samples by using 4 sources and 25 congeners. In PD1 data set, the sources were A1260 and A1016. As different from A1254 and A1260 were used since separation of them are more difficult than A1016 and A1260. Here, the aim is to resemble data set Lake Eymir sediments.

In this data set, two sources, A1254 and A1260, from Frame (1996) were used as in previous ones. The reason for selection of A1254 and A1260 as source is that these mixtures are generally used in transformers. Transformers are the major sources of PCB pollution both in Turkey and the world. Furthermore, these sources are observed as the results of congener specific analysis in the scope of Lake Eymir study. These two sources are used in simulation to display the real field conditions. Another reason for using these two sources that A1254 and A1260 mixtures includes much equivalent amount of congeners. Successful separation of these sources by PMF is more difficult than separation of two different sources having significantly profiles such as A1016 and A1260. Therefore, PMF can be used easily in the soils polluted by different sources.

In addition to these sources, two different sources are used in this artificial data set. These sources represent probable degradation and fate of PCBs in real environment. As different from PD1, one of the sources is A-35-2 profile from Bedard, (1996). It represents anaerobic degradation profile of A1260 taken from Woods Pond, Housatonic River, USA. Second source is volatilization A1260 profile from Chiarenzelli et. al. (1997).

When PCB congeners to be used in data set are decided, congeners distinguishing and defining four sources i.e. marker congeners, were selected. These are: #149 and #180 for A1260; #70, #99, #110 and #118 for A1254; #52, # 153, #180 and #187 for A-35-2 and #149 for VolA1260. According to these, numbers of chlorines, the congeners having high possible biological or physico-chemical degradation in the environment were also taken into consideration and a list of 25 congeners were prepared: #18, #28, #31, #33, #44, #49, #52, #70, #74, #87, #95, #99, #101, #110, #118, #132, #138, #149, #151, #153, #170, #180, #183, #187, #199.

Generation of Artificial Data Set by Monte Carlo Simulation:

A1260 and A1254 sources in Frame et al. (1996) are expressed as with percentages. Units of other sources are all converted to weight % in order to provide a consistent unit for all sources.

A-35-2 given by Bedard et al. (1996) was in mole percentage unit. Percentage in weight for each congener in the profile is calculated with the following formula:

$$X_a = \frac{MW_a \times n_a}{\sum_{a=1}^{k} MW_a \times n_a} \times 100 \qquad (4\text{-}1)$$

X_a= Percentage in weight for selected ath congener,

MW_a= Molecular weight of selected ath PCB congener,

n_a= mole of selected ath PCB congener,

k= total number of PCB congeners in the sample collected for A-35-2

Chiarenzelli et. al. (1997) presented the volatilized A1260 profile (VolA1260) in normalized form in a table. Therefore, the formula below is used to express all congeners in weight % base like the other three sources:

$$X_a = \left(\frac{100 - Y_a}{100}\right) \times Fr_a \qquad (4\text{-}2)$$

X_a= Percentage weight of selected ath PCB congener after volatilization

Y_a= Volatilization percentage of ath PCB congener normalized by Chiarenzelli et. al., (1997),

Fr_a= Percentage weight of ath PCB congener in A1260 from Frame et. al. (1996)

The data set size was selected as 40 samples by 25 congeners. Hence, the Monte Carlo Simulation was run for three times to form PD2 data set (Figure 4-15 and Table 4-6):

- Monte Carlo Simulation was done for the first 20 samples by using only A1254 and A1260 sources,
- Monte Carlo Simulation was done for the following 10 samples by using only A1260 and VolA1260 sources,
- Monte Carlo Simulation was done for the last 10 samples by using only A1260 and A-35-2 sources

In this artificial data set, the first twenty samples represent sediment PCB pollution due to PCB oils into transformer. The following 10 samples represent PCB pollution due to undergoing biological degradation after A1260 pollution. The last 10 samples represent surface soil/sediment samples polluted by PCBs which has undergone volatilization as a result of wetting/drying cycles. In conclusion, the PD2 data set was formed as 40 samples x 25 congeners. Figure

4-15 and Table 4-6 explain the formation of PD2. The PCB distribution of PD2 is displayed in Figure 4-16 and average and standard deviation values of PD2 are given in Appendix F.

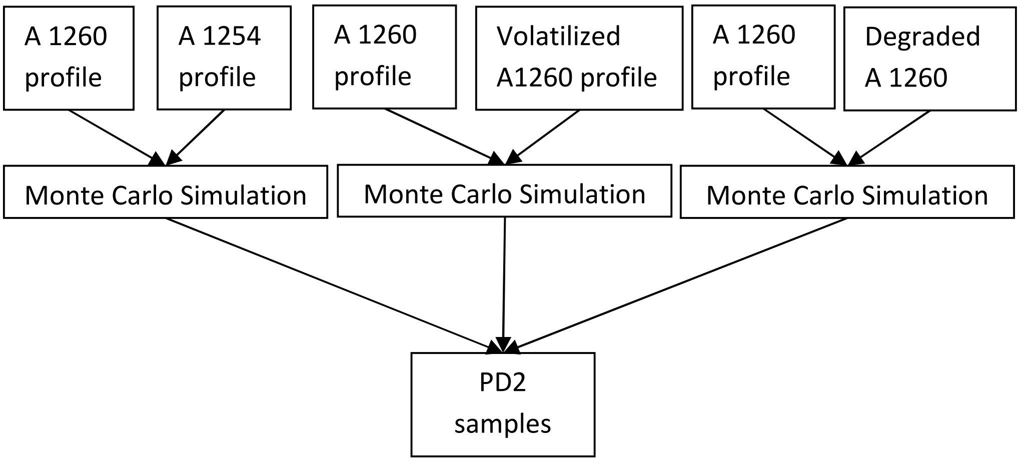


Figure 4-15 The Sources Used in PD2 Data Set

Table 4-6 Generated PD2 Data Set Prepared in 3 Stages

Sample Stages	**Sources Used**	**Purpose**
1-20	A1254 & A1260	Unchanged contaminated soil
21-30	VolA1260 & A1260	A1260 contaminated soil, undergone volatilization
31-40	A-35-2 & A1260	A1260 contaminated soil, undergone dechlorination

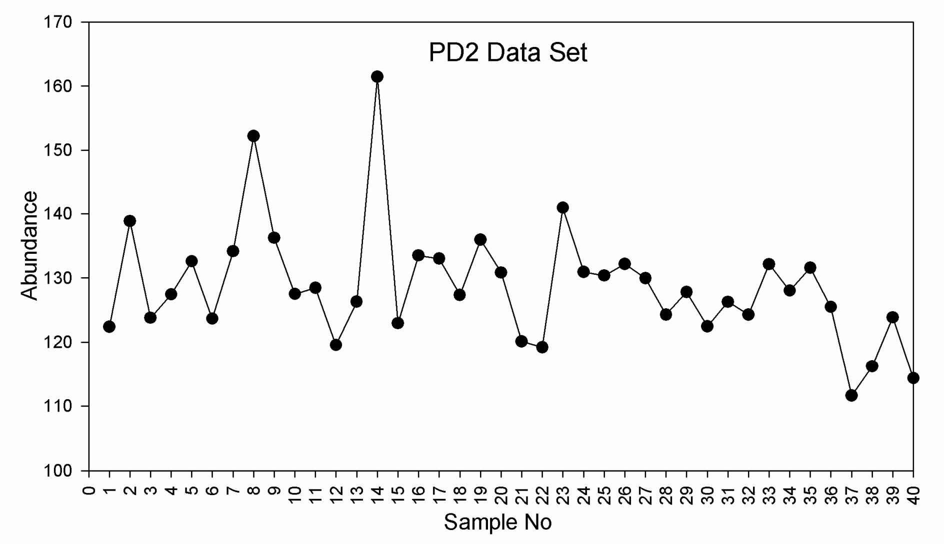


Figure 4-16 PCB Distribution of PD2 Data Set

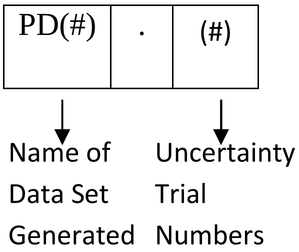
4.3.2 PMF Results on Artificial Data Sets

Three artificial data sets, from simple to complex, were used in this study. The basic aims to use PMF model in these data sets are:

- to understand the working principles of the PMF model,
- to identify the correct number and composition of sources in known artificial data inputs,
- to identify degradation profiles of PCBs,
- to investigate PMF sensitivity to the uncertainty data matrix as well as to parameters such as FPEAK, SEED vs. by systematically testing each parameters,
- to investigate the effect of the changes in the uncertainty matrix data values on the outcome of the model, specifically in the success of

identification of PCB patterns and improvement in the resolution of these profiles.

Three artificial data sets (PD0, PDI and PDII) from simple to more complex were prepared to understand the sensitivity of the PMF model. Different uncertainties have been tried on these data sets. Meaning of the number after point in data name denotes the name of the uncertainty trials. For example, PD2.1 means that data set is the most complex one (PD2) and uncertainty trial is the first one.



In the trials, uncertainties have been changed point by point. However, it will not be possible to change point by point for real data sets due to unknown inputs. In this study, the general aim is to understand the working principles and obtain an overall formulation on the use of the PMF model.

Detection limit in the model was taken as 0.0001 as declared in Frame et. al. (1996). To search for the sensitivity of uncertainty matrix and effects on model results of three data sets generated from simple to complex , the studies done are given below.

4.3.2.1 Simple Data Set: (PD0)

PD0 was used as input in the PMF model using 4 different uncertainty files. The discussions are made regarding goodness of fit parameters previously explained in Materials and Methods.

Different uncertainty matrices are experienced in the scope of PD0 trials. As seen from Table 4-7, relationships between Q_{theo} and global Q_{robust} were monitored for different uncertainty matrixes regarding fpeak values in one seed. Q_{robust} results of 4 different trials cannot catch the Q_{theo} (20).

Table 4-7 Q_{robust} Results of 4 Different Uncertainty Trials To Values of Seed and Fpeak

Trial No	**Changes in Uncertainty File**	**SEED**	**FPEAK**	**Q_{robust}**
1.	0,05*Data+DL	2	1	0,0107
2.	0,05*Data+DL	2	2	0,0937
3.	0,05*Data+DL $Std_{\#118}$= DL for last 5 samples	2	1	0,095
4.	Instead of uncertainty file, Ad-hoc eqn. is used (C1=0.0001 C2=0 C3=0.1)	2	1	0,3468

Q_{theo}= 20

In Table 4-8, second column displays percentages of 5 congeners used in two Aroclor sources from Frame et. al. (1996). Percentages of two sources are calculated by summing values of 5 congeners in A1254 and A1248 sources taken from the article and dividing sum of congeners in each source. Second and third rows in second column indicate A1248 and A1254, respectively (F1=29.46%, F2=70.54%). Not much difference is observed in explained variances of first and second trials having same starting random number (seed=2). As a result, third and forth trials give results close to Aroclor percentages.

Table 4-8 Explained Variance of F For each Factor and Aroclor Profiles

Trial Name/EV-F	F1, % (A1248)	F2, % (A1254)	Residuals, %
Actual	29.46	70.54	-
1.Trial, sd=2 fp=1	55.12	44.82	0.08
2. Trial, sd=2 fp=2	55.16	44.62	0.2
3. Trial, sd=2 fp=1	39.4	60.56	0.02
4. Trial, sd=2 fp=1	79.32	20.04	0.64

* Percentages of Aroclor Profiles in first and second sources

The profiles obtained from the results of F loading matrix present a discussion to evaluate the outputs of PMF model. The source percentages of Aroclors (A1248 and A1260) and F Loading matrix obtained from four uncertainty trials are displayed. As seen in Figure 4-17, when uncertainty trial 1 and 2 which have same seed (i.e. to start from different random numbers) and different rotations are compared (i.e. to start the rotations), it is observed that congener #118 in A1248 and congener #70 in A1254 can not separated by the model. For this reason, in the third trial, DL value is given to congener #118 for last 5 samples in the uncertainty matrix in order to increase the sensitivity of model for this congener. By this way, the poor prediction of congener #118 in the 1st and 2nd trials was tried to be eliminated. As doing changes in uncertainty matrix point by point, the strength of the PMF model in separation of similar factor is aimed to be understood closely resemble.

In the 3rd trial, the results of factor loadings, Aroclor 1248 (F1) and 1254 (F2), closely resemble Aroclor profiles (Figure 4-17). In the uncertainty matrix, the changes made for the last 5 samples also improve the fit of congener #70 in F loadings. This result is very important to indicate that a change in the uncertainty of one congener may affect modelling of another congener. When the 3rd trial of PD0 is compared with 1st and 2nd trials, there is a significant improvement in the modelling. This result shows the importance of the uncertainty file in the outcome of the model even though no change was made in the uncertainty matrix for this congener.

The last trial (PD0 4.) is formed by using Had-hoc Formula instead of preparation of uncertainty file. F loadings of this trial are similar to the results of the 1st and 2nd trials, but it is not successful in separation of sources as much as the 3rd trial.

G scores used for evaluation of PMF outputs are displayed in Figure 4-18. G scores show the contribution of each source to each sample. F Loadings and G scores should be parallel for the same source. For PD0, A1248 and A1254 are used as sources and first 5 samples are directly under the effect of A1248 and last 5 samples are under the effect of A1254 only. Therefore, G scores should display these sources as a result of modelling. As understood from Figure 4-18, 3rd trial indicates that G1 and G2 scores give parallel results with A1248 and A1254, respectively. However, G scores of 4th trial don’t agree with source profiles. This specifies that the sources could not separate when ad-Hoc formula is used in the uncertainty matrix.

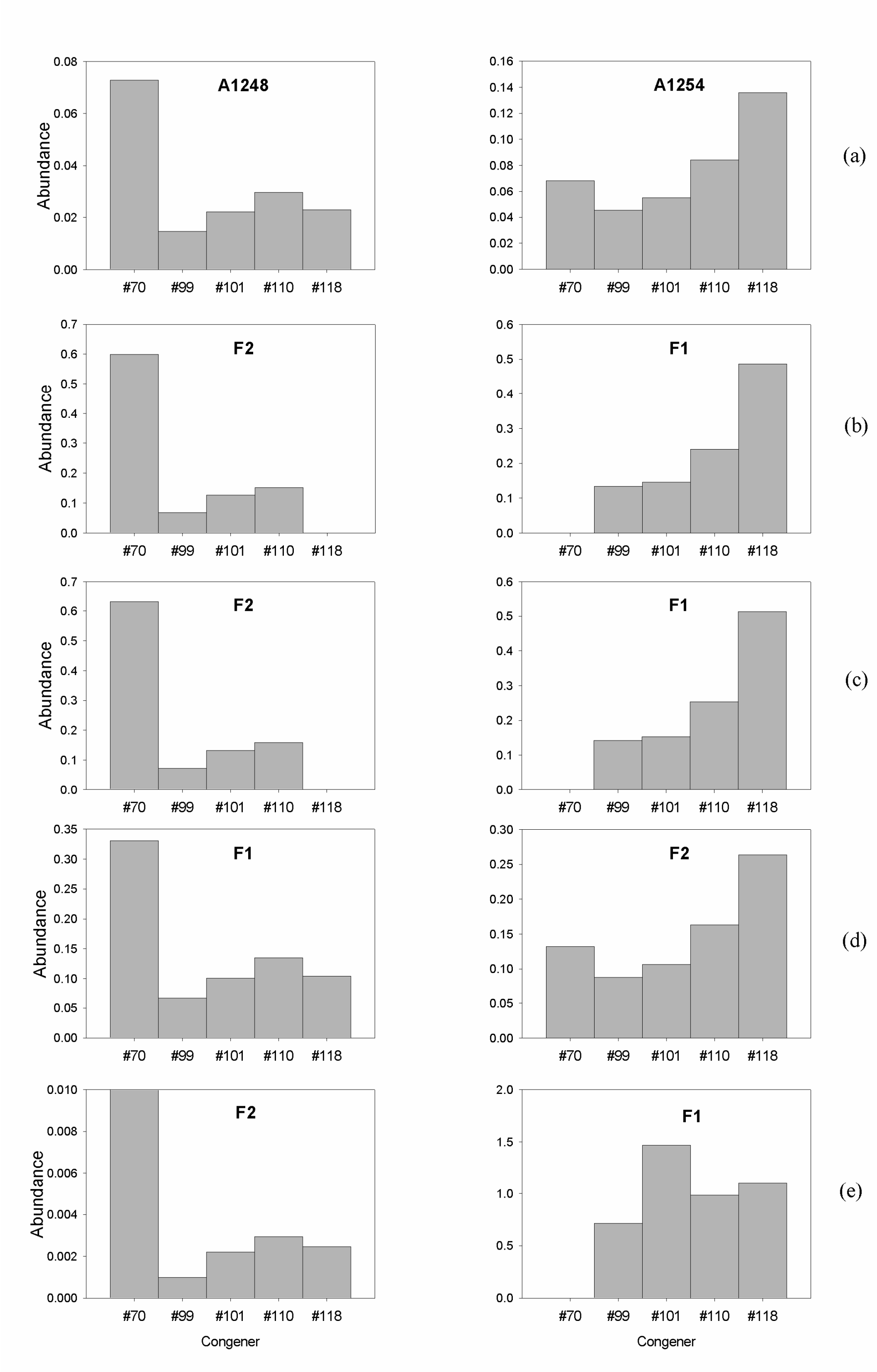
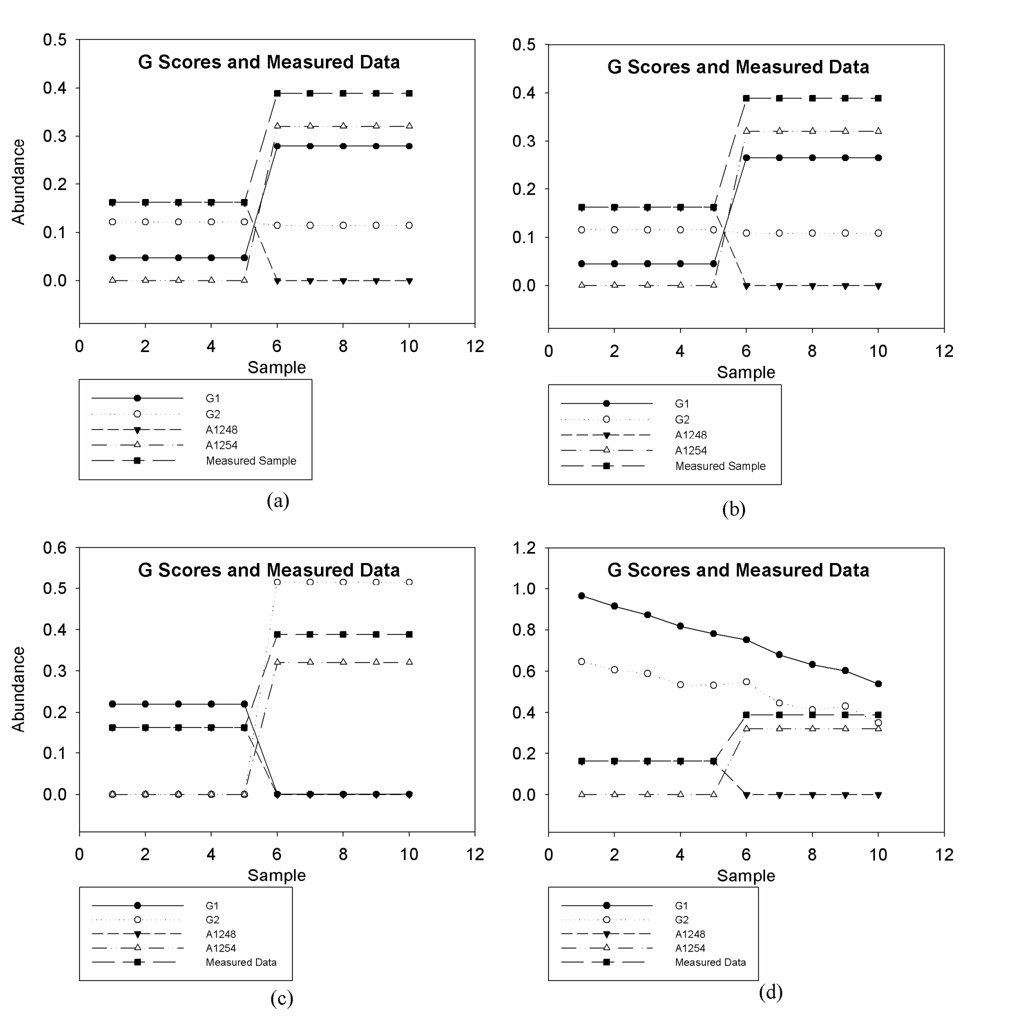


