

AQUEOUS HYDROCARBON SYSTEMS:
EXPERIMENTAL MEASUREMENTS AND
QUANTITATIVE STRUCTURE-PROPERTY
RELATIONSHIP MODELING

By

BRIAN J. NEELY

Associate of Science
Ricks College
Rexburg, Idaho
1985

Bachelor of Science
Brigham Young University
Provo, Utah
1990

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1996

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2007

AQUEOUS HYDROCARBON SYSTEMS:
EXPERIMENTAL MEASUREMENTS AND
QUANTITATIVE STRUCTURE-PROPERTY
RELATIONSHIP MODELING

Dissertation Approved:

K. A. M. Gasem

DISSERTATION ADVISOR

R. L. Robinson, Jr.

K. A. High

M. T. Hagan

A. Gordon Emslie

Dean of the Graduate College

Preface

The objectives of the experimental portion of this work were to (a) evaluate and correlate existing mutual hydrocarbon-water LLE data and (b) develop an apparatus, including appropriate operating procedures and sampling and analytical techniques, capable of accurate mutual solubility (LLE) measurements at ambient and elevated temperatures of selected systems. The hydrocarbon-water systems to be studied include benzene-water, toluene-water, and 3-methylpentane-water. The objectives of the modeling portion of this work were to (a) develop a quantitative structure-property relationship (QSPR) for prediction of infinite-dilution activity coefficient values of hydrocarbon-water systems, (b) evaluate the efficacy of QSPR models using multiple linear regression analyses and back propagation neural networks, (c) develop a theory based QSPR model, and (d) evaluate the ability of the model to predict aqueous and hydrocarbon solubilities at multiple temperatures.

I wish to thank Dr. K. A. M. Gasem for his support, guidance, and enthusiasm during the course of this work and Dr. R. L. Robinson Jr. for his guidance and editorial skills. I am also thankful for the service of Drs. K. High and M. Hagen on my graduate committee. Discussions with my fellow students, particularly Srimi Godavarthy, also provided encouragement and assistance in overcoming obstacles.

I gratefully acknowledge the sacrifice and support of my parents in the pursuit of my educational goals. My wife deserves many thanks for her tolerance and

understanding, and the unconditional love and support I received from my children during this endeavor always resulted in excellent stress alleviation.

Table of Contents

CHAPTER 1. OVERVIEW	1
1.1 Rationale	1
1.2 Objectives	2
1.3 Dissertation Organization	2
CHAPTER 2. LITERATURE REVIEW	4
2.1 Introduction	4
2.2 Benzene-Water System	4
Solubility of Benzene in Water	4
Solubility of Water in Benzene	5
2.3 Toluene – Water System	6
Solubility of Toluene in Water	6
Solubility of Water in Toluene	6
2.4 3-Methylpentane – Water System	7
Solubility of 3-Methylpentane in Water	7
Solubility of Water in 3-Methylpentane	7
2.5 Experimental Methods	7
2.6 Summary	9
CHAPTER 3. EXPERIMENTAL APPARATUS	10
3.1 Introduction	10
3.2 Description	11
Feed Section	11
Equilibration Section	13
Separation Section	13
CHAPTER 4. EXPERIMENTAL METHODS AND TECHNIQUES	16
4.1 Introduction	16
4.2 Sample Preparation	16
Organic Phase	17
Solvent Amount	17
4.3 Sample Collection	19
4.4 Instrument Calibration	19
4.5 Sample Analysis	26
4.6 Experimental Discussion	26
CHAPTER 5. MUTUAL SOLUBILITIES IN THE BENZENE – WATER, TOLUENE – WATER, AND 3-METHYLPENTANE-WATER SYSTEMS	29
5.1 Introduction	29
5.2 Experimental Measurements	30
Materials	30

Apparatus	31
Feed Section.....	32
Equilibration Section	32
Separation Section	33
Sampling Section	33
Methods and Procedures.....	35
5.3 Data Correlation and Evaluation.....	37
Solubility of Hydrocarbons in Water.....	37
Solubility of Water in Hydrocarbons.....	39
5.4 Results and Discussion	40
Benzene Solubility in Water	42
Water Solubility in Benzene	43
Toluene Solubility in Water.....	44
Water Solubility in Toluene.....	45
3-Methylpentane Solubility in Water.....	46
Water Solubility in 3-Methylpentane.....	47
5.5 Conclusions.....	48
5.6 Nomenclature.....	49
5.7 Literature Cited.....	50
CHAPTER 6. CONCLUSIONS AND RECOMMENDATIONS	74
6.1 Conclusions.....	74
6.2 Recommendations.....	75
CHAPTER 7. OVERVIEW	77
7.1 Rationale	77
7.2 Objectives	79
7.3 Dissertation Organization	79
CHAPTER 8. INFINITE-DILUTION ACTIVITY COEFFICIENT MODELS	81
8.1 Theoretical Models	82
Regular Solution Theory Models.....	82
Equation of State Models.....	83
8.2 Pure-Component Models	84
8.3 Group Contribution Models.....	86
8.4 Empirical Models.....	87
8.5 LSER Model	89
8.6 Computational Chemistry Models	90
Molecular Simulations.....	90
QSPR Modeling.....	90
CHAPTER 9. COMPUTATIONAL CHEMISTRY.....	92
9.1 Introduction.....	92
9.2 Historical Background	92
9.3 QSPR Model Development.....	94
Data Entry	96
Molecular Modeling.....	96
Descriptor Generation.....	97
Constitutional Descriptors	98
Topological Descriptors.....	98

Geometrical Descriptors	98
Charge Distribution Related Descriptors	99
Molecular Orbital Related Descriptors	100
Thermodynamic Descriptors.....	100
Solvation Descriptors.....	100
Constructed Descriptors.....	101
Feature Selection.....	101
Objective Feature Selection	102
Subjective Feature Selection.....	102
Model Construction	102
Multiple Linear Regression.....	103
Artificial Neural Networks	104
Model Validation	106
Predictive Ability	106
Monte Carlo Randomization.....	106
CHAPTER 10. NEURAL NETWORKS	108
10.1 Introduction.....	108
10.2 Historical Background	109
10.3 Neural Network Overview	112
Neural Network Construction.....	112
Example Neural Networks.....	115
Back Propagation.....	115
Radial Basis Function	117
Neural Network Architecture.....	117
Learning Algorithm	117
Network Topology	118
Data Type.....	119
CHAPTER 11. QUANTITATIVE STRUCTURE PROPERTY RELATIONSHIP MODELS.....	122
11.1 Introduction.....	122
11.2 Database Development	125
11.3 QSPR Model Development and Results	129
Type I Analysis.....	130
Type II Analysis.....	132
Type III Analysis	133
11.4 Discussion.....	134
QSPR Model Development.....	134
BG-EOS – QSPR Model.....	138
Model Comparisons	140
11.5 References.....	143
CHAPTER 12. CONCLUSIONS AND RECOMMENDATIONS	164
12.1 Conclusions.....	164
12.2 Recommendations.....	165
LITERATURE CITED	167
APPENDIX A. STANDARD OPERATING PROCEDURE	177
A.1 Initialization	177

Pressurization of the Backpressure Regulator	177
Isolation of the Apparatus	178
Pressurization of the System	178
Preparation for Sample Collection	179
Preparation of the Sample Bottles	180
A.2 Sampling	181
Sample Collection	181
A.3 Shut Down and Preventive Maintenance	182
Apparatus Shut Down	182
Preventive Maintenance	183
A.4 Analysis	183
Preparation for Analysis	183
Sample Analysis	184
APPENDIX B. CALIBRATION TECHNIQUE AND DATA	186
B.1 Aqueous Phase	186
Procedure	186
Material Balance	187
Sample Calculation	190
B.2 Organic Phase	191
Procedure	191
Material Balance	192
Sample Calculation	193
B.3 Calibration Results	194
APPENDIX C. PROPAGATED CALIBRATION ERROR	197
C.1 Introduction	197
C.2 Uncertainty	197
C.3 Uncertainty Associated with the Weight of Pure Solute	199
C.4 Uncertainty Associated with the Weight of Pure Solvent	199
C.5 Uncertainty Associated with the Weight Ratio of the Diluent	200
C.6 Uncertainty Associated with the Weight of Diluent	200
C.7 Uncertainty Associated with the Gas Chromatograph Area Ratio	200
C.8 Total Fractional Uncertainty	201
C.9 Sample Calculation (Toluene-Water)	202
APPENDIX D. SOLUBILITY CALCULATIONS AND DATA	206
D.1 Sample Calculation (Toluene-Water)	207
APPENDIX E. ORGANIC PHASE SAMPLE ANALYSIS CORRECTION	208
E.1 Introduction	208
E.2 Material Balance	209
E.3 Organic Phase Sample Calculations	210
E.4 Calibration Calculations	211
E.5 Water Correction Calculations	212
APPENDIX F. EXPERIMENTAL ERROR ANALYSIS	214
F.1 Introduction	214
F.2 Uncertainty Associated with the Solvent-to-Sample Weight Ratio	215
F.3 Uncertainty Associated with the Weight Ratio	218
F.4 Uncertainty Associated with the Temperature	220

Aqueous Phase	220
Hydrocarbon Phase	221
F.5 Total Fractional Uncertainty	222
F.6 Sample Calculations Toluene - Water	225
Toluene Solubility in Water.....	225
Water Solubility in Toluene.....	228
APPENDIX G. DATABASE INFORMATION	232
APPENDIX H. TYPE I ANALYSIS	264
APPENDIX I. TYPE II ANALYSIS	275
APPENDIX J. TYPE III ANALYSIS.....	285

List of Tables

Table 4-1. Solvent-to-Sample Ratios.....	19
Table 4-2. Gas Chromatograph Operating Conditions	20
Table 4-3. Gas Chromatograph Calibration Parameters	22
Table 5-1. Mutual Solubilities for Hydrocarbon-Water Systems with Error Estimates ..	53
Table 5-2. WRMS for Hydrocarbon-Water Systems.....	54
Table 5-3. Derivative Properties for Hydrocarbon Solubilities	54
Table 5-4. Derivative Properties for Water Solubilities	54
Table 9-1. General Classification of Theoretical Molecular Descriptors	97
Table 9-2. QSPR Model Types.....	103
Table 10-1. Supervised Neural Networks.....	120
Table 10-2. Unsupervised Neural Networks.....	121
Table 10-3. Nonlearning Neural Networks.....	121
Table 11-1. Numerical Analysis of the Database	147
Table 11-2. Database Hydrocarbon Structures	147
Table 11-3. Summary of Case Studies.....	148
Table 11-4. Functional Group Constructed Descriptors	149
Table 11-5. Constructed Descriptors	149
Table 11-6. Summary of Type I Results	150
Table 11-7. Summary of Type I Results after Outlier Elimination	150
Table 11-8. Summary of Type I, Type II, and Type III Results	151
Table 11-9. Descriptors Used in the Type III Analyses.....	152
Table 11-10. Molecules in the BG EOS Case Study	153
Table 11-11. Summary of Bader-Gasem EOS Case Study.....	154
Table 11-12. Model Comparisons.....	154
Table G-1. Molecular Structures Found in the Database.....	233
Table G-2. Organics in Water (Direct Measurements).....	241
Table G-3. Organics in Water (Indirect Measurements)	255

Table G-4. Water in Organic Measurements	260
Table H-1. Summary of Type I Results	264
Table H-2. Optimum Descriptors Used in the Type I Analyses	265
Table I-1. Summary of Type II Results	275
Table I-2. Optimum Descriptors Used in the Type II Analyses	276
Table J-1. Summary of Type III Results.....	285
Table J-2. Descriptors Used in the Type III Analyses.....	286

List of Figures

Figure 3-1. Schematic Diagram of the Continuous Flow Apparatus.....	12
Figure 4-1. Effect of Extractant Amount on Toluene Concentration	18
Figure 4-2. Benzene – Decane Calibration	22
Figure 4-3. Water – Ethanol Calibration (Benzene)	23
Figure 4-4. Toluene – 2,2,4-Trimethylpentane Calibration	23
Figure 4-5. Water – Ethanol Calibration (Toluene).....	24
Figure 4-6. 3-Methylpentane – Decane Calibration	24
Figure 4-7. Low Temperature Water – Ethanol Calibration (3-Methylpentane).....	25
Figure 4-8. High Temperature Water – Ethanol Calibration (3-Methylpentane)	25
Figure 5-1. Schematic Diagram of the Continuous Flow Apparatus.....	55
Figure 5-2. Solubility of Benzene in Water	56
Figure 5-3. Weighted Deviation in the Solubility of Benzene in Water.....	57
Figure 5-4. Percent Deviation in the Solubility of Benzene in Water	58
Figure 5-5. Solubility of Water in Benzene	59
Figure 5-6. Weighted Deviation in the Solubility of Water in Benzene.....	60
Figure 5-7. Percent Deviation in the Solubility of Water in Benzene	61
Figure 5-8. Solubility of Toluene in Water.....	62
Figure 5-9. Weighted Deviation in the Solubility of Toluene in Water	63
Figure 5-10. Percent Deviation in the Solubility of Toluene in Water	64
Figure 5-11. Solubility of Water in Toluene.....	65
Figure 5-12. Weighted Deviation in the Solubility of Water in Toluene	66
Figure 5-13. Percent Deviation in the Solubility of Water in Toluene	67
Figure 5-14. Solubility of 3-Methylpentane in Water	68
Figure 5-15. Weighted Deviation in the Solubility of 3-Methylpentane in Water	69
Figure 5-16. Percent Deviation in the Solubility of 3-Methylpentane in Water.....	70
Figure 5-17. Solubility of Water in 3-Methylpentane	71
Figure 5-18. Weighted Deviation in the Solubility of Water in 3-Methylpentane	72

Figure 5-19. Percent Deviation in the Solubility of Water in 3-Methylpentane.....	73
Figure 9-1. QSPR/QSAR Model Generation Flowchart.....	95
Figure 9-2. 2-D Representation of Cyclohexane	96
Figure 10-1. Schematic of a Neural Network	113
Figure 10-2. Abbreviated Notation for a Neural Network.....	114
Figure 10-3. Example of a Linear (left) and a Tan-Sigmoidal Activation Function (right)	114
Figure 11-1. Infinite-Dilution Activity Coefficients of CS3 Case Study Showing Outliers	155
Figure 11-2. Optimum Number of Descriptors for the CS3 Case Study.....	155
Figure 11-3. Infinite-Dilution Activity Coefficients of CS3 Case Study (Type I) using Ten Descriptors.....	156
Figure 11-4. Infinite-Dilution Activity Coefficients of CS1 Case Study (Type II).....	156
Figure 11-5. Infinite-Dilution Activity Coefficients of CS2 Case Study (Type II).....	157
Figure 11-6. Infinite-Dilution Activity Coefficients of CS3 Case Study (Type II).....	157
Figure 11-7. Infinite-Dilution Activity Coefficients of CS4 Case Study (Type II).....	158
Figure 11-8. Infinite-Dilution Activity Coefficients of CS5 Case Study (Type II).....	158
Figure 11-9. Contour Plot of Cross Validation RMSE of CS3.....	159
Figure 11-10. Infinite-Dilution Activity Coefficients of CS1 Case Study (Type III) ...	159
Figure 11-11. Infinite-Dilution Activity Coefficients of CS2 Case Study (Type III) ...	160
Figure 11-12. Infinite-Dilution Activity Coefficients of CS3 Case Study (Type III) ...	160
Figure 11-13. Infinite-Dilution Activity Coefficients of CS4 Case Study (Type III) ...	161
Figure 11-14. Infinite-Dilution Activity Coefficients of CS5 Case Study (Type III) ...	161
Figure 11-15. C_{12} Parameter of BG EOS Case Study.....	162
Figure 11-16. D_{12} Parameter of BG EOS Case Study	162
Figure 11-17. Infinite-Dilution Activity Coefficients of BG EOS Case Study	163
Figure H-1. Infinite-Dilution Activity Coefficients of CS1-A with Outliers	266
Figure H-2. R^2 Plot for Number of Parameters Determination (CS1-A Case Study)....	266
Figure H-3. Infinite-Dilution Activity Coefficients of CS1-A Case Study (Type I).....	267
Figure H-4. Infinite-Dilution Activity Coefficients of CS1-B with Outliers.....	267
Figure H-5. R^2 Plot for Number of Parameters Determination (CS1-B Case Study)....	268
Figure H-6. Infinite-Dilution Activity Coefficients of CS1-B Case Study (Type I)	268
Figure H-7. Infinite-Dilution Activity Coefficients of CS2 with Outliers.....	269

Figure H-8. R ² Plot for Number of Parameters Determination (CS2Case Study).....	269
Figure H-9. Infinite-Dilution Activity Coefficients of CS2Case Study (Type I)	270
Figure H-10. Infinite-Dilution Activity Coefficients of CS3 with Outliers.....	270
Figure H-11. R ² Plot for Number of Parameters Determination (CS3 Case Study).....	271
Figure H-12. Infinite-Dilution Activity Coefficients of CS3 Case Study (Type I)	271
Figure H-13. Infinite-Dilution Activity Coefficients of CS4 with Outliers.....	272
Figure H-14. R ² Plot for Number of Parameters Determination (CS4 Case Study).....	272
Figure H-15. Infinite-Dilution Activity Coefficients of CS4 Case Study (Type I)	273
Figure H-16. Infinite-Dilution Activity Coefficients of CS5 with Outliers.....	273
Figure H-17. R ² Plot for Number of Parameters Determination (CS5 Case Study).....	274
Figure I-1. R ² Plot for Number of Parameters Determination (CS1 Case Study)	277
Figure I-2. Infinite-Dilution Activity Coefficients of CS1 Case Study (Type II)	277
Figure I-3. Difference Plot of CS1 Case Study (Type II)	278
Figure I-4. R ² Plot for Number of Parameters Determination (CS2 Case Study)	278
Figure I-5. Infinite-Dilution Activity Coefficients of CS2 Case Study (Type II)	279
Figure I-6. Difference Plot of CS2 Case Study (Type II)	279
Figure I-7. R ² Plot for Number of Parameters Determination (CS3 Case Study)	280
Figure I-8. Infinite-Dilution Activity Coefficients of CS3 Case Study (Type II)	280
Figure I-9. Difference Plot of CS3 Case Study (Type II)	281
Figure I-10. R ² Plot for Number of Parameters Determination (CS4 Case Study)	281
Figure I-11. Infinite-Dilution Activity Coefficients of CS4 Case Study (Type II)	282
Figure I-12. Difference Plot of CS4 Case Study (Type II)	282
Figure I-13. R ² Plot for Number of Parameters Determination (CS5 Case Study)	283
Figure I-14. Infinite-Dilution Activity Coefficients of CS5 Case Study (Type II)	283
Figure I-15. Difference Plot of CS5 Case Study (Type II)	284
Figure J-1. Contour Plot of Cross Validation RMSE of CS1 Case Study	287
Figure J-2. Infinite-Dilution Activity Coefficients of CS1 Case Study (Type III).....	287
Figure J-3. Difference Plot of CS1 Case Study (Type III)	288
Figure J-4. Contour Plot of Cross Validation RMSE of CS2 Case Study	288
Figure J-5. Infinite-Dilution Activity Coefficients of CS2Case Study (Type III).....	289
Figure J-6. Difference Plot of CS2Case Study (Type III)	289
Figure J-7. Contour Plot of Cross Validation RMSE of CS3 Case Study	290
Figure J-8. Infinite-Dilution Activity Coefficients of CS3 Case Study (Type III).....	290

Figure J-9. Difference Plot of CS3 Case Study (Type III)	291
Figure J-10. Contour Plot of Cross Validation RMSE of CS4 Case Study	291
Figure J-11. Infinite-Dilution Activity Coefficients of CS4 Case Study (Type III).....	292
Figure J-12. Difference Plot of CS4 Case Study (Type III)	292
Figure J-13. Contour Plot of Cross Validation RMSE of CS5 Case Study	293
Figure J-14. Infinite-Dilution Activity Coefficients of CS5 Case Study (Type III).....	293
Figure J-15. Difference Plot of CS5 Case Study (Type III)	294
Figure J-16. Contour Plot of C_{12} from Bader-Gasem EOS Case Study.....	295
Figure J-17. C_{12} Parameter from Bader-Gasem EOS Case Study	295
Figure J-18. Difference Plot of C_{12} from Bader-Gasem EOS Case Study.....	296
Figure J-19. Contour Plot of D_{12} from Bader-Gasem EOS Case Study	296
Figure J-20. D_{12} Parameter from Bader-Gasem EOS Case Study	297
Figure J-21. Difference Plot of D_{12} from Bader-Gasem EOS Case Study	297
Figure J-22. Infinite-Dilution Activity Coefficients of Bader-Gasem EOS Case Study	298
Figure J-23. Difference Plot of Bader-Gasem EOS Case Study.....	298

Chapter 1. Overview

1.1 Rationale

Phase equilibrium data are essential for the proper design, operation, and simulation of many chemical processes. Processes such as distillation, adsorption, and liquid-liquid extraction are examples of prevalent applications. A common application is the removal of environmentally harmful organic substances from wastewater streams in refineries and petrochemical plants. Here, both sour water strippers [1], and liquid-liquid extraction processes (based on the equilibration of hydrocarbon-rich and water-rich liquid streams) are used [2]. Additionally, the development of environmental impact studies, such as the potential contamination of a body of water by a liquid hydrocarbon, is dependent on phase equilibrium data.

Several important processes involve hydrocarbon-water liquid-liquid phase equilibrium (LLE). When experimental data are unavailable, thermodynamic models for LLE are used to predict the phase equilibrium. The accuracy of these models is dependent on the quality, as well as the quantity, of the experimental data used in the model development. While sufficient literature data are available on LLE for select hydrocarbon-water systems at ambient or near ambient temperatures, a deficiency of data exists at elevated temperatures. Due to the lack of data for elevated temperatures and non-studied systems, experimental work such as this is needed.

1.2 Objectives

The objectives of this work are to (a) evaluate and correlate existing mutual hydrocarbon-water LLE data and (b) develop an apparatus, including appropriate operating procedures and sampling and analytical techniques, capable of accurate mutual solubility (LLE) measurements at ambient and elevated temperatures of selected systems. The hydrocarbon-water systems to be studied include benzene-water, toluene-water, and 3-methylpentane-water.

1.3 Dissertation Organization

This dissertation is organized in two sections. Section 1, Chapter 1-6, presents the experimental study, and Section 2, Chapter 7-12, addresses the modeling efforts. Further, the document is written in “manuscript style,” with Chapter 5 written in the form of a manuscript, complete with an independent set of tables, figures, nomenclature, and references. As a result, some introductory remarks for each chapter are repetitive. Following is an outline for Section 1, and a similar outline for Section 2 is given later.

Chapter 2 presents a review of the literature. Data for each system studied in this work is reviewed at ambient and elevated temperatures and pressures. The various analysis methods employed and the types of apparatuses used are also considered. A detailed description of the experimental apparatus is given in Chapter 3. The experimental methods and procedures are discussed in Chapter 4, including sample preparation, sample collection, instrument calibration, and sample analysis. In Chapter 5, results of the experimental measurements are presented and discussed, along with the expected uncertainty in the measured values. Chapter 6 provides a summary of this

work, followed by the appendices. In Appendix A, a standard operating procedure is given to accompany the apparatus. The calibration techniques and the calibration data are given in Appendix B, and the propagated error analysis used to interpret the calibration data is shown in Appendix C. A description of how the solubilities were calculated is given in Appendix D, and Appendix E shows how the water present in the ethanol feedstock was accounted for in the sample analysis. A complete propagated error analysis of the experimental solubility data is presented in Appendix F.

Chapter 2. Literature Review

2.1 Introduction

One objective of this work is mutual solubility measurements at ambient and elevated temperatures of selected hydrocarbon-water systems, which include benzene-water, toluene-water, and 3-methylpentane-water. The amount of liquid-liquid equilibrium data available in the literature varies among binary systems. While there exists an abundance of data for the benzene-water system with over 30 independent measurements at 298 K, literature data are limited for the toluene-water and 3-methylpentane-water systems. The majority of the available data were collected at the ambient temperature of 298 K with few studies measuring mutual solubilities at elevated temperatures and pressures. In general, a critical evaluation of the mutual solubility data is complicated by the limited availability of literature data and differences in experimental pressure conditions when collecting data at elevated temperatures.

2.2 Benzene-Water System

Solubility of Benzene in Water

Much of the literature data for the solubility of benzene in water at atmospheric pressure are in reasonable agreement. However, data by Kudchadker and McKetta [3], Krasnoshchekova and Gubergrits [4], Schwarz [5], and Sanemasa et al. [6] deviate significantly from other studies.

The reported mutual solubility data for the benzene-water system at elevated temperatures and pressures are in fair agreement. The studies of Anderson and Prausnitz [2], Tsonopoulos and Wilson [1], Chandler et al. [7] Jou and Mather [8], Marche et al. [9], Miller and Hawthorne [10], and Neely et al. [11], all of which are along the three-phase equilibrium curve, exhibit fair agreement, but measurements by Guseva and Parnov [12] show significant deviations from the other data.

Solubility of Water in Benzene

Due to the difficulty in quantifying the low concentration of water present in the benzene-rich phase, the reported values of water solubility in benzene at atmospheric pressure are only in fair agreement. The data of Englin et al. [13] and Bittrich et al. [14] are significantly lower than the values given by other studies. Hefter [15] and Chen and Wagner [16] noted that water solubility in benzene data collected since 1965 have a tendency to be higher than the overall average, with smaller standard deviations.

At temperatures below 325 K there is fair agreement, but at higher temperatures data tend to fall into two groups; agreement is fair within each group, but the two groups differ by approximately 25%. The studies of Anderson and Prausnitz [2], Tsonopoulos and Wilson [1], and Chandler et al. [7] comprise the group of lower solubility measurements, while the studies of Jou and Mather [8], Chen and Wagner [16], Umamo and Hayano [17], and Neely et al. [11] comprise the group of higher solubility measurements.

2.3 *Toluene – Water System*

Solubility of Toluene in Water

There is reasonable agreement among the large number of toluene-in-water solubility data at ambient temperature. The data of Krasnoshchekova and Gubergrits [18], Schwarz [5], Pierotti and Liabastre [19], and Sanemasa et al. [6] deviate significantly from other studies.

The reported mutual solubility data for the toluene-water system at elevated temperatures and pressures are in fair agreement. The studies of Anderson and Prausnitz [2], Chandler et al. [7], Jou and Mather [8], and Chen and Wagner [20], which are along the three-phase equilibrium curve, exhibit fair agreement, but the study of Miller and Hawthorne [10] produced somewhat lower solubility than the other data.

Solubility of Water in Toluene

There is good agreement among the moderate number of studies reporting the solubility of water in toluene at ambient temperature, despite the difficulty in quantifying the small amount of water present in the hydrocarbon. The data of Tarassenkow and Poloshinzewa [21, 22] and Roddy and Coleman [23] deviate significantly from other studies.

The studies of Marche et al. [24], Anderson and Prausnitz [2], Chandler et al. [7], Jou and Mather [8], and Chen and Wagner [20], which are along the three-phase equilibrium curve, exhibit fair agreement, but scatter in the data becomes more pronounced at higher temperatures.

2.4 3-Methylpentane – Water System

Solubility of 3-Methylpentane in Water

Four separate studies, all conducted at or near ambient temperatures, comprise the available literature data for the solubility of 3-methylpentane in water. The ambient temperature solubility measurements of McAuliffe [25], Price [26], and Rudakov and Lutsyk [27] show good agreement, but the data of Polak and Lu [28] are approximately 30% higher in solubility. According to Hefter [15], the hydrocarbon solubilities measured by Polak and Lu tend to be much higher in value compared to measurements of other investigators.

Solubility of Water in 3-Methylpentane

Polak and Lu [28] provide the only reported data, which were measured at and near ambient temperatures, for the aqueous solubility. While the lack of relevant data precludes a critical evaluation of Polak and Lu's data, their aqueous solubility data are normally consistent with other researchers in well-characterized systems [15].

2.5 Experimental Methods

Generally, two types of experimental methods are used to collect liquid-liquid equilibrium data; namely, static cells and continuous flow apparatus. A static cell is charged with a solution, agitated for some time period, and then allowed to gravimetrically separate before analysis. Those investigations using a static cell to achieve equilibrium include McAuliffe [25, 29], Karlsson [30], Franks [31], Franks et al. [32], Goldman [33], Guerrant [34], Anderson and Prausnitz [2], Polak and Lu [28],

Chandler et al. [7], Krasnoshchekova and Gubergrits [4], Ng and Chen [35], Schatzberg [36], Jou and Mather [8], and Marche et al. [9].

A continuous flow apparatus has a gravimetric separation cell, which is continuously charged with a well mixed, equilibrated saturated solution. The investigations by Wang and Chao [37], Chen [38], Chen and Wagner [16, 20, 39], Bennett [40], Miller and Hawthorne [10], Neely et al. [11], Ratzlaff [41], and Stevenson et al. [42] use continuous flow apparatus, similar to the apparatus presented in this study, to collect liquid-liquid equilibrium data. Advantages of continuous flow apparatus over a static equilibrium cell include more rapid measurement of phase compositions [42] and minimization of hydrocarbon thermal degradation [37, 42].

Various analytical techniques have been utilized for sample analysis. These include volumetric analysis (Guerrant [34] and Umamo and Hayano [17]), Karl Fischer titration (Polak and Lu [28], Karlsson [30], Tsonopoulos and Wilson [1], and Stevenson et al. [42]), ultraviolet spectrophotometry (Bradley et al. [43], Franks et al [32], and Arnold et al. [44]), high performance liquid chromatography (Marche, et al. [9]), and gas chromatography/mass spectroscopy (Miller and Hawthorne [10]). Gas chromatography is probably the most common analytical technique, and studies utilizing this method include Chen and Wagner [16, 20, 39], Chen [38], Bennett [40], Polak and Lu [28], McAuliffe [25, 29], Franks et al. [32], Anderson and Prausnitz [2], Tsonopoulos and Wilson [1], Chandler et al. [7], Stevenson et al. [42], Jou and Mather [8], Neely et al. [11], and Ratzlaff [41].

2.6 *Summary*

The availability of liquid-liquid equilibrium data at elevated temperatures is limited for most systems, with the exception of benzene-water. Due to the thorough investigation of the benzene-water system by several independent investigators, this system is often employed as a benchmark when testing an apparatus or analytical technique. There is a moderate amount of literature data available for the toluene-water system, and very few literature data are available for the 3-methylpentane-water system.

Static cells and continuous flow apparatus are the most common methods for collection of liquid-liquid equilibrium data. Various analytical techniques have been employed among the studies presented here with gas chromatography being the most utilized.