Figure 4-17 (a) A1248 and A1254 profiles, (b) Trial F Loading Results of PD0 1., (c) Trial F Loading Results of PD0 2., (d) Trial F Loading Results of PD0 3., (e) Trial F Loading Results of PD0 4



(a) (b) (c) (d)

Figure 4-18 Total PCB Values in Data Set and G Scores (a) PD0 1. Trial, (b) PD0 2. Trial, (c) PD0 3. Trial, (d) PD0 4. Trial

The goodness of fit parameters, COD and R^2, are tabulated for uncertainty trials of PD0 data set in Table 4-9. As understood from goodness of fit parameters in Table 4-9, model results of PD0 trials specify that trials express measured data set well, except for the 4th Trial. This shows that although input data can be predicted well, the source profiles may not have physical meaning, hence many parameters should be achieved before concluding that the output from the model is satisfactory and meaningful.

Table 4-9 COD and R2 Values for Different Uncertainty Trials

Trial No	**COD**	**R^2**
1. Trial	1 (all)	1
2.Trial	1 (all)	1
3. Trial	1 (all)	1
4. Trial	COD(#70) = -1936 COD(#99) = -49.73 COD(#101)= -165.2 COD(#110)= -31.96 COD (#118)= -11.40	0.72

Scaled residual matrix, another goodness of fit parameter, helps to comment on the model results. Hopke (2010) suggests that scaled residual values should be between +3 and -3 and should vary randomly. This was not the case for 1st, 2nd and 3rd trials as can be seen from Figure 4-19. The reason for this is that artificial data set has same sample values in two groups with 5 members, but random scaled residuals are seen in 4th trial due to random uncertainty given by the ad-hoc equation. However, this trial does not produce satisfactory results for other criteria regarding source compositions and contributions.

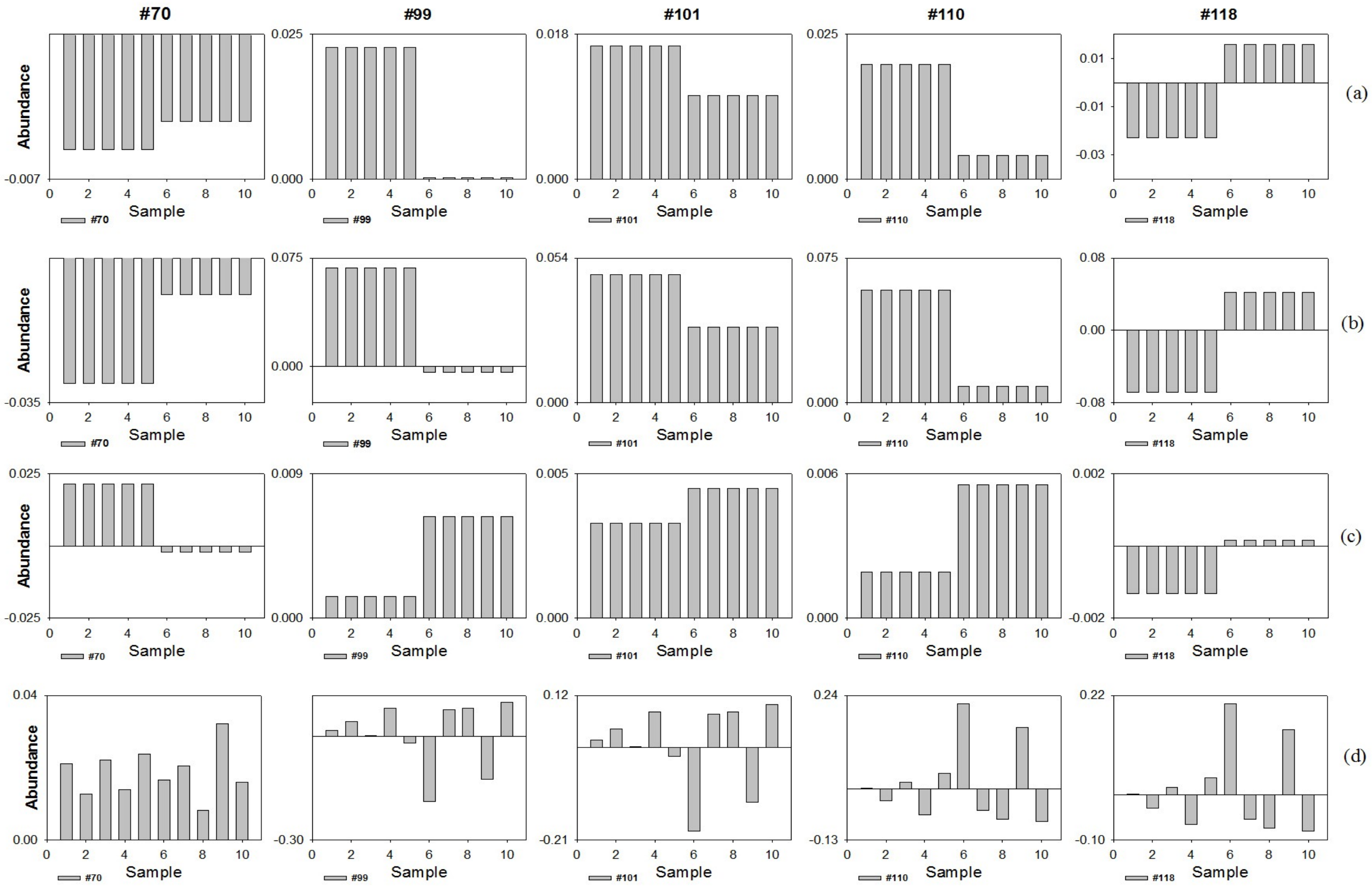


Figure 4-19 Scaled Residual Values for each PCB Congener: (a) PD0 1. Trial, (b) PD0 2. Trial, (c) PD0 3. Trial, (d) PD0 4. Trial

<u>*Assessment of PD0 Trials*</u>*:* Effects of different uncertainty files prepared for PD0 data set on PMF model results are evaluated. According to this;

- 3rd trial in different uncertainty files gives the best results. In this trial, the equation suggested by Hopke (2010) and selectively increased sensitivity for congener #118 chosen as a marker congener, are used.
- When uncertainty matrix is formed by using Hopke (2010)'s equation, smaller measured data values have relatively smaller uncertainty values and bigger measured data values have relatively bigger uncertainty values. This causes the model to behave very sensitive for small data and less sensitive for greater data. Therefore, the model becomes satisfactory as uncertainty value of the marker congener #118 in A1254 is decreased to 0.0001.
- Modeling of congener #70 is also affected when only the uncertainty of congener #118 in the uncertainty matrix is changed.

4.3.2.2 Complicated Data Set - I (PD1)

The values below detection limit are not considered in this data set and the congeners including "0" values in source profile are added. Only best result is given for this data set. As different from equation suggested by Hopke (2010), the equation expressed in Table 4-10 is applied in uncertainty matrix. In the profile, except for A1016 greater amount of high chlorinated PCB congeners are selected to be used in the Monte Carlo simulation. Therefore, after Monte Carlo simulation, lower chlorinated congener #18 specified by the equation "x_{ij}*0,01+DL" will have higher sensitivity than other congeners since values of congener #18 in data set is the smallest. As using equation "x_{1j}*0.1+DL" for congener #18, the model considers this congener to separate the data set. The optimum Q (87) is obtained by sd 4 and fp 1 (Table 4-10).

The explained variance results after minimizing Q for 3 factor trial is given in Table 4-11 and the results to control confirmation of goodness of fit parameters are provided in Table 4-12. The factor profiles of F Loadings are compared with Aroclor Profiles in Figure 4-20.

Table 4-10 Seed, Fpeak and Q_{robust} Results for Uncertainty Trial of PD1

Changes in Uncertainty File	Sd	Fp	Q_{robust}
x_{ij}*0,01+DL Std $_{\#18}$ = x_{ij}*0,1+DL	4	1	87

Q_{theo}= 70*26-26*3=1742

Table 4-11 Percentages of Explained Variance-F of The Model for PD1 Trial

Trial Name	A1016, %	A1260, %	WP9, %	Vol1260, %
Aroclor Profile	21	29	23	27
For Data Set				
	F1, %	F2, %	F3,%	Residual, %
PD1, sd=4 fp=1	22	46	32	0.5

Table 4-12 Results of Goodness of Fit Parameters for PD1

	COD %	R^2	Scaled Residuals
PD1 Trial	>99, $COD_{\#180}$ and $COD_{\#187}$ 53	1	<3

For validation of PMF model, PD1 is evaluated regarding to parameters explained in section 3.3.4. The results indicate that goodness of fit parameters are close to the values required. However COD values of PCB# 180 and #187 are nearly 50%. As seen in Figure 4-20, 3 factor model run success to separate 3 factors (9WP, A1016 and A1260). However, some congeners in the sources which factors resemble do not have same ratio. As a result, model will pay greater attention to measured data with the same formula and data set will be inspected in more detail.

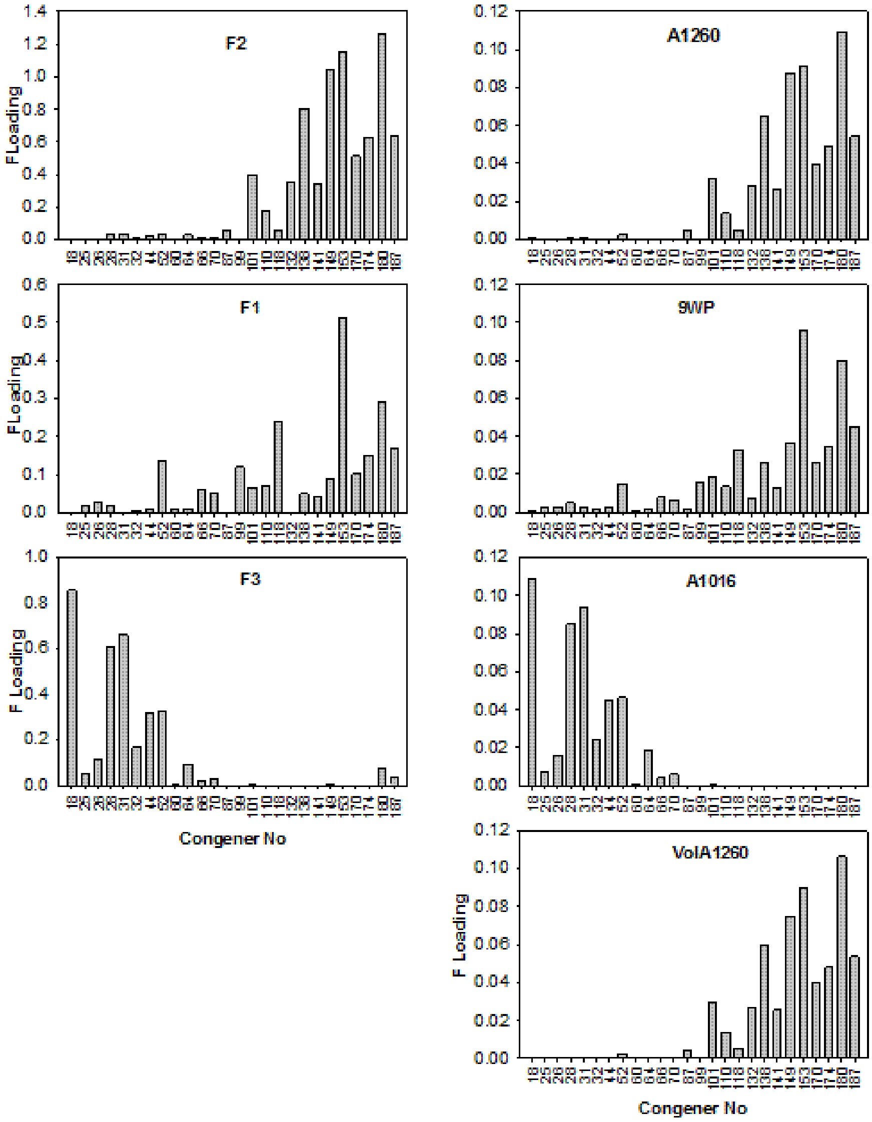


Figure 4-20 Comparison of Aroclor Profiles and Factor Loadings for PD1 Trial

Results show that overall, PMF could separate three of the four contributing sources to the data set. The unidentified source (VolA1260) has a profile significantly similar to Aroclor 1260. No further trials are performed with this data set so that a final simulated data set is prepared, reflecting, as much as possible, environmental sediment data expected to be obtained from a site such as Lake Eymir.

4.3.2.3 Complicated Data Set- II (PD2)

The third artificial data set generated for verification of PMF model is designed for 40 samples, 25 PCB congeners and 4 PCB sources. Detailed information about preparation of the artificial data set was given in section 4.3.1.

Different uncertainty files formed for PD2 data set and some model indicators of them are summarized in Table 4-13. As seen in Table 4-13, there are ten different group trials with PD2 artificial data set. In each of them only uncertainty matrix is changed while keeping data matrix unchanged. The changes done are explained in the last column of Table 4-13. In trials, the relationships and closeness between Q_{theo} and Q_{robust} are evaluated initially by looking at seed and fpeak values. Although artificial data set is generated from 4 sources with Monte Carlo Simulation, Q_{robust} outputs of various uncertainty files cannot reach to Q_{theo} (805) value for 4 factor results in the PMF model. In that case, the model is run with three sources. However, in 3 factor result, PD2.1 (When sd:1 and fp=-0.1, Q_{robust}=871) and PD2.x.2 (in all trials) reach to Q_{theo} (740) value.

Table 4-13 Trials and Changes in Standard Deviation Files and Summary Information about PMF

Trial No		***4- FACTOR RESULTS***			***3-FACTOR RESULTS***			**Changes Done in the Uncertainty Matrix**
		Sd	Fp	Q_{robust}	Sd	Fp	Q_{robust}	
PD2.1	1	1	1	53	1	-0.1	871	Constant 0.1 value is for all data points
	2	5	-1	22				
	3	**1**	**0**	**19**				
PD2.2	1	7	-1	138	2	-2	179	“x_{ij}*0.05+DL^{1}” equation is applied for all data points
	2	**7**	**0**	**26**				
PD2.X.1	1	3	1	59	**6**	**-2**	**5681**	Random numbers are given for data points.
	2	5	1	64				
	3	5	2	91				
PD2.X.2	1	2	-1	169	6	0	702	**For Sample between 1-20**: Congeners #70, #110, #118, #149 and #180 take DL, **For Sample between 21-30**: Congener #149 takes DL, **For Sample between 31-40**: Congeners #149 and #153 takes DL **For others**: 0.05*Data+DL/3
	2	**2**	**0**	**156**	6	0.5	726	
	3	2	0.8	168	8	-0.6	719	
	4				9	1	717	
FPD2.X.2	1	9	1	224	6	0.5	42119	As different from PD2.x.2, Uncertainty values for selected congeners in PD2.x.2 are applied to all samples 1-40, not to selected samples.
	2	**9**	**0**	**193**	9	1	42092	
PD2.X.3	**1**	**7**	**0**	**68**	8	0	54269	As different from PD.2.x.2, **For Sample between 1-40:** Congeners #18,#28,#31,#33,#44,#49,#52 takes 0.09 **For Sample between 1-20:** **Congener** #199 takes 0.09 **For Sample between 21-40:** **Congener** #99 takes 0.09
	2	7	-1	155				
	3	4	-2	248				
	4	8	-1	161				
PD2.X.5	1	5	1	38	3	-1	1164	1/x_{ij} equation is applied for all data points.
	2	5	0	7				

Table 4-13 (*Cont'd)* Trials and Changes in Standard Deviation Files and Summary Information about PMF

Trial No		***4- FACTOR RESULTS***			***3-FACTOR RESULTS***			**Changes Done in Uncertainty Matrix**
		Sd	Fp	Q_{robust}	Sd	Fp	Q_{robust}	
PD.2.X.6	1	5	0	0.2	5	0	3	(xijmax-xijmin)*1/xij equation are applied for all data points
	2	5	2	40	5	2	46	
PD.2.X.8	1	6	2	133	9	0	2727	1/xij*xijmax*xijmin equation are applied for all data points
	2	6	1	68	9	-2	2727	
	3	6	0	24				
PD.2.X.9	1	6	0.1	4468				1/x_{ij}*x_{ijmax}*x_{ijmin}*a equation are applied for all data points "a" is taken as 0.05
	2	6	0	4470				
CI.PD.2.X.10	1	4	-1	2715				As different form PD2.x.9, samples in congeners having high residuals in PD2.x.9 trial take "a" as 0.1
	2	4	0	2720				
CII.PD.2.X.10	1	9	0	1880				As different form PD2.x.9, samples in congeners having higher than 10 residuals in PD2.x.9 trial take "a" as 1 and those having between 3 &10 take "a" as 0.1.
CIII.PD.2.X.10	1	4	0	1911				As different form PD2.x.9, samples in congeners having higher than 20 residuals in PD2.x.9 trial take "a" as 0.4 , those having between 8 &10 take "a" as 0.2 and those having between 3 &8 take "a" as 0.1
$Q_{theoric}$		805			740			

[1] x_{ij}: each data in data set, DL: Method Detection Limit, Sd: Seed, Fp: Fpeak

First column of Table 4-14 indicates total percentage of Aroclor A1260, 1254, A-35-2 and VolA1260 source used in generation of PD2 artificial data set. Percentage amount of each factor calculated from explained variance of F loading matrix are given in Table 4-14. If explained variance of F loading which is an output of PMF model shows correct factor compositions, they are expected to be similar to percentage of four Aroclor percentages. Therefore, percentage distributions of Aroclor profiles are investigated with this data.

Table 4-14 Explained Variance Values of F Loading Matrix related to sources and Percentage of Aroclor Profiles among themselves

Aroclor Profiles %		**A-35-2**			**VolA1260**		**A1260**			**A1254**
		22.74			24.62		25.81			26.83
Trial No		***4- FACTOR RESULTS***					***3- FACTOR RESULTS***			
		F1	F2	F3	F4	Residual	F1	F2	F3	Residual
PD2.1	**1**	41.2	13.4	11.7	31.2	2.4	25.4	50.4	22.5	1.6
	2	25.1	21.1	2.8	50.2	0.8				
	3	23.7	22.3	16.4	36.9	0.7				
PD2.2	**1**	22.7	30.2	38.3	7.7	1.2	23.0	21.0	54.7	1.32
	2	23.0	26.5	30.3	19.7	0.5				
PD2.X.1	**1**	19.6	17.2	33.8	25.7	3.8	19.6	53.2	24.4	2.8
	2	21.7	30.3	24.8	18.4	4.9				
	3	22.4	30.4	24.7	16.6	5.9				
PD2.X.2	**1**	21.6	47.1	28.2	2.3	0.8	35.3	26.5	37.0	1.2
	2	22.9	34.4	23.6	18.3	0.9	49.2	20.3	29.3	1.2
	3	14.7	26.6	29.7	28.0	1.1	46.7	29.6	22.4	1.2
	4						29.2	20.4	49.2	1.2
FPD2.X.2	**1**	23.4	27.4	31.2	16.5	1.4	19.8	27.7	50.9	1.6
	2	22.7	20.4	33.6	22.1	1.12	19.9	27.8	50.6	1.6
PD2.X.3	**1**	23.8	14.9	23.3	37.1	0.9	19.5	32.5	14.9	33.1
	2	22.5	22.1	21.9	32.2	1.4				
	3	27.2	29.9	22.4	22.0	1.5				
	4	8.6	20.0	22.0	48.0	1.5				
PD2.X.5	**1**	23.5	18.8	30.3	22.6	4.9	48.9	22.1	25.1	3.9
	2	17.1	21.6	20.2	37.3	3.8				
PD2.x.6	**1**	21.0	32.3	19.1	16.0	11.6	36.6	25.5	26.1	11.8
	2	7.7	11.5	30.3	34.4	16.1	39.5	10.9	34.4	15.2
PD2.x.8	**1**	17.4	17.3	31.1	30.1	4.2	41.1	28.4	28.0	2.6
	2	15.2	26.0	23.4	32.5	2.9	48.9	25.6	22.7	2.8
	3	15.7	21.4	23.0	37.6	2.3				
PD2.x.9	**1**	9.0	37.3	19.2	33.6	0.8				
	2	13.1	25.6	22.7	37.8	0.8				
CI-PD2.x.10	**1**	48.0	4.5	22.2	24.5	0.9				
	2	37.9	13.0	22.6	25.7	0.7				
CII-PD2.x.10	**1**	25.9	37.9	22.5	13.0	0.7				
CIII-PD2.x.10	**1**	38.0	13.0	22.5	25.8	38.0				

F Loading, G scores and Explained Variance of F loading which are some of the goodness parameters are evaluated to understand success of trials. The discussions and results of uncertainty trials in the scope PD2 are presented in the tables between Table 4-15 and Table 4-27. Figures of F Loading results and their assessments are put in Figure 4-22 - Figure 4-26 to compare with Aroclor profiles in Figure 4-21. Additions to that, G scores graphs of trials in Figure 4-27 are compared with presence of added source in the samples to see the success of trials. For example for the first 20 samples A1254 is generated with using Monte Carlo Simulation. Therefore, for these samples, the factor A1254 should be observed in G score matrix. For other sources; they are VolA1260 between 21-30 samples, A-35-2 between 31-40 and A1260 between 1-40 samples. G score results are compared with first graph in Figure 4-27.

Table 4-15 The Results of F Loading, G Scores and Explained Variance of F Factor for PD2.1 (PD2.1.3-4F-SEED1fp0)

<table>
<tr><th colspan="2">F Loading Matrix</th><th>Explained Variance (%)</th><th>G Scores</th></tr>
<tr><td>F1</td><td>VolA1260</td><td>24</td><td>G scores are compatible with VolA1260. Samples between 21 and 30 seem to be affected from this factor. A few amount is observed for the samples between 1-20 and 31-40 in G score graph.</td></tr>
<tr><td>F2</td><td>A-35-2 (#180 and #153.not like in the profile)</td><td>22</td><td>G scores are compatible with A-35-2. Samples between 31 and 40 seem to be affected from this factor. A few amounts are observed for the samples between 1-20 in G score graph.</td></tr>
<tr><td>F3</td><td>A1260 (#180 and #149 not like in the profile)</td><td>16</td><td>G scores are compatible with A1260. Samples between 1 and 40 seem to be affected from this factor.</td></tr>
<tr><td>F4</td><td>A1254</td><td>37</td><td>G scores are compatible with A1254. Samples between 1 and 20 seem to be affected from this factor.</td></tr>
<tr><td colspan="4">Assessment: A small and constant uncertainty (0.1) gives good fit for 4 factors except for some congeners as seen in F profile in Figure 4-22. High percentage of A1254 calculated from EV-F also confirms this. For different fpeak rotations higher Q values are provided, but they don’t give equivalent F paths with artificial profiles. So fpeak 0 (without rotations) indicates better profiles. Although 0.1 uncertainty value is constant for all data values, it shows a good fit to data profiles. This is due to the fact that, 0.1 is small and between maximum and minimum values in the artificial data set (24 and 0.02 respectively). However, objective function is not close to theoretical Q. This can cause misunderstanding while deciding true factor numbers during modeling for environmental data sets. Therefore, new uncertainty files are tried.

Constant and much smaller 0.0001 uncertainty was also used as uncertainty in stdev file. The model didn’t run and get results for those small values. (Results of it are not shown).</td></tr>
</table>

Table 4-16 The Results of F Loading, G Scores and Explained Variance of F Factor for PD.2.2 (PD2.2.2-4F-sd7fp0)

<table>
<tr><th colspan="2">F Loading Matrix</th><th>Explained Variance (%)</th><th>G Scores</th></tr>
<tr><td>F1</td><td>A-35-2 (#180 and #153.not like in the profile)</td><td>23.01</td><td>G scores are compatible with A-35-2. Samples between 31 and 40 seem to be affected from this factor. A few amounts are observed for the samples between 1-20 in G score graph.</td></tr>
<tr><td>F2</td><td>VolA1260</td><td>26.52</td><td>G scores are compatible with VolA1260. Samples between 21 and 30 seem to be affected from this factor. A few amounts are observed for the samples between 1-20 and 31-40 in G score graph.</td></tr>
<tr><td>F3</td><td>A1254</td><td>30.25</td><td>G scores are compatible with A1254. Samples between 1 and 20 seem to be affected from this factor.</td></tr>
<tr><td>F4</td><td>A1260 (#180 and #149 not like in the profile)</td><td>19.74</td><td>G scores are compatible with A1260. Samples between 1 and 40 seem to be affected from this factor.</td></tr>
<tr><td colspan="4">Assessment: This uncertainty gives good fit for 4 factors except for some congeners as seen in profiles of F Loading in Figure 4-22. As different from PD2.1, percentage of A1254 source calculated from EV-F output is more close to real percentage in the source profile. Objective function, Qrobust (26) is not close to theoretical Q (805). Therefore, different rotations are tried with fpeak. Five times bigger Qrobust (138) is obtained by fpeak=-1. However, F loadings do not resemble to source profiles so results without rotations provide meaningful outputs. 3-factor results are tried in the model to see partitioning and increase in Qrobust. Qrobust (179) increases with 3-factor and fp=-2 rotation, but A1260 and Vol1260 cannot be separated.

Uncertainty equation suggested by Hopke provides different weighting (uncertainty) for each data point. But problem is for small data which will have smaller weighting than greater data, so the model tries to weight much for smaller values. To prevent it and to obtain Qrobust which is closer to theo. Q, smaller constant weightings are given for higher data points in the following trial.</td></tr>
</table>

Table 4-17 The Results of F Loading, G Scores and Explained Variance of F Factor for PD2.x.1 (PD2.x.1.1-3F-sd6fp-2)

<table>
<tr><th colspan="2">F Loading Matrix</th><th>Explained Variance (%)</th><th>G Scores</th></tr>
<tr><td>F1</td><td>A-35-2</td><td>19</td><td>G scores are compatible with A-35-2. Samples between 31 and 40 seems be affected from this factor.</td></tr>
<tr><td>F2</td><td>A1254 + A1260</td><td>53</td><td>G scores are compatible with A1254. Samples between 1 and 20 seems be affected from this factor. Since there are abundance in samples between 21-40 in G score graph, A1260 is also presence in this factor.</td></tr>
<tr><td>F3</td><td>VolA1260</td><td>24</td><td>G scores are compatible with VolA1260. Samples between 21 and 30 seem to be affected from this factor. A few amounts are observed for the samples between 1-20 and 31-40 in G score graph.</td></tr>
<tr><td colspan="4">Assessment: This uncertainty is prepared to see the effect of small and large values in uncertainty file. 3- factor outputs are evaluated in order not to partition PCB sources and determine the numbers of sources In 4-factor results (Q:91, sd:5 and fp:2). Since only A-35-2 and A1254 sources can be observed in F Loading for 4-factor result, G scores cannot help to distinguish sources in 4-factor output. In 3-factor result, as understood from % explained variance of F2, two sources (A1254 and A1260) is explained in one factor.

As a result, 3 factor F Loading and G scores results resemble to source profiles and artificial data set respectively, but one source is expected while running PMF model. Since 4 sources are not explained, some changes are added to uncertainty for some data points by using the advantages of PMF model which by uncertainties can be changed point by point in Standard deviation matrix. It will be easy to control weighting by changing from Hopke’s equation for selected data points.</td></tr>
</table>

Table 4-18 The Results of F Loading, G Scores and Explained Variance of F Factor for PD2.x.2 (PD2.x.2.4-4F-sd2-fp0)

<table>
<tr><th colspan="2">F Loading Matrix</th><th>Explained Variance (%)</th><th>G Scores</th></tr>
<tr><td>1</td><td>A-35-2</td><td>23</td><td>G scores are compatible with A-35-2. Samples between 31 and 40 seem to be affected from this factor. A few amounts are observed for the samples between 1-20 in G score graph.</td></tr>
<tr><td>2</td><td>A1254</td><td>34</td><td>G scores are compatible with A1254. Samples between 1 and 20 seem to be affected from this factor.</td></tr>
<tr><td>3</td><td>VolA1260</td><td>24</td><td>G scores are compatible with VolA1260. Samples between 21 and 30 seem to be affected from this factor. A few amounts are observed for the samples between 1-20 and 31-40 in G score graph.</td></tr>
<tr><td>4</td><td>A1260
(#149 not like in the profile)</td><td>18</td><td>G scores are compatible with A1260. Samples between 1 and 40 seem to be affected from this factor.</td></tr>
<tr><td colspan="4">Assessment: Some uncertainties are changed in the standard deviation file used in PD2.2 (Hopke’s equation). For some samples in artificial data set, detection limit is given as uncertainty. In this trial, effects of minus, zero and positive fpeaks are searched for optimum Q and goodness of fit parameters. Q values increase as decreasing and increasing fpeak values (2 and -2). However, the program gives “pulling down” message in some fpeaks and seeds and there is no calculation for these seeds and fpeaks. The reason of this error can be due to overflow, possibly caused by too strong pulling of the G factor element points determined by the model. This may be caused by too large value(s) in one lims, fpeak, and rotcom, Gkey values which are run parameters in PMF2DEF.INI file or too small uncertainty for a data point located on matrix in (a) row number.

Highest Qrobust for 4- factor is obtained with fp=-1, but F loadings results of this run are not similar with source profiles (Figure 4-23) and G scores of it are also not similar to raw data graph in three parts (Figure 4-27). As seen in Table 4-14, % Explained variance of F2 for fpeaks -1, 0 and 0.8, decrease respectively since the data which is subscribed from G factor is added to F factor in negative rotations and it is inverted in positive rotation. Qrobust which is close to theoretical Q is obtained by 3-factor runs. To improve factors and G scores and to obtain Qrobust close to theoretical Q with 4-factor runs, detection limit uncertainties given for selected congeners are applied to all samples.</td></tr>
</table>

Table 4-19 The Results of F Loading, G Scores and Explained Variance of F Factor for FPD2.x.2 (FPD2.x.2.2-3F-sd9-fp1)

F Loading Matrix		**Explained Variance (%)**	**G Scores**
1	Vol1260 or A1260	23	G scores are compatible with VolA1260. Samples between 21 and 30 seem to be affected from this factor. Higher amounts are observed in samples between 1-20 and 31-40 in G score graph indicates presence of A1260.
2	Vol1260 or A1260	20	G scores are compatible with VolA1260. Samples between 21 and 30 seem to be affected from this factor. Higher amounts are observed in the samples between 1-20 and 31-40 in G score graph indicates presence of A1260.
3	A1254 (#183, #187 and #199 in high amount seen)	34	G scores are compatible with A1254. Samples between 1 and 20 seem to be affected from this factor.
4	A-35-2	22	G scores are compatible with A-35-2. Samples between 31 and 40 seem to be affected from this factor. A few amount is observed for the samples between 1-20 in G score graph.
Assessment: The aim is to compare goodness of fit parameters with the parameters of PD2.x.2 in the same seed, fpeak and factor numbers in order to understand changes when detection limit uncertainties for selected congeners are given to all samples. Sd=2 and fp=0 trial which is best run for PD2.x.2, is not simulated since PMF model gives NDP error-divide by zero in cholesky due to overflow. Additional to that, other uncertainty trials, fp=0.8 and fp=-1, are not also simulated due to pulling down error. Therefore, sd=9 fp=1 and sd=9 fp=0 trials are evaluated. The results of the trial without rotation indicate factor separation better. Explained variances of factors in FPD2.x.2 are similar to that in PD2.x.2 (Table 4-18). However, G scores of FPD2.x.2 is not successful to separate A1260 and Vol1260 (Table 4-23) having same pathway. It indicates that giving uncertainty for a column (all samples) does not help separate the factors better. Like in PD2.x.2, pulling down and NDP errors are seen in different seeds and fpeaks for FPD2.x.2. They may show a problem of separation. To prevent and also to distinguish the factors in G scores, higher constant values are investigated for some congeners with lower data in new trial.			

Table 4-20 The Results of F Loading, G Scores and Explained Variance of F Factor for PD2.x.3 (PD2.x.3.1-4F-sd7fp0)

F Loading Matrix		**Explained Variance (%)**	**G Scores**
1	A-35-2 (#180 and #153.not like in the profile)	23.8	G scores are compatible with A-35-2. Samples between 31 and 40 seem to be affected from this factor. A few amount is observed for the samples between 1-20 in G score graph.
2	A1260(#180 and #149 not like in the profile)	14.9	G scores are compatible with A1260. Samples between 1 and 40 seem to be affected from this factor.
3	Vol1260	23.32	G scores are compatible with VolA1260. Samples between 21 and 30 seem to be affected from this factor.
4	A1254 (#183, #187 and #199 in high amount seen)	37.13	G scores are compatible with A1254. Samples between 1 and 20 seem to be affected from this factor.
Assessment: F Loading matrix describes PCB profiles like PD2.x.2, PD2.2 and PD2.1. However, expected resolution in G scores cannot be observed and Q values do not get close to theoretical Q values. 3-factor results have higher Qrobust when compared with PD2.x.2 and PD2.2. To approximate theoretical Q and to get high resolution G scores for goodness of fit parameters, new uncertainty is developed.			