Chapter 3. Experimental Apparatus

3.1 *Introduction*

A continuous flow apparatus, originally designed and constructed by Ratzlaff [11], was used to collect mutual solubility data of two liquid phases in equilibrium at elevated temperatures and pressures. In brief, the experimental apparatus may be described as consisting of a phase separation cell, which is located in a convection oven. A single inlet transports two well-mixed liquids to the separation cell. A backpressure regulator controls the pressure, using a pressurized nitrogen source, of the apparatus. The phases are separated gravimetrically in the phase separation cell, then exit the cell through outlets in the top and bottom of the cell, and are collected.

An earlier apparatus, constructed by Chen and Wagner [16], provided a basis for the design of the current apparatus, but a few improvements were made. A major improvement [11] is the accurate controllability of the hydrocarbon-water interface in the separation cell at elevated pressures, which minimizes the possibility of entrainment of either phase. Adequate phase separation then allows mutual solubilities at elevated temperatures and pressures to be measured. Minor improvements include the addition of tubing connections to aid in the maintenance of the apparatus and the addition of a line to act as a recycle stream directing flow from the waste collection cell to a recycle receptacle. The more costly 3-methylpentane was recycled with the aid of a separatory

funnel to separate the organic and aqueous phases. A more detailed description of the apparatus follows.

3.2 Description

Figure 3-1 presents a schematic drawing of the experimental apparatus. As shown, the apparatus consists of four sections: a feed section, an equilibration section, a separation section, and a sampling section. The feed section introduces two pure, partially miscible fluids at a constant flow rate to the equilibration section, where the two fluids are thoroughly mixed and allowed to come to equilibrium. After the equilibration section, the aqueous and hydrocarbon phases are separated in the separation section. Next, the separated phases are collected in the sampling section. The total volume of the apparatus is approximately 120 cm³. Abbreviations used below in the description of the apparatus correspond to those shown in Figure 3-1.

Feed Section

The feed section contains two reservoirs for the pure liquid feedstocks (hydrocarbon and water), and a LCD Analytical Type NSI-33R duplex miniPump (DP1). The duplex miniPump supplies the liquids at a constant total flow rate of 4.0 cm³/min with equal parts by volume (2.0 cm³/min) of hydrocarbon and water. The flow rate was varied to determine the effect on the solubilities, but no significant variation was observed with flow rates ranging from 1.5 to 4.5 cm³/min. The low flow rates reduce the formation of emulsions in the apparatus and allow sufficient time for the two liquids to reach equilibrium in the equilibration section. At the selected flow rate, the residence time in the system is approximately 30 minutes.

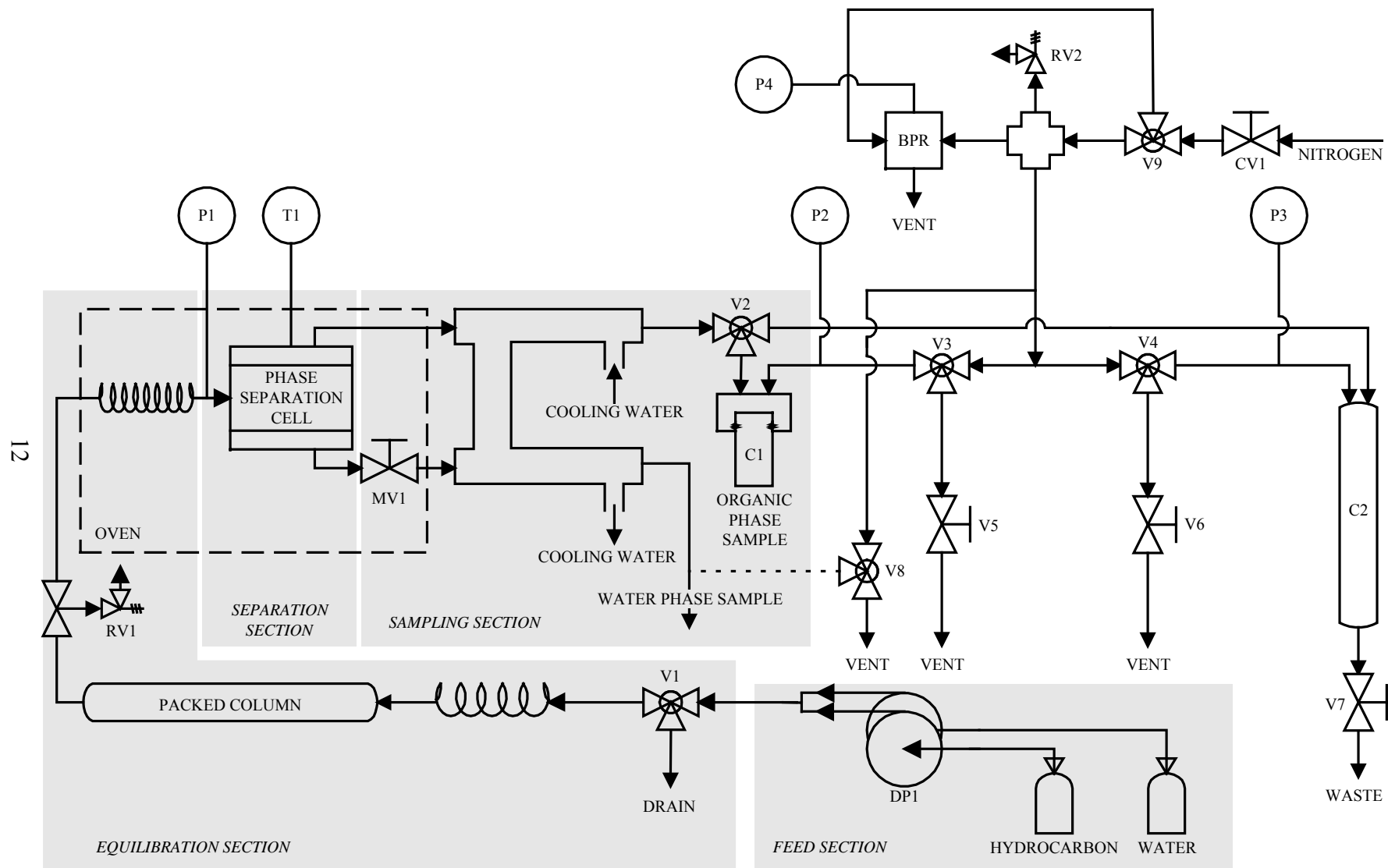


Figure 3-1. Schematic Diagram of the Continuous Flow Apparatus

Equilibration Section

The equilibration section facilitates the mixing necessary for the fluids to reach equilibrium at the selected experimental temperature. Immediately following the duplex miniPump is a Whitey three-way valve (V1), which acts as a bypass valve during startup, when the duplex miniPump is primed. The two fluids then pass through approximately 6.8 m of 0.318 cm-o.d. stainless steel tubing followed by 3.1 m of 0.835 cm-o.d. stainless steel tubing packed with 1.0 mm glass beads. Next, the liquid enters a 1.0 m section of 0.318 cm-o.d. stainless steel tubing before entering the oven.

A Hotpack Digimatic Model 213024 air oven, with a maximum temperature rating of 623 K, is employed to provide a suitable thermal environment. The oven temperature is controllable to within ± 0.1 K of the set point, as determined by the manufacturer. A J-type thermocouple, calibrated against a Minco platinum resistance thermometer that is NIST traceable, is used to measure the phase separation cell temperature. Once the fluid reaches the oven, it enters a 15.2 m section of 0.318 cm-o.d. stainless steel tubing, which allows thermal equilibration of the thoroughly mixed hydrocarbon-water mixture before entrance to the separation section.

Separation Section

The separation section consists of a phase separation cell, which is a 316 stainless steel Jerguson Model 12T40 Liquid Level Gage with an internal volume of 19 cm³, located inside the oven. After phase separation of the hydrocarbon-water mixture inside the cell, the aqueous phase exits from the bottom of the cell, and the less dense hydrocarbon phase exits from the top of the cell. Since a potential exists for phase

separation to occur in the aqueous and hydrocarbon phases due to adsorption of the solute on sample line surfaces [9], the separated phases exit the phase separation cell through 0.159 cm-o.d. stainless steel capillary tubing. This tubing minimizes dead volume, thus, minimizing the effects of phase separation on sample composition.

Sampling Section

After exiting the phase separation cell, the aqueous phase passes through an Autoclave Engineering micrometering valve (MV1), which due to the valve packing material limits the maximum operating temperature of the apparatus to 505 K. This valve, which is located inside the oven, controls the flow of the aqueous phase from the phase separation cell. By controlling the aqueous phase effluent rate, the hydrocarbon-water interface level is controlled near the center of the cell, which minimizes the possibility of entrainment. Each phase passes through a water-cooled heat exchanger 20.3 cm in length prior to being collected. Tap water is used on the shell side (0.635 cm-o.d. stainless steel tubing) to effectively cool each phase to room temperature before collection, which aids in the prevention of sample volatilization.

Since the presence of a vapor phase would interfere with the collection of the hydrocarbon phase from the top of the equilibrium cell, elevated pressures (above the three-phase pressure) in the apparatus are established using pressurized nitrogen gas to create a backpressure on the system. A Grove Mity Mite S-91XW backpressure regulator (BPR) is used to control the pressure in the high-pressure sampling cell (C1). To protect against overpressure, a spring-loaded Nupro relief valve is placed on each possible source of pressure. One relief valve (RV1) is located on the liquid mixture feed line, upstream of the oven, and a second relief valve (RV2) is located on the nitrogen stream line. The

hydrocarbon phase sample is collected in a glass bottle placed in a 300 cm³, sightless, high-pressure sampling cell (C1), which is pressurized by nitrogen gas. Cell pressure is measured at the feed port of the phase separation cell with a Sensotec STJE pressure transducer and 450D readout. The maximum pressure of the system is limited by the pressure transducer, which has a pressure limit of 13.8 MPa (2000 psia). The relief valves are set at 12.4 MPa (1800 psia).

A Whitey three-way valve (V2) is located between the phase separation cell and C1. This valve diverts the flow of the hydrocarbon phase sample to a 400 cm³, sightless, high-pressure, collection cell (C2), which allows continuous flow through the system at elevated pressures while changing the sample bottles in C1. The blanket of nitrogen gas also pressurizes C2. The nitrogen gas may be vented when C1 and C2 are isolated from the system by the utilization of Whitey three-way valves. The valve V3 is used to isolate C1 and the valve V4 is used to isolate C2. When isolated, C1 or C2 may be depressurized to atmospheric pressure while maintaining a constant elevated pressure inside the apparatus.

Chapter 4. Experimental Methods and Techniques

4.1 Introduction

The experimental procedure includes the following: preparation and collection of the samples, calibration of the instrument used for sample analysis, and analysis of the samples. A gas chromatograph (GC) equipped with a thermal conductivity detector (TCD) was used for the sample analysis. Calibration of the GC was by either a serial dilution technique or the use of external standards. A weighted-least-squares regression was used to model the calibration data and generate a calibration curve, which was utilized during sample analysis to determine the concentration of the sample. Problems encountered during this experimental work are also discussed.

4.2 Sample Preparation

Sample preparation includes selection of an appropriate solvent, for extraction of the hydrocarbon in the aqueous phase and homogenization of the organic phase, and determination of an optimum solvent amount for the aqueous and organic phase samples. Solvent amounts were determined using solutions of known concentrations representative of expected experimental sample concentrations. Prior to sample collection, a known weight of solvent was added to the 1 oz sample bottles in preparation for sample analysis. The bottles were obtained from Alltech and used open caps with Teflon liners.

Aqueous Phase

The purpose of the solvent used for aqueous phase analysis was to extract the hydrocarbon from the equilibrium water-rich phase. This provided a water-free sample for analysis. By excluding water from the sample analysis, a reproducible analysis was achieved. Depending on the system of interest, the solvent used in the aqueous phase samples was either decane or 2,2,4-trimethylpentane. For the systems with benzene and 3-methylpentane, decane was the solvent of choice, and for the system with toluene, 2,2,4-trimethylpentane was employed as the solvent. To avoid interference with the gas chromatograph analysis of the hydrocarbons of interest, the retention time of the solvents were used as the basis of their selection.

Organic Phase

Unlike the solvent for water phase analysis (which extracted the hydrocarbon), the solvent used in the organic phase analysis was chosen to homogenize the hydrocarbon sample to a single phase, which allows the analysis of water in the presence of the hydrocarbon. Since water is soluble in ethanol and good peak separation exists between water, ethanol, and the hydrocarbon of interest in the gas chromatograph analysis, ethanol was used as the homogenizing solvent in the organic phase samples.

Solvent Amount

Determination of the optimum amount of solvent involved the preparation of known concentration solutions. For each system studied, two calibration solutions were prepared for each phase. Lower and higher concentration solutions were prepared to represent expected concentrations at the lowest and highest experimental temperatures,

respectively. Incremental amounts of solvent were added to the initial solutions. After each addition, the effect of solvent addition on the measured mole fractions was determined by GC analysis. The calculated mole fraction of the solution gradually increased until only minor changes were observed with further solvent addition.. A leveling in the calculated mole fraction indicated that an optimum amount of solvent had been used. Determination of the amount of solvent, 2,2,4-trimethylpentane, used in the aqueous phase of the toluene-water system is shown in Figure 4-1, where a solvent to sample mass ratio of 0.15 was chosen.

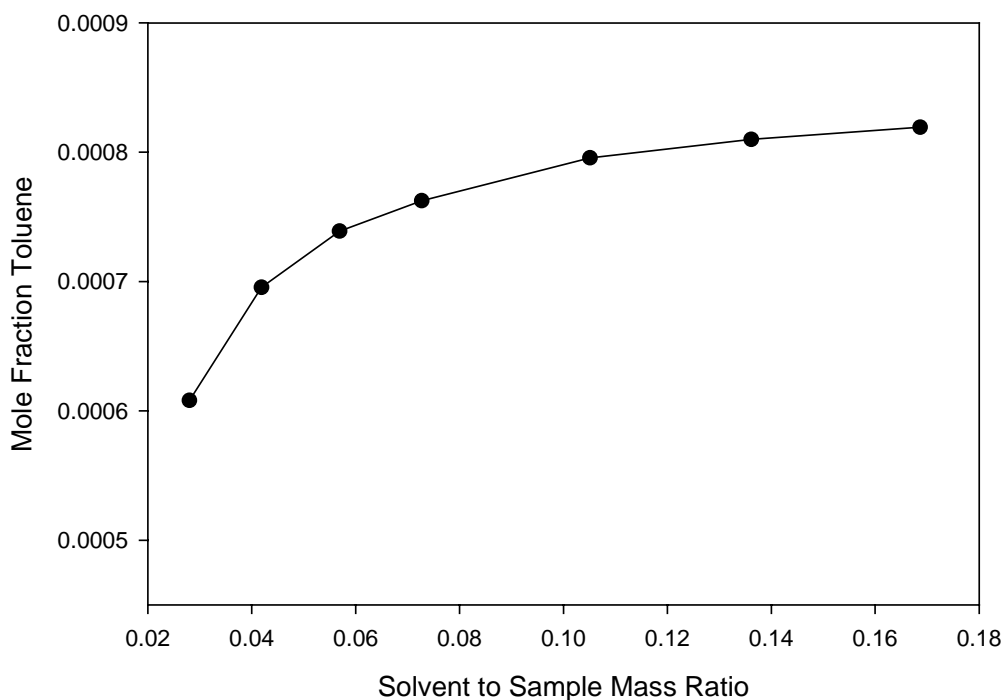


Figure 4-1. Effect of Extractant Amount on Toluene Concentration

The solvent-to-sample mass ratios used in this study are presented in Table 4-1. .

Table 4-1. Solvent-to-Sample Ratios

Benzene – Water	
Solvent-to-Sample Mass Ratio	
Aqueous Phase	0.40
Organic Phase	0.70
Toluene - Water	
Aqueous Phase	0.15
Organic Phase	0.35
3-Methylpentane - Water	
Aqueous Phase	0.22
Organic Phase	0.75

4.3 Sample Collection

In order to minimize sample contact with the atmosphere, the samples were collected after addition of solvent to the sample bottles. Three successive samples of each phase were collected at each temperature. In general, simultaneous collection of the organic and aqueous phases occurred at a pressure slightly above the three-phase equilibrium pressure.

4.4 Instrument Calibration

A recently refurbished Hewlett-Packard 5890A gas chromatograph (GC), equipped with a thermal conductivity detector (TCD) and a Hewlett-Packard 3396A integrator, was used for sample analysis. High purity helium was used as the carrier gas, and the column used to separate injected samples was a 3.6-m x 0.32-cm-o.d. stainless steel column packed with GasChrom 254, supplied by Alltech. The GC was calibrated by either a serial dilution technique or the use of external standards (see Appendix B for details). The range of calibration standards encompassed the experimental concentration

range, and these calibration standards were prepared gravimetrically and analyzed with the GC. Operating conditions, which were optimized for each calibration standard and sample analysis, are listed in Table 4-2. .

Table 4-2. Gas Chromatograph Operating Conditions

Benzene - Water System		
Variable	Water Phase	Benzene Phase
Detector Temperature, °C	300	300
Injector Temperature, °C	250	140
Initial Oven Temperature, °C	225	130
Initial Time, min.	4.0	5.0
Final Oven Temperature, °C	225	225
Final Time, min.	0.0	2.0
Rate, °C/min.	0.0	40.0
Total Gas Flow, cm ³ /min.	30.0	30.0

Toluene - Water System		
Variable	Water Phase	Toluene Phase
Detector Temperature, °C	300	140
Injector Temperature, °C	250	250
Initial Oven Temperature, °C	225	130
Initial Time, min.	2.5	5.0
Final Oven Temperature, °C	225	225
Final Time, min.	0.0	2.0
Rate, °C/min.	0.0	40.0
Total Gas Flow, cm ³ /min.	30.0	30.0

3-Methylpentane - Water System		
Variable	Water Phase	3-Methylpentane Phase
Detector Temperature, °C	300	300
Injector Temperature, °C	250	150
Initial Oven Temperature, °C	225	130
Initial Time, min.	5.1	4.0
Final Oven Temperature, °C	225	180
Final Time, min.	0.0	2.5
Rate, °C/min.	0.0	40.0
Total Gas Flow, cm ³ /min.	30.0	30.0

Temperature programming of the GC was employed to provide the most accurate and reproducible analysis in a reasonable amount of time. A sample volume of 0.003 cm³ (3 μL) was injected into the GC.

The calibration curves were prepared from the solute-to-solvent weight ratio as a function of the solute-to-solvent area ratio, and the data were regressed using the nonlinear weighted-least-squares Marquardt method [45]. Uncertainty in the weight ratio, which was used to weight each datum in the regressions, was determined from an analysis of propagated error. Details of this analysis are provided in Appendix C.

Each calibration curve was expressed as a power law function, as follows:

$$WR = \alpha AR^{\beta} \quad (4-1)$$

where WR is the weight ratio, AR is the area ratio and α and β are regressed parameters. The parameters from the regression of the calibration data are listed in Table 4-3, and the calibration curves are presented in Figures 4-2 - 4-8. Overall, the calibration curves demonstrated good reproducibility, but any datum with a deviation greater than two and one half times the standard deviation was not included in the regression. Since the range of the water – ethanol calibration for the 3-methylpentane system extended beyond the linear response region of the gas chromatograph, two calibration curves corresponding to the expected WRs at lower temperatures and higher temperatures, respectively, were generated to describe the calibration data. Details concerning the calibration technique and the calibration data are provided in Appendix B.

Table 4-3. Gas Chromatograph Calibration Parameters

BENZENE – WATER SYSTEM		
Calibration	α	β
Benzene – Decane	0.9086	1.134
Water – Ethanol	1.027	1.170

Toluene – Water System		
Toluene – 2,2,4-Trimethylpentane	0.2214	1.020
Water – Ethanol	1.572	1.227

3-Methylpentane – Water System		
3-Methylpentane – Decane	1.530	1.369
Water – Ethanol (High Temperature)	0.1203	1.230
Water – Ethanol (Low Temperature)	29690	4.308

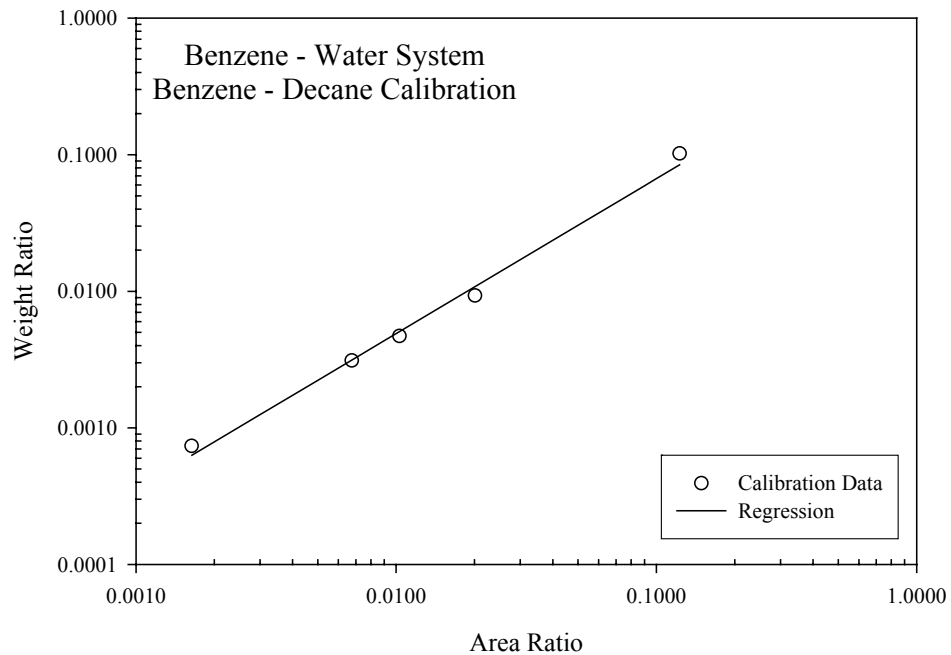


Figure 4-2. Benzene – Decane Calibration

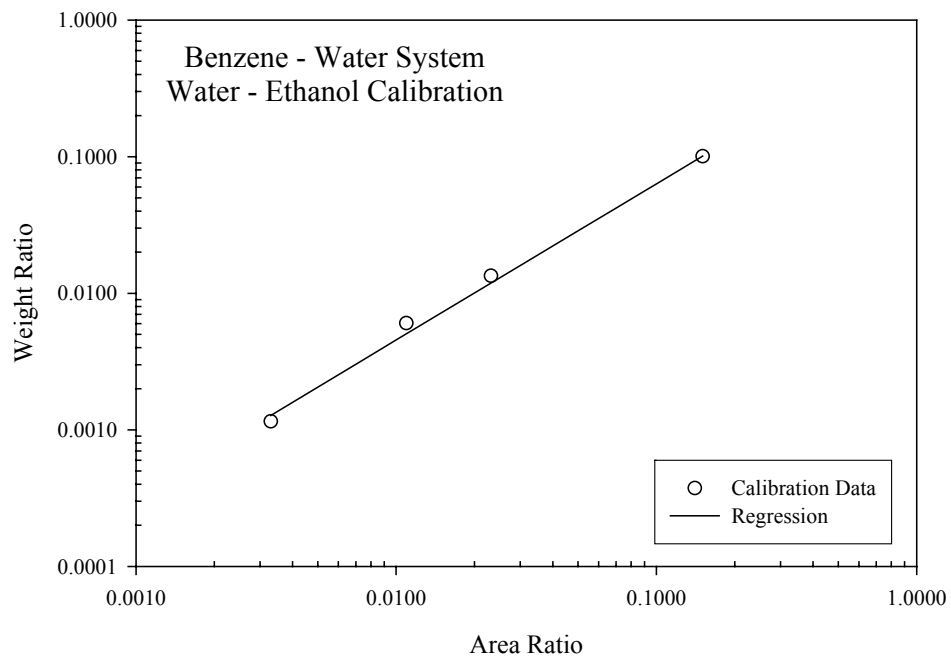


Figure 4-3. Water – Ethanol Calibration (Benzene)

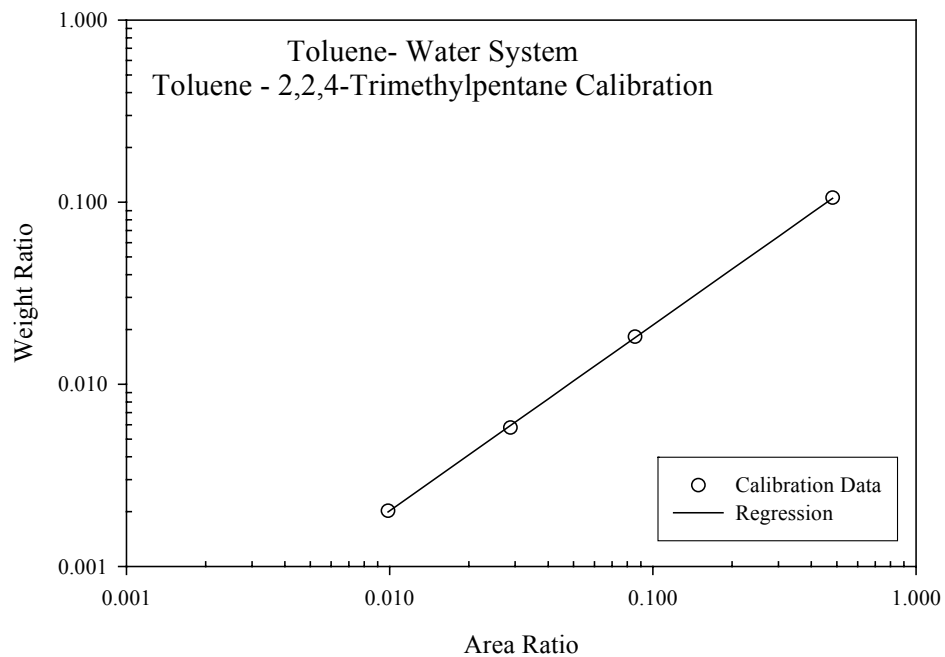


Figure 4-4. Toluene – 2,2,4-Trimethylpentane Calibration

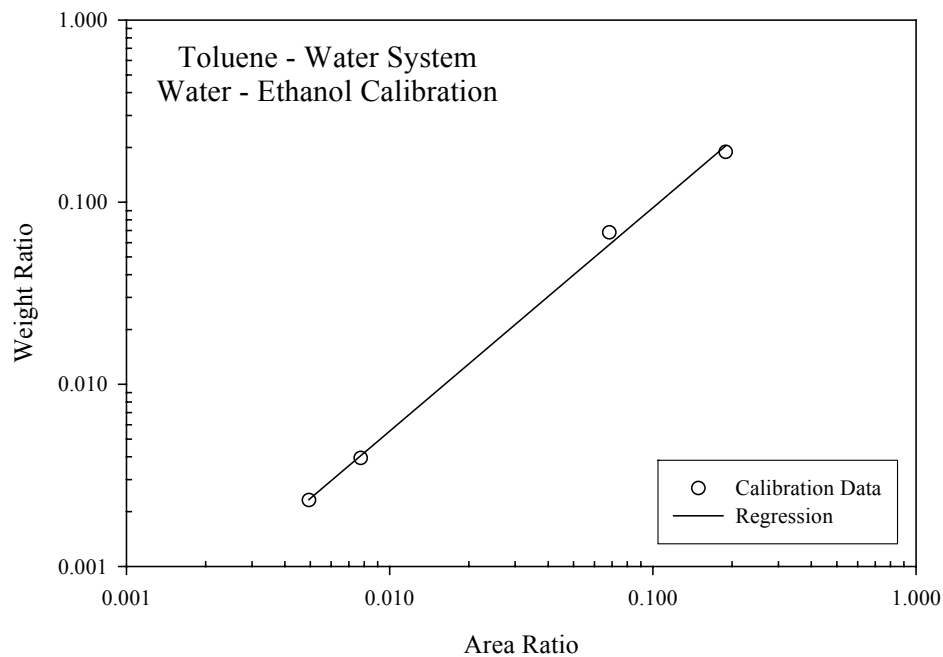


Figure 4-5. Water – Ethanol Calibration (Toluene)

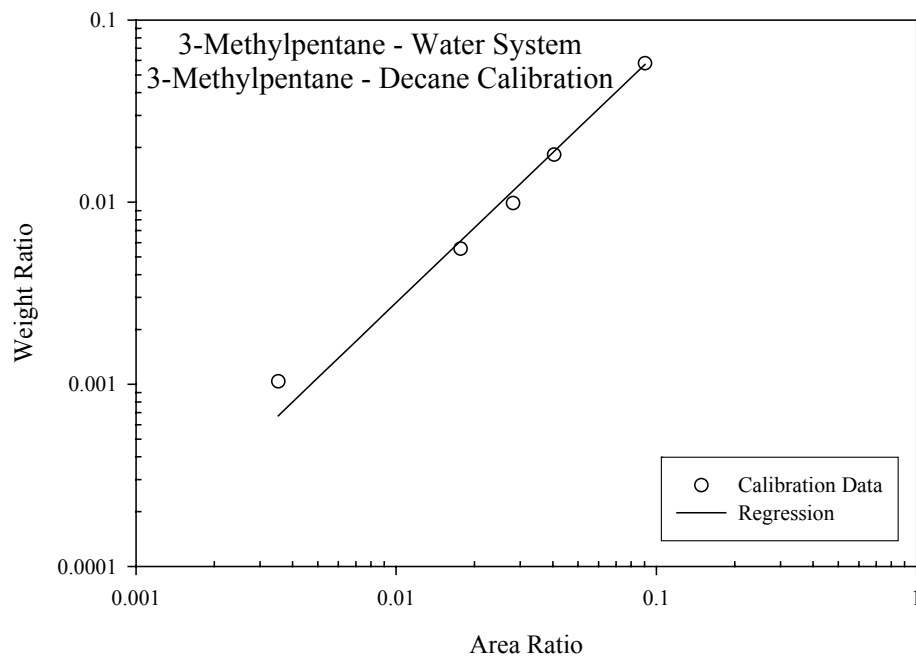


Figure 4-6. 3-Methylpentane – Decane Calibration

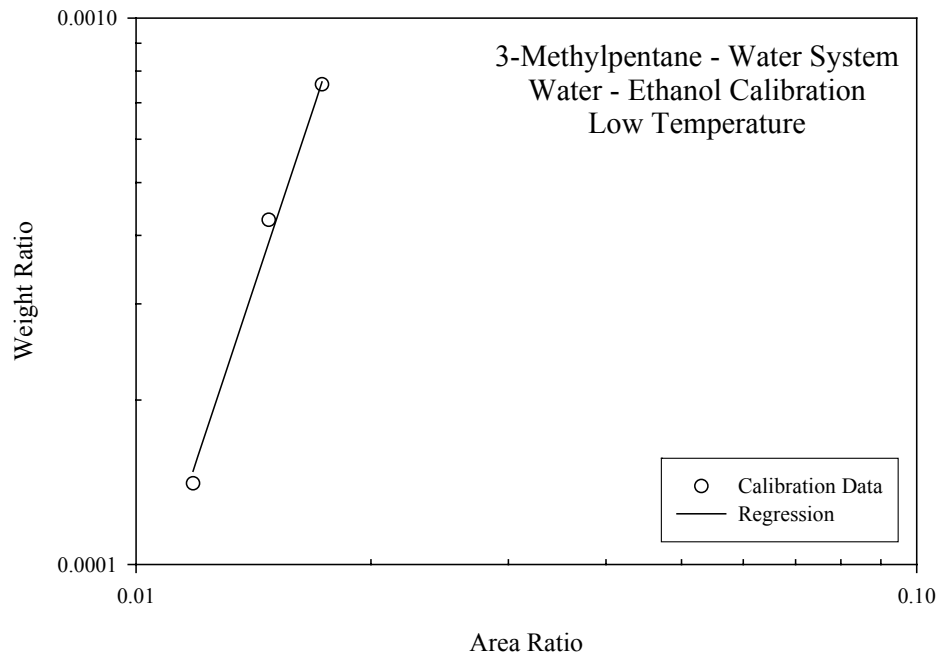


Figure 4-7. Low Temperature Water – Ethanol Calibration (3-Methylpentane)

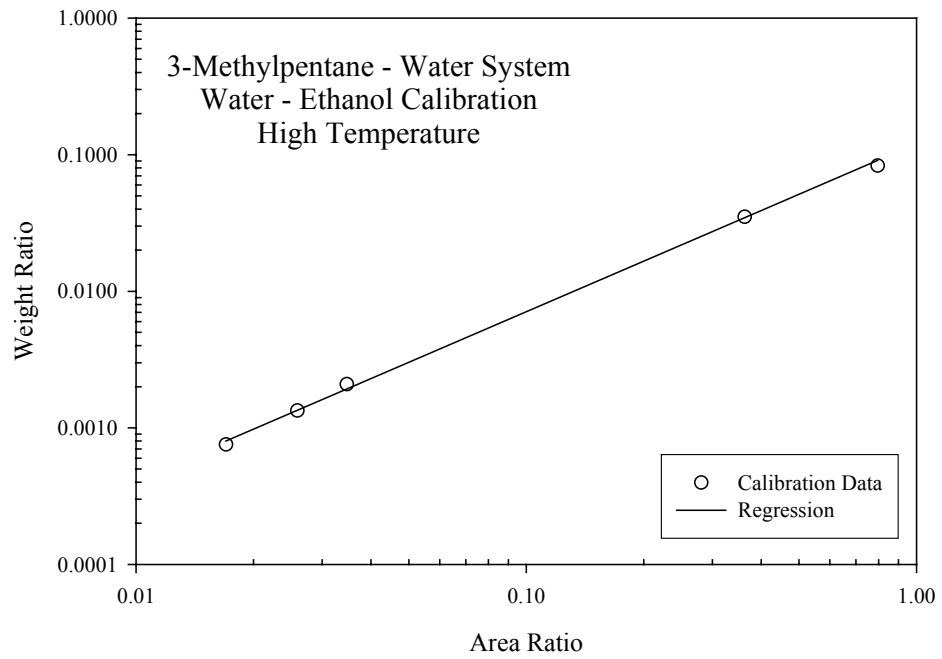


Figure 4-8. High Temperature Water – Ethanol Calibration (3-Methylpentane)

4.5 Sample Analysis

Sample analyses were performed using the same GC conditions as the calibrations. The amount of each component in a sample was given in the GC output as an area, which refers to the integrated portion under each individual curve for each component peak in the analysis. The solute-to-solvent area ratio was found by dividing the solute area by the solvent area, and this value was used with the calibration curve, Equation (4-1), to calculate the solute-to-solvent weight ratio. The weight ratio, area ratio, and molecular weight of the solute and solvent were used to calculate the mole fraction of the sample, and these solubility calculations are explained in Appendix D.