Table 4-21 The Results of F Loading, G Scores and Explained Variance of F Factor for PD2.x.5 (PD2.x.5.2-4F-sd5-fp0)

"F Loading Matrix		**Explained Variance (%)**	**G Scores**
1	Vol1260 or A1260(#180 and #149 not like in the profile)	17	G scores are compatible with VolA1260. Samples between 21 and 30 seem to have been affected from this factor. Higher amounts are observed in the samples between 1-20 and 31-40 in G score graph indicates presence of A1260.
2	A-35-2	22	G scores are compatible with A-35-2. Samples between 31 and 40 seem to have been affected from this factor. A few amount is observed for the samples between 1-20 in G score graph.
3	Vol1260 or A1260	20	G scores are compatible with VolA1260. Samples between 21 and 30 seem to have been affected from this factor. Higher amounts are observed in the samples between 1-20 and 31-40 in G score graph indicates presence of A1260.
4	A1254(#183, #187 and #199 in high amount seen)	37	G scores are compatible with A1254. Samples between 1 and 20 seem to be affected from this factor.
Assessment: This uncertainty file is prepared to give higher uncertainty to small data and lower uncertainty to higher data. Sd=5 results which provide higher Qrobust to reach theoretical Q are investigated without rotation and with fp=1 in 4 factor runs. Results with rotation fp=1 having Qrobust (38) can not show any factors and congeners in each factor loading cannot distribute as in pathways. This is due to the reason that positive rotations provide G score abundance of F loading as subscribing. As seen in Table 4-28, COD values do not show perfect results and Q values are not reasonable. To improve uncertainty equation, new trial is prepared.			

.

Table 4-22 The Results of F Loading, G Scores and Explained Variance of F Factor for PD2.x.6 (PD2.x.6.2-4F-sd-fp0)

F Loading Matrix		**Explained Variance (%)**	**G Scores**
1	A-35-2 (lower chlorinated congeners not like in the profile).	21	G scores are compatible with A-35-2. Samples between 31 and 40 seem to be affected from this factor. A few amount is observed in the samples between 1-30 in G score graph.
2	A1254(#183, #187 and #199 in high amount seen)	32	G scores are compatible with A1254. Samples between 1 and 20 seem to be affected from this factor. A few amounts are observed in the samples between 21-40 in G score graph.
3	Vol1260	19	G scores are compatible with VolA1260. Samples between 21 and 30 seem to be affected from this factor. Higher amounts are observed in the samples between 1-20 and 31-40 in G score graph indicates presence of A1260.
4	A1260	16	G scores are compatible with A1260. Samples between 1 and 40 seem to be affected from this factor.
Assessment: Different from PD2.x.5, data is multiplied by difference between maximum and minimum values. So, higher uncertainty levels which all data is multiplied with same ratio are obtained. 2, 3 and 4 factor results are evaluated. They indicate that increase of uncertainty do not provide better factor fits. For factor results, optimum result is obtained by seed=5 and fp=0. As understood from goodness of fit parameter in Table 4-28, there is no improvement in Q values and provided better fit for F Loading and G scores by this trial.			

Table 4-23 The Results of F Loading, G Scores and Explained Variance of F Factor for PD2.x.8 (PD2.x.8.3-4F-sd5-fp0)

F Loading Matrix		**Explained Variance (%)**	**G Scores**
1	A1260 (#180 and #149 not like in the profile)	16	G scores are compatible with A1260. Samples between 1 and 40 seem to have been affected from this factor. Higher amounts in samples between 21-30 indicates similarity of Vol1260 pathway.
2	Vol1260	21	G scores are compatible with VolA1260. Samples between 21 and 30 seem to have been affected from this factor. Greater amounts are observed in the samples between 1-20 and 31-40 in G score graph indicates presence of A1260.
3	A-35-2	23	G scores are compatible with A1254. Samples between 1 and 20 seem to be affected from this factor. A few amount is observed in the samples between 21-40 in G score graph.
4	A1254(#183, #187 and #199 seen in high amount)	37	G scores are compatible with A1254. Samples between 1 and 20 seem to be affected from this factor. A few amount is observed in the samples between 31-40 in G score graph.
Assessment: The aim of this trial is to keep uncertainty between maximum and minimum data while forming a Standard deviation file which is inversely proportional to magnitude of data. As seen from F Loadings and G scores (Figure 4-25 and Figure 4-27), they are not much better than PD2.x.2 and PD2.2. Additionally, theoretical Q cannot be reached or as discussed by Öztürk 2010, it is not in the range of 50%. Therefore, new trial is formed to keep uncertainties in 5 percentage range by multiplying by data inversely. In 3 factor results with/without rotations, Qrobust is 2.5 times of Qtheo which is reasonable.			

Table 4-24 The Results of F Loading, G Scores and Explained Variance of F Factor for PD2.x.9 (PD2.x.9.2-4F-sd6-fp0)

F Loading Matrix		**Explained Variance (%)**	**G Scores**
1	A1260 (#180 and #149 not like in the profile)	13	G scores are compatible with A1260. Samples between 1 and 40 seem to be affected from this factor.
2	Vol1260	26	G scores are compatible with VolA1260. Samples between 21 and 30 seem to be affected from this factor. Higher amounts can be observed in the samples between 1-20 and 31-40 in G score graph indicates presence of A1260.
3	A-35-2	23	G scores are compatible with A1254. Samples between 1 and 20 seem to be affected from this factor. A few amount is observed in the samples between 21-40 in G score graph.
4	A1254(#183, #187 and #199 seen in high amount)	38	G scores are compatible with A1254. Samples between 1 and 20 seems be affected from this factor. A few amounts are observed in the samples between 31-40 in G score graph.
Assessment: "a" value (0.05) multiplied with equation in PD2.x.9 decreases uncertainty values for all data, so uncertainty values shows a better fit to data. Q values are better than values of other trials. Drawback of this trial is some congeners in residual matrix have Standard deviation values higher than 3 (Table 4-28). Therefore, higher amounts are given in new uncertainty trial for sample points having stdev higher 3 in these congeners.			

Table 4-25 The Results of F Loading, G Scores and Explained Variance of F Factor for CI-PD2.x.10 (CI-PD2.x.10.2-4F-sd4-fp0)

F Loading Matrix		**Explained Variance (%)**	**G Scores**
1	A1254(#183, #187 and #199 seen in high amount)	38	G scores are compatible with A1254. Samples between 1 and 20 seem to be from this factor. A few amount is observed in the samples between 31-40 in G score graph.
2	A1260 (#180 and #149 not like in the profile)	13	G scores are compatible with A1260. Samples between 1 and 40 seem to be affected from this factor.
3	A-35-2	23	G scores are compatible with A1254. Samples between 1 and 20 seem to be affected from this factor. A few amounts are observed in the samples between 21-40 in G score graph.
4	Vol1260	26	G scores are compatible with VolA1260. Samples between 21 and 30 seem to be affected from this factor. Higher amounts can be observed for the samples between 1-20 and 31-40 in G score graph indicates presence of A1260.
Assessment: In this trial, for the congeners having values higher than 3 in residual matrices, different “a” value (0.1) is assigned in uncertainty file. 0.1 (a) is higher than previous 0.05 value since high residuals are decreased with high Standard deviation in uncertainty file. As a result, lower Q theo (2720) is obtained, however some congeners (149, 132 and 180) have standard deviation higher than 3. As understood from % of explained variance of A1254 (%38), it is still not separated from A1260 significantly.			

Table 4-26 The Results of F Loading, G Scores and Explained Variance of F Factor for CII-PD2.x.10 (CII-PD2.x.10.1-4F-sd4-fp0)

F Loading Matrix		**Explained Variance (%)**	**G Scores**
1	A1260 (#180 and #149 not like in the profile)	26	G scores are compatible with A1260. Samples between 1 and 40 seem to be affected from this factor.
2	A-35-2	38	G scores are compatible with A-35-2. Samples between 31 and 40 seem to be affected from this factor. A few amounts are observed in the samples between 1-20 in G score graph.
3	A1254(#183, #187 and #199 seen in high amount)	22	G scores are compatible with A1254. Samples between 1 and 20 seem to be affected from this factor. A few amounts are observed in the samples between 31-40 in G score graph.
4	Vol1260	13	G scores are compatible with VolA1260. Samples between 21 and 30 seem to have been affected from this factor. Higher amounts can be observed for the samples between 1-20 and 31-40 in G score graph indicates presence of A1260.
Assessment: In this trial, the congeners having a value higher than 3 in residual matrix are separated into three group. First group contains residuals between ±3, second one contains residuals between -10 and -3 and between 10 and 3 and third group consist of residuals higher than 10 and lower than -10. “a” is equal to 1 for third group since higher uncertainty is provided for higher scaled residuals. “a” is equal to 0.1 for the second group and is equal to 0.05 for the first group. The aim is to see the effects of point by point change in uncertainty file on Q value and resolution of F and G factors. Qrobust (1880) is 2.5 times than theo. Q, so it is close enough to. Qtheo. Scaled residuals are in ±3 scale. Scaled residuals are very small (close to zero), this can be a problem. The reason can be due to “1” a value which is too high. Therefore, another trial is prepared for higher values in scaled residuals in PD2.x.9.			

Table 4-27 The Results of F Loading, G Scores and Explained Variance of F Factor for CIII-PD2.x.10 (CIII-PD2.x.10.1-4F-sd4-fp0)

F Loading Matrix		**Explained Variance (%)**	**G Scores**
1	A1254(#183, #187 and #199 seen in high amount)	38	G scores are compatible with A1254. Samples between 1 and 20 seem to be affected from this factor. A few amounts are seen samples between 31-40 in G score graph.
2	A1260 (#180 and #149 not like in the profile)	13	G scores are compatible with A1260. Samples between 1 and 40 seem to be affected from this factor.
3	A-35-2	23	G scores are compatible with A-35-2. Samples between 31 and 40 seem to be affected from this factor. A few amounts are observed in the samples between 1-20 in G score graph.
4	Vol1260	26	G scores are compatible with VolA1260. Samples between 21 and 30 seem to be affected from this factor. The amount observed for the samples between 1-20 and 31-40 in G score graph indicates presence of A1260.
Assessment: Source contributions and source compositions are similar in CIIPD2.x.10 where goodness of fit parameters indicates similar results. As different from that trial, here, Qrobust is found higher.			

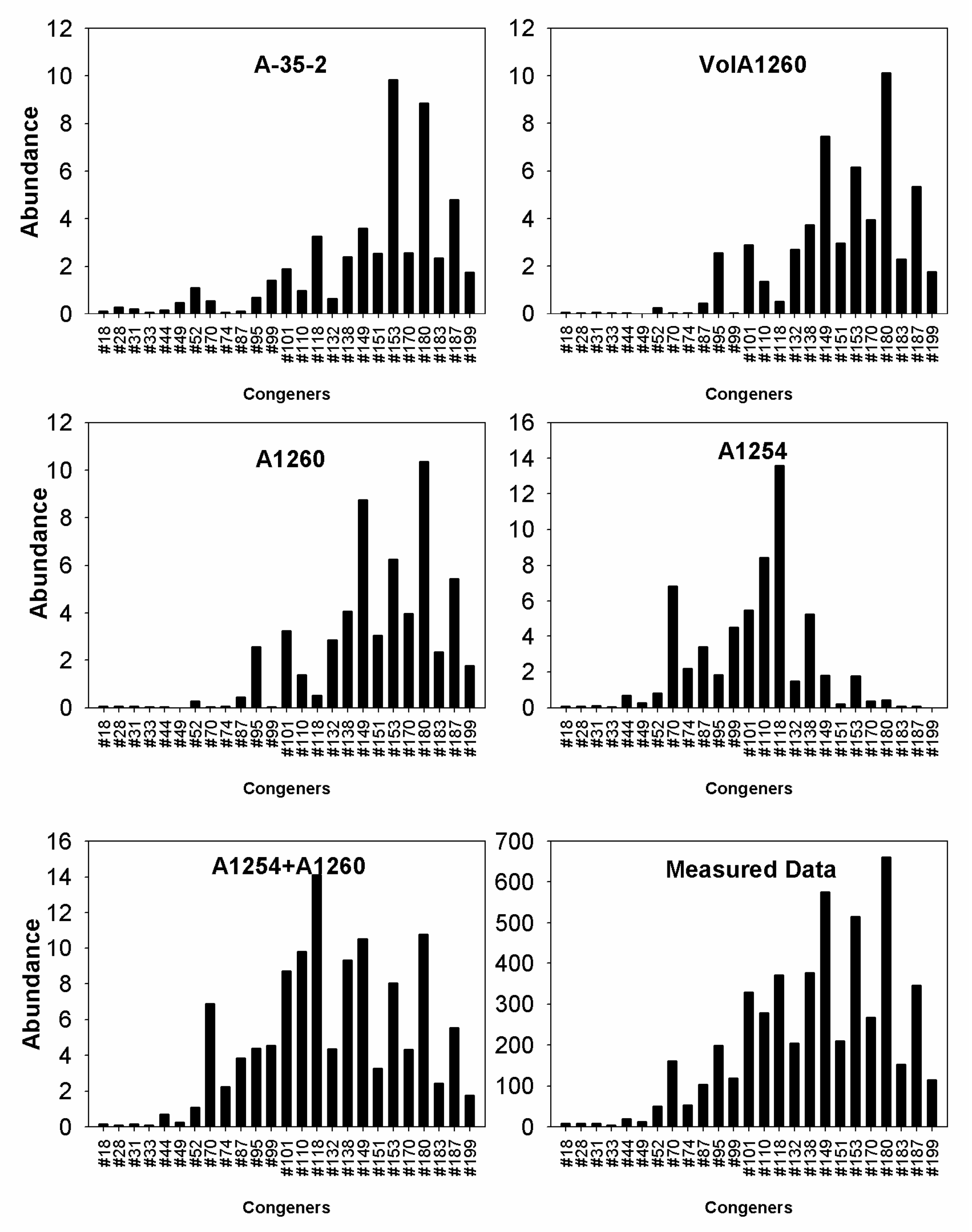
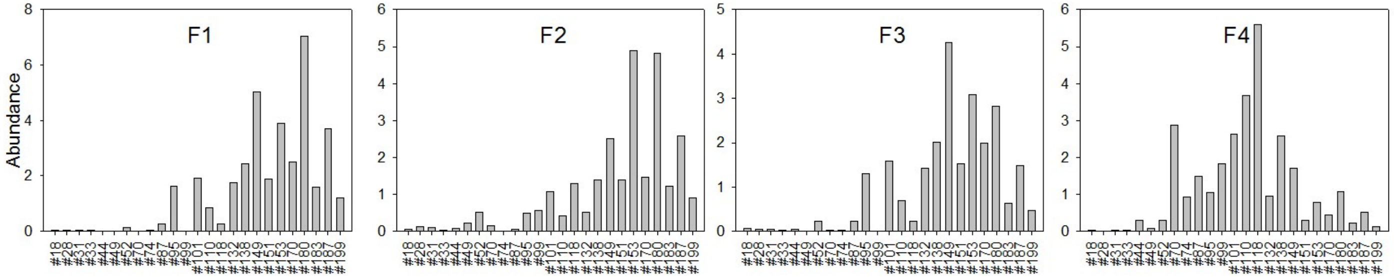
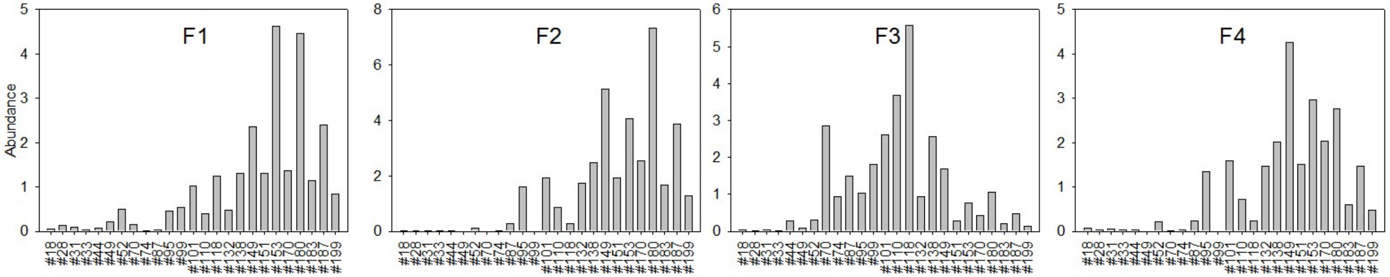


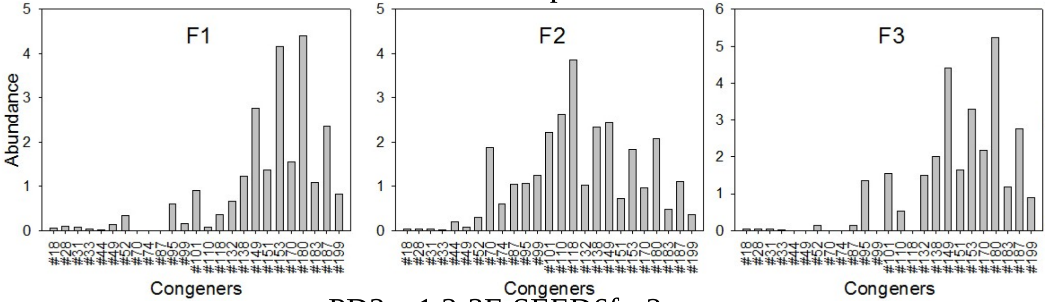
Figure 4-21 Aroclor and Measured Data Profiles