Ethanol, used as the solvent in the organic phase, is hygroscopic and contained small amounts of water. This additional water could cause the area ratio in the organic phase samples and, in effect, the calculated water solubility in the organic phase, to be slightly higher than the equilibrium value. A strategy for the accounting of additional water is described in Appendix E.

Complete description of the sample preparation, sample collection, and sample analysis is given as part of the standard operating procedures in Appendix A, and details concerning the instrument calibration are provided in Appendix B.

4.6 Experimental Discussion

During initial work on the selected systems, the integrator used with the GC failed. After numerous attempts to repair the integrator, a refurbished model was purchased. With the “new” integrator in place, the GC was found to be unresponsive during analyses. Again, considerably effort was expended in attempting to repair the GC

on site, but eventually, it was sent to a company specializing in GC refurbishment. The GC was returned along with chromatographs of selected test injections, and the instrument appeared to be in satisfactory working condition. However, results of the analyses of organic phase samples from the benzene-water system showed large deviations from both literature values and previous work done with the same experimental apparatus and GC [11, 41].

Typically, speculation on the origin of the experimental discrepancies centers on deviations from the established operating procedures, or problems arising with the operation or condition of the experimental apparatus. Experimental work then consisted of exactly replicating all procedures involved in this work, which includes calibration, sampling, and analysis tasks. Time was also spent on insuring that the experimental apparatus was in good working order. Since results obtained during this period were consistent but still inaccurate compared to literature data, the conclusion of this investigation was the existence of a systematic error, which remained unidentified.

Given the amount of time already devoted to these problems, and the need to proceed with other research, abandonment of the experiment appeared to be the most feasible option. However, one final measure was investigated. Originally, ethanol-water mixtures were serially diluted with ethanol to construct a calibration curve for use in the analysis of the organic phase samples, which also contain a large amount of the hydrocarbon of interest. The final investigation centered on whether or not the injected analytical sample, which contained hydrocarbon, was truly representative of the employed calibration standards. Individual external standards were prepared using a solvent to sample ratio similar to that expected experimentally, and a calibration curve

was constructed. After using this new calibration method, satisfactory results were obtained. There seems to be a difference in the GC since the refurbishment, perhaps due to all tubing and chambers being free of fouling. Previously, the peak areas obtained during repeat analyses of the same sample showed a low variability as contrasted by the greater variability now observed. A possible explanation for the variability and change in calibration method is the determination of peak areas by the GC being affected by the high molar vapor expansion of the water and ethanol in the detection chamber. This expansion in the detection chamber could change the relative detected amounts of the components in the injected sample.

Chapter 5. Mutual Solubilities in the Benzene – Water, Toluene – Water, and 3-Methylpentane-Water Systems

5.1 Introduction

Phase equilibrium data are essential for the proper design, operation, and simulation of many chemical processes, which include such applications as distillation, adsorption, and liquid-liquid extraction. An example of a common application is the removal of environmentally harmful organic substances from wastewater streams in refineries and petrochemical plants. Here, both sour water strippers [1], and liquid-liquid extraction processes (based on the equilibration of hydrocarbon-rich and water-rich liquid streams) are used [2]. Increasing environmental concerns have also led to the use of supercritical water, instead of organics, as the solvent in some reaction processes and extraction methods, which include the destruction of hazardous wastes in supercritical water [3] and chemical processing in supercritical and near critical water [4]. Additionally, the development of environmental impact studies, such as the potential contamination of a body of water by a liquid hydrocarbon, is dependent on phase equilibrium data.

When experimental data are unavailable, thermodynamic models for liquid-liquid equilibrium (LLE) are used to predict the phase equilibrium. The accuracy of these models is dependent on the quality, as well as the quantity, of the experimental data used in the model development. While sufficient literature data exist on LLE for many

hydrocarbon-water systems at ambient or near ambient temperatures, a deficiency exists for data at elevated temperatures.

The objectives of this work are to measure and correlate mutual hydrocarbon-water LLE data for the benzene-water, toluene-water, and 3-methylpentane-water systems over a temperature range from ambient conditions to approximately 500 K.

5.2 *Experimental Measurements*

In order to address the need for LLE data on selected systems, a continuous flow apparatus [5] was utilized to obtain liquid-liquid mutual solubilities at temperatures from ambient to 500 K and pressures up to 13.8 MPa. Mutual solubility data have been measured at temperatures from ambient to near the three-phase critical end points [6], 540 K and 560 K, for the hydrocarbon-water systems, benzene-water and toluene-water, respectively. At temperatures and pressures below the three-phase critical end point, three phases (liquid hydrocarbon, liquid water, and vapor) exist. Two phases exist at conditions above this point consisting of a liquid water phase and a vapor phase where the liquid hydrocarbon phase and vapor phase now have identical properties. The well-documented system, benzene-water, was used to validate the proper operation of the apparatus, as well as the sampling and analytical techniques employed.

Materials

Benzene (99.9+%) and toluene (99.8%) were supplied by Aldrich Chemical Co. and the 3-methylpentane (99+%) was supplied by Acros. Additional chemicals used as solvents in the phase analyses included 2,2,4-trimethylpentane (99+%) and decane (99+%) supplied by Aldrich Chemical Co. No further purification of these chemicals

was attempted. Ethanol (USP grade, Absolute-200 Proof), supplied by Pharmco Products, was dehydrated and stored over 4A molecular sieves from Fisher Chemical Company for at least two months prior to use. The Oklahoma State University School of Chemical Engineering supplied the nanopure, deionized water. High purity helium (99.997%) and ultra-high purity nitrogen (99.999%) were obtained from Sooner Airgas, Inc.

Alltech screw top bottles (1 oz.) with open-hole caps and Teflon liners were used for sample collection. Hamilton 10 cm³ syringes, from Alltech, were used for preparation of the calibration standards and were thoroughly rinsed between uses with acetone (99.9%) from Pharmco Products. Hamilton 0.01 cm³ (10 μ l) syringes, from Alltech, were used to inject the calibration standards and experimental samples into the gas chromatograph. Pharmco Products provided the ACS grade methanol (99.99%), used at times in combination with acetone (99.7%) to clean the apparatus and various glassware.

Apparatus

A detailed diagram of the experimental apparatus is presented in Figure 5-1. The apparatus consists of four sections: a feed section, an equilibration section, a separation section, and a sampling section. The feed section introduces two pure, partially miscible fluids at a constant flow rate to the equilibration section, where the two fluids are thoroughly mixed and allowed to equilibrate. After the equilibration section, the aqueous and hydrocarbon phases are separated in the separation section. Next, the separated phases are collected in the sampling section. The total volume of the apparatus is approximately 120 cm³. Abbreviations used in the description of the apparatus correspond to those given in Figure 5-1.

Feed Section

The feed section contains two reservoirs for the pure liquid feedstocks, a hydrocarbon and water, and a LCD Analytical Type NSI-33R duplex miniPump (DP1). The duplex miniPump supplies the liquids at a constant total flow rate of 4.0 cm³/min with equal parts (2.0 cm³/min) of hydrocarbon and water. The flow rate was varied to determine the effect on the solubilities, but no significant variation was observed with flow rates ranging from 1.5 to 4.5 cm³/min. Flow rates in this range reduce the formation of emulsions in the apparatus and allow sufficient time for the two liquids to reach equilibrium in the equilibration section. At the selected flow rate, the residence time of the system is approximately 30 minutes.

Equilibration Section

The equilibration section facilitates the mixing necessary for the fluids to reach equilibrium at the selected experimental temperature. Immediately following the duplex miniPump is a Whitey three-way valve (V1), which acts as a bypass valve when the duplex miniPump is primed. The two fluids then pass through approximately 6.8 m of 0.318 cm-o.d. stainless steel tubing followed by 3.1 m of 0.835 cm-o.d. stainless steel tubing packed with 1.0 mm glass beads. Next, the liquid enters a 1.0 m section of 0.318 cm-o.d. stainless steel tubing before entering the oven.

A Hotpack Digimatic Model 213024 air oven, with a maximum temperature rating of 623 K, is utilized to provide a suitable thermal environment. The oven temperature is controllable to within ± 0.1 K of the set point, as determined by the manufacturer. A J-type thermocouple, calibrated against a Minco platinum resistance thermometer that is NIST traceable, is used to measure the phase separation cell

temperature. Once the fluid reaches the oven, a 15.2 m section of 0.318 cm-o.d. stainless steel tubing allows the thermal equilibration of the thoroughly mixed hydrocarbon-water mixture before entrance to the separation section.

Separation Section

The separation section consists of a phase separation cell, which is a 316 stainless steel Jerguson Model 12T40 Liquid Level Gage with an internal volume of 19 cm³, located inside the oven. After phase separation of the hydrocarbon-water mixture inside the cell, the aqueous phase exits from the bottom of the cell, and the less dense hydrocarbon phase exits from the top of the cell. Since a potential exists for phase separation to occur in the aqueous and hydrocarbon phases due to adsorption of the solute on sample line surfaces [7], the separated phases exit the phase separation cell through 0.159 cm-o.d. stainless steel capillary tubing. This tubing minimizes dead volume, thus, minimizing the effects of phase separation on sample composition.

Sampling Section

After exiting the phase separation cell, the aqueous phase passes through an Autoclave Engineering micrometering valve (MV1), which due to the valve packing material limits the maximum operating temperature of the apparatus to 505 K. This valve, which is located inside the oven, controls the flow of the aqueous phase from the phase separation cell. By controlling the aqueous phase effluent rate, the hydrocarbon-water interface level is controlled near the center of the cell, which minimizes the possibility of entrainment. Each phase passes through a water-cooled heat exchanger 20.3 cm in length prior to being collected. Tap water is used on the shell side (0.635 cm-

o.d. stainless steel tubing) to effectively cool each phase to room temperature before collection, which aids in the prevention of sample volatilization.

Since the presence of a vapor phase would interfere with the collection of the hydrocarbon phase from the top of the equilibrium cell, elevated pressures above the three-phase pressure in the apparatus are established using pressurized nitrogen gas to create a backpressure on the system. A Grove Mity Mite S-91XW backpressure regulator (BPR) is used to control the pressure in the high-pressure sampling cell (C1). To protect against overpressure, a spring-loaded Nupro relief valve is placed at each possible source of pressure. One relief valve (RV1) is located on the liquid mixture feed line, upstream of the oven, and a second relief valve (RV2) is located on the nitrogen streamline. The hydrocarbon phase sample is collected in a glass bottle placed in a 300 cm³, sightless, high-pressure sampling cell (C1), which is pressurized by nitrogen gas. Cell pressure is measured at the feed port of the phase separation cell with a Sensotec STJE pressure transducer and 450D readout. The maximum pressure of the system is limited by the pressure transducer, which has a pressure limit of 13.8 MPa (2000 psia). The relief valves are set at 12.4 MPa (1800 psia).

A Whitey three-way valve (V2) is located between the phase separation cell and C1. This valve diverts the flow of the hydrocarbon phase sample to a 400 cm³, sightless, high-pressure, collection cell (C2), which allows continuous flow through the system at elevated pressures while changing the sample bottles in C1. The blanket of nitrogen gas also pressurizes C2. The nitrogen gas may be vented when C1 and C2 are isolated from the system by sequencing of the Whitey three-way valves. The valve V3 is used to isolate C1 and the valve V4 is used to isolate C2. When isolated, C1 or C2 may be

depressurized to atmospheric pressure while maintaining a constant elevated pressure inside the apparatus.

Methods and Procedures

The analyses of equilibrium phase samples proceeded as follows. First, a known amount of solvent, by weight, was added to the sample bottles. For the organic phase analysis, the sample was mixed with the solvent, ethanol in an approximate weight ratio of 0.7, 0.3, and 0.8 for the benzene-water, toluene-water, and 3-methylpentane-water systems, respectively. The ethanol functioned as a homogenizing cosolvent to provide a single-phase sample for analysis. The ethanol contained a small amount of water, which was accounted for in the sample analysis.

The water phase sample was mixed with a known weight of solvent, decane, for the benzene-water and 3-methylpentane-water systems and 2,2,4-trimethylpentane for the toluene-water system. In order to avoid interference with the gas chromatograph analysis of the hydrocarbons of interest, the retention time of the solvents were used as the basis of their selection. The solvent-to-sample weight ratio was 0.4 for the benzene-water system and 0.2 for the toluene-water and 3-methylpentane-water systems. The solvent was used in the water phase to extract the hydrocarbon from the water, thus providing a water free sample. A more reproducible analysis was achieved by excluding water from the analysis, since aqueous solutions are difficult to analyze accurately by gas chromatography.

In order to avoid reopening the sample bottles and thus, minimize sample contact with the atmosphere, the samples were collected after addition of the solvent to the sample bottles. At each temperature, samples of each phase, organic and water, were

collected simultaneously at a pressure slightly above the three-phase equilibrium pressure. Three samples of each phase were collected for analysis at each experimental equilibrium condition.

Sample volumes of 0.003 cm^3 ($3 \mu\text{l}$) were analyzed using a Hewlett-Packard 5890A gas chromatograph (GC), equipped with a thermal conductivity detector (TCD) and a Hewlett-Packard 3396A integrator. The GC column used was a 3.6-m x 0.32-m stainless steel packed GasChrom 254, supplied by Alltech, and high purity helium was used as the carrier gas.

The GC was calibrated by one of two techniques. A serial dilution technique was utilized for the aqueous phase and, for the hydrocarbon phase, individually-prepared external standards were employed. Calibration data were used to generate calibration curves, which represented the solute-to-solvent weight ratio as a function of the solute-to-solvent area ratio. The calibration data were regressed using a nonlinear weighted least squares Marquardt method [8]. The weighting of each datum was determined by an analysis of propagated error. Each calibration curve was expressed empirically as a power law equation:

$$WR_i = \alpha AR_i^\beta \quad (5-1)$$

where, WR is the weight ratio, AR is the area ratio, and α and β are regressed parameters. This expression was utilized in the sample analysis to determine the solute-to-solvent weight ratio. The following mass balance relation used the solute-to-solvent weight ratio, WR, solvent-to-sample weight ratio, SSR, and the molecular weights of the solute, MW_1 , and solvent, MW_2 , to calculate the mole fraction of the solute in the sample by:

$$x_1 = \frac{[(WR)(SSR)]/MW_1}{\left\{ \frac{[(WR)(SSR)]}{MW_1} \right\} + \left\{ \frac{1}{MW_2} \right\}} \quad (5-2)$$

A complete description of the apparatus and the operating procedures and techniques is given by Neely [9] in Chapters 3 and 4.

5.3 Data Correlation and Evaluation

Correlations for the solubility of liquid hydrocarbons in water and for the solubility of water in liquid hydrocarbons were developed from published solubility data and calorimetric data for enthalpy and specific heat of solution.

Solubility of Hydrocarbons in Water

Benson and Kraus [10] and Wilhem, et al. [11] have discussed the merits of different empirical equations for correlating the temperature dependence of solubility data. The following form, correlating the mole fraction of hydrocarbon as a function of temperature, was selected for hydrocarbon solubility in water:

$$\ln x_{hc} = A + BT_{r,hc}^{-1} + CT_{r,hc}^{-2} \quad (5-3)$$

where x_{hc} is the hydrocarbon mole fraction and $T_{r,hc}$ is the temperature (absolute) of the system divided by the critical temperature of the hydrocarbon, hc . The constants, A , B , and C in Equation (5-3) were obtained by non-linear regression, minimizing the weighted sum of squares (WSS) in the calculated solubilities. All our measured solubility data for the three systems of interest were included in the preliminary regressions, however, any data point with a weighted deviation greater than 2.5 times the standard deviation was not included in the analysis. A final regression was performed on the reduced data set

resulting in the final parameters. Since error estimates for the data points vary with temperature, regressions were weighted by the expected experimental error for each data point, as determined through the analysis of propagated error. The objective function, WSS, employed is given by:

$$WSS = \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{\sigma_i} \right)^2 \quad (5-4)$$

where n is the number of data points, y_i is the predicted data value, \hat{y}_i is the measured data value, and σ_i is the error estimate.

Knowledge of solvation processes and available calorimetric data can be used to assess the quality of the correlation. For very dilute hydrocarbon-water systems, the temperature dependence of the solubility can be expressed by the Gibbs-Duhem equation [12]:

$$\left(\frac{\partial \ln x_i}{\partial T} \right)_p \cong \frac{\Delta \bar{H}_i}{RT^2} \quad (5-5)$$

where the heat of solution, $\Delta \bar{H}_i$, is the excess enthalpy of component i , expressed as the difference between the partial molar enthalpy of component i in solution and the pure molar enthalpy of component i . The heat capacity of solution, $\Delta \bar{C}_{p_i}$, is defined as:

$$\left(\frac{\partial \Delta \bar{H}_i}{\partial T} \right)_p = \Delta \bar{C}_{p_i} \quad (5-6)$$

This excess heat capacity is the difference between the partial molar heat capacity of component i in solution and the pure molar heat capacity of component i . The derivative properties can be calculated from Equation (5-3) as

$$\Delta H_{soln, hc} = -RT \left[\frac{B}{T_{r, hc}} + \frac{2C}{T_{r, hc}^2} \right] \quad (5-7)$$

and

$$\Delta C_{p, soln, hc} = \frac{2RC}{T_{r, hc}^2} \quad (5-8)$$

For many hydrocarbon-water systems, a minimum hydrocarbon solubility exists where $\Delta H_{soln, hc} = 0$. The corresponding temperature, $T_{min, hc}$, can be estimated from Equation (5-7) as

$$T_{min, hc} = \frac{-2CT_{c, hc}}{B} \quad (5-9)$$

where $T_{c, hc}$ is the critical temperature of the hydrocarbon. Derivative data are very sensitive to the solubility measurements, but they can be compared to calorimetric data to provide some insight into the quality of the correlations, not to compare enthalpy effects from solubility and calorimetric measurements.

Solubility of Water in Hydrocarbons

Based on derivative properties and the current understanding of the solvation process of water in hydrocarbons [13], the data for the water solubility in hydrocarbons were correlated by an equation expressing the mole fraction of water in hydrocarbon, x_w , as a function of temperature, as follows:

$$\ln x_w = A + B \ln T_{r, w} \quad (5-10)$$

where $T_{r, w}$ is the temperature (absolute) of the system divided by the critical temperature of water, $T_c = 647.1$ K. The constants, A and B , in Equation (5-10) were obtained by non-linear regression in a fashion similar to the hydrocarbon solubility. All solubility

measurements were included in the preliminary regressions; however, data points with a weighted deviation greater than 2.5 times the standard deviation were not used in the final regressions to establish the expressions in Equations 5-3 and 5-9. During the regressions, the weighting of the solubility measurements was determined through the analysis of propagated error.

Using Equation (5-10) in a similar fashion as in the hydrocarbon solubility section, the derivative properties, enthalpy of solution and heat capacity of solution of water, can be calculated, respectively, as:

$$\Delta H_{soln,w} = RBT \quad (5-11)$$

and

$$\Delta C_{soln,w} = RB \quad (5-12)$$

The solvation process of water dissolving into a non-polar hydrocarbon liquid phase is primarily described as a process of breaking hydrogen bonds [14]. Typical hydrogen bond energies are in the range of 21 to 29 kJ·mole⁻¹ at 298 K, which should correspond to the value of $\Delta H_{soln,w}$.

5.4 Results and Discussion

The mutual solubility data and error estimates of the benzene-water, toluene-water, and 3-methylpentane-water systems are reported in Table 5-1, the weighted-root-mean-square (WRMS) error of the solubility data are given in Table 5-2, and the correlation parameters and derivative property values are given in Table 5-3 and Table 5-4 for the hydrocarbon and water solubilities, respectively. Ideally, both phases were collected at the same time, but occasionally additional experimental data were collected

for only one phase, which accounts for differences in operating temperatures and pressures seen in Table 5-1. Figures 5-2 through 5-19 present the mutual solubility data graphically, which includes plots of the solubility, weighted deviation, and percent deviation for each hydrocarbon or water solubility.

Error bars representing the uncertainty in the solubility measurements have been omitted from the solubility graphs since they do not extend beyond the symbols. By error propagation analysis of the three systems studied, the maximum uncertainty is 4% at a mole fraction of 0.0027 (0.00011 absolute error) and 8% at a mole fraction of 0.0052 (0.00044 absolute error) in the water phase and organic phase measurements, respectively. The higher uncertainty associated with the organic phase measurements are indicative of the difficulty in accurately analyzing aqueous samples by gas chromatography (GC). The contributing factors to the uncertainty in the mole fractions include the solvent and sample weights, the GC analysis, and the temperature. For both hydrocarbon and water solubility, the solvent and sample weights account for less than 1% of the total uncertainty. The GC analysis and temperature account for approximately 84% and 16%, respectively, of the total hydrocarbon solubility uncertainty and 97% and 3%, respectively, of the total water solubility uncertainty. The weighted deviations for the water-in-hydrocarbon systems were generally observed to be much less than 1.0, which differed from the deviations associated with the hydrocarbon-in-water systems. This may be indicative of an overestimation of the propagated error magnitude. The calibration procedure used previously was changed for the water-in-hydrocarbon systems, and the propagated error estimate may need further refinement.

Benzene Solubility in Water

Benzene solubility measurements are presented in Figure 5-2. At temperatures near ambient, an abundance of data exists, which allows for detailed comparisons; however, the system has not been investigated as thoroughly at temperatures greater than 375 K. The evaluations of Hefter [15] and Wagner [13] were utilized extensively in evaluating the quality of the data.

Equation (5-3) was employed to correlate the benzene solubility measurements. Upon analysis, the measurement taken at 490.8 K was not included in the determination of the equation parameters due to a weighted deviation greater than 2.5 times the standard deviation; these weighted deviations are shown graphically in Figure 5-3. As estimated by error propagation, the solubility measurements have a maximum uncertainty of 4.0% at a mole fraction of 0.0027 (0.00011 absolute error) and an average uncertainty of 1.6%, which is shown in Figure 5-4. The WRMS error of the solubility data is 0.72.

At temperatures less than 400 K, the solubility measurements agree within 10% of the broad range of literature data. Generally, in the higher temperature range, the measurements agree within 10% of the more recent results reported by Jou and Mather [16], Chandler et al. [17], Chen and Wagner [18], Anderson and Prausnitz [2], and Marche et al. [7]. Deviations greater than 10%, however, were observed at higher temperatures in comparison with the recent studies of Ratzlaff [19] and Miller and Hawthorne [20].

From Equation (5-7), the heat of mixing at 298.15 K is 1.25 kJ/mole. This value agrees more favorably with the calorimetric heat of solution reported by Reid et al. [21] of 0.80 kJ/mole than the values reported by Gill et al. [22] of 2.08 kJ/mole and De Lisi et

al. [23] of 2.34 kJ/mole. The second derivative property, the specific heat of solution, calculated using Equation (5-8) is 306 J/mole-K. While this value is in good agreement with the value of 301 J/mole-K reported by Clarke and Glew [24], there is disagreement with the values of 373, 351, and 225 J/mole-K found by Makhatadze and Privalov [25], Wauchope and Haque [26], and Gill et al. [22], respectively. Using Equation (5-9), the temperature at which the minimum solubility of benzene in water occurs is 294 K, which is reasonably consistent with the value of 289.0 K reported by Gill et al. [22].

Water Solubility in Benzene

The solubility of water in benzene is shown in Figure 5-5. An abundance of solubility data for water in benzene exists in the literature at atmospheric pressures; however, Hefter [15] notes the more recent studies tend toward slightly higher solubilities than previous studies and with considerably smaller measurement uncertainty.

Correlation of the water solubility measurements employed Equation (5-10). As estimated by error propagation, the solubility measurements have a maximum uncertainty of 4.1% at a mole fraction of 0.083 (0.0034 absolute error) and an average uncertainty of 3.2%. The WRMS error of the solubility data is 0.61. The weighted and percent deviations are shown graphically in Figure 5-6 and Figure 5-7, respectively.

The solubility measurements from this study agree within 10% of most literature data. Particularly good agreement is noted between this study and the work conducted by Anderson and Prausnitz [2]. Over a mid-temperature range, deviations approaching 20% are observed with the recent data of Jou and Mather [16], Chandler et al. [17], and Chen and Wagner [18], but better agreement is seen as the temperature increases. At higher temperatures, large differences are seen with the work of Ratzlaff [19].

From Equation (5-11), the heat of solution is determined to be 21.9 kJ/mole, at 298.15 K, which is in good agreement with the value of 20.7 and 23.3 kJ/mole reported by De Lisi et al. [23] and Chen and Wagner [18]. This value supports the theory stated by Franks [14] that liquid water dissolving into a non-polar hydrocarbon liquid phase is essentially a process of breaking hydrogen bonds, which possess energy in the 21-29 kJ/mole range. Using Equation (5-12), the heat capacity of solution at 298.15 K is estimated to be 73.4 J/mole-K, which is consistent with the value of 78.3 J/mole-K at 298.15 K reported by Chen and Wagner [18].

Toluene Solubility in Water

An abundance of data exists at near ambient temperatures, but at temperatures greater than 325 K, the system has not been investigated as thoroughly, and the solubility measurements are presented in Figure 5-8.

Equation (5-3) was utilized to correlate the toluene solubility measurements. The maximum uncertainty of the solubility measurements, as estimated by error propagation, was 3.3% at a mole fraction of 0.00011 (0.0000035 absolute error) with an average uncertainty of 1.8%. The weighted-root-mean-square (WRMS) error of the solubility data is 0.68. Figure 5-9 and Figure 5-10 provide a graphical representation of the weighted and percent deviations, respectively.

At temperatures less than 400 K, the solubility measurements agree within 10% of the broad range of literature data. In the higher temperature range, the measurements are in reasonable agreement with the results of Miller and Hawthorne [20] and Anderson and

Prausnitz [2], but the work of Jou and Mather [16] and Chandler et al. [17] have large observable deviations.

The heat of solution at 298.15 K was determined using Equation (5-7). The 0.751 kJ/mole value obtained is in poor agreement with the values reported by Gill et al. [22] of 1.73 kJ/mole and De Lisi et al. [23] of 1.80 kJ/mole. The second derivative property, the specific heat of solution, calculated with Equation (5-8) is 324 J/mole-K. This value is in fair agreement with the values of 351 and 363 J/mole-K reported by Gill et al. [22] and Chen and Wagner [27], respectively, but the value of 461 J/mole-K given by Makhatadze and Privalov [25] is much higher. Using Equation (5-9), the temperature at which the minimum solubility of toluene in water occurs is 296 K, which is consistent with the values of 291.6 and 297.3 K reported by Gill et al. [22] and Chen and Wagner [27].

Water Solubility in Toluene

The solubility of water in toluene is shown in Figure 5-11, and an abundance of literature solubility data for water in toluene was observed at near ambient temperatures.

Equation (5-10) was used to correlate the water solubility measurements, which have a maximum uncertainty of 4.8% at a mole fraction of 0.022 (0.0011 absolute error) and an average uncertainty of 3.4%. The WRMS error of the solubility data is 0.36. The weighted and percent deviations are shown graphically in Figure 5-12 and Figure 5-13, respectively.

The solubility measurements from this study agree within 10% of almost all literature data over the entire temperature range with a few exceptions. These exceptions include the recent data of Anderson and Prausnitz [2], which approach deviations of

approximately 15% at higher temperatures, and Chandler et al. [17], which deviate by 30% at high temperatures.

From Equation (5-11), the heat of solution is determined to be 22.4 kJ/mole, at 298.15 K, which is consistent with the values of 30.9 and 23.9 kJ/mole reported by De Lisi et al. [23] and Chen and Wagner [27], respectively. Again, this value supports the theory [14] that liquid water dissolving into a non-polar hydrocarbon liquid phase is essentially a process of breaking hydrogen bonds. Using Equation (5-12), the heat capacity of solution at 298.15 K is estimated to be 75.1 J/mole-K, with no available literature for comparison.

3-Methylpentane Solubility in Water

3-Methylpentane solubility measurements are presented in Figure 5-14. Limited data exist at near ambient temperatures, and the system has not been investigated at temperatures greater than 300 K.

Correlation of the 3-methylpentane solubility measurements employed Equation (5-3), however the measurement taken at 491.9 K was not included in the determination of the equation parameters due to a weighted deviation greater than 2.5 times the standard deviation, and these weighted deviations are shown graphically in Figure 5-15. The solubility measurements have a maximum uncertainty of 3.7% at a mole fraction of 0.0000026 (0.000000097 absolute error) and an average uncertainty of 2.8%, which is shown in Figure 5-16. The weighted-root-mean-square (WRMS) error of the solubility data is 0.52.

At a temperature of 300 K, the solubility measurement agrees within 3% of the measurements reported by Rudakov and Lutsyk [28], McAuliffe [29], and Price [30], but the work of Polak and Lu [31] shows deviations approaching 40%. However, according to the critical data review by Hefter [15], other hydrocarbon solubility measurements made by Polak and Lu tend to be approximately 30% higher than other reported values.

From Equation (5-7), the heat of solution at 298.15 K is -0.407 kJ/mole, and the second derivative property, the specific heat of solution, calculated with Equation (5-8) is 411 J/mole-K. Using Equation (5-9), the temperature at which the minimum solubility of 3-methylpentane in water occurs is 299 K. Literature data are unavailable for comparison.

Water Solubility in 3-Methylpentane

Limited data exist at near-ambient temperatures, and the system has not been investigated at temperatures greater than 300 K. The solubility of water in toluene is shown in Figure 5-17.

Correlation of the water solubility measurements employed the use of Equation (5-10); however, the measurements taken at 350.9 and 432.4 K were not included in the determination of the equation parameters due to a weighted deviation greater than 2.5 times the standard deviation, and these weighted deviations are shown graphically in Figure 5-18. The solubility measurements have a maximum uncertainty of 8.4% at a mole fraction of 0.0052 (0.00044 absolute error) and an average uncertainty of 5.0%, which is presented in Figure 5-19. The WRMS error of the solubility data is 0.10.

The solubility measurements from this study show reasonable agreement with the data reported by Polak and Lu [31]. As shown by Hefter [15], a decided lack of accuracy exists in Polak and Lu's hydrocarbon solubility values, but their water solubility data are more consistent with those of other researchers.

From Equation (5-11), the heat of solution is determined to be 29.3 kJ/mole, at 298.15 K, which falls within the range of hydrogen bond energy. Using Equation (5-12), the heat capacity of solution at 298.15 K is estimated to be 98.2 J/mole-K, with no available literature for comparison.

5.5 Conclusions

A continuous flow apparatus was utilized to measure mutual solubilities at temperatures ranging from ambient to 500 K, which is near the three-phase critical end point of the benzene-water and toluene-water systems. The well-documented system, benzene-water, was used to validate the proper operation of the apparatus, including the sampling and analytical techniques employed. Generally, adequate agreement was observed for the benzene-water, toluene-water, and 3-methylpentane-water systems with literature data. A propagated error analysis of the three systems studied calculated the maximum uncertainty as 4% at a mole fraction of 0.0027 (0.00011 absolute error) and 8% at a mole fraction of 0.0052 (0.00044 absolute error) in the water phase and organic phase measurements, respectively.

Enthalpies of solution for the hydrocarbon solubility in water estimated from experimental measurements are in reasonable agreement with available calorimetric measurements from the literature. The enthalpies of solution for the water solubility in

the hydrocarbons were within the range of the hydrogen bonding energies and were consistent with available literature data.

5.6 Nomenclature

α, β	constants in calibration correlations
WR	weight ratio
SSR	solvent-to-sample weight ratio
AR	area ratio
MW	molecular weight
A, B, C	constants in solubility correlations
x	mole fraction
T	temperature (K)
P	pressure
H	enthalpy
R	ideal gas constant
$\Delta\bar{H}_i$	heat of solution (kJ/mole)
$\Delta\bar{C}_p$	specific heat of solution (J/mole-K)
WRMS	weighted-root-mean-square
\hat{y}_i	measured data value
y_i	predicted data value
σ_i	error estimate
n	number of data points

Subscripts

hc	hydrocarbon
c	critical property
w	water
r	reduced
soln	solution
1	component
2	component
i	component

5.7 *Literature Cited*

1. Tsonopoulos, C.; Wilson, G. M., *AIChE J.* **1983**, 29, 990-9.
2. Anderson, F. E.; Prausnitz, J. M., *Fluid Phase Equilib.* **1986**, 32, 63-76.
3. Shaw, R. W.; Brill, T. B.; Clifford, A. A.; Eckert, C. A.; Franck, E. U., *Chem. Eng. News* **1991**, 69, 26-39.
4. Katritzky, A. R.; Allin, S. M.; Siskin, M., *Acc. Chem. Res.* **1996**, 29, 399-406.
5. Neely, B. J.; Ratzlaff, D. W.; Wagner, J.; Robinson, J., R.L.; Gasem, K. A. M., *Integrated Petroleum Environmental Consortium*, Albuquerque, NM, 2000.
6. Roof, J. G., *J. Chem. Eng. Data* **1970**, 15, 301-3.
7. Marche, C.; Delepine, H.; Ferronato, C.; Jose, J., *J. Chem. Eng. Data* **2003**, 48, 398-401.
8. Chandler, J. P.; Jackson, L. W., *Computer Program* **1981**, Stillwater, OK, Department of Computing and Information Sciences, Oklahoma State University.
9. Neely, B. J., Ph.D. Dissertation, Oklahoma State University: Stillwater, 2006.
10. Benson, B. B.; Krause, D., Jr., *J. Chem. Phys.* **1976**, 64, 689-709.
11. Wilhem, E.; Batino, R.; Wilcock, R., *J. Chem. Rev.* **1977**, 77, 219.
12. Smith, J. M.; Van Ness, H. C.; Abbott, M. M., *Introduction to Chemical Engineering Thermodynamics*, 5; McGraw-Hill: New York, NY, 1996.
13. Wagner, J., *GPSA Engineering Data Book, Revitalization and Maintenance, Water-Hydrocarbon Mutual Solubility Data*, Presented to Gas Processors Association, Tulsa, OK, 1999.
14. Franks, F., *Nature* **1966**, 210, 87-8.
15. Hefter, G. T., 37, D. G. Shaw; Pergamon Press, Oxford, UK, 1989.
16. Jou, F.-Y.; Mather, A. E., *J. Chem. Eng. Data* **2003**, 48, 750-2.
17. Chandler, K.; Eason, B.; Liotta, C. L.; Eckert, C. A., *Ind. Eng. Chem. Res.* **1998**, 37, 3515-8.
18. Chen, H. P.; Wagner, J., *J. Chem. Eng. Data* **1994**, 39, 475-9.
19. Ratzlaff, D. W., M.S. Thesis, Oklahoma State University: Stillwater, OK, 1999.
20. Miller, D. J.; Hawthorne, S. B., *J. Chem. Eng. Data* **2000**, 45, 78-81.
21. Reid, D. S.; Quickenden, M. A. J.; Franks, F., *Nature* **1969**, 224, 1293-4.
22. Gill, S. J.; Nichols, N. F.; Wadso, I., *J. Chem. Thermodynamics* **1976**, 8, 445-52.
23. De Lisi, R.; Goffredi, M.; Turco Liveri, V., *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases* **1980**, 76, 1660-2.

24. Clarke, E. C. W.; Glew, D. N., *Trans Faraday Soc* **1966**, 62,
25. Makhataдзе, G. I.; Privalov, P. L., *J. Chem Thermo* **1988**, 20, 405-12.
26. Wauchope, R. D.; Haque, R., *Canadian J. of Chem* **1972**, 50, 133-8.
27. Chen, H. P.; Wagner, J., *J. Chem. Eng. Data* **1994**, 39, 679-84.
28. Rudakov, E. S.; Lutsyk, A. I., *Zh. Fiz. Khim.* **1979**, 53, 1298-300.
29. McAuliffe, C., *J. Phys. Chem.* **1966**, 70, 1267-75.
30. Price, L. C., *Am. Assoc. Petrol. Geol. Bull.* **1976**, 60, 213-44.
31. Polak, J.; Lu, B. C. Y., *Can. J. Chem.* **1973**, 51, 4018-23.
32. Guseva, A. N.; Parnov, E. I., *Vestn. Mosk. Khim.* **1963**, 18, 76-9.
33. Connolly, J. F., *J. Chem. Eng. Data* **1966**, 11, 13-6.
34. Brown, R. L.; Wasik, S. P., *J. Res. Natl. Bur. Stds. A.* **1974**, 78, 453-60.
35. Mackay, D.; Shiu, W. Y., *Can. J. Chem. Eng.* **1975**, 53, 239-41.
36. May, W. E.; Wasik, S. P.; Freeman, D. H., *Anal. Chem.* **1978**, 50, 997-1000.
37. Bittrich, H. J.; Gedan, H.; Feix, G., *Z. Phys. Chem. Leipzig* **1979**, 260, 1009-13.
38. Banerjee, S.; Yalkowsky, S. H.; Valvani, S., *Environ. Sci. Technol.* **1980**, 14, 1227-9.
39. Umamo, S.; Hayano, I., *Kogyo Kagaku Zasshi* **1957**, 60, 1436-7.
40. Guerrant, R. P., M.S. Thesis, Pennsylvania State University: University Park, PA, 1964.
41. Moule, D. C.; Thurston, W. M., *Can. J. Chem.* **1966**, 44, 1361-7.
42. Roddy, J. W.; Coleman, C. F., *Talanta* **1968**, 15, 1281-6.
43. Karlsson, R. J., *Chem. Eng. Data* **1973**, 18, 290-2.
44. Goldman, S., *Can. J. Chem.* **1974**, 52, 1668-80.
45. Kirchnerova, J.; Cave, G. C. B., *Can. J. Chem.* **1976**, 54, 3909-16.
46. Singh, R. P.; Sah, R., *Indian J. Chem.* **1978**, 16A, 692-4.
47. Pierotti, R. A.; Liabastre, A. A., Presented to U. S. Nat. Tech. Inform. Serv., 1972.
48. Sanemasa, I.; Araki, M.; Deguchi, T.; Nagai, H., *Chem. Lett.* **1981**, 225-8.
49. Korenman, I. M.; Aref'eva, R. P., Determination of the solubility of liquid hydrocarbons in water, U.S. Patent No. 553524, 1977.
50. Sanemasa, I.; Arakawa, S.; Piao, C. Y.; Deguchi, T., *Bull. Chem. Soc. Jpn.* **1985**, 58, 3033-4.
51. Andrews, L. J.; Keefer, R. M., *J. Am. Chem. Soc.* **1949**, 71, 3644-7.
52. Klevens, H. B., *J. Am. Chem. Soc.* **1950**, 72, 3780-5.

53. Morrison, T. J.; Billett, F., *Journal of the Chemical Society, Abstracts* **1952**, 3819-22.
54. McAuliffe, C., *Nature* **1963**, 200, 1092-3.
55. Sutton, C.; Calder, J. A., *J. Chem. Eng. Data* **1975**, 20, 320-2.
56. Rossi, S. S.; Thomas, W. H., *Environ. Sci. Technol.* **1981**, 15, 715-16.
57. Gross, P. M.; Saylor, J. H., *J. Am. Chem. Soc.* **1931**, 53, 1744-51.
58. Bohon, R. L.; Claussen, W. F., *J. Am. Chem. Soc.* **1951**, 73, 1571-8.
59. Englin, B. A.; Plate, A. F.; Tugolukov, V. M.; Pryanishnikova, M. A., *Khim. Tekhnol. Topl. Masel* **1965**, 10, 42-6.
60. Uspenskii, S. P., *Neftyanoe Khozyaistvo* **1929**, 17, 713-7.
61. Rosenbaum, C. K.; Walton, J. H., *J. Am. Chem. Soc.* **1930**, 52, 3568-73.
62. Caddock, B. D.; Davies, P. L., *Journal of the Institute of Petroleum* **1960**, 46, 391-6.
63. Glasoe, P. K.; Schultz, S. D., *J. Chem. Eng. Data* **1972**, 17, 66-8.
64. Hogfeldt, E.; Bolander, B., *Acta. Chem. Scand.* **1964**, 18, 548-52.
65. Johnson, J. R.; Christian, S. D.; Affsprung, H. E., *J. Chem. Soc., Inorg., Phys., Theoret.* **1966**, 77-8.
66. Tarassenkow, D. N.; Poloshinzewa, E. N., *Zh. Obshch. Khim.* **1931**, 1, 71-9.
67. Wing, J.; Johnston, W. H., *J. Am. Chem. Soc.* **1957**, 79, 864-5.
68. Jones, J. R.; Monk, C. B., *Journal of the Chemical Society, Abstracts* **1963**, 2633-5.
69. Tarassenkow, D. N.; Poloshinzewa, E. N., *Chem. Ber.* **1932**, 65, 184-6.

Table 5-1. Mutual Solubilities for Hydrocarbon-Water Systems with Error Estimates

Aqueous Phase					Organic Phase				
Benzene-Water System									
T	P	x_{benzene}	Absolute Error	% Error	T	P	x_{water}	Absolute Error	% Error
(K)	(MPa)	(-10^4)	(-10^4)		(K)	(MPa)	(-10^2)	(-10^2)	
299.1	0.194	4.13	0.057	1.3	299.0	0.204	0.32	0.012	3.8
324.3	0.139	4.68	0.034	0.73	324.3	0.139	0.64	0.020	3.2
350.2	0.271	6.61	0.055	0.83	350.2	0.271	1.26	0.046	3.7
376.2	0.443	10.0	0.01	1.0	376.2	0.443	2.17	0.067	3.1
400.3	0.638	15.2	0.099	0.65	400.3	0.638	3.83	0.10	2.6
431.4	1.944	27.2	1.1	4.0	431.4	1.944	8.33	0.34	4.1
461.8	3.426	48.5	1.7	3.5	461.8	3.426	14.4	0.48	3.3
490.8	6.873	66.0*	0.73	1.1	490.8	6.873	25.5	0.051	2.0
Toluene-Water System									
T	P	x_{toluene}	Absolute Error	% Error	T	P	x_{water}	Absolute Error	% Error
(K)	(MPa)	(-10^4)	(-10^4)		(K)	(MPa)	(-10^2)	(-10^2)	
297.8	0.115	1.07	0.035	3.3	298.5	0.112	0.3	0.0081	2.8
324	0.199	1.31	0.0092	0.70	324.3	0.201	0.55	0.023	4.1
350.6	0.197	1.78	0.014	0.77	350.9	0.358	1.15	0.052	4.5
376.1	0.167	2.77	0.018	0.66	376.6	0.478	2.22	0.11	4.8
401.6	0.792	4.49	0.084	1.9	401.5	0.716	4.11	0.18	4.3
431.9	1.261	7.46	0.22	2.9	431.7	1.45	7.79	0.23	2.9
461	2.04	14.5	0.60	2.4	461.8	3.076	15.4	0.35	2.3
490.4	3.8	25.3	0.57	2.4	491.4	4.733	25.2	0.47	1.8
3-Methylpentane-Water System									
T	P	$x_{3\text{-MP}}$	Absolute Error	% Error	T	P	x_{water}	Absolute Error	% Error
(K)	(MPa)	(-10^4)	(-10^4)		(K)	(MPa)	(-10^2)	(-10^2)	
298.3	0.188	0.026	0.00097	3.7	295.5	0.168	0.031	0.0017	5.6
324.5	0.275	0.031	0.0010	3.2	299.1	0.194	0.038	0.0027	7.0
351.2	0.402	0.047	0.0010	2.2	324.3	0.139	0.11	0.0079	7.4
377.2	0.478	0.077	0.00091	1.2	350.9	0.486	0.29*	0.0077	2.7
401.2	0.709	0.14	0.0044	3.2	376.4	0.575	0.52	0.044	8.4
432.2	1.924	0.26	0.0081	3.1	400.3	0.638	1.17	0.020	1.7
462.1	3.352	0.6	0.018	3.0	432.4	2.144	2.51*	0.0935	3.7
491.9	7.248	2.18*	0.061	2.8	491.5	5.86	13	0.43	3.3

* Value was not used in the regression of the solubility parameters since the weighted deviation was greater than 2.5 times the standard deviation

Table 5-2. WRMS for Hydrocarbon-Water Systems

SYSTEM	WRMS	
	Aqueous Phase	Organic Phase
Benzene-Water	0.72	0.61
Toluene-Water	0.68	0.36
3-Methylpentane- Water	0.52	0.10

Table 5-3. Derivative Properties for Hydrocarbon Solubilities

Solute	Parameters in Equation (5-3)			ΔH	ΔC_p	T_{min}
	A	B	C	KJ/mol at 298.15 K	J/mol-K at 298.15 K	K
Benzene	11.09	-19.79	5.176	1.25	306	294
Toluene	10.71	-19.81	4.951	0.75	324	296
3-Methylpentane	11.74	-29.14	8.641	-0.41	411	299

Table 5-4. Derivative Properties for Water Solubilities

Solute	Parameters in Equation (5-10)		ΔH	ΔC_p
	A	B	KJ/mol at 298.15 K	J/mol-K at 298.15 K
Benzene	1.029	8.824	21.9	73.4
Toluene	1.122	9.035	22.4	75.1
3-Methylpentane	1.215	11.81	29.3	98.2