PD2.1.3-4F-sd1-fp0

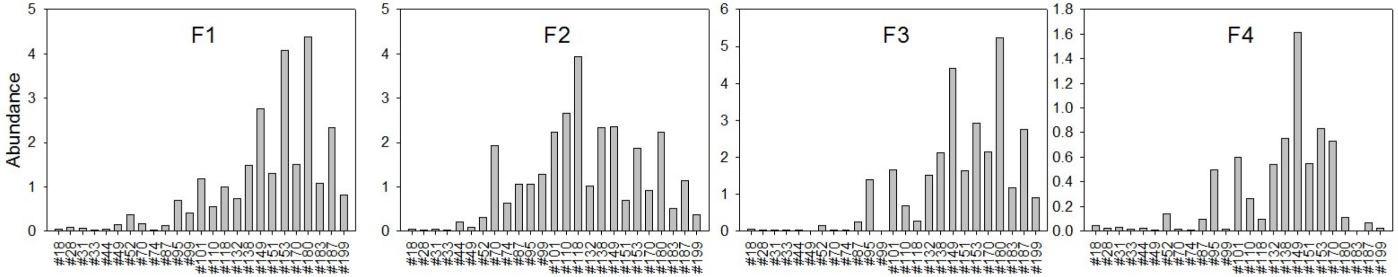


PD2.2.2-4F-sd7-fp0

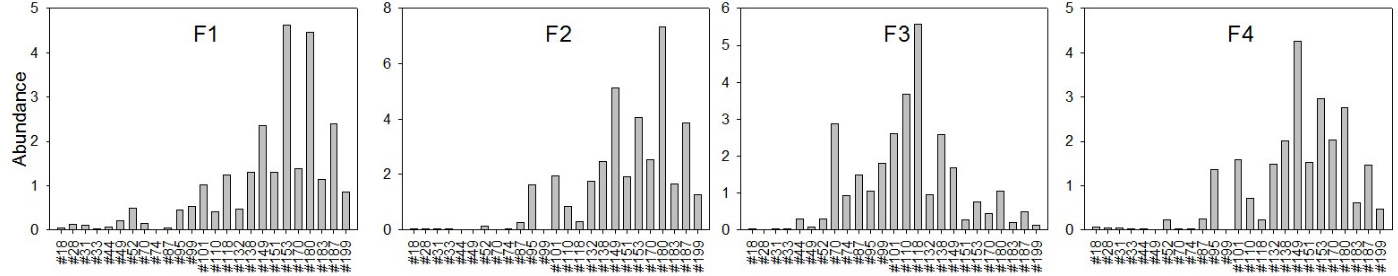


PD2.x.1.3-3F-SEED6fp-2

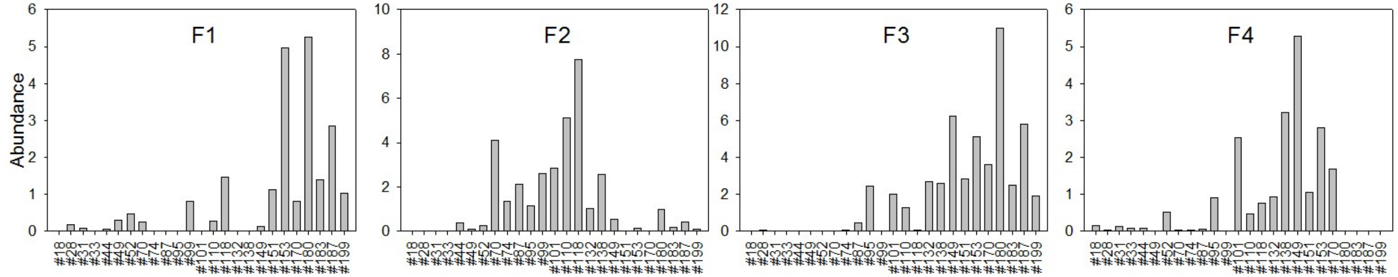
Figure 4-22 Aroclor Profiles and Results of Different Uncertainties (PD2.1, PD2.2 and PD.2.x.1)



PD.2.x.2--4F-sd2-fp-1

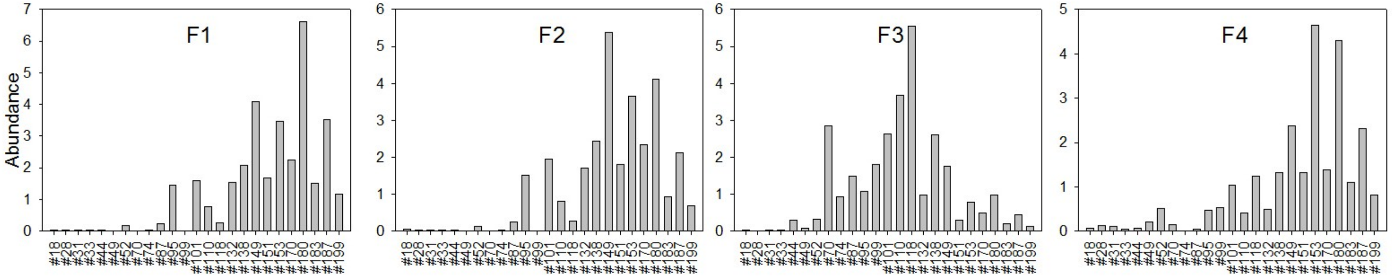


PD.2.x.2--4F-sd2-fp0

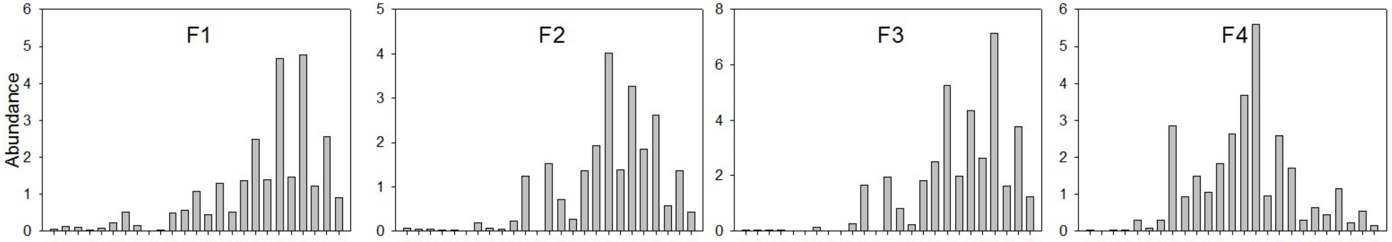


PD.2.x.2--4F-sd2-fp0.8

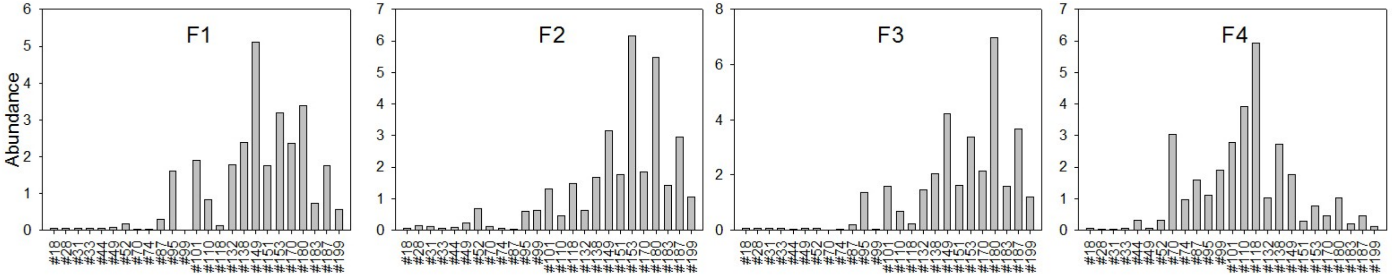
Figure 4-23 Aroclor Profiles and Results of Different Uncertainties (PD.2.x.2)



FPD2.x.2.2-4F-sd9-fp0

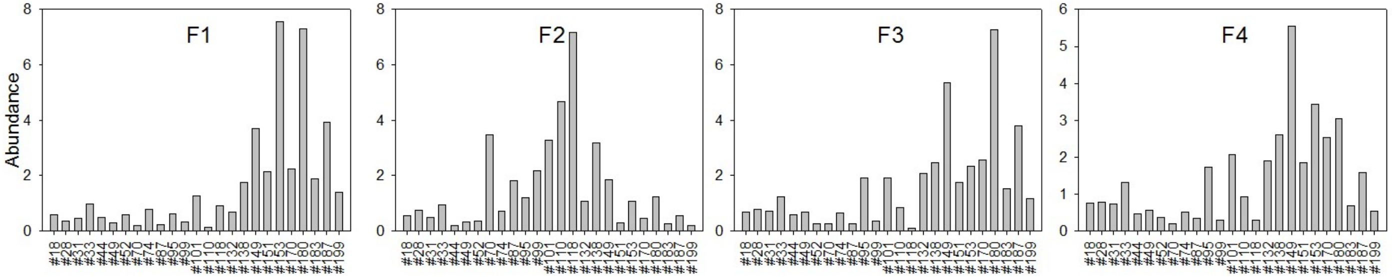


PD2.x.3.1-4F-sd7fp0

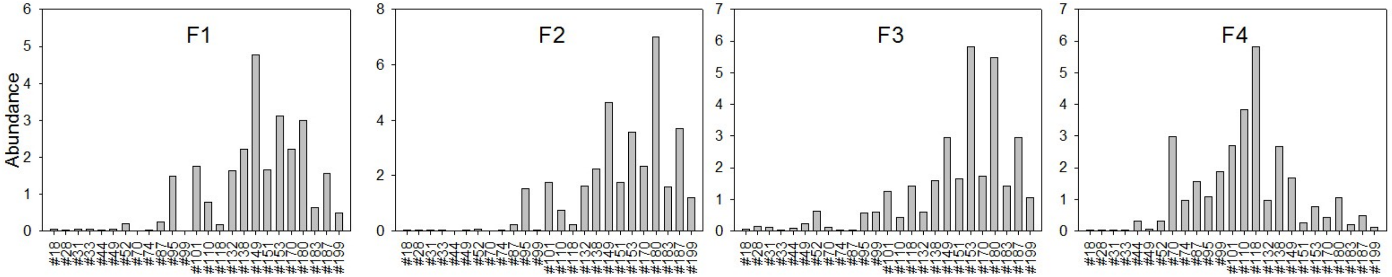


PD2.x.5.1-4F-sd5-fp0

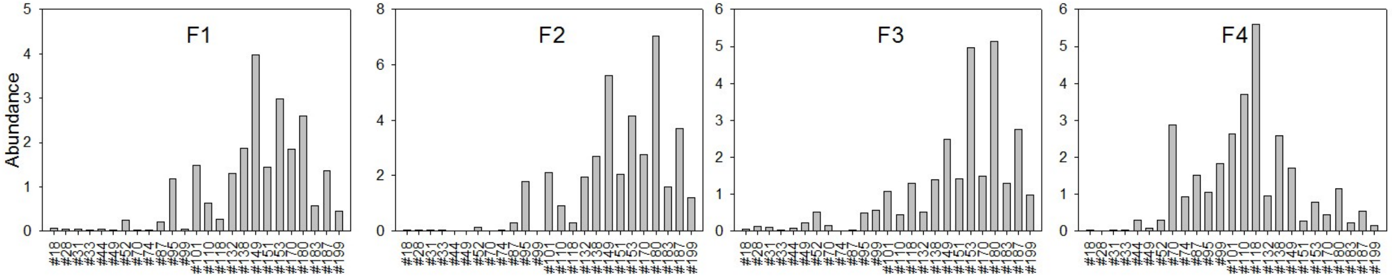
Figure 4-24 Aroclor Profiles and Results of Different Uncertainties (FPD.2.x.2, PD2.x.3 and PD2.x.5)



PD2.x.6.1-4F-sd5-fp0

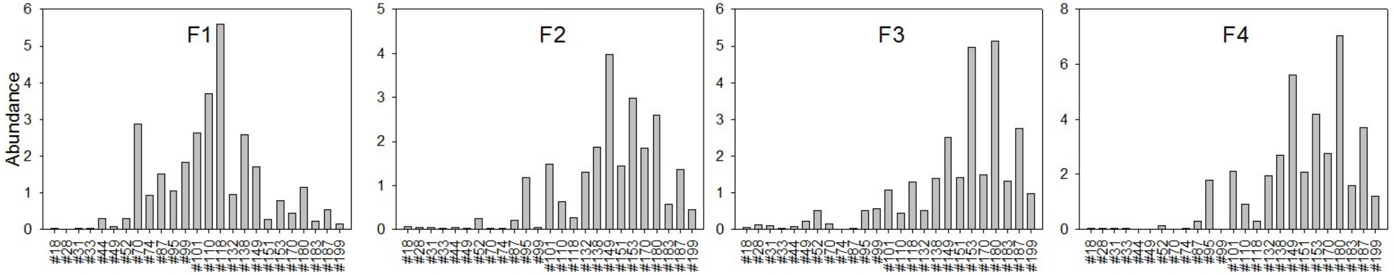


PD2.x.8.3-4F-sd6-fp0

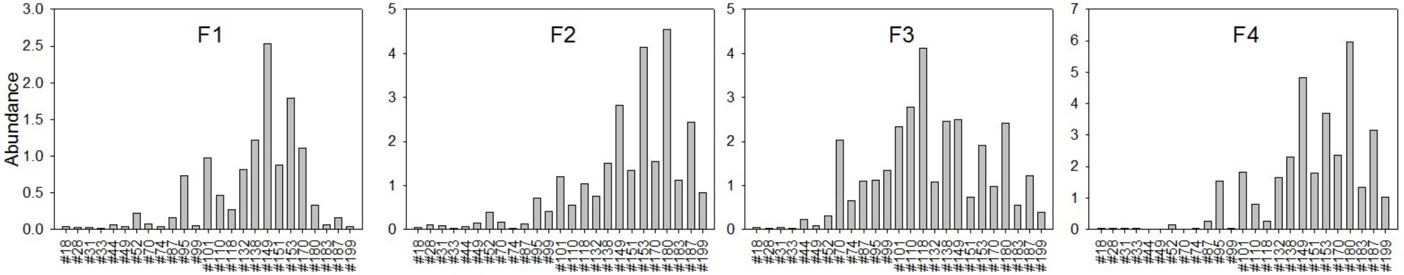


PD2.x.9.2-4F-sd6-fp0

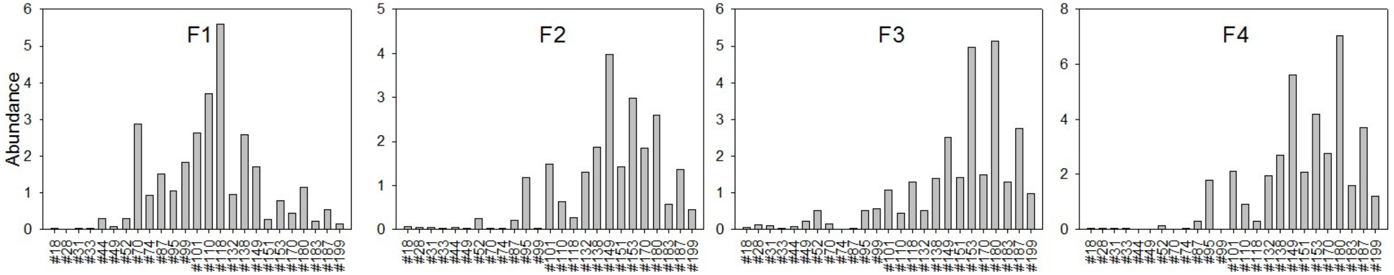
Figure 4-25 Aroclor Profiles and Results of Different Uncertainties (PD.X.6, PD.X.8 and PD.X.9)



CI-PD2.x.10.2-4F-sd4-



CII-PD2.x.10.1-4F-sd9-fp0



CIII-PD2.x.10.1-4F-sd4-fp0

Figure 4-26 Aroclor Profiles and Results of Different Uncertainties (CI-PD2.x.10, CII-PD2.x.10 and CIII-PD2.x.10)