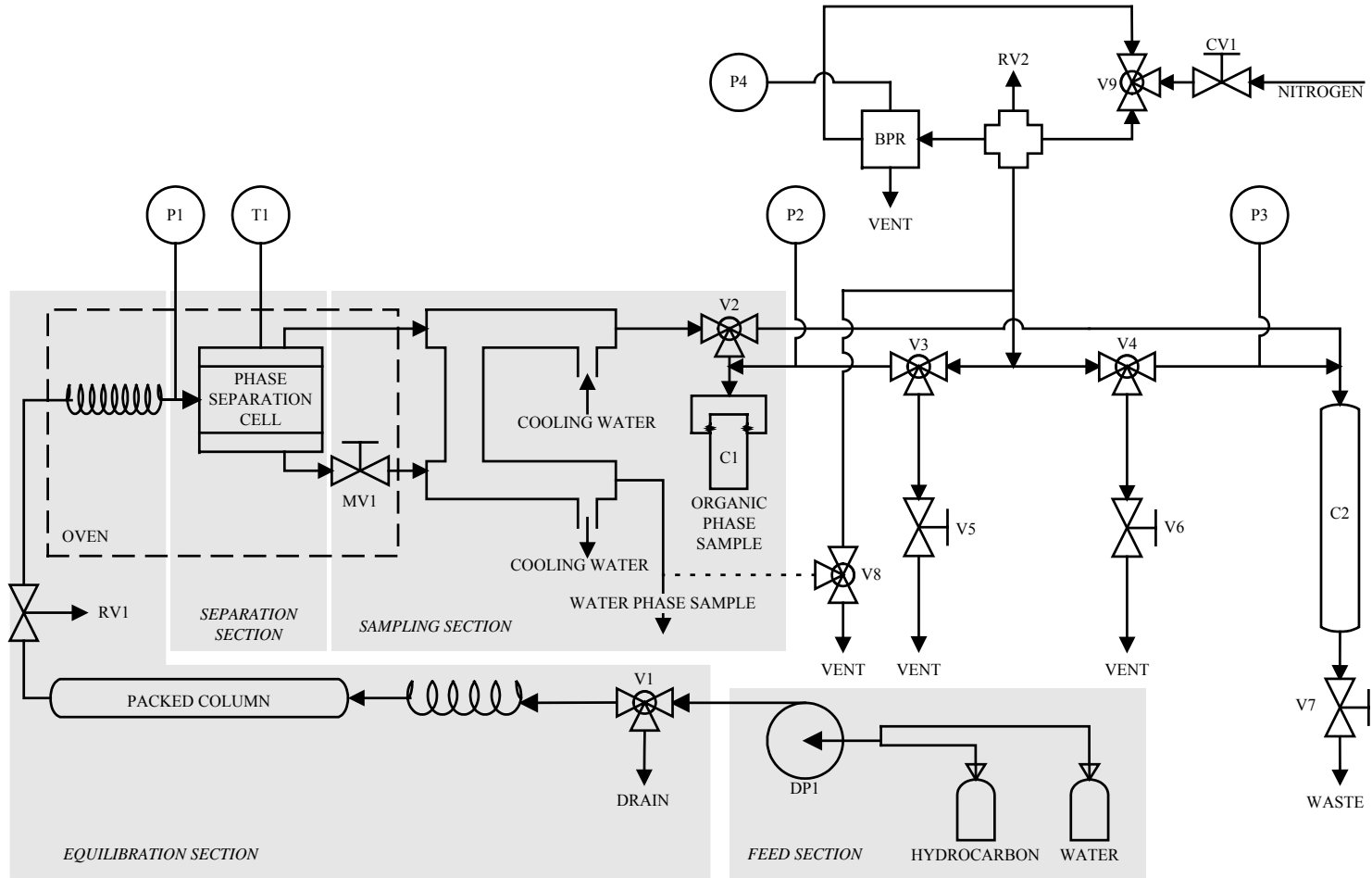


Figure 5-1. Schematic Diagram of the Continuous Flow Apparatus

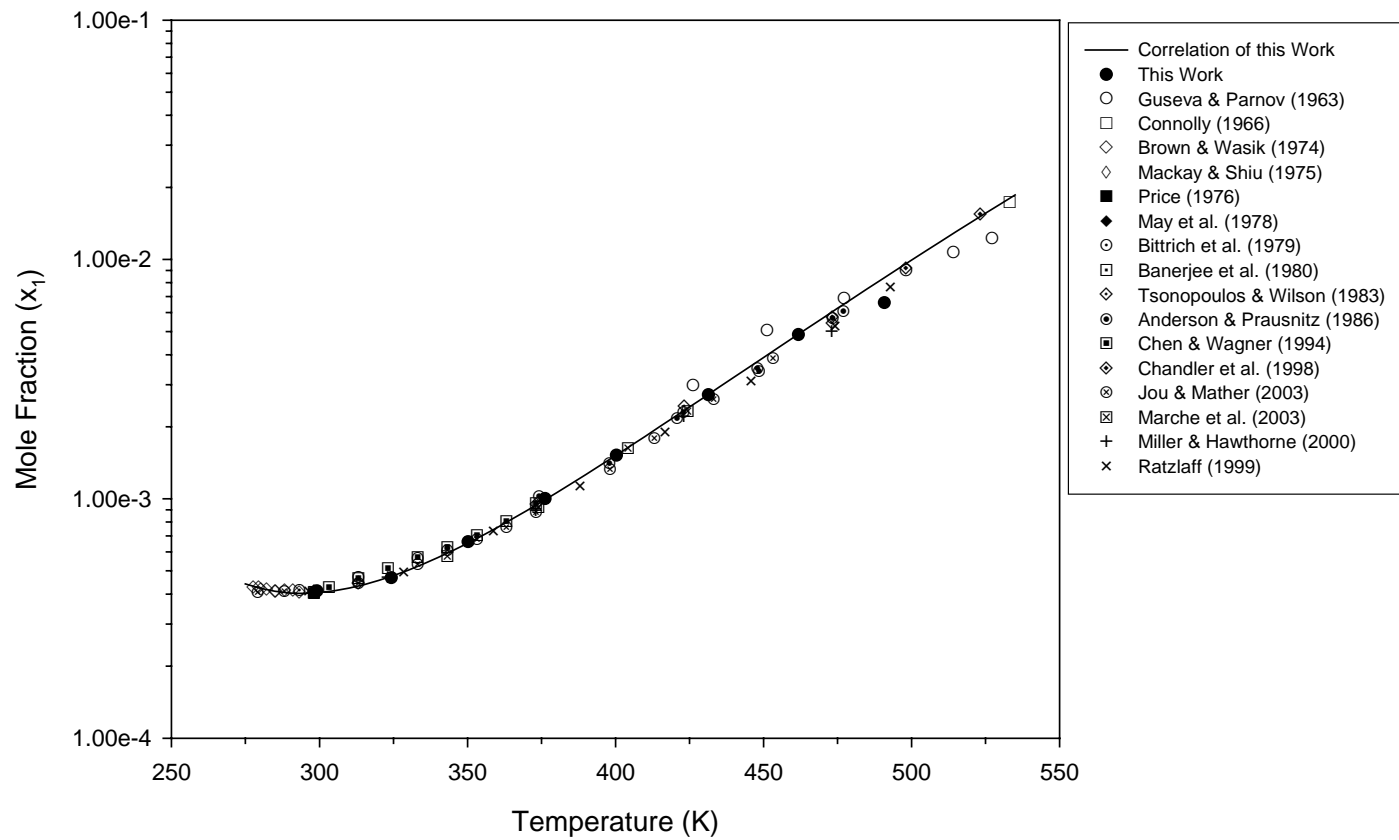


Figure 5-2. Solubility of Benzene in Water

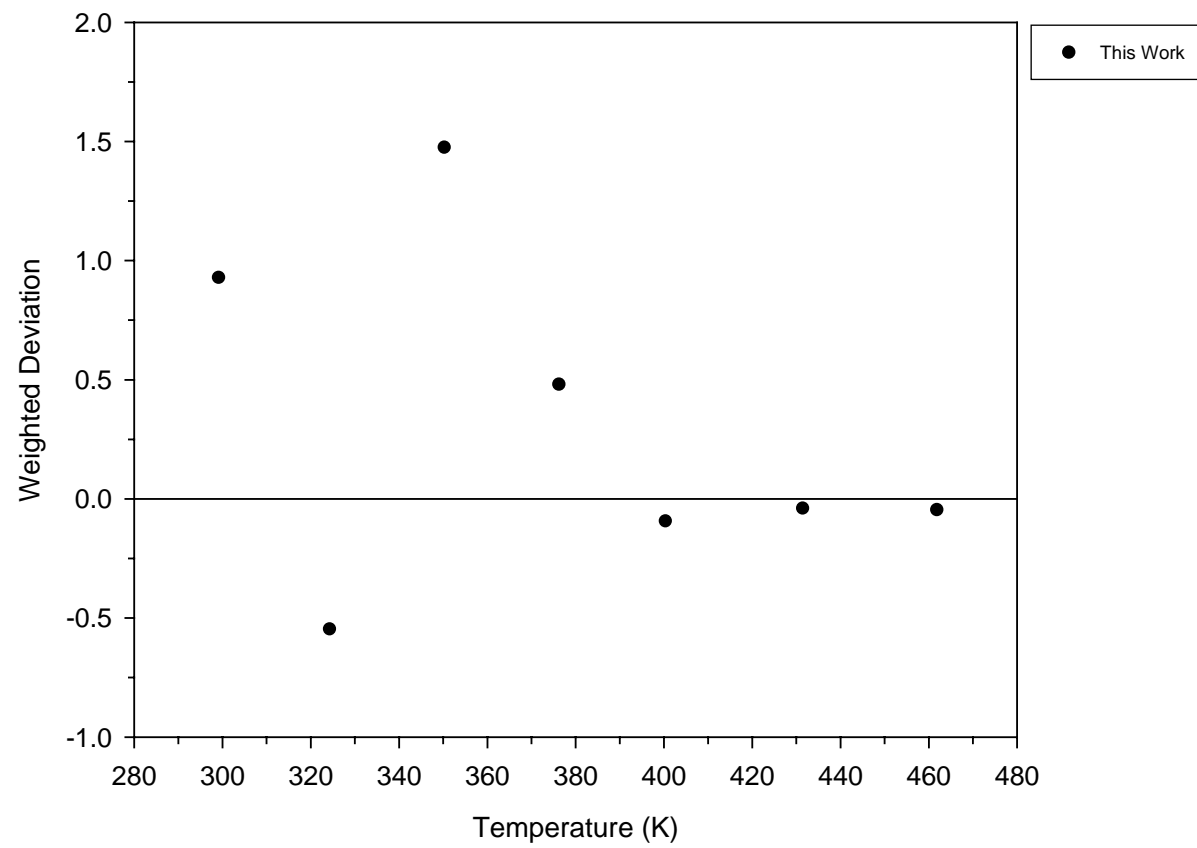


Figure 5-3. Weighted Deviation in the Solubility of Benzene in Water

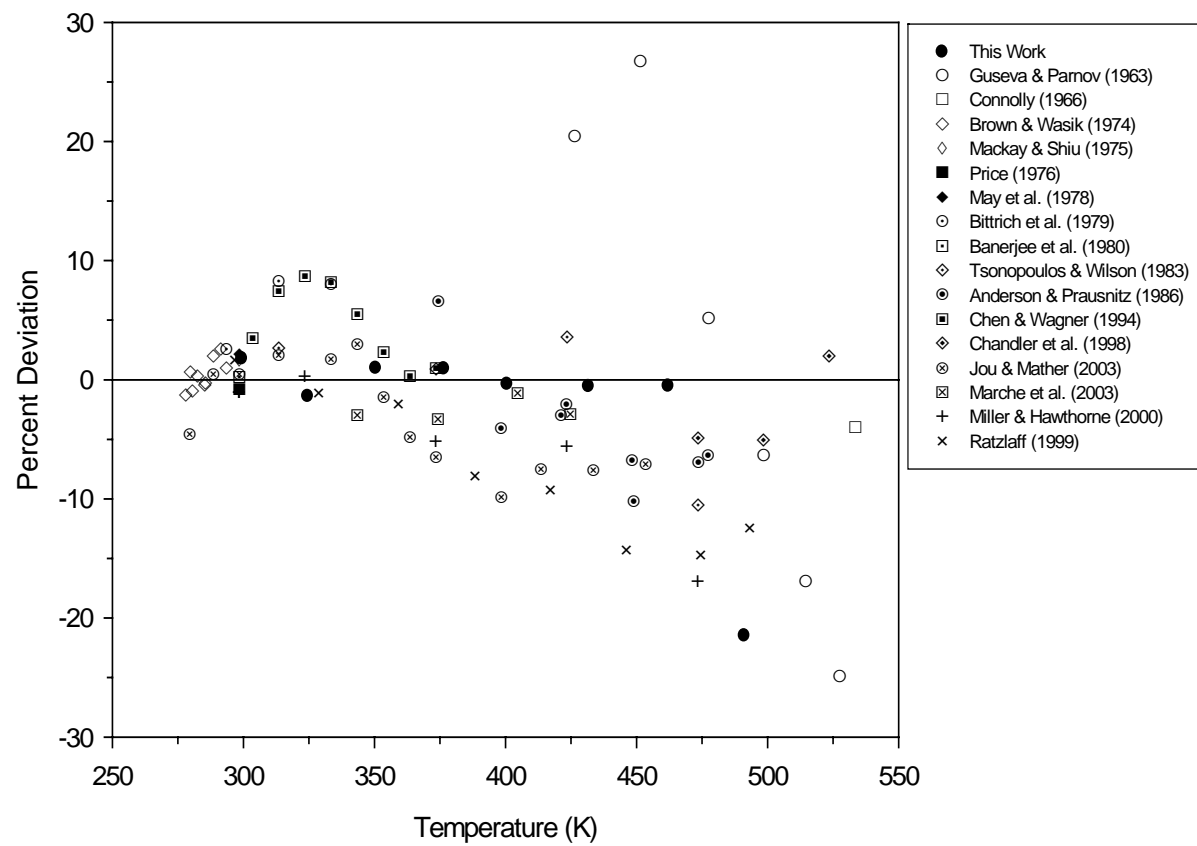


Figure 5-4. Percent Deviation in the Solubility of Benzene in Water

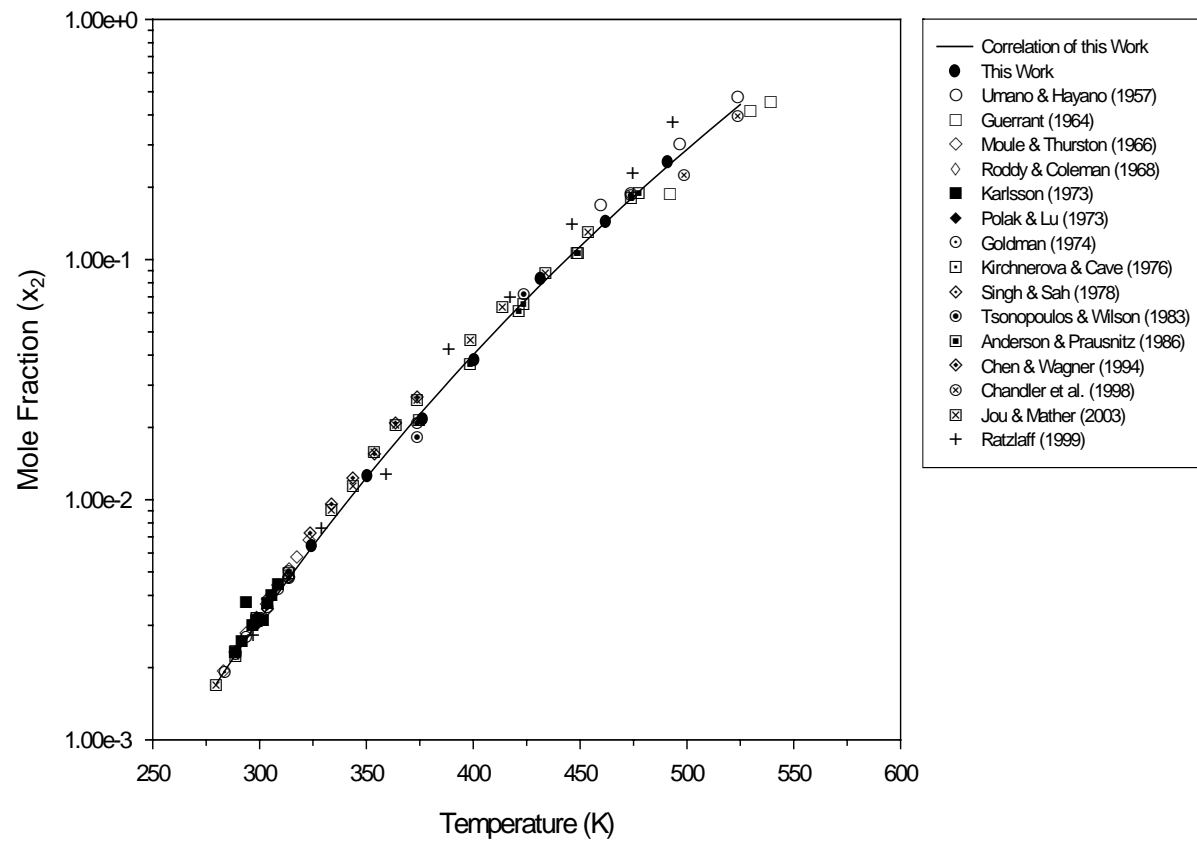


Figure 5-5. Solubility of Water in Benzene

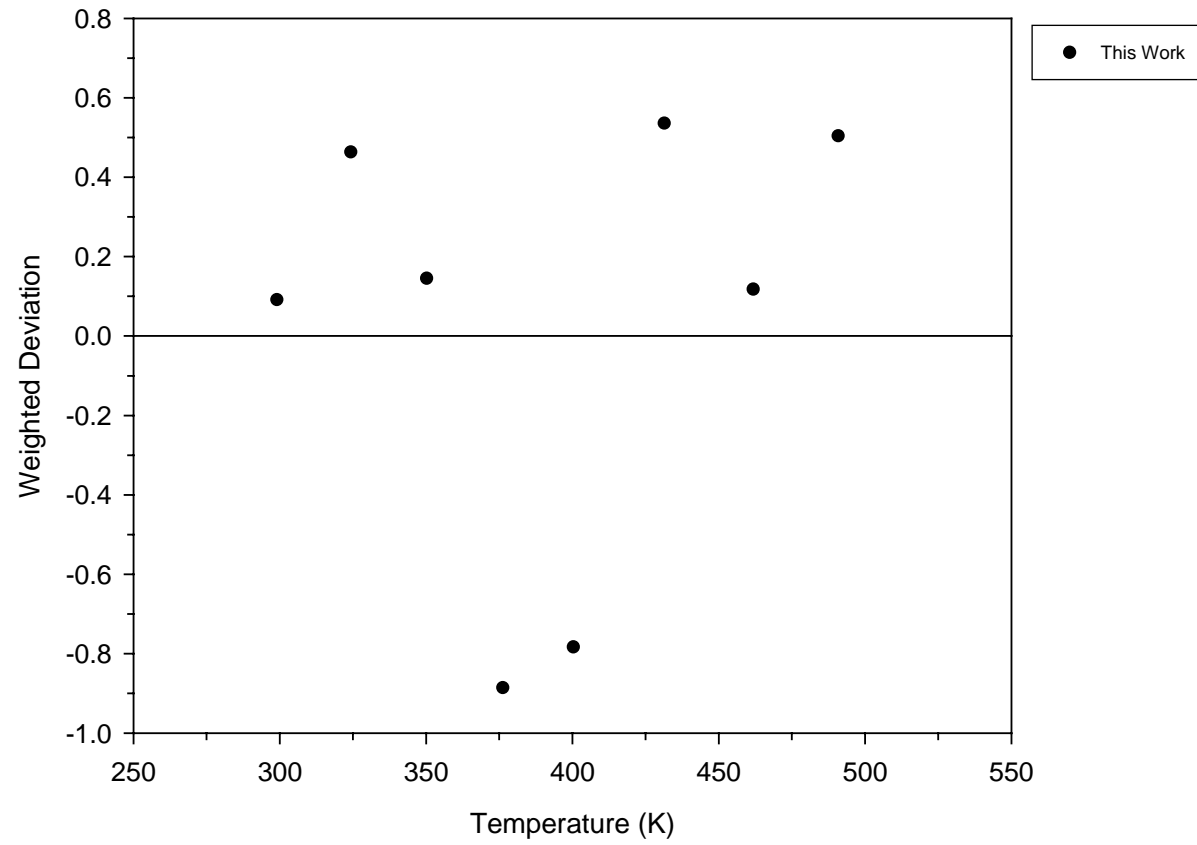


Figure 5-6. Weighted Deviation in the Solubility of Water in Benzene

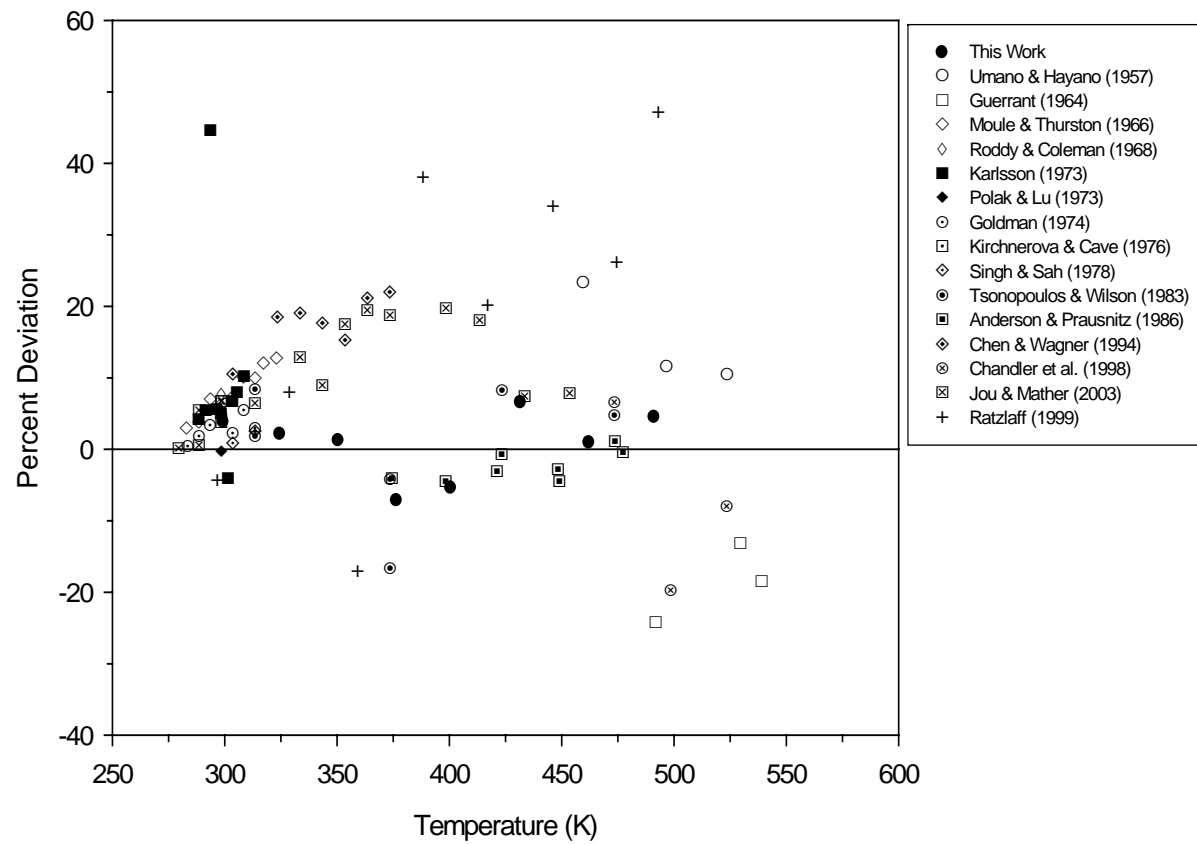


Figure 5-7. Percent Deviation in the Solubility of Water in Benzene

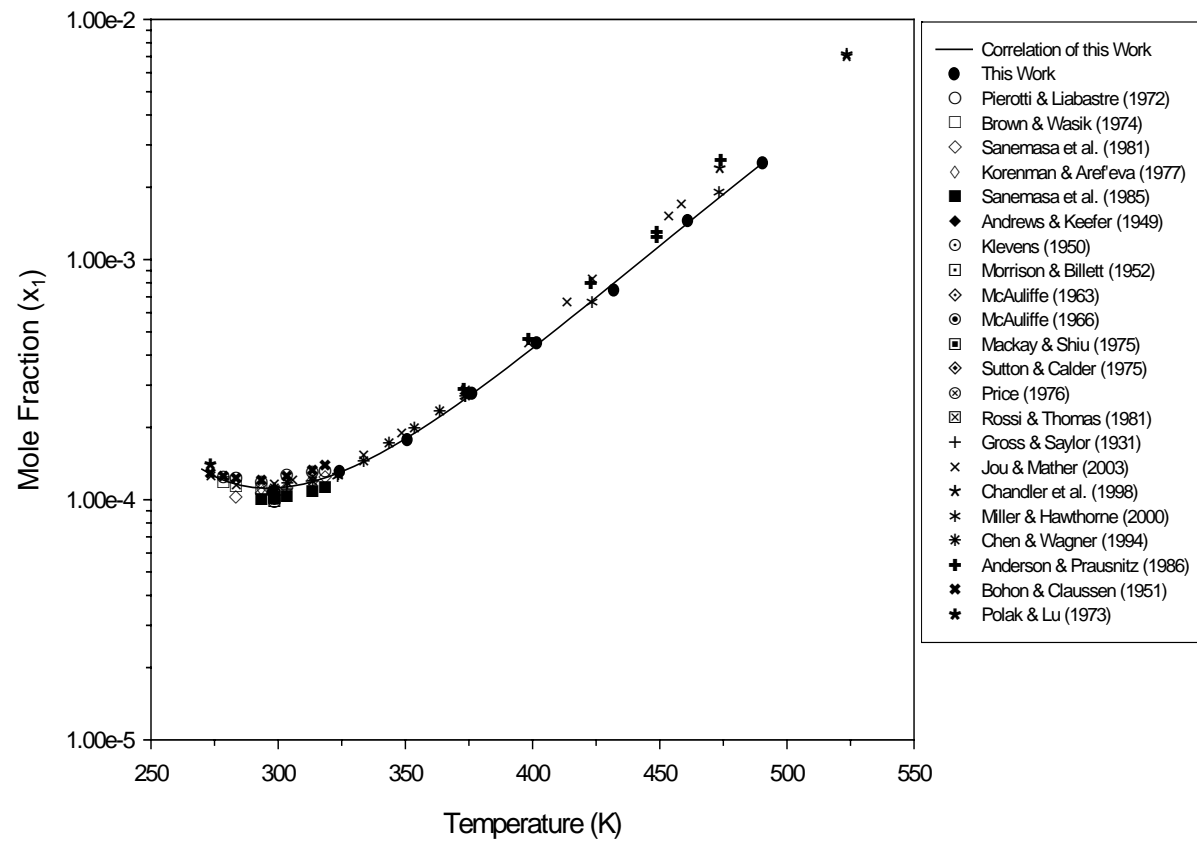


Figure 5-8. Solubility of Toluene in Water

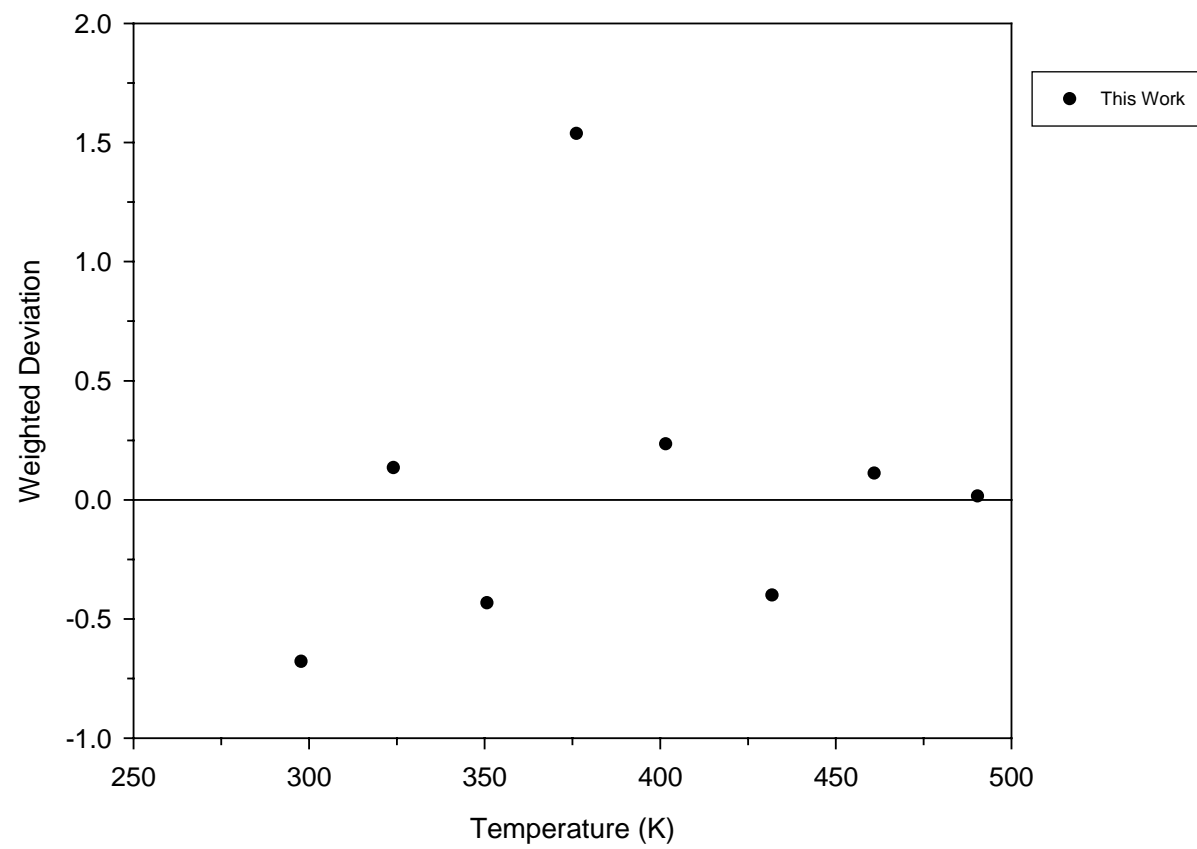


Figure 5-9. Weighted Deviation in the Solubility of Toluene in Water

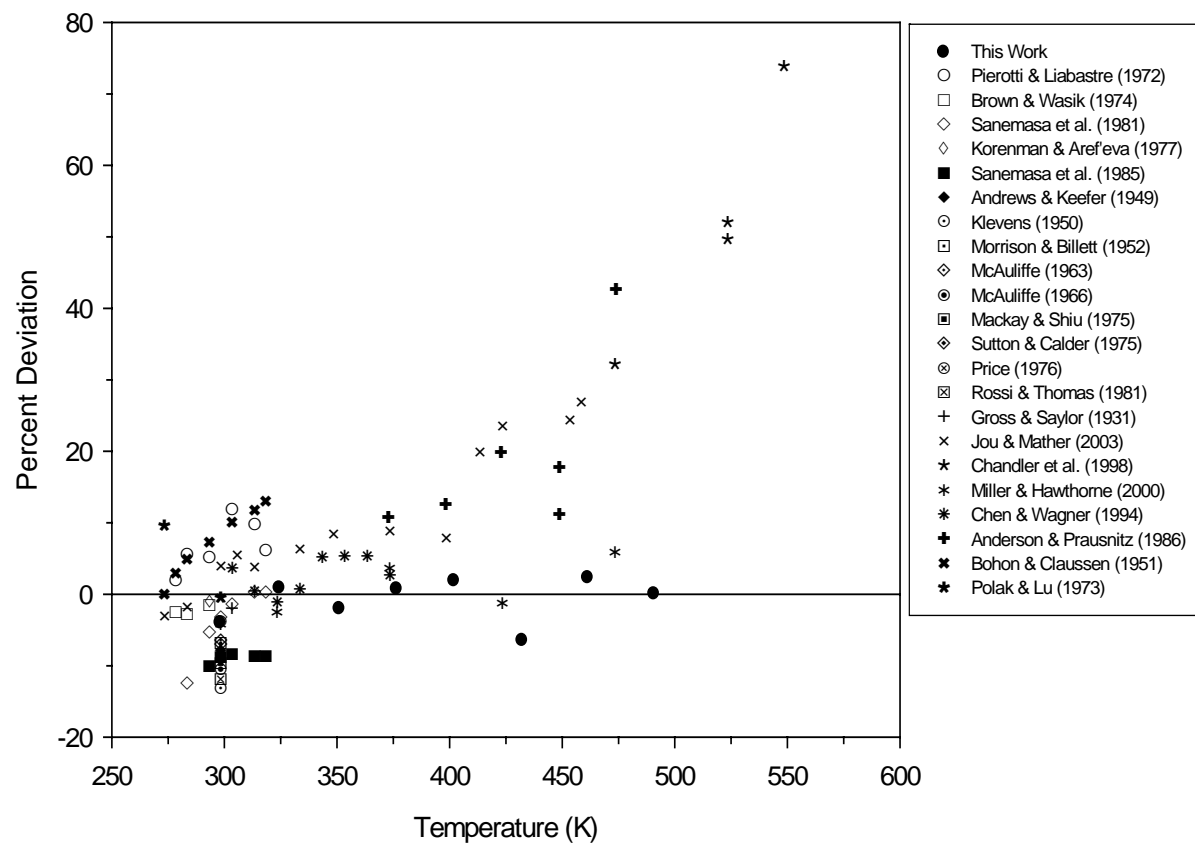


Figure 5-10. Percent Deviation in the Solubility of Toluene in Water

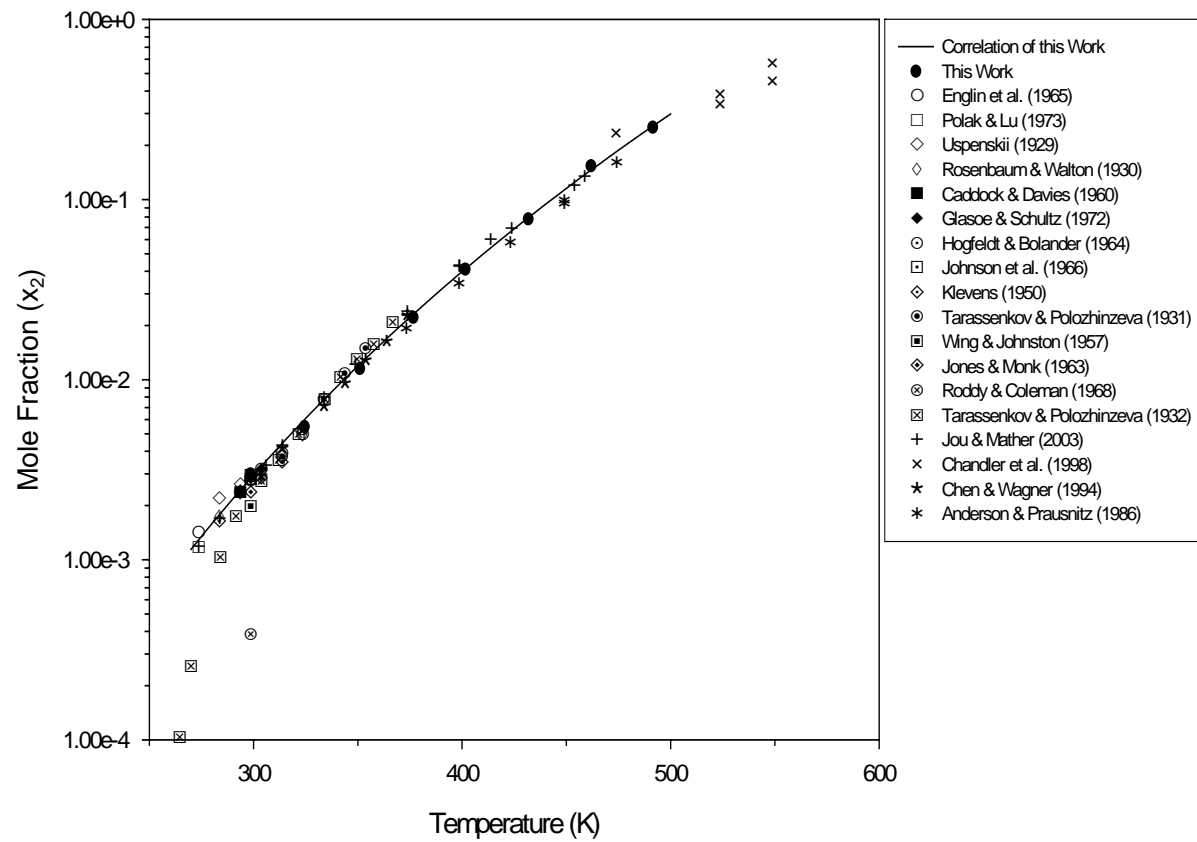


Figure 5-11. Solubility of Water in Toluene

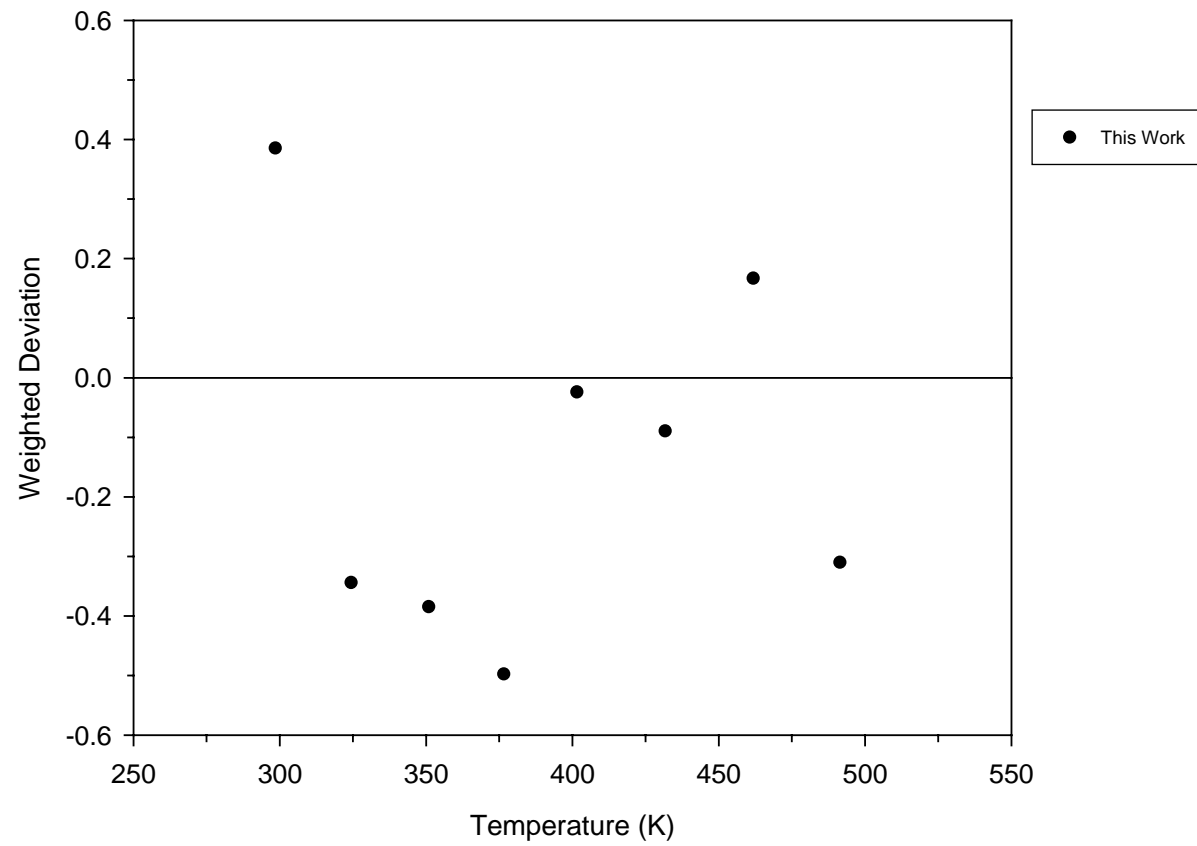


Figure 5-12. Weighted Deviation in the Solubility of Water in Toluene

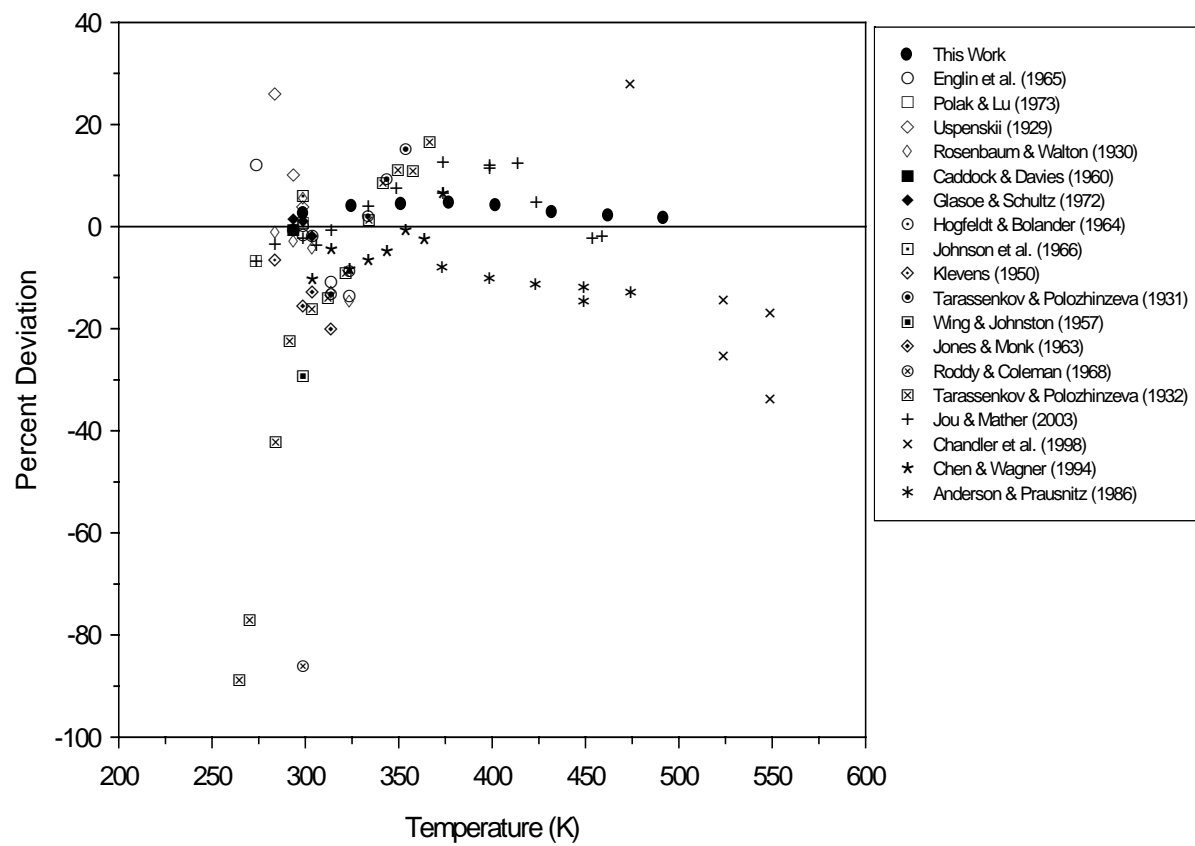


Figure 5-13. Percent Deviation in the Solubility of Water in Toluene

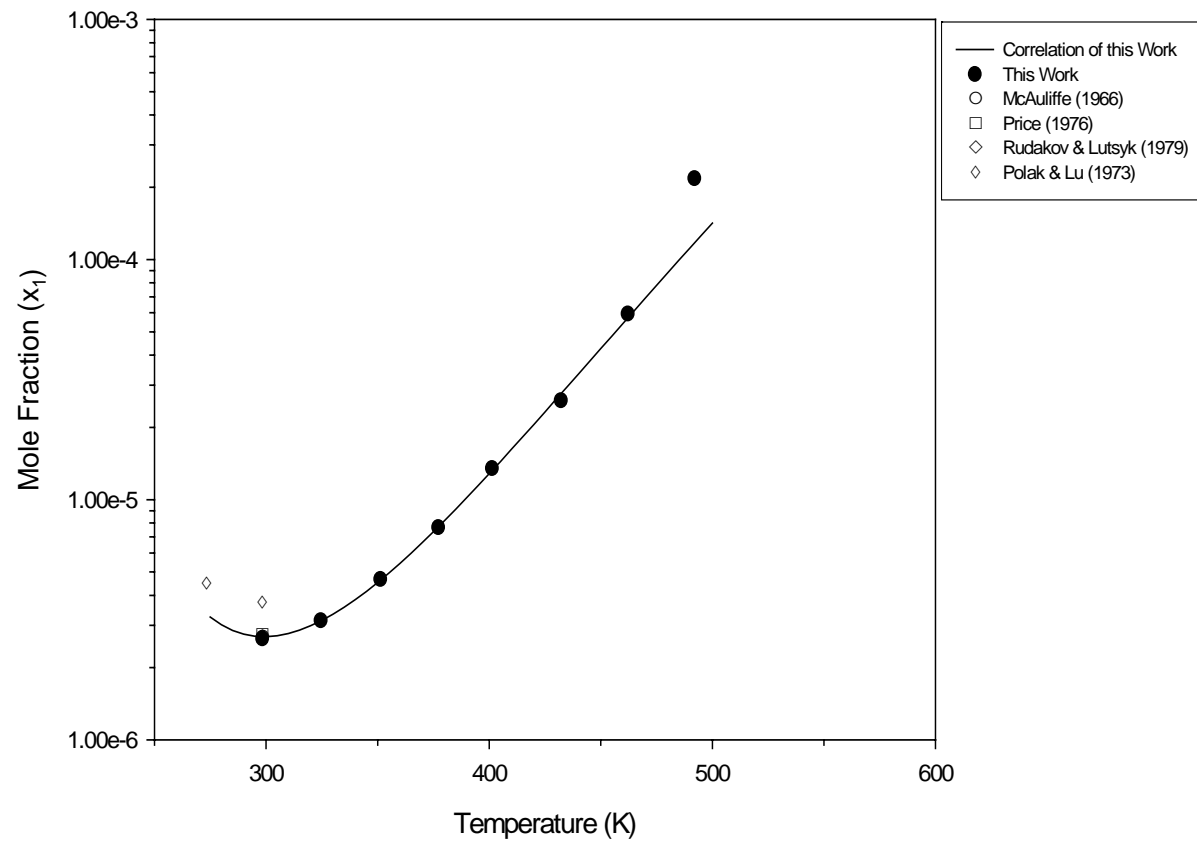


Figure 5-14. Solubility of 3-Methylpentane in Water

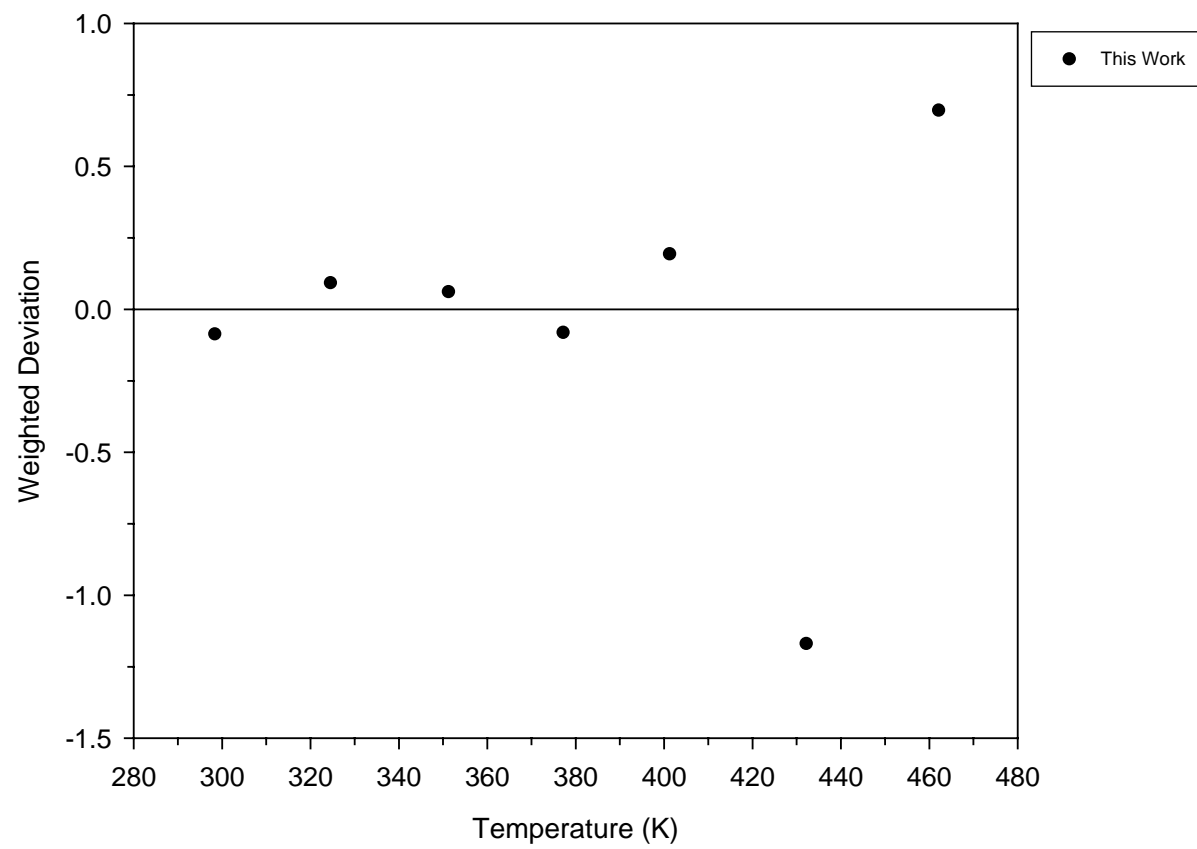


Figure 5-15. Weighted Deviation in the Solubility of 3-Methylpentane in Water

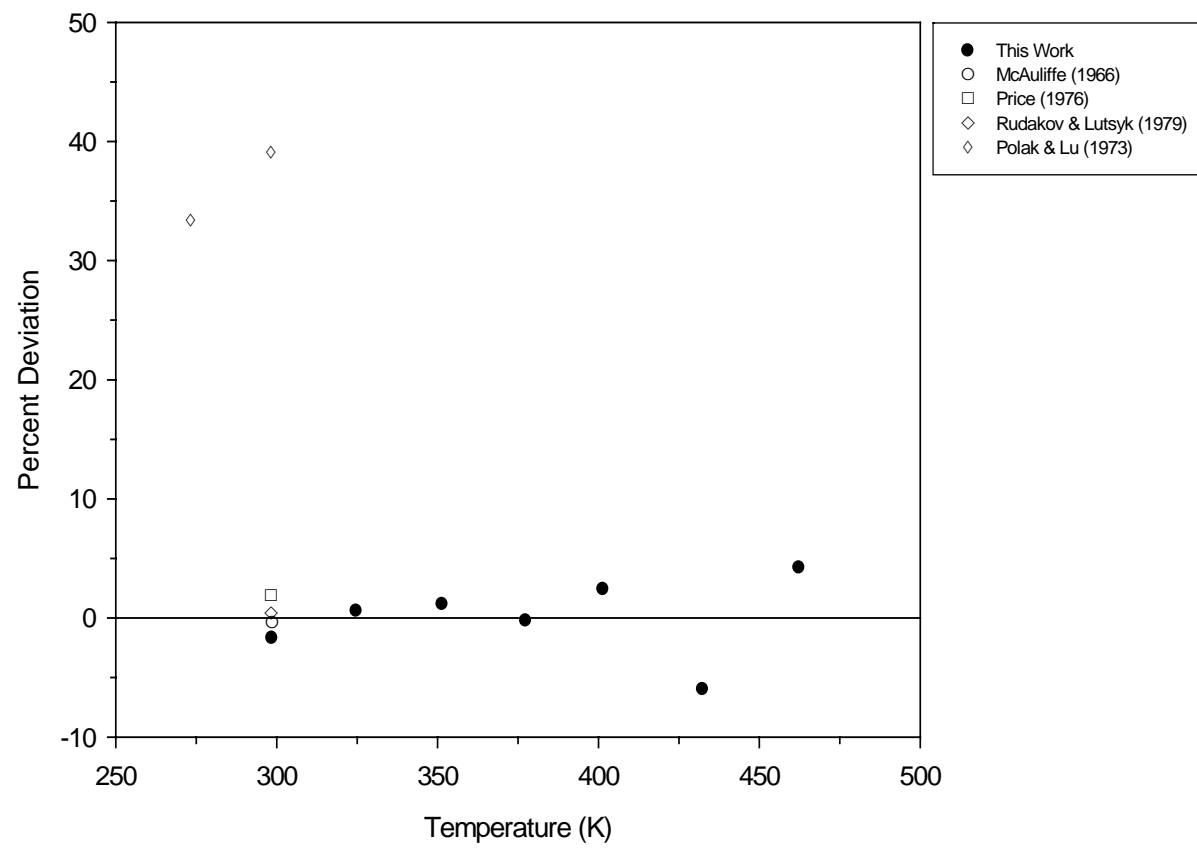


Figure 5-16. Percent Deviation in the Solubility of 3-Methylpentane in Water

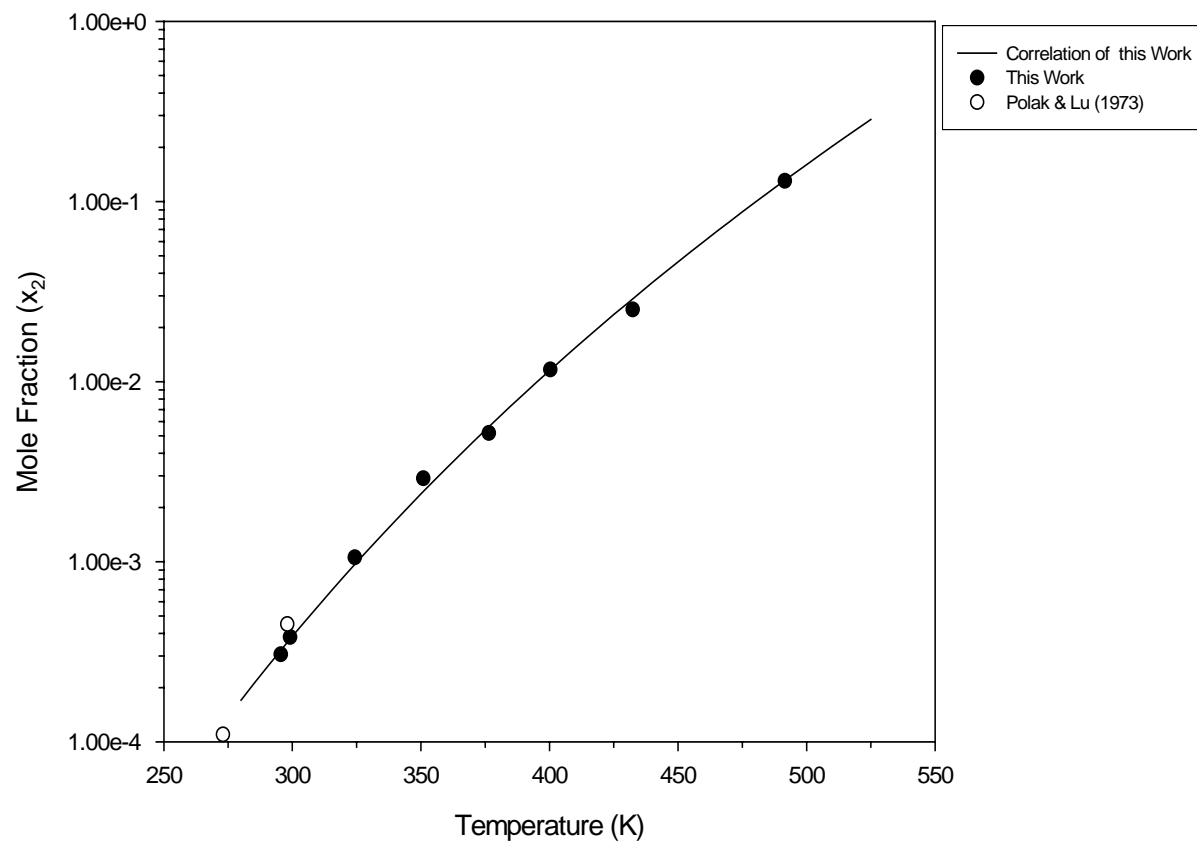


Figure 5-17. Solubility of Water in 3-Methylpentane

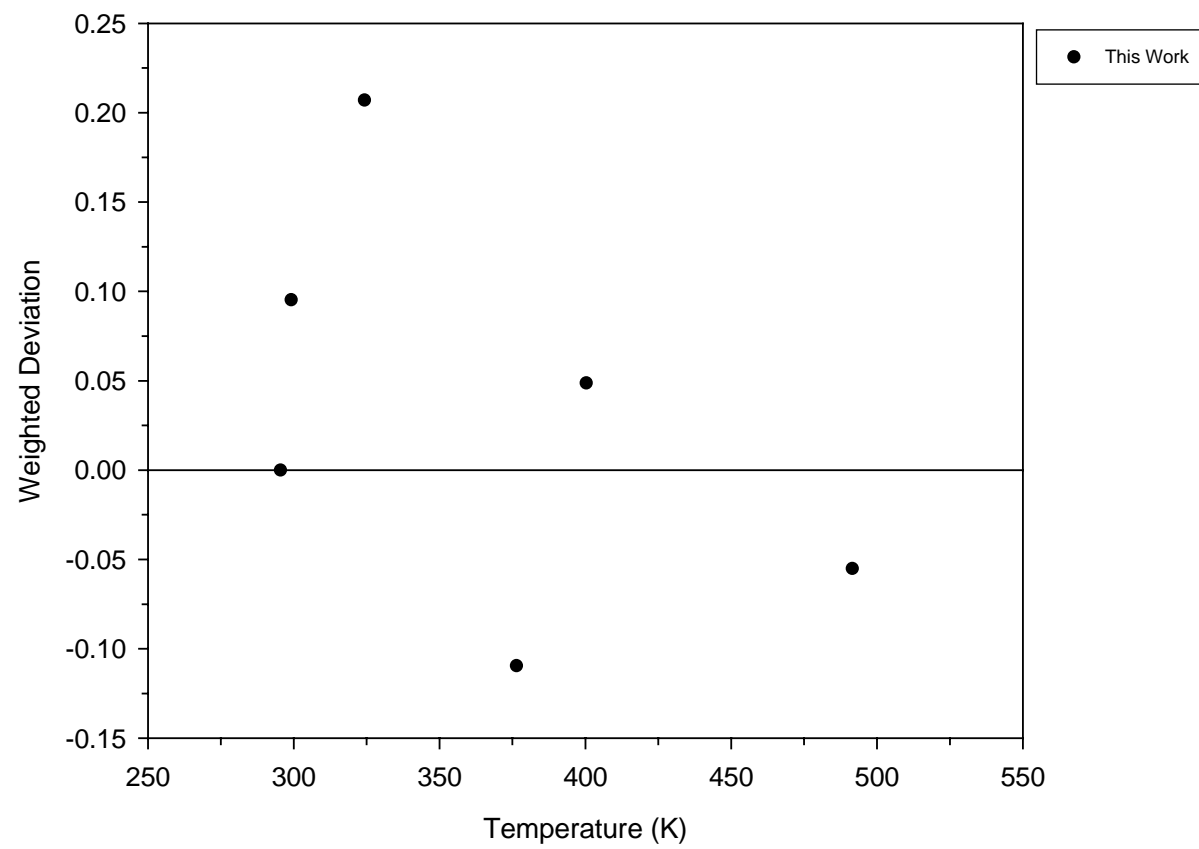


Figure 5-18. Weighted Deviation in the Solubility of Water in 3-Methylpentane

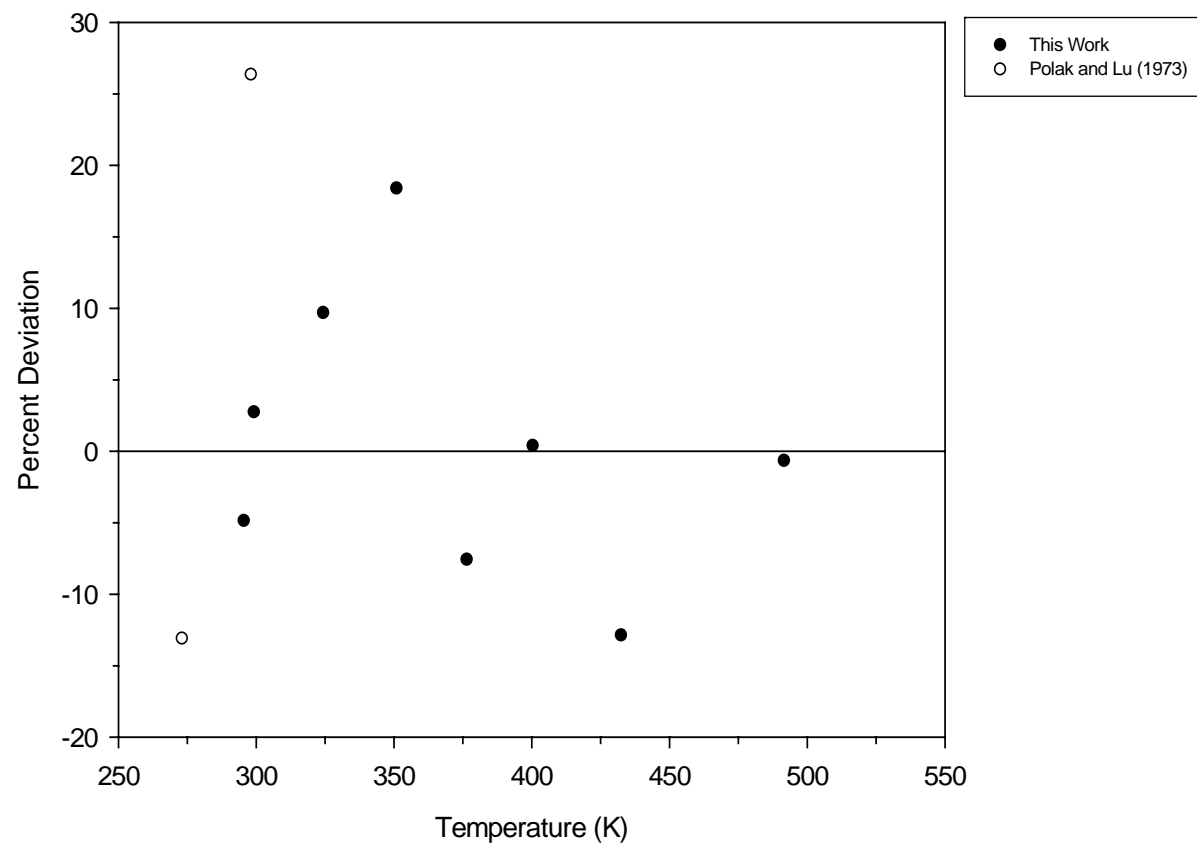


Figure 5-19. Percent Deviation in the Solubility of Water in 3-Methylpentane

Chapter 6. Conclusions and Recommendations

The objectives of the experimental portion of this work were to (a) develop an apparatus, including appropriate operating procedures and sampling and analytical techniques, capable of accurate mutual solubility (LLE) measurements at ambient and elevated temperatures of selected systems, (b) perform experimental measurements for the hydrocarbon-water systems, benzene-water, toluene-water, and 3-methylpentane-water, and (c) evaluate and correlate mutual hydrocarbon-water LLE data on these systems.

6.1 *Conclusions*

Based on the results of the present work, the following conclusions may be made:

1. A continuous flow apparatus was utilized successfully to investigate liquid-liquid equilibrium systems at temperatures from ambient to 500 K and at pressures from ambient to 13.8 MPa. The well-characterized benzene-water system provided a means to validate the operational status of the apparatus and the efficacy of the sampling methods and analytical techniques.
2. Although challenges existed in the operation of the gas chromatograph near its lower detection limit, standard operating procedures were developed, along with improved sampling and analytical techniques, which produced consistent and accurate data.

3. Mutual solubilities were measured for the systems benzene-water, toluene-water, and 3-methylpentane-water. The measurements were made near the three-phase equilibrium curve from ambient temperature to near the three-phase critical end point for the benzene and toluene systems. Due to the limited or nonexistent amount of data available at temperatures greater than ambient, the solubility measurements presented here provide a significant addition to the general body of knowledge of liquid-liquid equilibrium systems.

4. An error analysis was performed, which determined that approximately 90% of the total error in the solubility measurements was accounted for by the gas chromatograph analysis. The maximum expected uncertainty was approximately 4% and 8% in the water phase and organic phase measurements, respectively.

6.2 Recommendations

The apparatus, along with the sampling methods and analytical techniques, has been utilized successfully in obtaining mutual solubility data at elevated temperatures. Additional measurements should be made for systems where elevated temperature data do not exist, and different hydrocarbon molecules should be studied to elucidate structural effects on solubility. Systems should include families of six-carbon and seven-carbon molecules with differing bonding arrangements such as double, triple, or ring bonds and functional groups such as straight chain and branched alcohols, aldehydes, and esters. While the investigation of new systems may be of interest, there are recommendations concerning the apparatus and analytical technique employed.

1. Manual recycling was employed during this study, but the addition of a recycle stream to the apparatus would reduce chemical costs, generate less waste, and generally be more environmentally suitable. This could be accomplished easily with the use of a large glass carboy as a recycle stream receptacle/feed tank and rearrangement of the feed section tubing from the current location to the carboy.

2. The gas chromatograph currently in use showed variations in peak area determinations, requiring a large number of replicate analyses at each experimental condition. A solution to the effort involved in sample analysis would be the use of an autosampler to automate the analytical procedure.

3. Since the gas chromatograph is employed for analyses near its lower detection limit, different analytical techniques should be investigated. Two examples of established techniques that could replace the GC are high performance liquid chromatography and fluorescence monitoring, which would also have the added benefit of being an “online” analysis. A third potential technique would require the samples to be at supercritical conditions. At these conditions, the hydrocarbon-water system is totally miscible, and the use of a standard gas chromatograph with a thermal conductivity detector or flame ionization detector could be used. Solvents would not be necessary for the miscible hydrocarbon-water mixtures, and the calibration procedure would be simplified. The use of such a technique in an “online” fashion may also be possible.

Chapter 7. Overview

7.1 Rationale

Water, which is the most common industrial solvent, plays an important role in many areas including separation processes, distillation units, chromatographic systems, waste treatment, and environmental concerns [46-52]. With growing application of biotechnologies, there also exists an increased need for knowledge of the phase equilibria of aqueous systems in those processes [53]. When experimental data are unavailable, thermodynamic models, such as group contribution methods, are used to predict phase equilibria. The accuracy of these models in predicting infinite-dilution activity coefficients (γ^∞) of aqueous systems is questionable. Moreover, model development is hampered by a lack of (a) γ^∞ data at temperatures above 300 K, and (b) γ^∞ data for water-in-hydrocarbon systems.

Due to the unique molecular structure of water and its attendant physical characteristics, including hydrogen bonding, systems containing hydrocarbons and water often exhibit strong nonideality when compared to systems comprised only of hydrocarbons. The activity coefficient, γ , is a function that quantifies the extent of nonideality present in a system. When a component of a hydrocarbon-water binary system is sufficiently dilute, the infinite-dilution activity coefficient, γ^∞ , is reflective of only intermolecular solute-solvent and solvent-solvent interactions without the additional

complication of solute-solute interactions. Insight into the chemical and physical forces present in an aqueous system is provided by these coefficients.

While several experimental methods exist for the determination of infinite-dilution activity coefficients, these methods often suffer serious limitations [54-56] and are time consuming. Models for the prediction or calculation of infinite-dilution activity coefficients would be useful and are represented by examples from theoretical regular solution theory models [57-63], theoretical equation of state models [64], pure component models [65-67], group contribution models [68-70], empirical models [71-76], the linear solvation energy relationship (LSER) model [77, 78], computational chemistry models [79-84], and quantitative structure-property relationship (QSPR) models [81, 82, 84-86]. These models generally do not provide satisfactory predictions, and early QSPR studies were limited by the involvement of only single temperature data of one component of the aqueous systems.

The molecular structure of a chemical substance determines its chemical and physical properties. Continuing investigations have centered on elucidation of the relationship between physical properties and molecular structure. As the computational capability has improved, such research has revolved around developing free energy relationships by molecular mutation using Monte Carlo (MC) simulators [87]. Although this approach remains attractive, Monte Carlo is being replaced in many applications by QSPR models. The QSPR approach often provides predictions for chemical and physical properties of as-yet-unmeasured or unknown compounds based on structure information. High quality predictions are obtained using these descriptors since structure-property mapping is at an atomic level rather than at a functional group level. QSPR models will

be influential in enabling advances in chemical design, where a key challenge is the development of tools permitting the rapid identification, then creation of unique molecules for a targeted application. Over the last ten years, QSPRs have played an increasingly important role in drug screening and discovery [88], and applications are appearing in areas outside the pharmaceutical industry. While standard methodologies for chemical design result in a discovery phase of research and development requiring from two to three years, QSPR methodologies are estimated to result in a reduction of this phase to three to six months.

7.2 Objectives

The objectives of this work are to (a) develop a quantitative structure-property relationship (QSPR) for prediction of γ_i^∞ values of hydrocarbon-water systems, (b) evaluate the efficacy of QSPR models using multiple linear regression analyses and back propagation neural networks, (c) develop a theory based QSPR model, and (d) evaluate the ability of the model to predict aqueous and hydrocarbon solubilities at multiple temperatures.

7.3 Dissertation Organization

The modeling section of this dissertation is composed of Chapters 7-12. Following the present introduction, Chapters 8-10 present reviews of pertinent material and literature concerning infinite-dilution activity coefficients, quantitative structure property relationships, and neural networks, respectively. These chapters include information and procedures used in this study. In Chapter 11, which is written in the

form of a manuscript, complete with an independent set of tables, figures, and references, the experimental results are presented and discussed. Chapter 12 provides a summary of this work, and following this chapter the appendices appear. In Appendix G, molecular structure illustrations and infinite-dilution activity coefficient values of the molecules in the database are provided. During the course of QSPR model development, different types of models were employed where the descriptor selection and model selection were either “linear” or “nonlinear.” The QSPR model types known as Type I employ linear methods for both descriptor and model selection and Type III utilizes nonlinear methods for both descriptor and model selection. Type II models are a hybrid approach where descriptor selection is linear-based and model development is nonlinear. Supporting material, such as summary tables, descriptor sets, data outliers, and plots for determination of number of descriptors, contour, difference, and results, are given in Appendices H, I, and J for the Type I, II, and III analyses, respectively.

Chapter 8. Infinite-Dilution Activity Coefficient Models

Water, which is the most common industrial solvent, plays an important role in many different areas including separation processes, distillation units, chromatographic systems, waste treatment, and environmental concerns [46-52]. With growing application of biotechnologies, there also exists an increased need for phase equilibria of aqueous systems in those processes [53].

Due to the unique molecular structure of water and its attendant physical characteristics, including hydrogen bonding, systems containing hydrocarbons and water often exhibit strong nonideality when compared to systems comprised only of hydrocarbons. The activity coefficient, γ , is a parameter that quantifies the amount of nonideality present in a system. When a component of a hydrocarbon-water binary system is sufficiently dilute, the infinite-dilution activity coefficient, γ^∞ , is reflective of only intermolecular solute-solvent and solvent-solvent interactions without the additional complication of solute-solute interactions. Insight into the chemical and physical forces present in an aqueous system is provided by these coefficients.

While several experimental methods exist for the investigation of infinite-dilution activity coefficients, these methods often suffer serious limitations [54-56] and are time consuming. Models for the prediction or calculation of infinite-dilution activity coefficients would be useful and are represented by examples from theoretical models, pure component models, group contribution models, empirical models, the LSER model, and computational chemistry models.

8.1 Theoretical Models

Theoretical models include those developed using regular solution theory and equations of state (EOS).

Regular Solution Theory Models

Regular solution theory was originally conceived by Hildebrand and co-workers in the 1920s and is based, in part, on van Laar's observation that a solution at constant temperature and volume will not have zero excess entropy of mixing. Scatchard [57] and Hildebrand and Wood [58] developed a generalized method of deriving an expression for the excess Gibbs energy without the use of the van der Waals equation of state (EOS). The assumptions employed by Scatchard as listed by Malanowski and Anderko [89] were:

1. The mutual energy of two molecules depends only on the distance between them and their relative orientation and not on the nature of the other molecules between or around them and not on the temperature.
2. The distribution of the molecules is random, i.e., it is independent of temperature and the nature of the other molecules present.
3. The change of volume on mixing at constant pressure is zero.

By applying these assumptions, the estimations of activity coefficients were limited to mixtures composed of non-polar and similarly sized and shaped molecules.