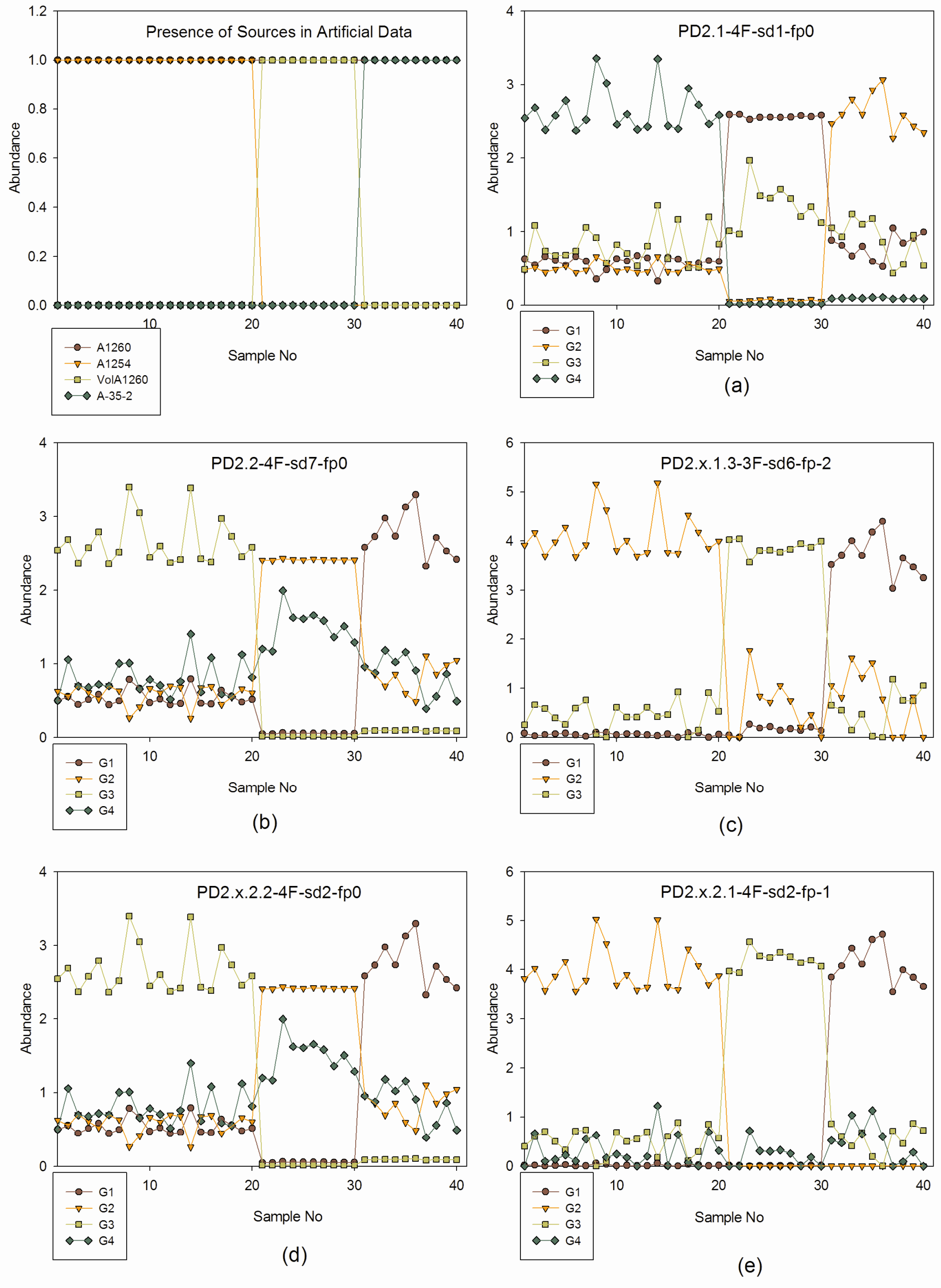


Figure 4-27 G Score Results

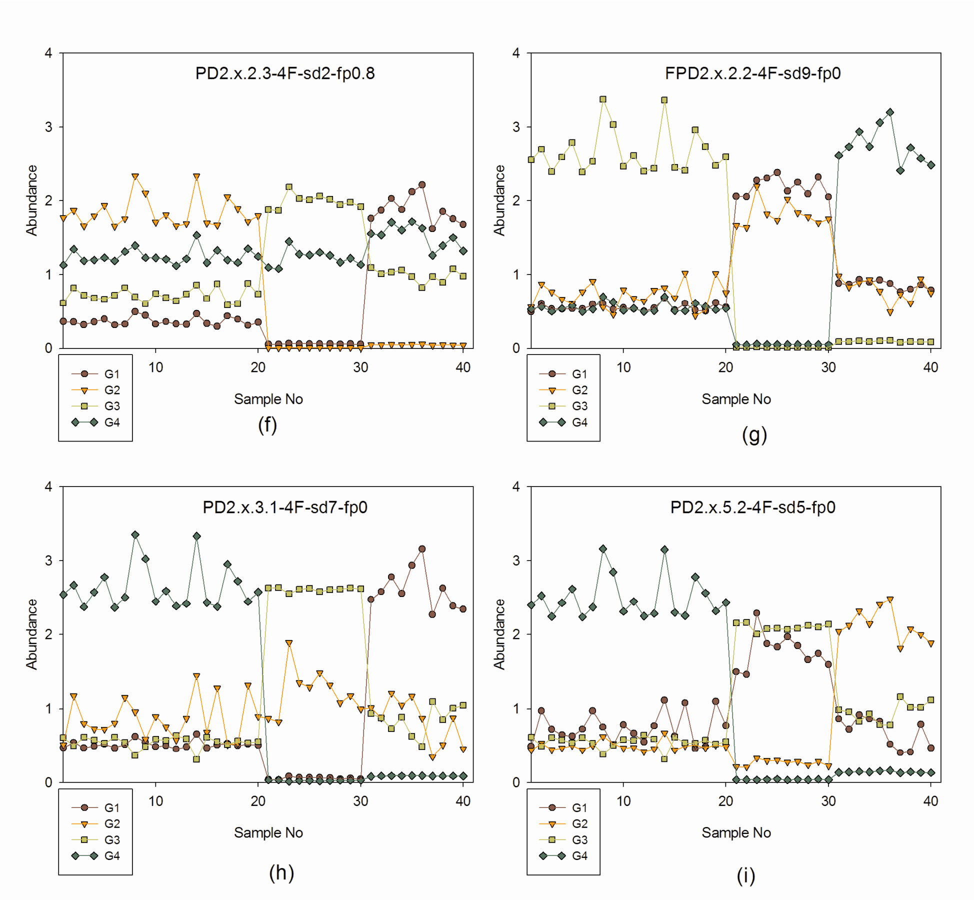


Figure 4-27 (*Cont'd)* G Score Results

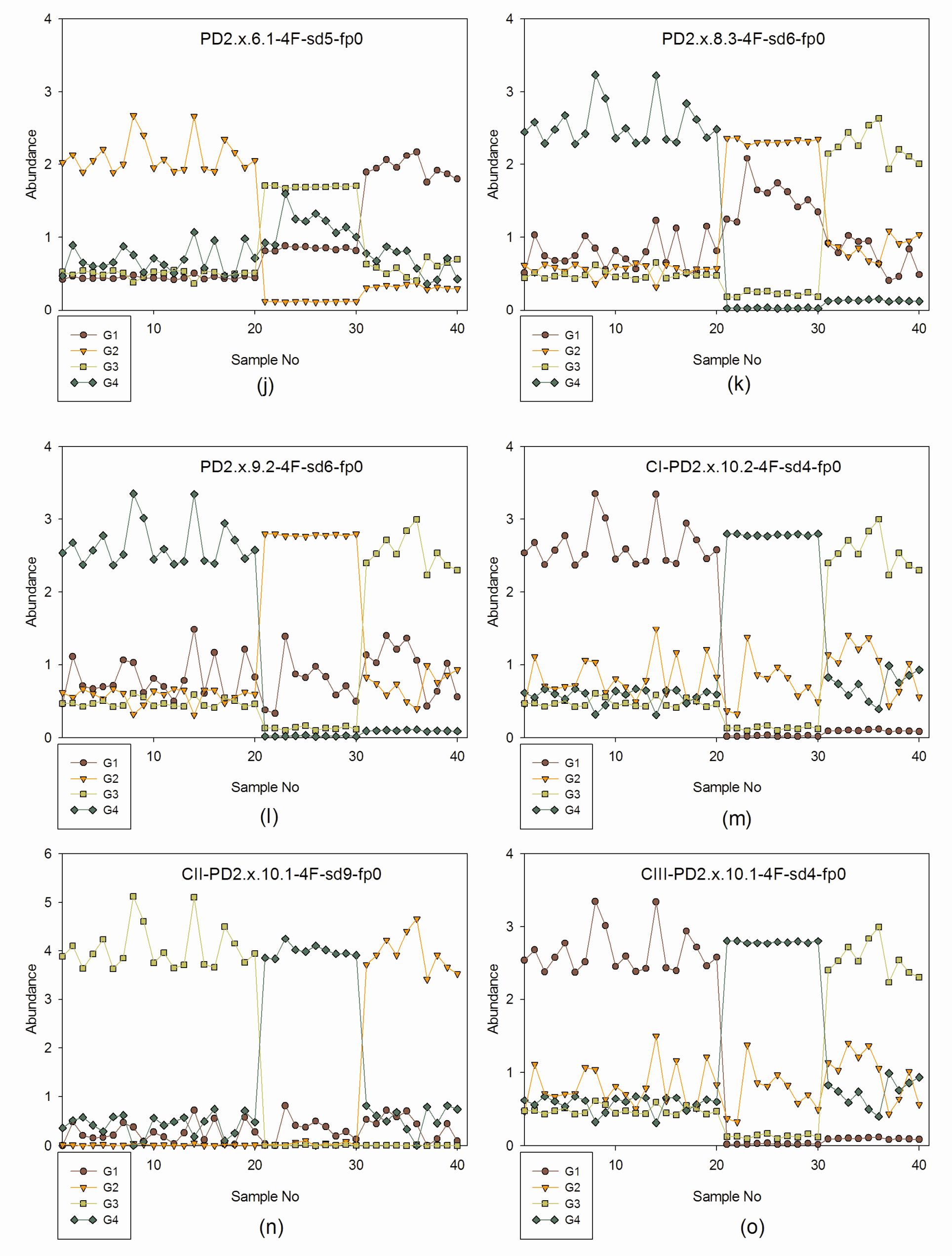


Figure 4-27 (Cont'd) G Score Results

General Assessment of the Study Using Artificial Data Sets:

Several important goals are reached by uncertainties prepared for artificial data sets from simple to complex:

- Operating principles of PMF model was investigated and applied using various data sets.
- The effect of uncertainty matrix on PMF results was evaluated.
- The uncertainty matrix was tested by several methods used for uncertainty data matrix in the literature.
- The same values for all data points in uncertainty matrix are not enough to obtain sources with good resolution.
- Q robust can be changed with rotations (fpeak). However, fpeak rotation makes factor separation worse while adding factors to F or G matrix.
- Before running the PMF model, the importance of measured data matrix, physicochemical properties of PCB congeners and fate of PCBs in the environment should be considered in developing uncertainty files.
- Giving small values to all cells of the uncertainty matrix theoretically means that the data set is very accurate; however, the results of this study has shown that it is not possible to obtain a result from the PMF model.
- To separate the source profiles which resemble each other highly, G scores can be used. For the real data set, when preparing the uncertainty matrix, Hopke's equation, the newly developed equation, laboratory bias and sampling errors should also be considered. Generally, optimum Q robust close to Q_{theo} is obtained in 3 factor

solutions. Additionally, if possible, uncertainty matrix value should be specific to each sample.

- If explained variance of A1254 is higher than 30%, it seems that A1254 and A1260 cannot perfectly be separated.
- When dealing with validation of PMF model, optimum Q value, F factors, G scores and scaled residuals should be considered. According to these determinations, the trials of PD2.x.2 (Hopke’s equation) and PD2.x.10 (newly developed equation) provide more realistic source profiles.

Table 4-28 Goodness of Fit Parameters Results of Uncertainty Trials

Trial No		**4-FACTOR RESULTS**				**3-FACTOR RESULTS**			
		COD %	***R^2***	***GvsG***	***Scaled Residual***	***COD %***	***R2***	***G vs. G***	***Scaled Residual***
PD.2.1	1	>0.85	1	G4 vs. G2	All random and Std<3	>0.84	1		Std$_{\#180}$=3.3
	2	>0.94	1	G4 vs. G2	All random and Std<3				
	3	>0.95	1	G4 vs. G2	All random and Std<3				
PD.2.2	1	>0.90	1		All random and Std<3	#18=89% others >98	1	G1 vs. G2	All random Std<3
	2	>0.95	1	G3 vs. G2 G1 vs. G2 G1 vs. G3	All random and Std<3				
PD2.X.1	1	COD$_{\#49}$=34 others >62%	1	G4 vs. G2 G1 vs. G2 G1 vs. G4	All random and Std<3	>0.90	1		Std$_{\#101}$=6. Std$_{\#110}$=12, Std$_{\#138}$=8, Std$_{\#149}$=17, Std$_{\#180}$=8
	2	>-0.4	1	G3 vs. G4	All random and Std<3				
	3	>-3.31	1	G3 vs. G4	All random and Std<3				
PD2.X.2	1	>0.96	1	G1 vs. G3 G1 vs. G2	All random and Std<3	#18=89% others >97%	1	G3 vs. G2	Std$_{\#118}$=4 random Std<3
	2	>0.90	1	G1 vs. G3 G1 vs. G2	All random and Std<3	#18=89% others >97%	1	G3 vs. G2	Std$_{\#118}$=4 random Std<3
	3	>0.89	1	G1 vs. G3 G1 vs. G2	All random and Std<3	#18=89% others >97%	1	G3 vs. G1	Std$_{\#118}$=4 random Std<3
	4					#18=89% others >97%	1	G1 vs.G2	Std$_{\#118}$=4 random Std<3
FULL PD2.X.2	1	COD$_{\#18}$=0.86 others >0.96	1	G1 vs. G4	All random and Std<3	COD$_{\#18}$=0.87 others >0.96	1	G1 vs.G2	Std$_{\#70}$=3,8, Std$_{\#110}$=187 Std$_{\#118}$=148 random and Std<3
	2	COD$_{\#18}$=0.87 others>0.96	1	G3 vs. G4	All random and Std<3(some congeners are close to 0)	COD$_{\#18}$=0.87 ve>0.96	1	G1 vs. G2	Std$_{\#70}$=3 , Std$_{\#110}$=186 Std$_{\#118}$=146 random Std<3

Table 4-28 (*Cont’d*) Goodness of Fit Parameters Results of Uncertainty Trials

Trial No		**4-FACTOR RESULTS**				**3-FACTOR RESULTS**			
		COD %	R^2		***COD %***	R^2		***COD %***	R^2
PD2.X.3	1	$COD_{\#18}$=93% others >98%	1	G3 vs. G1	All random and Std<3	>-0.28	0.72	Random	$Std_{\#118}$=42, $Std_{\#52}$=3.9, $Std_{\#74}$=5.7, $Std_{\#99}$=7.4, $Std_{\#110}$=31, $Std_{\#118}$=97, $Std_{\#149}$=496, $Std_{\#153}$=173
	2	$COD_{\#18}$=88% others >96%	1	Accumulated in the edge of axis	All random and Std<3				
	3	$COD_{\#18}$=87% others >96%	1	Accumulated in the edge of axis	All random and Std<3				
	4	$COD_{\#18}$=87% others >96%	1	All same	All random and Std<3				
	1	$COD_{\#74}$=-3.59 others >0.46	1	G4 vs.G2 G1 vs. G2 G1 vs. G4	All random and Std<3	>0.88	1	Random	All random Std<3
PD2.X.5	2	$COD_{\#33}$=0.07 $COD_{\#49}$=0.55 others >0.83	1	G4 vs. G2 G2 vs. G3	Std<3 Excess (-) values in scaled residual and all Std values are close to “0”.				
PD.2.x.6	1	> -51	1	Accumulated in the edge of axis	Std<3 and Std of all close to “0”	Values until -268	1	all	All random Std<3
	2	> -817	1	Accumulated in the edge of axis	Std<3	Values until -498	1	all	All random Std<3

Table 4-28 (*Cont'd*) Goodness of Fit Parameters Results of Uncertainty Trials

Trial No		**4-FACTOR RESULTS**				**3-FACTOR RESULTS**			
		COD %	R^2		***COD %***	R^2		***COD %***	R^2
PD.2.x.8	1	Some COD's have minus values	1	All have linearity	Std<3	>0.88	1	Accumulated in the edge of axis	Std$_{\#153}$=3.5, Std$_{\#149}$=7, Std$_{\#180}$=9
	2	COD$_{\#74}$=0.46 >0.75	1	All have linearity	Std<3	COD$_{\#49}$=0.52 >0.80	1	Accumulated in the edge of axis	Std$_{\#149}$=6.8, Std$_{\#180}$=10
	3	COD$_{\#33}$=0.7 >0.86	1	All have linearity (not randomly distributed)	Std<3				
PD2.X.9	1	>0.93	1	not randomly distributed	Std$_{\#95}$=5.6, Std$_{\#118}$=3.1, Std$_{\#132}$=4.7, Std$_{\#149}$=9.9, Std$_{\#151}$=3.7, Std$_{\#170}$=10.3				
	2	>0.95	1	not randomly distributed	Std$_{\#95}$=5.5, Std$_{\#118}$=3.2, Std$_{\#132}$=4.7, Std$_{\#149}$=9.7, Std$_{\#151}$=3.7, Std$_{\#170}$=10.4				
CI-PD2.X.10	1	COD$_{\#33}$=0.8 >0.93	1	not randomly distributed	Std$_{\#149}$=6, Std$_{\#170}$=5.6				
	2	>0.95	1	not randomly distributed	Std$_{\#132}$=3.1, Std$_{\#149}$=6, Std$_{\#170}$=5.5				
CII-PD2.X.10	1	>0.95	1	G3 vs. G2	Std<3				
CIII-PD2.X.10	1	>0.95	1	not randomly distributed	Std<3				

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

In the first part of this study, PCB pollution in the region between Lake Eymir and TEDAS was investigated. The sediment/soil sampling, laboratory studies for analysis were performed. PCB concentrations observed were in the range of below detection limit and 73.6 ng/g as $\Sigma_{Ar1016+1260}$ PCB and below detection limit and 5.4 ng/g as Σ_7 PCB. The CMB model was applied to the PCB data set and results indicate Aroclor 1254 and 1260 to be dominant PCB mixtures as pollution sources. These mixtures are known to be used in transformers and electrical equipment, which is consistent with the historical information regarding the area. The predicted vs. measured graphs indicated that although acceptable fit was obtained for many samples, congeners #118 and #138 were typically underestimated and hence not predicted well.

In the second part of this study, the FORTRAN code of the CMB model, as applied by Imamoglu (2001) was converted to MS Excel visual basic application. The major advantages of the modified CMB model software are; (i) automatic testing of combination of potential source profiles given in the input file, (ii) sorting of results from the best fit to the worst, according to the goodness of fit criteria, (iii) user friendly interface with capability of automatic plotting of the best results. These modifications enable the users to use program effectively, quickly and easily.

In the last part of this study, the verification of the PMF model was performed using Monte Carlo simulated data sets in varying complexity. Artificial data sets were created as a first stage of the PMF modeling effort. They were used to understand the effects of values in the uncertainty matrix on the results and to develop some suggestions before using on environmental PCB data set. The aim of the PMF model in this study is four fold: (i) to understand PMF model, (ii) to identify the composition of PCB sources and contribution of each source on samples, (iii) to identify any underlying degradation profiles and (iv) to investigate the effect of the changes in the uncertainty matrix data values on the outcome of the model, specifically in the success of identification of PCB patterns and improvement in the resolution of these profiles. The preparation of the uncertainty matrix is a critical and very important stage of modeling with PMF and increasing its sensitivity. So different uncertainty files were developed for the same simulated data set. As a result, two general formulations to be used for real data sets, developed and already existing in the literature, $\sigma_{ij}=0.05*X_{ij}+DL/2$ and $\sigma ij=a*X_{ij}^{max}*X_{ij}^{min}*1/X_{ij}$, respectively, were seen to be effective in obtaining PCB sources and degradation patterns with satisfactory resolution.