Weimer and Prausnitz [59] extended the applicability of regular solution theory to mixtures containing polar components, using the assumption that polar component energy of vaporization could be separated into two parts, one concerning non-polar

interactions (dispersion or non-polar) and the other concerning the dipole-dipole effect (induction or polar). The effects of molecular size and shape were accounted for by the use of a correction term based on the Flory-Huggins expression. The infinite-dilution activity coefficient can then be expressed as

$$\ln \gamma_1^\infty = \gamma_1^{R\infty} + \gamma_1^{C\infty} \quad (8-1)$$

which consists of two contributions, residual and combinatorial. For a non-polar component ($x_1 \rightarrow 0$), the infinite-dilution activity coefficient is given by

$$\ln \gamma_1^\infty = \frac{v_1}{RT} [(\lambda_2 - \lambda_1)^2 + \tau_2^2 - 2\psi_{12}] + \left[\ln \frac{v_1}{v_2} + 1 - \frac{v_1}{v_2} \right] \quad (8-2)$$

where v is the molar liquid volume, R is the universal gas constant, T is the temperature, λ is the non-polar solubility parameter, τ is the polar solubility parameter, and ψ is the induction (polar) energy parameter. In Equation (8-2) the first term is the residual or interaction contribution and the second term is the combinatorial or size contribution.

Following this extension of the regular solution theory, other investigators [60-63] made analogous modifications which further extended the theory to polar and associating components by assuming that factors contributing to the energy of vaporization (including dispersion, induction, orientation, and hydrogen bonding) act independently and are additive. In general, these models are adequate for activity coefficient prediction for limited classes of non-polar mixtures.

Equation of State Models

An example of an EOS based model is that of Bader and Gasem [64], which uses a cubic equation of state to correlate infinite-dilution activity coefficients of hydrocarbon-

water systems. A general form for the prediction of the infinite-dilution activity coefficient, γ_1^∞ , written for component 1, is given as

$$\ln \gamma_1^\infty = \ln \gamma_1^{E\infty} + \ln \gamma_1^{I\infty} + \ln \gamma_1^{II\infty} \quad (8-3)$$

where $\gamma_1^{E\infty}$ is the excess activity coefficient, which accounts for the deficiency of the EOS in dealing with polar components, and $\ln \gamma_1^{I\infty}$ and $\ln \gamma_1^{II\infty}$ are terms accounting for repulsion and attraction, respectively. Using the general form given in Equation (8-3) with the Peng-Robinson EOS [90], results in the following equations:

$$\ln \gamma_1^{E\infty} = \ln \left(\frac{v_2}{v_1} \right) + \left(\frac{v_2}{v_1} \right) - C_5 \quad (8-4)$$

where v is the molar volume and C_5 is a regressed parameter,

$$\ln \gamma_1^{I\infty} = \frac{(b_1 - b_2)(1 + D_{12}) - b_2}{b_2} \left[\frac{p_2 v_2}{RT} - 1 \right] - \left[\frac{p_2 v_2}{RT} - 1 \right] + \ln \left[\frac{p_1 (v_1 - b_1)}{p_2 (v_2 - b_2)} \right] \quad (8-5)$$

where b is the co-volume parameter, D_{12} is the EOS binary interaction parameter, p is the system pressure, R is the gas constant, and T is the temperature and,

$$\begin{aligned} \ln \gamma_1^{II\infty} = & \frac{a_1}{b_1 RT} \frac{1}{2\sqrt{2}} \ln \left[\frac{v_1 + (1 + \sqrt{2})b_1}{v_1 + (1 - \sqrt{2})b_1} \right] \\ & + \frac{a_2}{b_2 RT} \frac{1}{2\sqrt{2}} \ln \left[\frac{v_2 + (1 + \sqrt{2})b_2}{v_2 + (1 - \sqrt{2})b_2} \right] \left[\frac{(b_1 + b_2)(1 + D_{12}) - b_2}{b_2} - \frac{2\sqrt{a_1 a_2} (1 - C_{12})}{a_2} \right] \end{aligned} \quad (8-6)$$

where a is the energy parameter and C_{12} is an EOS binary interaction parameter.

8.2 Pure-Component Models

Inspired by the modifications of the regular solution theory, Thomas and Eckert [65] developed a model, the modified separation of cohesive energy density (MOSCED),

for predicting γ^∞ from pure component parameters, which could be applied to mixtures containing polar and hydrogen bonding entities. Distinguishing this model from previous work is the applicability to binary systems composed of more than one polar molecule and in the prediction of activity coefficients at both concentration ends of a binary system. The general form of the MOSCED model written for component 1 is

$$\ln \gamma_1^\infty = \frac{v_1}{RT} \left[(\lambda_2 - \lambda_1)^2 + \frac{q_1^2 q_2^2 (\tau_2 - \tau_1)^2}{\psi_2} + \frac{(\alpha_2 - \alpha_1)(\beta_2 - \beta_1)}{\xi_2} \right] + \left[\ln \left(\frac{v_1}{v_2} \right)^{C_1} + 1 - \left(\frac{v_1}{v_2} \right)^{C_1} \right] \quad (8-7)$$

where λ is the dispersion parameter, q is the induction parameter, τ is the polar parameter, α is the acidity parameter, β is the basicity parameter, ψ is the polar asymmetry parameter, ξ is the hydrogen bonding asymmetry factor, and C_1 is an adjustable parameter.

The polar, acidity, and basicity parameters for the MOSCED model were correlated with a limited database of activity coefficients. With the advent of spectroscopic measurements [77, 91-94], the acidity-basicity of the hydrogen bond and the dipolarity-polarizability parameters (solvatochromic parameters) could be determined. A modified MOSCED model incorporating these new measurements was developed [66], and a new model, solvatochromic parameters for activity coefficients estimation (SPACE), was developed [67], which also took advantage of the expanded database of solvatochromic parameters and removed the adjustable parameters present in MOSCED.

8.3 Group Contribution Models

Group contribution methods are particularly attractive since they can be employed to estimate activity coefficients and other excess thermodynamic properties of liquid mixtures when experimental data are unavailable. The underlying assumption of these models is the additivity of contributions made by molecular functional groups. A contribution made by a group is assumed independent of any other group contributions within that same molecule.

Two methods have been used extensively: the analytical solution of groups (ASOG) [68] and UNIQUAC functional group activity coefficient (UNIFAC) [69]. While the two methods share a common basis, the equation used for representing the Gibbs excess energy of a mixture differs. The Wilson equation is used in ASOG, and UNIQUAC is used in UNIFAC.

In the UNIFAC-93 model, the activity coefficient is expressed in two parts. First is the combinatorial (entropic) term, which accounts for differences in molecular size and shape and is calculated from pure component properties. Second is the residual (group interaction contribution) term, which represents the intermolecular forces calculated from mixture properties. This model is expressed for component i as

$$\ln \gamma_i = \ln \gamma_i^{comb} + \ln \gamma_i^{resid} \quad (8-8)$$

The combinatorial part is a function of the molecular size and shape of the mixture components

$$\ln \gamma_i^{comb} = f(r, q, x) \quad (8-9)$$

where r is the molecular volume parameter, q is the molecular surface parameter, and x is the liquid mole fraction. The residual part is a function of temperature dependent interaction parameters

$$\ln \gamma_i^{resid} = f(q, T, a, b, c) \quad (8-10)$$

where T is the system temperature and a , b , and c are functional group interaction parameters. Detailed treatment of Equations (8-9) and (8-10) are given by Gmehling et al. [70].

Early versions of UNIFAC failed to determine the differences between isomers and to account for group proximity effects. These models often provide poor predictions when molecules of a binary system vary greatly in size and they fail to address hydrogen-bonding behavior. Additionally, these models are dependent on the quality of the structural parameters derived from experimental data.

8.4 Empirical Models

In contrast to theoretical models, several empirical models have been developed to correlate infinite-dilution activity coefficients in aqueous systems. These models provide a means of estimating activity coefficients and other excess thermodynamic properties in the absence of experimental data.

Pierotti et al. [71] developed a scheme for the prediction of γ^∞ in water and other solvents, which was based on a number of empirical group interaction parameters for each solute series in a given solvent. This approach assumes that γ^∞ can be taken as the sum of contributions from individual interactions between pairs of structural groups in the solute and solvent molecules. These interactions are dependent on the number, type,

and configuration of the groups within the respective molecular structures. For various aromatic hydrocarbon-aqueous systems, Tsonopoulos and Prausnitz [72] developed a similar scheme.

Medir and Giralt [73] developed a correlation for aqueous systems based on the first order molecular connectivity, number of carbon atoms, surface area, acentric factor, dipole moment, and total electronic energy. For aliphatic and two aromatic families, monocyclic and polynuclear, the best correlating factor was the first order molecular connectivity and dipole moment, which is given as

$$\ln \gamma_1^\infty = C_1 + C_2 {}^1\chi_1^9 + C_3 D_{M1}^2 \quad (8-11)$$

where ${}^1\chi_1^9$ is the first order molecular connectivity of the organic solute, D_{M1} is the dipole moment of the organic solute, and $C_1 - C_3$ are regressed parameters. Using solute molar refraction, R_{M1} , as an input parameter, a similar correlation was developed by Dutt and Prasad [74], which is given as

$$\ln \gamma_1^\infty = C_1 + C_2 R_{M1} \quad (8-12)$$

Other simple correlations between structural features and hydrocarbon solubility in water have been developed, which include those based on solute accessible surface areas [75] and number of solute carbon atoms [76]. While these methods provide greater accuracy in predicting γ^∞ values than group contribution methods, the correlations involve no more than two structural features and are limited to a small set of hydrocarbon families.

8.5 LSER Model

A linear solvation energy relationship (LSER) model was first developed by Taft et al. [77] in response to the limitations inherent in the early empirical models. This model attempted to generalize solvation using the assumption that solute-solvent interactions are due to independent and additive nonspecific dipolarity/polarizability effects and specific hydrogen bonding interactions. These effects are quantified by solvatochromic parameters as determined by spectroscopic or chromatographic experimentation.

Sherman et al. [78] improved the early LSER model by the addition of a Flory-Huggins term and the use of the saturation fugacity rather than the gas-liquid partition coefficient. The final form of this model for the prediction of a hydrocarbon γ^∞ (component 2) in water is given as

$$\ln \gamma_2^\infty = I \log L^{16} + p\pi_2^* + a\alpha_2 + b\beta_2 + h \left[\ln \left(\frac{v_2}{v_1} \right)^{0.75} + 1 - \left(\frac{v_2}{v_1} \right)^{0.75} \right] - \ln f_2^\circ + 16.10 \quad (8-13)$$

where I , p , a , b , and h are parameter coefficients, L^{16} is the partition coefficient of the solute between a gas and hexadecane, π^* is the dipolarity/polarizability effect, α is the hydrogen bonding donation effect, β is the hydrogen bonding acceptance effect, v is the component molar volume, and f° is the standard-state fugacity. While encouraging results have been obtained for aqueous systems with this model, the main disadvantage remains the availability and accuracy of solvatochromic data.

8.6 Computational Chemistry Models

Computational chemistry methods include both molecular simulation and quantitative structure-property relationship (QSPR) models.

Molecular Simulations

Molecular simulations for the prediction of aqueous activity coefficients include free energy perturbation simulations [79], and Widom insertion approaches [80]. The main disadvantages of molecular simulation are the computational expense and time involved. As computer technology advances are made, these disadvantages will become less severe. The accuracy of the prediction is also highly dependent on the ability of the intermolecular force potentials to adequately characterize the molecular interactions of the hydrocarbon-water system.

QSPR Modeling

The underlying assumption of QSPR is that the physical properties of a given molecule are described completely by its chemical structure. With a given set of data for a thermo-physical property of interest (e.g., γ^∞), a quantitative relationship can be constructed between this property and molecular structure, which may then be used in the prediction of this property for other molecules based solely on their molecular structures.

Nelson and Jurs [81] first used a QSPR for the aqueous solubility of organic compounds. The QSPR employed a nine-variable regression model with the structures represented by topological, geometrical, and electronic descriptors. A later study by Mitchell and Jurs [82] provided a QSPR for γ^∞ of hydrocarbons in water. This study

developed a 12-variable regression model with the molecular structures represented by topological, geometrical, and electronic descriptors.

Recent work by He and Zhong [83] revisited the database used by Mitchell and Jurs [82] and incorporated descriptors based on molecular connectivity indices. While the predictive capability of both models was similar, He and Zhong only required six descriptors, and these descriptors are more easily calculated than the descriptors utilized by Mitchell and Jurs.

Huibers and Katritzky [84] correlated the aqueous solubility of hydrocarbons with molecular structure using a minimum number of geometrical, topological, and constitutional descriptors. By minimizing the number of descriptors included in the QSPR, the contribution of each individual descriptor is more easily observed. A three-parameter model for the hydrocarbon solubility was given as

$$-\log S_w = -0.13 + 0.0437MV - 0.258^0BIC + 0.0523PNSA \quad (8-14)$$

where MV is the molecular volume, 0BIC is the structural information content of 0th order, and $PNSA$ is the atomic charge weighted partial negative surface area. Of the three descriptors, the molecular volume was determined to be the most important in correlating solubility and structure.

For good predictive capability, QSPR models need an adequate database containing the physical property of interest for a wide variety of molecular structures. Additionally, current QSPR models lack versatility since activity coefficient data are mostly for a single temperature, commonly 298 K, and only involve one of the limiting activity coefficients of a binary system.

Chapter 9. Computational Chemistry

9.1 *Introduction*

For a homologous series of alkanes [95], the boiling point temperature increases in a regular fashion as a function of carbon number. While this was considered common knowledge in organic chemistry, the possibility also existed that other similar regularities might exist between molecular structures and other physical properties or activity. From this early observation, quantitative structure-activity relationship (QSAR) and quantitative structure-property relationship (QSPR) models have been developed.

These theoretical models rely on the assumption that there exists a quantifiable relationship between the thermophysical property, chemical affinity, or biological activity and molecular structure. With the utilization of QSAR and QSPR, elucidation of the information contained within the molecular structure is obtainable. This chapter will discuss the history and the development of a general quantitative structure-property relationship (QSPR).

9.2 *Historical Background*

Early work in this area of computational chemistry was centered on QSAR, the relation between structure and a chemical activity, which was usually biological in nature. The methodology of QSAR originated around 1900 with the independent studies of Meyer and Overton [96, 97], and their observations relating the potency of anesthetics

to lipophilicity. During this same time period, the concept of biological receptors was emerging. This concept states that interactions between drug molecules and certain proteins (receptors) are constrained to act in a specific manner to elicit a desired effect. Additionally, Fischer [98] determined that there was a governing stereospecificity of the drug-receptor interaction, which means a particular structure of the drug molecule was required for a particular receptor response. From this research, the ability to create drugs with specific benefits was realized by maintaining the stereospecificity and completing alterations to the base molecular drug structure, which would result in different therapeutic benefits.

QSPR is closely related to QSAR, and the general methodology of both models is similar. Historically, Hammett [99] is credited with the first application of QSPR, which concerned the relationship between the structures of various substituted benzenes and rate and equilibrium constants of chemical reactions. These relationships, which are termed linear free energy relationships (LFER), were extended by Taft [100] to aliphatic molecules.

The use of QSAR as a practical tool for drug design was founded in the 1960s with the introduction of two extrathermodynamic methods. The first method developed by Hansch and Fujita [101] was based on a LFER between biological activities and the disassociative, hydrophobic, and steric properties of congeneric drug molecules. This was accomplished by using the physicochemical properties of chemical substituents on a common parent molecule. The second method developed by Free and Wilson [102] was based on the theory that biological activity resulted from the addition of contributions from various substituents groups at multiple substituents positions. Both methods

utilized multiple linear regression to determine the combination of substituents resulting in a maximum activity for congeneric molecules. The disadvantage of these methods is the need for experimental data to describe the substituents group contributions rather than utilizing theoretical descriptions.

With the advent of improved computational capability, multivariate chemometric techniques [103], and better molecular descriptors [104, 105], there has been a substantial increase in the number of QSAR and QSPR studies. The QSAR methodology is used extensively in the pharmaceutical industry in computer aided drug design [106-111], and a partial list of properties described by QSPR includes pure-fluid boiling points [112, 113], vapor pressures [114], and critical pressures [115]; hydrocarbon solubilities in water [116]; refractive indexes of polymers [117]; drug activity [106]; and protein and ligand characteristics [118, 119].

9.3 QSPR Model Development

A general flowchart presenting the main components involved in a QSAR/QSPR study is given as Figure 9-1.

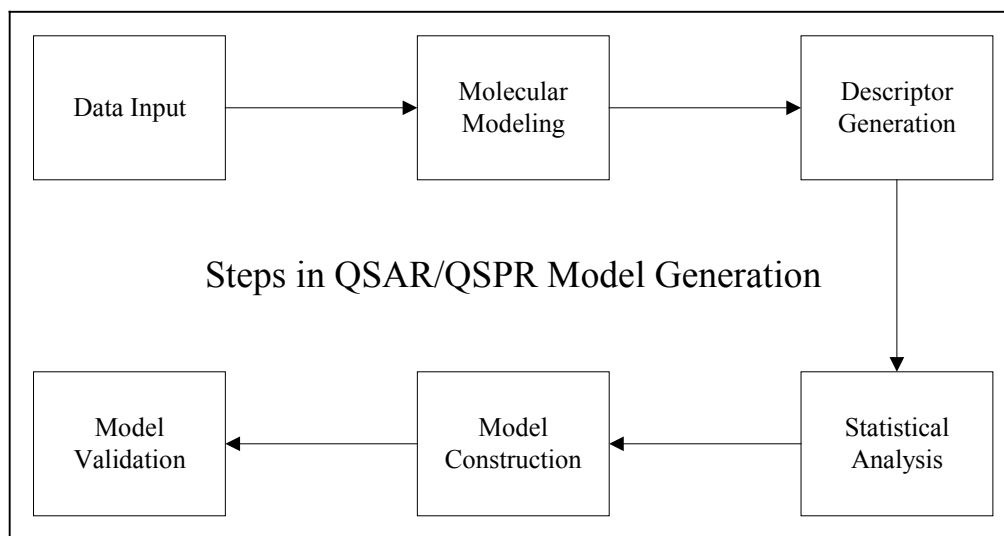


Figure 9-1. QSPR/QSAR Model Generation Flowchart

The data input component concerns the entry of molecular structures in a suitable topological representation, along with the property of interest. Each molecular structure is next submitted for conformational analysis in the molecular modeling component. Next, descriptors are generated for each molecule based on the topological and conformational representation, and these descriptors are then analyzed in the statistical analysis component to determine the best subset of descriptors to use in describing the property of interest. A model based on the descriptor subset is developed in the model construction component, and then these models are validated in the model validation component using an external prediction set. A brief overview of each component follows.

Data Entry

Initially, a database containing molecular structures and the associated property of interest for each molecule is required. The structures may be represented as a two-dimensional (2-D) stick-figure drawing, as shown for cyclohexane in Figure 9-2.

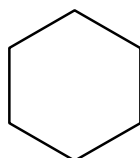


Figure 9-2. 2-D Representation of Cyclohexane

While this type of representation enables easy recognition of the molecule, the disadvantage is that the sketched figures are not amenable to automated database searches (data mining).

As an alternative, simplified molecular input line entry specification (SMILES) provides a general purpose chemical nomenclature, which is based on the representation of a valence model, and a universal data format exchange, which is not limited to a specific computer language or program [120]. For example, the SMILES notation for cyclohexane would be C1CCCCC1.

Molecular Modeling

Molecular modeling software is employed to generate a three-dimensional (3-D) structure from the 2-D structures obtained in the previous step. The 3-D structures are required for generation of geometric descriptors; however, molecular structures can have different conformations depending on the types of bonds present, which would affect the geometric descriptors. A semi-empirical molecular orbital routine such as MOPAC [121]

can be used to find the minimum energy conformation, which allows for accurate assessment of the geometric descriptors.

Descriptor Generation

A major component in the development of QSPR is the generation of molecular structure descriptors, which can describe the entire molecular structure or any structural fragment. As given by Karelson [105], Table 9-1 presents the classes and subclasses of theoretical descriptors.

Table 9-1. General Classification of Theoretical Molecular Descriptors

Class	Subclass
Constitutional descriptors	Counts of atoms or bonds
Topological descriptors	Atomic weight based descriptors
	Topological (connectivity indices)
	Information theoretical descriptors
Geometrical descriptors	Topochemical descriptors
	Distance-related descriptors
	Surface area related descriptors
	Volume related descriptors
Charge distribution related descriptors	Molecular steric field descriptors
	Atomic partial charges
	Molecular electrical moments
	Molecular polarizabilities
	Molecular electrostatic field descriptors
Molecular orbital related descriptors	Frontier molecular orbital energies
	Bond orders
	Fukui's reactivity indices
Thermodynamic descriptors	Thermodynamic functions
	Boltzmann factor weighted descriptors
Solvation descriptors	Electrostatic energy of solvation
	Dispersion energy of solvation
	Free energy of cavity formation
	Hydrogen bonding descriptors
	Entropy of solvation
	Theoretical linear solvation energy descriptors

Brief descriptions of these descriptor classes and the class of constructed descriptors along with specific descriptor examples follow, but a complete treatment of descriptors may be found elsewhere [105, 122].

Constitutional Descriptors

The chemical composition of the molecule is represented by the constitutional descriptors, which do not involve the geometry or electronic structure of the molecule. These descriptors are the easiest to conceptualize and calculate; however, the interpretation of individual molecular properties is frequently cumbersome. Some examples of this descriptor class include total number of atoms, absolute and relative numbers of a specific atom, absolute and relative number of a specific functional group or substituents, and molecular weight.

Topological Descriptors

Information about the atoms and bonds present in a molecule can be formalized by the application of the mathematical field termed graphs. These descriptors, or topological indices, are developed from the graph invariants obtained after formalization of the atomic structure and connectivity [123]. Examples of topological descriptors include the Wiener index [124], Kier shape indices [125], Kirchoff number [126], and bonding information content index of k^{th} order [127].

Geometrical Descriptors

These descriptors are derived from the geometric structure of the molecule as determined by the 3-dimensional coordinates of the atomic nuclei and masses. Principal moments of inertia characterize the molecular mass distribution and the degree of

different rotational transitions, and these descriptors depend only on the atomic coordinates and masses. Other widely used descriptors of this class based on 2-D projections of the 3-D structure include molecular surface area [128], solvent-accessible surface area [129], molecular volume [130], and solvent-excluded volume of the molecule [131].

Charge Distribution Related Descriptors

All chemical interactions may be classified as either electrostatic (polar) or orbital (covalent) in nature, according to current molecular structure theory. Since the mechanism and rate of most chemical reactions and the physicochemical properties are determined from electron densities or charges, descriptors of this class are widely utilized as reactivity indices or as measures of intermolecular interactions. Empirical schemes [132, 133], which involve the concept of atomic electronegativity, quantum chemical theory [121], and schemes involving analysis of physical observations predicted from wave function have been used to calculate atomic partial charges [134, 135]. Using the atomic partial charges, simple electrostatic descriptors including the minimum and maximum partial charges [121], minimum and maximum partial charges for a particular atom type [121], and a polarity factor [136] can be calculated.

In order to characterize the interactions between polar molecules, Jurs et al. [137, 138] developed the charged partial surface area (CPSA) descriptors, which are calculated from contributions from the partial atomic charges and the molecular solvent-accessible surface area.

By characterizing either the hydrogen bonding donor ability (HDSA1) or hydrogen bonding acceptor ability (HASA1) [137, 138] of the molecule through the

summation of solvent-accessible areas of potential atomic donors/acceptors, possible intermolecular hydrogen bonding interactions can be delineated.

Molecular Orbital Related Descriptors

Molecular quantum chemistry calculations provide for the development of a new source of descriptors that can, in principal, characterize almost any molecular geometric and electronic property and define intermolecular interactions. This class of descriptors is composed of charge distribution-related descriptors, valency-related descriptors [139], quantum mechanical energy-related descriptors, and quantum mechanically calculated molecular solvation descriptors. Examples of these descriptors include energies of the highest occupied (HOMO) and lowest unoccupied (LUMO) molecular orbital, maximum bond order for a given pair of atomic species, HOMO-LUMO energy gap [140], and Born solvation energy.

Thermodynamic Descriptors

The use of the total molecular partition function, Q , and constitutive components, electronic, translational, rotational, and vibrational, enables the construction of the thermodynamic descriptors [141-143]. Some examples of this class of descriptors include thermodynamic heat of formation at 300 K, translational enthalpy of the molecule at 300 K, and vibrational entropy of the molecule at 300 K.

Solvation Descriptors

Descriptors characterizing solvation effects from chemical structures and properties are constructed from physical models, which describe solvation phenomena arising from the creation of a solvation cavity and the insertion of a solute molecule into

the cavity. Examples of these descriptors include free energy of the solute cavity formation [144], molecular van der Waals volume [145], and Hildebrand's solubility parameter [146].

Constructed Descriptors

When the theoretical descriptors available in commercial QSPR software provide inadequate representation of the molecular structural information, new descriptors can be constructed. While the effect of these descriptors is of limited influence in liquid property modeling, the effect becomes significant when investigating solid-state properties such as melting point. Functional group descriptors, which are based on the concept that each functional group contributes a positive or negative increment to the total molecular property, are used infrequently, but their addition usually alters molecular properties by changing the polarizability and dipole moment of the molecule.

Feature Selection

Generally in a QSPR study, the number of descriptors generated is larger than can be realistically employed in model construction. Similar to over-specification of variables in a process unit design problem, redundant descriptors lessen the performance of the QSPR model and can lead to erroneous predictions. Thus, the large initial set of descriptors is reduced to provide a small set, which retains sufficient information about the molecular structure as it affects the property to be predicted. Reduction is accomplished by objective feature selection and subjective feature selection.

Objective Feature Selection

Reduction of the descriptor pool by objective feature selection is limited to the use of only the independent variables (descriptors). The selection of descriptors for deletion from the initial set is completed through the utilization of pairwise correlations, test of identical values, and vector space descriptor analysis.

Subjective Feature Selection

Once the descriptor set has been reduced by objective feature selection, subjective feature selection is used to further reduce the set of descriptors. Subjective feature selection uses the dependent variable values (property of interest) along with the independent variables (descriptors) values. Many statistical and computational techniques exist for subjective feature selection, including multiple linear regression analysis (MLR), simulated annealing (SA), principal component analysis (PCA) [147], partial least squares (PLS) [148], genetic algorithms (GA) [149], artificial neural networks (ANN) [150], support vector machines (SVM) [151], local learning (LL), self organizing maps (SOM) [152], cluster analysis (CA) [153], factor analysis (FA) [103], and discriminant analysis (DA) [154].

Model Construction

With a reduced descriptor set, a QSPR model is produced either with a statistical or neural network approach. The statistical approach is usually either a MLR or PLS analysis. The resultant QSPR models can be classified as one of three types depending on the combination of feature selection and model type as shown in Table 9-2.

Table 9-2. QSPR Model Types

Type	Feature Selection	Model Selection
Type I	MLR (linear)	MLR (linear)
TYPE II	ANN (nonlinear)	MLR (linear)
Type III	ANN (nonlinear)	ANN (nonlinear)

Type I models are completely linear, Type II models are a hybrid, and Type III models are completely nonlinear. Generally, Type III models show the best performance in QSPR modeling [155].

Multiple Linear Regression

A multiple linear regression model is developed, which relates the molecular structures to the property of interest through a linear combination of the descriptors. The general form of the correlation is:

$$y = \beta_0 + \sum_{i=1}^N \beta_i x_i \quad (9-1)$$

where, y is the property of interest, β_0 is the intercept, N is the number of molecular descriptors in the correlation, β_i is the coefficient for descriptor i , and x_i is the molecular descriptor. The multi-parameter regression that maximizes the predicting ability is determined using the following strategy [136] (adapted from CODESSA documentation).

1. All orthogonal pairs of descriptors i and j (with $R_{ij}^2 < R_{\min}^2$) are found in a given data set. R_{ij}^2 denotes the correlation coefficient between descriptor i and j , and R_{\min}^2 is set to the recommended value of 0.1.
2. The property analyzed is treated by using the two-parameter regression with the pairs of descriptors, obtained in Step 1. The pairs with highest regression

correlation coefficients, R_{ij}^2 , are chosen for performing higher-order regression treatments.

3. For each descriptor pair, obtained in the previous step, a non-collinear descriptor, k is added, and the respective three-parameter regression treatment is performed. If the Fisher criterion at a given probability level, F , is smaller than that for the best two-parameter correlation, the latter is chosen as the final correlation. Otherwise, the three-parameter correlations with highest regression correlation coefficients are chosen for the next step.
4. For each descriptor set, chosen in the previous step, an additional non-collinear descriptor scale is added, and the respective $(n+1)$ parameter regression treatment is performed. If the Fisher criterion at the given probability level, F , is smaller than for the best two-parameter correlation, the latter is chosen. Otherwise, descriptor sets with highest regression correlation coefficients are chosen, and this step is repeated with $n = n + 1$.
5. The final product of the above steps is a linear relationship between molecular structure and the property of interest containing n parameters.

Artificial Neural Networks

Several types of ANNs are employed in QSPR models, with feed-forward and back-propagation neural networks being commonly employed. More detail concerning ANNs is provided in the following chapter. The purpose of the ANN is to create an association between the structural descriptors and the property of interest. The level of agreement between the input and output of the property of interest in an ANN occurs

through the adjustment of weights, which affect individual inputs, and biases, which affect the net input.

In order to create this association, the ANN is trained using a partial set of structural descriptors and properties (training set), which occurs in two stages. The first stage involves learning the general features of the training set, and the second stage involves learning the individual characteristics of the molecules in the training set. Property prediction of new molecular structures is enhanced by the first stage of training, but the second stage leads to memorization of the training set molecules (over-training). Over-training may be effectively avoided by the use of a cross-validation set in addition to the training set. The cross-validation set is another partial set of structural descriptors and properties, which differ from those in the training set. During training, the property of interest is periodically predicted for the cross-validation set. The error in the predictions will reach a minimum and then begin to rise with additional training [156]. When the cross-validation set reaches the error minimum, the optimum stopping point to cease training has been reached.

Other considerations in the application of ANNs in QSPR models include the following [136]:

1. Random generation of the training set, cross-validation set, and prediction set (for validation) from the reduced set of descriptors should be made in the proportions of 70%, 10%, and 20%, respectively.
2. The ratio of the molecules to the number of descriptors in the training set should be greater than two [157].

3. The number of hidden layers in the ANN is determined empirically. The determination of the best architecture involves starting with a small number and gradually increasing the number of hidden layers.
4. The error observed in the cross-validation predictions should vary smoothly over the training period of the ANN.
5. The starting weights and biases of the ANN should be assigned in a random fashion.