Future studies are recommended to utilize the uncertainty equations put forth by this study in the PMF model to investigate the PCB source compositions and contributions in real environmental data sets. It can be difficult to decide the values in uncertainty matrix by changing the uncertainty data point by point. Therefore use of the aforementioned equations are recommended. However, in this study we also try to pay attention to the values in residual matrix. By this way, the values in uncertainty matrix can be changed for higher or lower values of residual matrix to keep scaled residuals between -3vand +3 as discussed in the literature. However, experience obtained from this study indicates that an excess amount of point by point substitution in uncertainty data may lead to user-defined, hence biased results rather than actual ones. So attention should be paid to this process in real data sets.

In this study, PCBs have been measured in the ppb level. Therefore the remaining particles in glasses and equipments may lead to great variations in the analysis results. The equipments used for extraction and cleanup methods of PCBs should be cleaned well and higher attention should be paid to the cleanliness of the working environment during laboratory studies.

The result obtained from the reef roof sample indicates the probability of absorption of PCBs in the reef roofs. In the future studies, more reef roof samples can be used to understand whether reef roots can absorb PCBs or not.

This study which is done around the Lake Eymir was a part of a research project (Project No: 104I126) funded by TÜBİTAK (The Scientific and Technological Research Council of Turkey). In the scope of this research project, other contaminated sites around Turkey have also been investigated. PMF model studied by artificial data sets and the modified CMB model software presented in this study are also used in the other contaminated sites covered under this research project to understand the sources and contributions of PCB contamination.

REFERENCES

Akduman, N. (2009). *Investigation of PCB Pollution: Method Validation and Case Studies from Ankara.M.S. Thesis.* Ankara. Turkey: Department of Environmental Engineering, Middle East Technical University.

Amato, F., Pandolfi, M., Escrig, A., Querol, X., Alastuey, A., Peya, J., et al. (2009). Quantifying road dust resuspension in urban environment by Multilinear Engine: A comparison with PMF2. *Atmospheric Environment 43* , 2770–2780.

Anttila, P., Paatero, P., Tapper, U., & Jarvinen, O. (1995). Source Identification of Bulk Wet Deposition in Finland by Positive Matrix Factorization. *Atmospheric Environment 29(14)* , 1705-1718.

ATSDR. (2000). *Toxicological Profile for Polychlorinated Biphenyls (PCBs).* Georgia: Agency for Toxic Substances and Disease Registry.

Bedard, D. L., & May, R. J. (1996). Characterization of Polychlorinated Biphenyls in the Sediments of Woods Pond:Evidence for Microbial Dechlorination of Aroclor 1260 in Situ. *Environ. Sci. Technol. 30* , 237-245.

Begum, B. A., Biswas, S. K., Hopke, P. K., & Khaliquzzaman, M. (2005). Investigation of sources of atmospheric aerosol at a hot spot area in Dhaka, Bangladesh. *Journal of the Air & Waste Management Association 55(2)* , 227-240.

Breivik, K., Sweetman, A., Pacyna, J. M., & Jones, K. C. (2002). Towards a global historical emission inventory for selected PCB congeners - a mass balance approach. *The Science of the Total Environment 290* , 181-198.

Brown, S. G., Frankel, A., & Hafner, H. R. (2007). Source apportionment of VOCs in the Los Angeles area using positive matrix factorization. *Atmospheric Environment 41(2)* , 227-237.

Bzdusek, P. A. (2005). *PCB or PAH Sources and Degradation in Aquatic Sediments Determined By Factorization. Ph.D. Dissertation.* Milwaukee, WI:

Department of Civil Engineering and Mechanics, University of Wisconsin-Milwaukee.

Bzdusek, P. A., & Christensen, E. R. (2006a). Comparison of a new variant of PMF with other receptor modeling methods using artificial and real sediment PCB data sets. *Environmetrics 17* , 387–403.

Bzdusek, P. A., Christensen, E. R., Lee, C. M., Pakdeesusuk, U., & Freedman, D. L. (2006b). PCB Congeners and Dechlorination in Sediments of Lake Hartwell, South Carolina, Determined from Cores Collected in 1987 and 1998. *Environ. Sci. Technol. 40,* , 109-119.

Bzdusek, P., & Christensen, E. R. (2004). Source Apportionment of Sediment PAHs in Lake Calumet, Chicago: Application of Factor Analysis with Nonnegative Constraints. *Environ. Sci. Technol.* , 38, 97-103.

Cetin, B., Yatkin, S., Bayram, A., & Odabasi, M. (2007). Ambient Concentrations and Source Apportionment of PCBs and Trace Elements around an Industrial Area in Izmir, Turkey. *Chemosphere* , 69, 1267–1277.

Cheng, M. D., & Hopke, P. (1986). Investigation on the Use of Chemical Mass Balance Receptor Model: Numerical Computations. *Chenzometrics and Intelligent LaboratorySystems* , 1, 33-50.

Chiarenzelli, J. R., Scrudato, R. J., & Wunderlich, M. L. (1997). Volatile loss of PCB Aroclors from subaqueous sand. *Environ. Sci. Technol.* , 31 (2), 597.

Christensen, E. R., Li, A., Ab Razak, I. A., Rachdawong, P., & Karls, J. F. (1997). Sources of polycyclic aromatic hydrocarbons in sediments of the Kinnickinnic River, Wisconsin. *Journal of Great Lakes Research* , 23 (1) 61-73.

Christensen, E. R., Rachdawong, P., Karls, J. F., & Van Camp, R. P. (1999). PAHs in sediments: Unmixing and CMB modeling of sources. *Journal of Environmental Engineering-Asce 125(11)* , 1022-1032.

Chueinta, W., Hopke, P. K., & Paatero, P. (2000). Investigation of sources of atmospheric aerosol at urban and suburban residential areas in Thailand by positive matrix factorization. *Atmospheric Environment 34(20)* , 3319-3329.

Chung, Y. S., Kim, S. H., Moon, J. H., Kim, Y. J., Lim, J. M., & Lee, J. H. (2005). Source identification and long-term monitoring of airborne particulate matter PM2.5/PM10) in an urban region of Korea. *Journal of Radioanalytical and Nuclear Chemistry 267(1)* , 35-48.

Du, S., & Rodenburg, L. A. (2007). Source identification of atmospheric PCBs in Philadelphia/ Camden using positive matrix factorization followed by the potential source contribution function. *Atmospheric Environment 41* , 8596–8608.

Du, S., Belton, T. J., & Rondenburg, L. A. (2008). Source Apportionment of Polychlorinated Biphenyls in the Tidal Delaware River. *Environ. Sci. Technol. 42* , 4044–4051.

EPA. (2009, 03 24). *Polychlorinated iphenyls*. Retrieved 04 25, 2010, from Environmental Protection Agency: www.epa.gov

Erickson, M. (1997). *Analytical Chemistry of PCBs.* Baco Raton, New York: CRC-Lewis Publishers.

Frame, G. M., Cochran, J. W., & Bowadt, S. S. (1996). Complete PCB Congener Distributions for 17 Aroclor Mixtures Determined by 3 HRGC Systems Optimized for Comprehensive, Quantitative, Congener-Specific Analysis. *J. High Resol. Chroinatogr.* , 19, 657-668.

Gedik, K., & İmamoğlu, İ. (2010). As Assessment of the Spatial Distribution of Polychlorinated Biphenyl Contamination in Turkey. *CLEAN 38(2)* , 117-128.

Gouin, T., Mackay, D., Webster, E., & Wania, F. (2000). Screening Chemicals for Persistence in the Environment. *Environ. Sci. Technol.* , 34, 881-884.

Hansen, L. G. (1999). *The Ortho Side of PCBs, Occurence and Disposition.* London: Kluwer Academic.

Henry, R. C., Lewis, C. W., Hopke, P. K., & Williamson, H. J. (1984). Review of Receptor Model Fundamentals. *Atmospheric Environment* , 18 (8), 1507-1515.

Honda, T., Wada, M., & Nakashima, K. (2008). Concentration and characteristics of polychlorinated biphenyls in the sediments of sea and river in Nagasaki Prefecture, Japan. *Journal of Health Science* , 54(4), 400-408.

Hopke, P. A. (2010, 1 12). *Guide to Positive Matrix Factorization. Tutorial Document.* Retrieved 06 19, 2010, from PMF-Guidance: http://people.clarkson.edu/~phopke/PMF-Guidance.htm

Hopke, P. K. (1991). An introduction to receptor modeling. *Chemometrics and Intelligent Laboratory Systems* , 10, 21-43.

Hopke, P. K. (2003). Recent Developments in Receptor Modeling. *Journal of Chemometrics 17* , 255-265.

Imamoglu, I. (2001). *PCB sources and degradation in river sediments determined by receptor modeling. PhD thesis.* Milwaukee, WI, USA: University of Wisconsin- Milwaukee.

Imamoğlu, I., Gedik, K., Demircioglu, F., & Akduman, N. (June 2010). *Investigation of PCB pollution in Turkey and Fate of PCBs.* Ankara: TUBITAK 104I126, Project Final Report.

Imamoglu, İ., Li, K., & Christensen, E. R. (2002). PCB sources, transformations, and contributions in recent Fox River, Wisconsin sediments determined from receptor modeling. *Water Science and Technology* , 36, 3449-3462.

Ishikawa, Y., Noma, Y., Mori, Y., & Sakai, S. (2007). Congener profiles of PCB and a proposed new set of indicator congeners. *Chemosphere* , 67, 1838–1851.

Jensen, S. (1966). Report of a New Chemical Hazard. *New Science* , 32, 612.

Johnson, G. W., Jarman, W. M., Bacon, C. E., Davis, J. A., Ehrlich, R., & Risebrough, R. W. (2000). Resolving polychlorinated biphenyl source fingerprints in suspended particulate matter of San Francisco Bay. *Environ. Sci. Technol. 34(4)* , 552-559.

Johnson, G. W., Quensen, I. J., Chiarenzelli, J. R., & Hamilton, M. C. (2005). *Chapter 10: Polychlorinated Biphenlys. In: Environmental Forensics: Contaminant Specific Guide.* R.D. Morrison and B. L. Murphy.

Johnson, G., Quensen III, J. F., Chiarenzelli, J. R., & Hamilton, M. C. (2005). Polychlorinated Biphenyls. In B. L. Murphy, & R. Morrison, *Environmental Forensics: Contaminant Specific Guide* (pp. 187-225).

Juvela, M., Lehtinen, K., & Paatero, P. (1996). The Use of Positive Matrix Factorization in the Analysis of Molecular Line Spectra. *Mon. Not. R. Astron. Soc. 280* , 616-626.

Lee, E., Chan, C. K., & Paatero, P. (1999). Application of positive matrix factorization in source apportionment of particulate pollutants in Hong Kong. *Atmospheric Environment 33* , 3201-3212.

Li, A., Kiljang, J., & Scheff, P. (2003). Application of EPA CMB8.2 Model for Source Apportionment of Sediment PAHs in Lake Calumet, Chicago. *Environ. Sci. Technol.* , 37, 2958-2965.

Li, K., Christensen, E. R., Van Camp, R. P., & Imamoglu, İ. (2001). PAHs in Dated Sediments of Ashtabula River, Ohio, USA. *Environmental Science and Technology* , 35, 2896-2902.

Ozturk, F. (2009). *Investigation of Short and Long Term Trends in The Eastern Mediterranean Aerosol Composition. PhD Thesis.* Ankara, Turkey: Middle East Technical University.

Paatero, P. (1997). Least squares formulation of robust nonnegative factor analysis. *Chemometrics and Intelligent Laboratory Systems (37)* , 23–35.

Paatero, P. (2002). User's Guide for Positive Matrix Factorization programs PMF2 and PMF3, Part 2: reference.

Paatero, P., & Tapper, U. (1994). Positive Matrix Factorization - a Nonnegative Factor Model with Optimal Utilization of Error-Estimates of Data Values. *Environmetrics 5(2)* , 111-126.

Pakdeususuk, U. (2002). *Assessment of In Situ Reductive Dechlorination of Polychlorinated Biphenyls (PCBs) as a Monitored Natural Attenuation Process in Contaminated Sediment from Lake Hartwell.* South Carolina: Ph.D. Deparment of Environmental Engineering and Science University of Clemson University.

Poirot, R. L., Wishinski, P. R., Hopke, P. K., & Polissar, A. V. (2001). Comparative application of multiple receptor methods to identify aerosol sources in northern Vermont. *Environmental Science & Technology 35(23)* , 4622-4636.

Polissar, A. V., Hopke, P. A., & Poirot, R. L. (2001). Atmospheric Aerosol over Vermont: Chemical Composition and Sources. *Environ. Sci. Technol. 35* , 4604-4621.

Polissar, A. V., Hopke, P. K., & Paatero, P. (1998). Atmospheric aerosol over Alaska - 2. Elemental composition and sources. *Journal of Geophysical Research-Atmospheres 103(D15)* , 19045-19057.

Rachdawong, P. (1997). *Receptor models for source attribution of PAHs and PCBs in Lake Michigan sediments. Ph.D. Dissertation.* Milwaukee: Rachdawong, P. (1997). Receptor models for source attribution o Department of Civil Engineering and Mechanics, University of Wisconsin-Milwaukee.

Song, X. H., Polissar, A. V., & Hopke, P. K. (2001). Sources of fine particle composition in the northeastern US. *Atmospheric Environment 35(31)* , 5277-5286.

Su, M., Christensen, E. R., Karls, J. F., Kosuru, S., & Imamoglu, i. (2000). Apportionment of polycyclic aromatic hydrocarbon sources in lower Fox River, USA, sediments by a chemical mass balance model. *Environmental Toxicology and Chemistry* , (19), 6 1481-1490 .

Su, M.-C. (1997). *Apportionment of sources of PAHs and PCDD/Fs in aquatic sediments by a chemical mass balance model. Ph.D. Dissertation.* Su, M.-C. (1997) Apportionment of sources of PAHs and PCDD/Fs in aquatic sedimentsMilwaukee, WI.: Su, M.-C. (1997) Apportionment of sources of PAHs and PCDD/Fs in aquatic sediments byUniversity of Wisconsin-Milwaukee.

UNEP. (1999). *Guidelines for the Identification of PCBs and Materials Containing PCBs, UNEP Chemicals.*

UNIDO. (2006). *Turkey's Draft National Implentation Plan For The Stockholm Convention on Persistent Organic Pollutants.* UNIDO POPs Project. Project No. GF/TUR/03/008.

Willis, R. (2000). Workshop on UNMIX and PMF as Applied to PM2.5, National Exposure Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, NC.

Xie, Y. L., & Berkowitz, C. M. (2006). The use of positive matrix factorization with conditional probability functions in air quality studies: An application to hydrocarbon emissions in Houston, Texas. *Atmospheric Environment 40(17)* , 3070-3091.

Yeniova, M. (1998). *Biological and Environmental Monitoring of Polychlorinated Biphenyls. Ph.D. Thesis.* Ankara, Turkey: Ankara University.

APPENDIX A

LIST OF PCB CONGENERS

Table A-1 List of PCB Congeners

#	Structure	#	Structure	#	Structure	#	Structure	#	Structure
MonoCB		41	234-2	84	236-23	HexaCB		HeptaCB	
1	2-	42	23-24	85	234-24	**128**	**234-234**	**170**	**2345-234**
2	3-	43	235-2	86	2345-2	129	2345-23	**171**	**2346-234**
3	4-	**44**	**23-25**	**87**	**234-25**	130	234-235	172	2345-235
DiCB		45	236-2	88	2346-2	131	2346-23	173	23456-23
4	2-2	46	23-26	89	234-26	**132**	**234-236**	174	2345-236
5	23-	47	24-24	90	235-24	133	235-235	175	2346-235
6	2-3	48	245-2	91	236-24	134	2356-23	176	2346-236
7	24-	**49**	**24-25**	92	235-25	135	235-236	**177**	**2356-234**
8	2-4	50	246-2	93	2356-2	136	236-236	178	2356-235
9	25-	51	24-26	94	235-26	137	2345-24	179	2356-236
10	26-	**52**	**25-25**	**95**	**236-25**	**138**	**234-245**	**180**	**2345-245**
11	3-3	53	25-26	96	236-26	139	2346-24	181	23456-24
12	34-	54	26-26	97	245-23	140	234-246	182	2345-246
13	3-4	55	234-3	98	246-23	141	2345-25	**183**	**2346-245**
14	35-	56	23-34	**99**	**245-24**	142	23456-2	184	2346-246
15	4-4	57	235-3	100	246-24	143	2345-26	185	23456-25
TriCB		58	23-35	**101**	**245-25**	144	2346-25	186	23456-26
16	23-2	59	236-3	102	245-26	145	2346-26	**187**	**2356-245**
17	**24-2**	60	234-4	103	246-25	146	235-245	188	2356-246
18	**25-2**	61	2345-	104	246-26	147	2356-24	189	2345-345
19	26-2	62	2346-	**105**	**234-34**	148	235-246	190	23456-34
20	23-3	63	235-4	106	2345-3	**149**	**236-245**	**191**	**2346-345**
21	234-	64	236-4	107	235-34	150	236-246	192	23456-35
22	23-4	65	2356-	108	234-35	**151**	**2356-25**	193	2356-345
23	235-	66	24-34	109	2346-3	152	2356-26	OctaCB	
24	236-	67	245-3	**110**	**236-34**	**153**	**245-245**	**194**	**2345-2345**
25	24-3	68	24-35	111	235-35	154	245-246	**195**	**23456-234**
26	25-3	69	246-3	112	2356-3	155	246-246	196	2345-2346
27	26-3	**70**	**25-34**	113	236-35	**156**	**2345-34**	197	2346-2346
28	**24-4**	71	26-34	114	2345-4	157	234-345	198	23456-235
29	245-	72	25-35	115	2346-4	**158**	**2346-34**	**199**	**23456-236**
30	246-	73	26-35	116	23456-	159	2345-35	200	2346-2356
31	**25-4**	**74**	**245-4**	117	2356-4	160	23456-3	201	2345-2356
32	26-4	75	246-4	**118**	**245-34**	161	2346-35	202	2356-2356
33	**34-2**	76	345-2	119	246-34	162	235-345	203	23456-245
34	35-2	77	34-34	120	245-35	163	2356-34	204	23456-246
35	34-3	78	345-3	121	246-35	164	236-345	**205**	**23456-345**
36	35-3	79	34-35	122	345-23	165	2356-35	NonaCB	
37	34-4	80	35-35	123	345-24	166	23456-4	**206**	**23456-2345**
38	34-5	81	345-4	124	345-25	167	245-345	207	23456-2346
39	35-4	PentaCB		125	345-26	168	246-345	**208**	**23456-2356**
TetraCB		**82**	**234-23**	126	345-34	**169**	**345-345**	DecaCB	
40	23-23	83	235-23	127	345-35			**209**	**23456-23456**

APPENDIX B

THE CODES OF MONTE CARLO SIMULATION

```
Sub MonteCarloden()

Dim Seed As Double, P As Double, R As Double

Dim x As Double, Erf As Double, CV As Double, Ai As Double, D As Double

Dim i As Integer, j As Integer, k As Integer, M As Integer, Sam As Double, Sa
As Integer, Com As Double, Co As Double

Dim jk(100, 100) As Double, Mean(100) As Double, Vr(100, 100) As Double

Dim l As Double, Tmen As Double, Dfnm As Double, Rij As Double,
GenData(100, 100) As Double

Dim Sample As Integer, Source As Integer, Cong As Integer

Dim DFNMD(100) As Double

Source = UserForm1.txtsource

Sample = UserForm1.txtsample

Cong = UserForm1.txtcongener

For i = 1 To Source

   For j = 1 To Sample

   ’Generate Random number

   Randomize

   Seed = Rnd
```

```
 '\Generate Random number
 Sheets("MonteCarlo").Range("Q4:v73").Cells(j, i) = Seed
 'Calculate and make positive inside of Erf
 P = 2 * Sheets("MonteCarlo").Range("Q4:v73").Cells(j, i) - 1
 R = Abs(P)
'  Sheets("Xij").Range("c5:f54").Cells(j, i) = R
     'Calculate x values which make value of erf(x) same with seed
         x = -0.0005
         Do
         x = x + 0.0005
         Erf = Application.WorksheetFunction.Erf(x)
         'Erf = Application.Run("ATPVBAEN.XLA!Erf", 0, x)
         Loop Until (R - Erf) < 0.001
         '/ Calculate x values which make value of erf(x) same with seed
'        Sheets("Sheet2").Range("f17:ı66").Cells(j, i) = x
     'If P < 0 Then
         '   x = -x
         'End If
     For k = 1 To Cong
     CV = Sheets("MonteCarlo").Range("I4:n29").Cells(k, i)
     Ai = Sheets("MonteCarlo").Range("b4:g29").Cells(k, i)
     D = Ai + CV * Ai * (Sqr(2) * x)
     jk(j, k) = D + jk(j, k)
```

```
        Next k
    Next j
Next i
For j = 1 To Sample
    For k = 1 To Cong
    Sheets("MonteCarlo").Range("z4:ay73").Cells(j, k) = jk(j, k) ’Generated Data
    GenData(k, j) = jk(j, k)
    Next k
Next j
’Calculation of MEAN
For l = 1 To Cong
    Tmen = 0
    For M = 1 To Sample
        Tmen = Tmen + GenData(l, M)
    Next M
    Mean(l) = Tmen / Sample
’    Sheets("Mean").Range("g4:g28").Cells(l) = Mean(l)
’/ Calculation of MEAN
    Dfnm = 0
    Sam = 0
    Do
        Sam = Sam + 1
```

```
        Dfnm = Dfnm + (GenData(l, Sam) - Mean(l)) ^ 2

    Loop Until Sam = Sample

    DFNMD(l) = Dfnm

Next l

 For Co = 1 To Cong

    For Com = 1 To Cong

    Rij = 0

        For Sam = 1 To Sample

         Rij = Rij + ((GenData(Co, Sam) - Mean(Co)) * (GenData(Com, Sam) -
Mean(Com)))

        Next Sam

    Vr(Co, Com) = Rij / (DFNMD(Co) * DFNMD(Com)) ^ 0.5

    Sheets("Vr").Range("c4:z27").Cells(Co, Com) = Vr(Co, Com)

    Next Com

 Next Co

        End Sub

Sub initialize()

UserForm1.Show

UserForm1.Hide

End Sub

Private Sub CommandButton1_Click()

Call MonteCarlo.MonteCarloden

End Sub
```

APPENDIX C

THE CODES OF COEFFICIENT OF DETERMINATION

```
Option Explicit

Public sample As Integer, congener As Integer, source As Double

Sub COD()

Dim i As Integer, j As Integer

Dim Initialdata As Double, Predicteddata As Double, Residualdata(100, 100) As Double, sum As Double

Dim Xavg As Double, Residualavg(100) As Double, Varianceres(50) As Double, Varianceact(100) As Double

Dim COD As Double, XavgDim(100) As Double



’inputfile

sample = UserForm1.txtsample

congener = UserForm1.txtcongener

source = UserForm1.txtsource

’\inputfile

Call TRANSPOSE.TRANSPOSE

Call Multiply.Multiply

For i = 1 To sample  ’Residual bulmak için
```

```
    For j = 1 To congener
        Initialdata = Sheets("Correction of MV and
Final").Range("b2:AA71").Cells(i, j)
        Predicteddata = RESULT(i, j)

        Residualdata(i, j) = Initialdata - Predicteddata
’       Range("b144:aa213").Cells(i, j) = Residualdata
    Next j
Next i
For i = 1 To congener ’ActualAvg
    sum = 0
    For j = 1 To sample
        sum = sum + Sheets("Correction of MV and
Final").Range("b2:AA71").Cells(j, i)
    Next j
    Xavg = sum / sample
    XavgDim(i) = Xavg
Next i
For i = 1 To congener     ”ResidualAvg
    sum = 0
    For j = 1 To sample
        sum = sum + Residualdata(j, i)
    Next j
    Residualavg(i) = sum / sample
```

```
’   Range("ac4:bb4").Cells(i) = Residualavg
Next i


For i = 1 To congener     ’Variance for residual data
    sum = 0
    For j = 1 To sample
’        Residualavg = Range("ac4:bb4").Cells(i)
’        Residualdata = Range("b144:aa213").Cells(j, i)
        sum = sum + (Residualdata(j, i) - Residualavg(i)) ^ 2
    Next j
    Varianceres(i) = sum / (sample - 1) ’sample’ın TAKE ONE MINUS!!
’   Range("bd4:cc4").Cells(i) = Varianceres
Next i


For i = 1 To congener    ’Variance for actualdata
    sum = 0
    For j = 1 To sample
’        XavgDim(i) = Range("ac2:bb2").Cells(i)
’       Initialdata = Sheets("Correction of MV and
Final").Range("b2:AA71").Cells(j, i)
        sum = sum + (Sheets("Correction of MV and
Final").Range("b2:AA71").Cells(j, i) - XavgDim(i)) ^ 2
    Next j
    Varianceact(i) = sum / (sample - 1) ’sample’ın TAKE ONR MINUS!!!
```

```
’   Range("bd2:cc2").Cells(i) = Varianceact

Next i

For i = 1 To congener

’  Varianceact = Range("bd2:cc2").Cells(i)

’  Varianceres = Range("bd4:cc4").Cells(i)

COD = (Varianceact(i) - Varianceres(i)) / Varianceact(i)

Worksheets("FACTORs").Range("x2:aw2").Cells(i) = COD

Next i

End Sub
```

APPENDIX D

THE MEASURED AND PREDICTED CONGENER PROFILE OF 10 SAMPLES

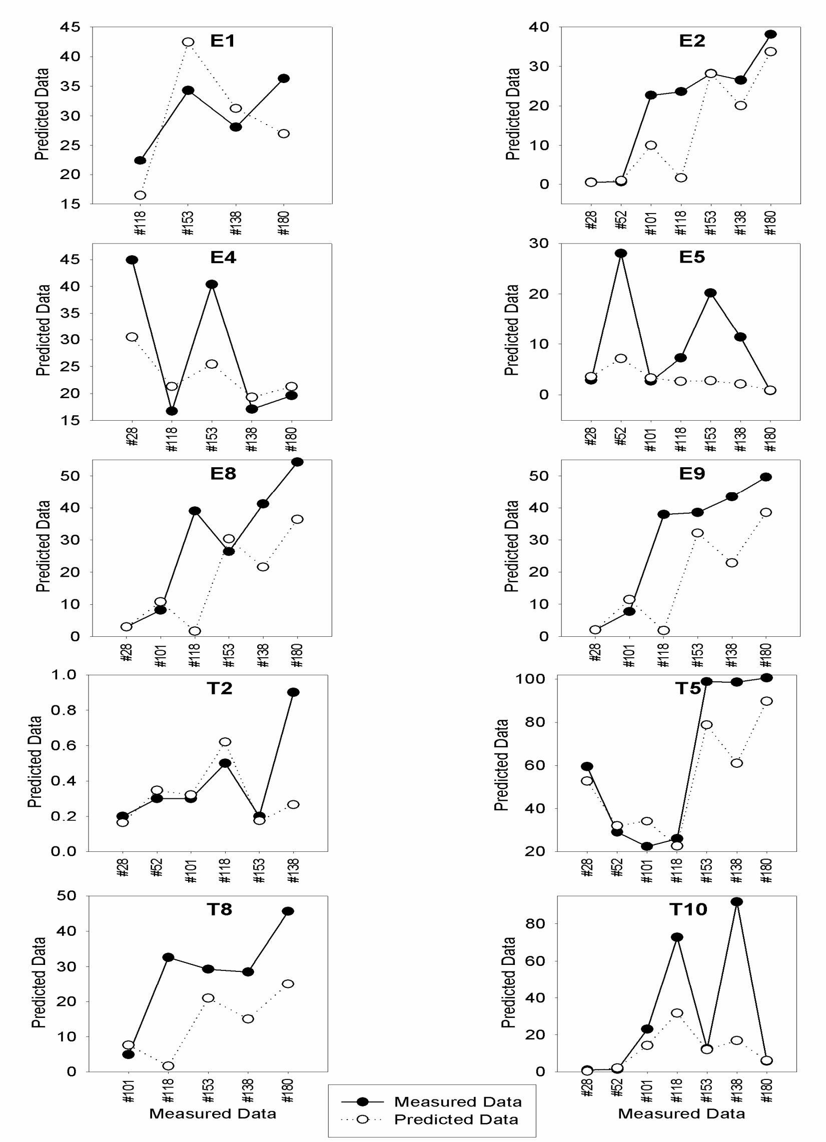


Figure D-1 The Measured and Predicted Congener Profile of 10 Samples

APPENDIX E

PD0 DATA SET WITH AVERAGE AND STANDARD DEVIATION VALUES

Table E-1 PD0 Data Set with Average and Standard Deviation Values

Samp. #/Con. #	**#70**	**#99**	**#101**	**#110**	**#118**	**Std.**	**Avg.**
1	0.07280	0.01470	0.02220	0.02970	0.02290	**0.02300**	**0.03200**
2	0.07280	0.01470	0.02220	0.02970	0.02290	**0.02300**	**0.03200**
3	0.07280	0.01470	0.02220	0.02970	0.02290	**0.02300**	**0.03200**
4	0.07280	0.01470	0.02220	0.02970	0.02290	**0.02300**	**0.03200**
5	0.07280	0.01470	0.02220	0.02970	0.02290	**0.02300**	**0.03200**
6	0.06830	0.04530	0.05490	0.08420	0.13590	**0.03600**	**0.07800**
7	0.06830	0.04530	0.05490	0.08420	0.13590	**0.03600**	**0.07800**
8	0.06830	0.04530	0.05490	0.08420	0.13590	**0.03600**	**0.07800**
9	0.06830	0.04530	0.05490	0.08420	0.13590	**0.03600**	**0.07800**
10	0.06830	0.04530	0.05490	0.08420	0.13590	**0.03600**	**0.07800**
Std.	**0.00200**	**0.01600**	**0.01700**	**0.02900**	**0.06000**		
Avg.	**0.07100**	**0.03000**	**0.03900**	**0.05700**	**0.07900**		

Std.= Standard Deviation, Avg.= Average

APPENDIX F

AVERAGE AND STANDARD DEVIATION VALUES OF SAMPLES AND CONGENERS OF PD1 DATA SET

Table F-1 Average and Standard Deviation Values of Samples of PD1 Data Set

#	Std.	Avg.	#	Std.	Avg.	#	Std.	Avg.	#	Std.	Avg.	#	Std.	Avg.
1	0.0651	0.0709	16	0.0714	0.0791	31	0.0644	0.0694	46	0.0689	0.0769	61	0.0663	0.0749
2	0.0674	0.0721	17	0.0664	0.0707	32	0.0694	0.0756	47	0.0703	0.0796	62	0.0608	0.0650
3	0.0746	0.0802	18	0.0659	0.0737	33	0.0676	0.0722	48	0.0754	0.0832	63	0.0619	0.0671
4	0.0637	0.0681	19	0.0644	0.0707	34	0.0644	0.0706	49	0.0653	0.0690	64	0.0655	0.0728
5	0.0709	0.0779	20	0.0699	0.0795	35	0.0622	0.0672	50	0.0715	0.0795	65	0.0711	0.0804
6	0.0739	0.0840	21	0.0661	0.0728	36	0.0660	0.0717	51	0.0661	0.0702	66	0.0749	0.0790
7	0.0646	0.0699	22	0.0765	0.0806	37	0.0629	0.0679	52	0.0701	0.0793	67	0.0679	0.0738
8	0.0666	0.0716	23	0.0675	0.0738	38	0.0646	0.0686	53	0.0721	0.0775	68	0.0689	0.0744
9	0.0692	0.0782	24	0.0661	0.0705	39	0.0676	0.0746	54	0.0708	0.0745	69	0.0623	0.0680
10	0.0760	0.0803	25	0.0678	0.0744	40	0.0705	0.0790	55	0.0659	0.0741	70	0.0670	0.0752
11	0.0679	0.0758	26	0.0634	0.0682	41	0.0631	0.0697	56	0.0724	0.0780			
12	0.0627	0.0681	27	0.0665	0.0709	42	0.0651	0.0725	57	0.0740	0.0809			
13	0.0677	0.0757	28	0.0672	0.0720	43	0.0726	0.0796	58	0.0679	0.0714			
14	0.0722	0.0786	29	0.0694	0.0742	44	0.0695	0.0772	59	0.0682	0.0744			
15	0.0729	0.0781	30	0.0685	0.0758	45	0.0719	0.0761	60	0.0626	0.0666			

Table F-2 Average and Standard Deviation Values of Congeners of PD1 Data Set

Con. #	Std.	Avg.	Con. #	Std.	Avg.
#18	0.0233	0.1431	**#99**	0.0001	0.0010
#25	0.0012	0.0088	**#101**	0.0050	0.0718
#26	0.0025	0.0191	**#110**	0.0022	0.0319
#28	0.0137	0.1049	**#118**	0.0008	0.0122
#31	0.0150	0.1151	**#132**	0.0045	0.0644
#32	0.0038	0.0291	**#138**	0.0100	0.1443
#44	0.0072	0.0555	**#141**	0.0042	0.0601
#52	0.0075	0.0632	**#149**	0.0131	0.1882
#60	0.0004	0.0055	**#153**	0.0146	0.2105
#64	0.0020	0.0217	**#170**	0.0064	0.0923
#66	0.0004	0.0052	**#174**	0.0079	0.1130
#70	0.0006	0.0078	**#180**	0.0087	0.2335
#87	0.0007	0.0102	**#187**	0.0044	0.1170

Std.= Standard Deviation, Avg.= Average

APPENDIX G

AVERAGE AND STANDARD DEVIATION VALUES OF SAMPLES AND CONGENERS OF PD2 DATA SET

Table G-1 Average and Standard Deviation Values of Samples of PD2 Data Set

#	Std.	Avg.	#	Std.	Avg.	#	Std.	Avg.
1	4.2415	4.8953	**21**	5.898291767	4.803876622	**31**	5.9182	5.0506
2	4.7512	5.5535	**22**	5.8570	4.7669	**32**	5.8277	4.9701
3	4.2429	4.9512	**23**	6.8355	5.6369	**33**	6.1591	5.2858
4	4.3921	5.0970	**24**	6.3776	5.2367	**34**	5.9881	5.1202
5	4.5970	5.3033	**25**	6.3497	5.2146	**35**	6.1313	5.2626
6	4.2376	4.9459	**26**	6.4419	5.2873	**36**	5.8734	5.0193
7	4.5831	5.3664	**27**	6.3355	5.1978	**37**	5.3039	4.4652
8	5.3392	6.0868	**28**	6.0849	4.9709	**38**	5.4871	4.6468
9	4.7908	5.4493	**29**	6.2349	5.1112	**39**	5.8163	4.9541
10	4.3677	5.0998	**30**	6.0034	4.8972	**40**	5.4148	4.5729
11	4.4256	5.1377						
12	4.1176	4.7810						
13	4.3268	5.0519						
14	5.5765	6.4567						
15	4.2280	4.9172						
16	4.5540	5.3401						
17	4.6779	5.3201						
18	4.4360	5.0925						
19	4.6361	5.4374						
20	4.4918	5.2333						

Table G-2 Average and Standard Deviation Values of Congeners of PD2 Data Set

Con. #	Std.	Avg.	Con. #	Std.	Avg.
#18	0.0311	0.1990	**#110**	4.1489	6.9849
#28	0.1217	0.1974	**#118**	6.8036	9.2706
#31	0.0716	0.2213	**#132**	0.9378	5.1123
#33	0.0179	0.1127	**#138**	1.4368	9.4328
#44	0.3536	0.5044	**#149**	2.9962	14.3781
#49	0.2073	0.3119	**#151**	1.5728	5.2447
#52	0.4003	1.2416	**#153**	4.1188	12.8748
#70	3.7515	4.0343	**#170**	1.9101	6.6990
#74	1.2179	1.3197	**#180**	5.0722	16.5363
#87	1.7951	2.5793	**#183**	1.2405	3.8180
#95	0.8379	4.9535	**#187**	2.7795	8.6644
#99	2.2588	2.9788	**#199**	0.9723	2.8579
#101	1.8351	8.2144			