Model Validation

The final component in the development of a QSAR/QSPR model is the validation of the newly constructed model. Validation of the model is generally accomplished by demonstration of the predictive ability of the model; however, Monte Carlo randomization testing is an additional validation technique that is sometimes utilized.

Predictive Ability

Using a previously prepared prediction set consisting of structural descriptors and properties taken from the reduced set of descriptors, the property of interest is predicted using the new model. The model should be capable of predictions at a desired level of accuracy, and the prediction error should be comparable to that observed for the training and cross-validation sets.

Monte Carlo Randomization

A potential danger in QSPR model development is the possibility of creating a model by chance correlation between the structural descriptors and property of interest.

To avoid this situation, the sequence of model construction is repeated using the same structural descriptors, but the dependent variables (property of interest) are randomly assigned to the molecular structures in the training set. If a chance correlation did exist between the descriptors and dependent variable of a particular molecular structure, the randomization of the dependent variable insures the original correlation will not be possible. The predictive results from the original and randomized training set models are compared. If the predictive results from the randomized training set model do not exceed the performance of the original predictive results, then the original model is considered to represent a relationship between the descriptors and the dependent variables that is not based on a chance correlation.

Chapter 10. Neural Networks

10.1 *Introduction*

Artificial neural networks (ANN), commonly referred to as neural networks (NN), were inspired by the recognition that the manner in which the human brain computes differs from the computations of a conventional computer. The concept of neurons as structural constituents in the brain was introduced by Ramon y Cajal [158], a pioneer in neurology.

While events in a computer chip happen in the nanosecond range, neural events occur in the millisecond range, which are approximately six orders of magnitude slower. The brain compensates for the relative slowness of each neuron by using extremely large numbers of interconnected neurons. One estimate for the human cortex placed the number of neurons on the order of 10 billion, and the number of synapses (interconnections) on the order of 60 trillion [159]. This neural architecture creates an extremely efficient structure, which has an energetic efficiency of approximately 10^{-16} Joule/operation/second as compared to 10^{-6} Joule/operation/second for a computer [160].

Using the efficient brain structure, which has the characteristics of a complex, nonlinear, and parallel computer as a model, investigators have developed NNs for applications to such fields as pattern recognition, optimization, coding, process control, drug discovery, and molecular design. This chapter will present a brief historiography and a general overview of neural networks.

10.2 *Historical Background*

The modern era of neural network theory began in the early 1940s with the seminal work of McCulloch and Pitts [161], which demonstrated the ability of a simple neural network, which was based on binary processing units termed neurons, to compute any arithmetic or logical function.

Hebb [162] provided a major contribution to the theoretical aspects of neural networks in 1949 with the presentation of an explicit statement of a physiological learning rule for synaptic modification. Specifically, as an organism learns different tasks, the connectivity of the brain is altered, and the results of these changes were called neural assemblies. Additionally, Hebb introduced a concept of learning, which states that repeated activation of one neuron by another interconnected neuron increases the effectiveness of the synaptic connection. This work served to inspire many investigators who were developing learning paradigms.

One concern arising at this time was the design of a reliable network with neurons that might be considered unreliable components. This potential problem was resolved by the work of von Neumann [163], which employed the concept of redundancy. Soon after this, Rosenblatt [164] developed a new approach to pattern recognition problems with the use of perceptrons, which were basically a one layer neural network. In 1960, a neural network consisting of adaptive linear elements (ADLINE), based on the least mean-square algorithm (LMS), was introduced by Widrow and Hoff [165] for signal processing study. The difference in these early network architectures lies in the training procedure. During the 1960s, the perceptron type networks enjoyed great popularity, but in 1969, Minsky and Papert [166] used extensive mathematical demonstrations to prove there are

fundamental limits to the computational capability of perceptrons. They also theorized that the extension of perceptrons to a many layer network would not provide any computational benefit.

The 1970s are often recognized as a “decade of dormancy.” According to Cowan [167], factors influencing this lack of scholarship during this period include the following:

1. A lack of supporting technology (computers) for adequate experimentation.
2. The criticism of Minsky and Papert convinced funding agencies to cease support of neural network projects and offered little encouragement for investigators to continue working in this field.
3. The analogy between neural networks and lattice spins, the spin-glass model, was not developed by Sherrington and Kirkpatrick [168] until 1975.

While there was a decided lack of research concerning neural networks during this time, one important concept to emerge was self-organizing maps using competitive learning [169, 170].

The 1980s began a renaissance for interest and study of neural networks and associated areas. In 1982, Hopfield [171] demonstrated a stable network capable of storing information, which was developed from using the Lyapunov (energy) function to analyze and understand the computations of networks with symmetric synaptic connections. These types of networks are known as Hopfield networks and are demonstrative of the isomorphism between the Hopfield network and the Ising model

(spin systems) used in physics. While Hopfield is noted for the concept of information storage, this work was based on the pioneering studies of Cragg and Tamperley [172, 173], Cowan [167], Grossberg [174, 175], Amari [176], Wilson and Cowan [177], Little and Shaw [178], and Anderson et al [179].

Shortly after this in 1986, the popular back-propagation algorithm for training multilayer networks was introduced by Rumelhart et al. [180], and later that year, a landmark book by Rumelhart and McClelland [181] covering back-propagation learning was published. Two other researchers, Parker [182] and LeCun [183], were independently investigating the back-propagation (BP) algorithm at about the same time.

During the 1980s other notable advances in the neural network field of study included simulated annealing [184], Boltzmann learning [185], the principle of maximum information preservation [186], and the design of layered feed forward networks using radial basis functions [187].

In the early 1990s, quantitative structure-activity relationship/quantitative structure-property relationship (QSAR/QSPR) studies utilizing BP neural networks to investigate complex relationships between molecular structure and physiochemical properties or biological activities began to appear in the literature with the early works of Aoyama et al. [188, 189], Aoyama and Ichikawa [190, 191], and de Saint Laumer et al. [192]. By the mid-1990s, an estimated 90% of neural networks employed some variant of BP [193]. Applications of BP can be found in a partial listing of such diverse areas of study as pharmacology [190, 191, 194-200], toxicology [201-204], carcinogenicity [204, 205], mutagenicity [206-209], n-octanol/water partition coefficients [210-212], aqueous solubility [213-215], activity coefficients [82, 83, 216], heat capacity [217], melting point

[218], normal boiling point [208, 217, 219-225], critical properties [208, 218, 225, 226], and density [217].

10.3 Neural Network Overview

The attributes of power and ease of use have led to the widespread usage of neural networks in such diverse areas as finance, medicine, pharmaceuticals, engineering, geology, chemistry, and physics. While linear techniques often fail to model complex functions adequately, the nonlinear nature of NNs provides a powerful and sophisticated technique capable of modeling complex functions. Additionally, NNs may be utilized readily with a lower depth of knowledge than would be required with more traditional nonlinear statistical techniques. This section will provide an overview of neural network construction, example architecture, and general architectural examples (more detailed presentations are available in a number of textbooks (*e.g.* [227-232])).

Neural Network Construction

A neural network attempts to replicate a biological network of neurons. As applied to an artificial neuron, the nature of the biological neuron is described as follows:

1. The neuron is the recipient of a number of inputs.
2. Inputs are conducted to the neuron via a connection, which has an associated weight (or strength). In a biological neuron these weights would correspond to a synaptic efficiency.
3. The neuron has a single threshold value, which is subtracted from the weighted summed inputs to provide an activation value of the neuron.

- This activation value is processed with an activation function to produce the neuronal output.

A collection of interconnected artificial neurons composes a NN, where the neurons are arranged in a minimum of two layers, an input and output layer. Figure 10-1 presents a schematic diagram of a feed-forward two layer NN, which consists of a nonlinear hidden layer and a linear output layer.

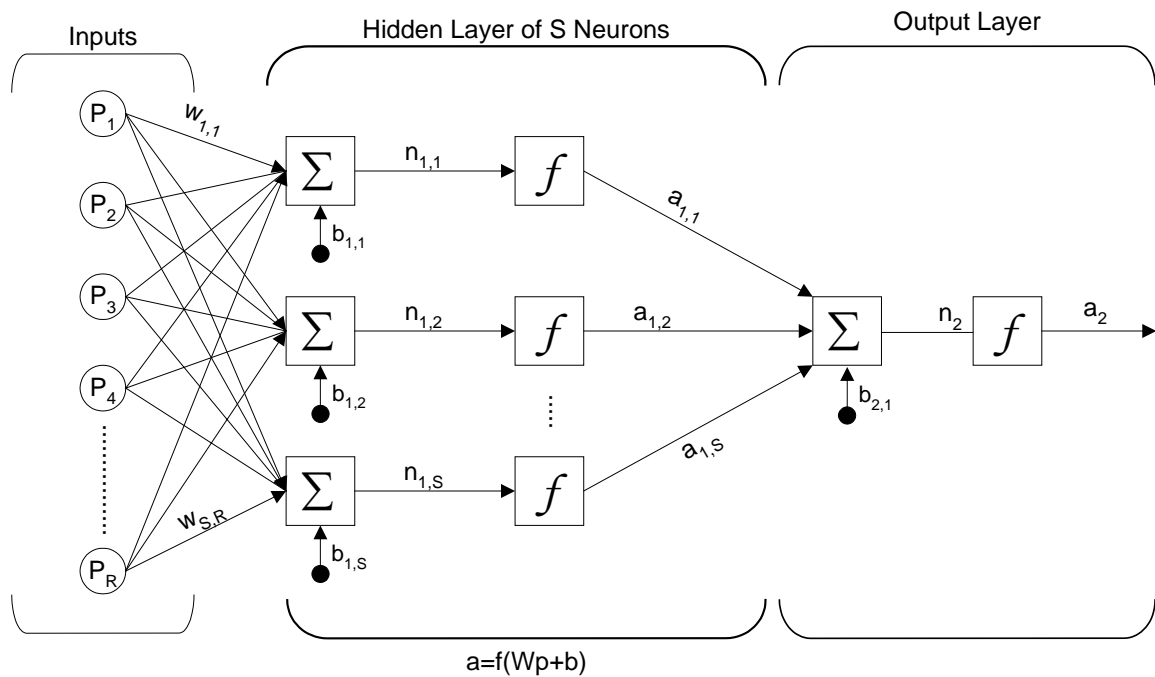


Figure 10-1. Schematic of a Neural Network

In this schematic, the NN is provided an input set ($P_1, P_2 \dots P_R$) and produces an output based on the relationships between the weights ($w_{i,j}$) and biases ($b_{i,j}$).

Since the notation used in Figure 10-1 quickly becomes cumbersome with a large number of neurons, an abbreviated notation has been developed [233] to simplify the representation of the neuronal architecture, as shown in Figure 10-2.

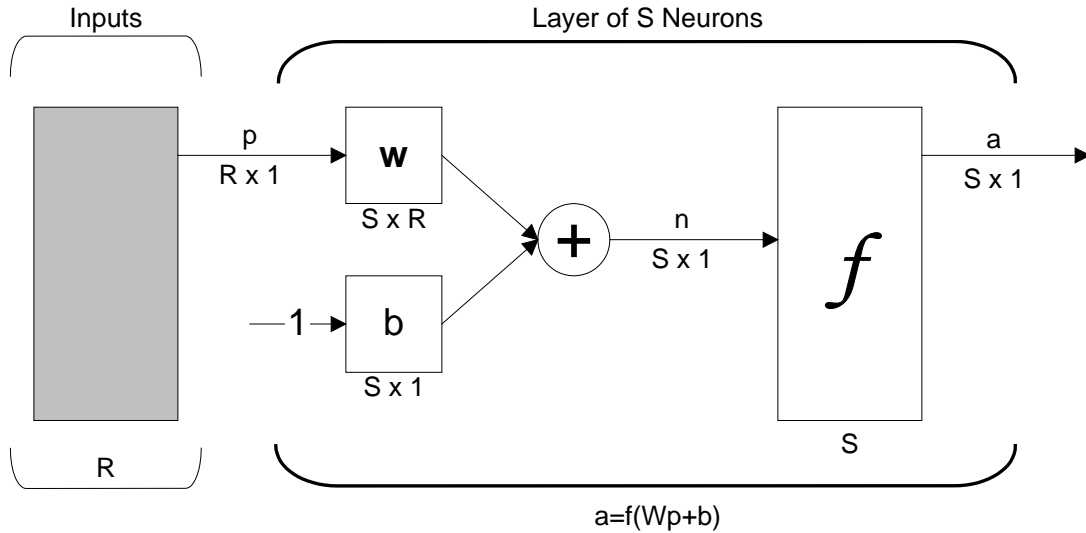


Figure 10-2. Abbreviated Notation for a Neural Network

For the example in Figure 10-2, the NN is composed of an input vector p with a number of R values, and S hidden layer neurons. The activation function is represented by f , which can be any continuous function, including typical examples such as linear, hard limit, soft limit, and sigmoidal. When representing nonlinear functions, most NNs will employ a tan-sigmoidal hidden layer function and a linear output layer function, which are shown graphically in Figure 10-3.

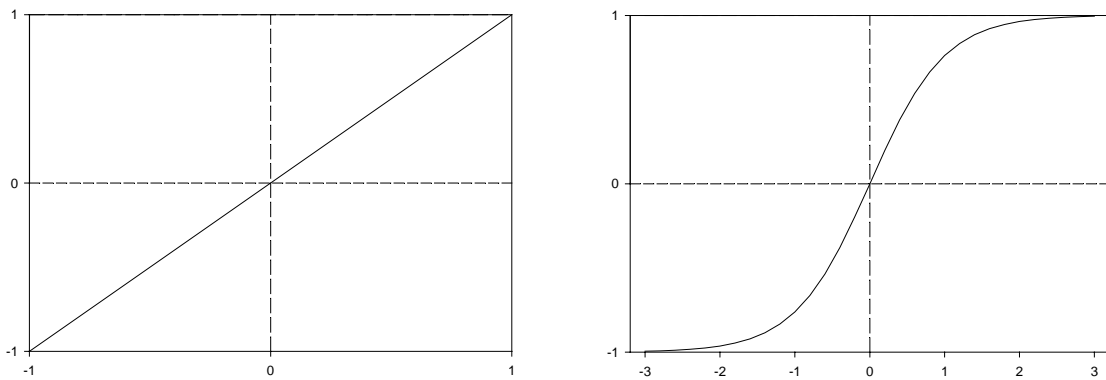


Figure 10-3. Example of a Linear (left) and a Tan-Sigmoidal Activation Function (right)

Mathematically, the output (a_i) of each neuron in Figure 10-2 may be expressed as

$$a_i = f(\mathbf{w}p + b) \quad (10-1)$$

where \mathbf{w} is a matrix of the weights, b is the bias, and f is the activation function.

During training of the NN, the values for the weights and biases are determined as a part of a nonlinear optimization problem, where an objective function is minimized while values of the weights and biases are varied. A sum of squared errors is typically utilized as an objective function and is expressed as:

$$E_D = \sum_{i=1}^n (t_i - a_i)^2 \quad (10-2)$$

where a_i is the network response (output) and t_i is the target value. Many different methods have been employed for the optimization of the objective function, including gradient descent, line searches, conjugate gradient, quasi-Newton, and Levenberg-Marquardt.

Example Neural Networks

While a large number of structurally different NNs exist, back propagation (BP) is one of the most popular choices. A variant of BP is a radial basis function (RBF) network. These two networks will be described briefly.

Back Propagation

The back propagation network is a feed forward multi-layered network employing supervised learning for adjustment of the weights. Once calculation results are passed from each layer until an output is generated, the error between the output and target values is calculated. This error is returned to the network, which then uses the

information to adjust the weights. After sufficient iteration of this process, a predefined tolerance is eventually met.

While the BP is capable of very good accuracy, there are disadvantages to this technique.

1. The exact number of hidden layers and neurons in the hidden layers is unknown, and several repetitions using different BP architectures are used to determine an optimum network structure.
2. Since optimization of the objective function involves the calculation of a gradient vector to move along the error surface, there are difficulties in knowing what size step to employ for each move of the gradient vector. Small steps will drastically increase the time until a solution is reached, and large steps may over step the solution.
3. Most importantly, BP networks are prone to over-learning, where the error associated with the input set is minimized, and predictions are poor using a new data set with the trained network.

Generalization is the ability of a network to accurately predict the values of a new data set, and an analogy can be drawn between polynomial curve fitting and generalization. When using a polynomial equation to fit data, a low order polynomial may not be flexible enough to fit the data accurately, but a higher order polynomial may become highly convoluted in order to fit the data at the expense of representing the underlying function. Similar to this, NNs with more weights are able to model complex functions with low error, but they are prone to over-learning. If fewer weights are used, the resulting model may have improved generalization capabilities, but relation to the

underlying function present in the input data is decreased. While using different BP network structures can result in better generalizations, regularization strategies, such as Bayesian methods, are available to alleviate over-learning.

Radial Basis Function

An alternative to the BP network is the use of a RBF network, which shares the same general architecture as a BP with a similar flow of information. The two networks differ in the choice of a hidden layer activation function, which is a fixed Gaussian function in a RBF and a general nonlinear function in a BP. While the speed of the network is improved from that of BP, the RBF can be complicated to train, and the network will learn incorrect patterns as quickly as correct patterns.

Neural Network Architecture

Appearing frequently in the literature, many types of networks, which include both newly developed networks and variants of known networks, have been applied to a variety of research interests. Taxonomic categories, such as learning algorithm, network topology, and data type, may be applied to organize these networks.

Learning Algorithm

Supervised and unsupervised learning are the two main types of learning algorithms utilized in NNs. Frequently, the type of learning algorithm applied may be difficult to classify.

With supervised learning, the NN is given the expected results or target values. During training, the NN weights are adjusted in an attempt to match the output values with the target values. After training, the NN is validated by providing a new set of input

values and observing the difference between the resulting output values and the correct target values. Supervised learning algorithms are further classified as auto-associative, where the input values and target values are equal, and hetero-associative, where the target values differ from the input values.

In unsupervised learning, the NN is not provided with any target values during training; however, input and target values are often equal; thus, these networks perform as auto-associative networks [234, 235]. Normally, these types of algorithms are employed for data compression, and two widely used algorithms are vector quantization (or "Kohonen network") [236, 237], and Hebbian learning [231]. Another example of unsupervised learning is Kohonen's self-organizing (feature) map [238], which combines competitive learning with dimensionality reduction by separation of clusters on an *a priori* grid.

Network Topology

Feedforward and feedback comprise the two types of network topology. The connections between neural units in a feedforward NN do not form cycles. This enables the NN to quickly respond to an input. Feedforward networks are trained with a wide variety of conventional numerical methods such as conjugate descent gradients, and Levenberg-Marquardt, but these methods do not guarantee a global optimum solution. To avoid local minima, conventional methods are employed with a variety of random starting points, or more complicated methods, simulated annealing and genetic algorithms, may be utilized to find a global optimum directly.

In a feedback NN, cycling occurs in the connections between neural units, which may produce slow responses to inputs. The number of cycles may be large, and these networks tend to be more difficult to train than feedforward NNs.

Data Type

Neural networks differ in the type of input data used, and the two types are categorical variables and quantitative variables. Categorical variables take on a number of possible values or symbolic values (classifications such as male, or female) with several members of each category present in the network. When supervised learning with categorical target values, or unsupervised learning with categorical outputs is used, these types of NNs are known as classification networks.

Quantitative variables involve the measurement of some attribute of an object, such as boiling points of compounds. These measurements should reflect analogous relationships among the input values of the objects. When supervised learning with quantitative target values is employed, these types of NNs are known as regression networks. Some variables can be treated as either categorical or quantitative, such as number of children or any binary variable. Organized tables of taxonomically classified example NNs, which were adapted from Sarle [234], are provided in Tables 10-1 - 10-3.

Table 10-1. Supervised Neural Networks

Supervised Neural Networks	
Feedforward	
Linear	
Hebbian	- [162, 228]
Perceptron	- [164, 166, 228, 239]
Adaline	- [165, 228]
Higher Order	- [227]
Functional Link	- [240]
MLP: Multilayer perceptron	- [227, 228, 241]
Backprop	- [242]
Cascade Correlation	- [228, 243]
Quickprop	- [244]
RPROP	- [245]
RBF networks	- [227, 246, 247]
OLS: Orthogonal Least Squares	- [248]
CMAC: Cerebellar Model Articulation Controller	- [249, 250]
Classification only	
LVQ: Learning Vector Quantization	- [228, 251]
PNN: Probabilistic Neural Network	- [228, 252-254]
Regression only	
GNN: General Regression Neural Network	- [255-257]
Feedback	- [231, 258]
BAM: Bidirectional Associative Memory	- [228, 237]
Boltzman Machine	- [185, 228]
Recurrent time series	
Backpropagation through time	- [259]
Elman	- [260]
FIR: Finite Impulse Response	- [261]
Jordan	- [262]
Real-time recurrent network	- [263]
Recurrent backpropagation	- [228, 264]
TDNN: Time Delay NN	- [265]
Competitive	
ARTMAP	- [266]
Fuzzy ARTMAP	- [267, 268]
Gaussian ARTMAP	- [269]
Counterpropagation	- [228, 236, 270, 271]
Neocognition	- [228, 272]

Table 10-2. Unsupervised Neural Networks

Unsupervised Neural Networks- [231]	
Competitive	
	Vector Quantization
	Grossberg - [273]
	Kohonen - [274]
	Conscience - [275]
	Self-Organizing Map
	Kohonen - [228, 238]
	GTM: - [276]
	Local Linear - [277]
	Adaptive Resonance Theory
	ART 1 - [228, 278, 279]
	ART 2 - [228, 280]
	ART 2-A - [266]
	ART 3 - [281]
	Fuzzy ART - [282]
	DCL: Differential Competitive Learning - [237]
Dimension Reduction - [283]	
	Hebbian - [162, 228]
	Oja - [284]
	Sanger - [285]
	Differential Hebbian - [237]
Autoassociation	
	Linear autoassociator - [179, 228]
	BSB: Brain State in a Box - [179, 228]
	Hopfield - [171, 228]

Table 10-3. Nonlearning Neural Networks

Nonlearning Neural Networks	
Hopfield - [231]	
Various networks for optimization - [286]	

Chapter 11. Quantitative Structure-Property Relationship Models

11.1 *Introduction*

Water, which is the most common industrial solvent, plays an important role in many areas including separation processes, distillation units, chromatographic systems, waste treatment, and environmental concerns [1-7]. With growing application of biotechnologies, there also exists an increased need for phase equilibria of aqueous systems in those processes [8]. When experimental data are unavailable, thermodynamic models, such as group contribution methods, are used to predict phase equilibrium. The accuracy of these models in predicting infinite-dilution activity coefficients (γ^∞) of aqueous systems is questionable. Moreover, model development is hampered by a lack of (a) γ^∞ data at temperatures above 300 K, and (b) γ^∞ data for water-in-hydrocarbon systems.

Due to the unique molecular structure of water and its attendant physical characteristics, including hydrogen bonding, systems containing hydrocarbons and water often exhibit strong nonideality when compared to systems comprised only of hydrocarbons. The activity coefficient, γ , is a parameter that quantifies the amount of nonideality present in a system. When a component of a hydrocarbon-water binary system is sufficiently dilute, the infinite-dilution activity coefficient, γ^∞ , is reflective of only intermolecular solute-solvent and solvent-solvent interactions without the additional

complication of solute-solute interactions. Insight into the chemical and physical forces present in an aqueous system is provided by these coefficients.

The solubility of any solute in a given solvent may be described in terms of the activity coefficients (γ_i) at a given temperature and pressure. For a given temperature and pressure, the mole fraction of a solute (x_i) can be expressed as follows, when the hydrocarbon is at low concentration:

$$x_i = \frac{(p - p_j^\circ)}{P_i^\circ \gamma_i^\infty - P_j^\circ} \quad (11-1)$$

where p is system pressure, p° is the pure vapor pressure and γ_i^∞ is the infinite-dilution activity coefficient. The subscripts i and j indicate the solute and the solvent, respectively. In deriving this relation, we assume low-pressure operations, where ideal-gas behavior applies to the vapor phase.

While several experimental methods exist for the investigation of infinite-dilution activity coefficients, these methods often suffer serious limitations [9-11] and are time consuming. Models for the prediction of infinite-dilution activity coefficients would be useful and are represented by examples from theoretical regular solution theory models [12-18], equation-of-state models [19], pure-component models [20-22], group contribution models [23-25], empirical models [26-31], the LSER model [32, 33], and computational chemistry models [34-39]. These models generally do not provide satisfactory predictions, and quantitative structure-property relationship (QSPR) studies have been limited to the modeling of only single-temperature data of one component of the aqueous systems.

The molecular structure of a chemical compound determines its chemical and physical properties. Continuing investigation has centered on elucidation of the relationship between physical properties and molecular structure. As computational capabilities have improved, research has revolved around developing free energy relationships for property prediction by molecular mutation using Monte Carlo (MC) simulators [40]. Although this approach remains attractive, Monte Carlo is being replaced in many applications by QSPR models. The QSPR approach often provides predictions for chemical and physical properties of as-yet-unmeasured and yet-to-be-synthesized compounds based on structure information. High quality property predictions are obtained using these descriptors since structure-property mapping is at an atomic level rather than at a functional group level. QSPR models will be influential in enabling advances in chemical design, where a key challenge is the development of tools permitting the rapid design of unique molecules. Over the last ten years, QSPR have played an increasingly important role in drug screening and discovery [41], and its applications are appearing in areas outside the pharmaceutical industry. While standard methodologies for chemical design result in a discovery phase of research and development typically require from two to three years, QSPR methodologies are estimated to require only three to six months.

The objectives of this work were to (a) develop a quantitative structure property relationship (QSPR) for prediction of γ_i^∞ values of hydrocarbon-water systems, (b) evaluate the efficacy of QSPR models, using multiple linear regression analyses and back propagation neural networks, and (c) evaluate the ability of the model to predict aqueous and hydrocarbon solubility at multiple temperatures.

11.2 Database Development

The database, which was developed from 96 journal literature sources dating from 1927 to 1995, consists of 1400 infinite-dilution activity coefficients (IDAC's) at temperatures ranging from 283.15 K to 373.15 K for a diverse set of structural classifications [42]. As a result of a literature search, this database was the most recent and extensive compilation available at the time of development. Data available consist of both hydrocarbon-in-water and water-in-hydrocarbon IDAC's. The water-in-hydrocarbon data were collected with direct measurement methods, and the hydrocarbon-in-water data were collected by either direct measurement or indirect measurement methods, and both datasets were classified with reference to experimental method. Examples of direct measurements are gas-liquid chromatography method (GLC), headspace GLC method, gas-stripping method, liquid-liquid chromatography method, differential ebulliometry method, and differential static method. Included under the general title of GLC methods are stationary phase GLC, non-steady-state GLC, and relative GLC. The indirect measurements include extrapolations to infinite dilution of vapor-liquid equilibrium data and calculations from other thermodynamic data, such as liquid-liquid equilibrium data and gas-liquid partition coefficient data.

Where provided by the source material, the database also contains uncertainty ("error") estimates. These error estimates were used to form error bars by taking the data point value \pm the error estimate. Table 11-1 and Table 11-2 provide a numerical analysis of the database and a list of the different hydrocarbon structures in the database, respectively.

The database was used as the basis for six case studies for this investigation. The first three case studies, CS1-A, CS1-B, and CS2, consisted of all available data, including error estimates, for each of the three sections of the database; direct, indirect, and water-in-hydrocarbon, respectively. The fourth case study, CS3, used data from the entire database, but did not include the error estimates due to a software limitation on the number of allowed values. The fifth and sixth case studies, CS4 and CS5 respectively, involved only matched hydrocarbon-in-water and water-in-hydrocarbon data. For example, a measurement of hexane in water, whether from the direct or indirect set, must have a corresponding measurement of water in hexane for inclusion in both CS4 and CS5. While the number of data points is the same in both cases, CS4 uses only the hydrocarbon molecular structures to represent both the hydrocarbon-in-water and water-in-hydrocarbon data, and CS5 uses the molecular structure of water to represent the water-in-hydrocarbon data. Regardless of the type of measurement, whether hydrocarbon-in-water or water-in-hydrocarbon, the hydrocarbon molecular structure was used exclusively in the other case studies, CS1-A, CS1-B, CS2, and CS3. A summary of the case studies is available in Table 11-3. After the initial step in QSPR model development, the CS1-A and CS1-B case studies were combined to form a case study, CS1, comprised of all hydrocarbon-in-water data.

The molecular structures included in the database were prepared in the following manner:

1. Molecular structures were drawn and optimized using the MMX molecular mechanics force field module available in ChemDraw Ultra [43].
2. 2D structures were generated using ChemDraw Ultra.

3. Chem3D Pro [44] was employed to generate 3D molecular structures from exported 2D structures.
4. These structures were initially optimized using the MOPAC [45] module available in Chem3DUltra.
5. The “pre-optimized” structures were submitted to the AMPAC 6.0 [46] program for further geometry refinement and for the calculation of molecular orbital parameters. The AM1 parameterizations were used to calculate the quantum-chemical molecular descriptors.
6. Output from AMPAC was used in CODESSA [47] to calculate various molecular descriptors.

In addition to a small number of constructed descriptors (described below), over 1400 descriptors from such categories as constitutional, topographical, geometric, electrostatic, quantum chemical, and thermodynamic [48] were generated for each molecular structure and are briefly described as follows:

1. Constitutional Descriptors: These simple descriptors reflect only the molecular composition of the compound without using the geometric or electronic structure of the molecule e.g., number of atoms, number of bonds, number of rings, and molecular weight.
2. Topological Descriptors: These descriptors provide the atomic connectivity in the molecule, which include molecular connectivity indices, substructure counts, molecular weights, weighted paths, molecular distance edge descriptors, kappa indices, electro topological state indices, and many other graph invariants [49, 50].
3. Geometric Descriptors: These descriptors are calculated to encode the 3D aspects of the structures and include such descriptors as moments of inertia, solvent-accessible surface area, length-to-breadth ratios, shadow areas, and gravitational index [51, 52].

4. **Electrostatic Descriptors:** These descriptors are calculated to encode aspects of the structures that are electron related, which include partial atomic charges, HOMO energies, LUMO energies, and dipole moment.
5. **Quantum Chemical Descriptors:** These descriptors represent quantum-chemically calculated charge distributions in the molecules. These descriptors may be used to describe interactions between molecules either by a classical point-charge electrostatic model [53] or summation of absolute or squared partial charges [54-56]. The descriptors also provide the value of the partial charge on the atoms in the molecule (e.g., dH_{\min} represents the minimum partial charge on a hydrogen atom). Additionally, these descriptors relate to the strength of intramolecular interactions and characterize the stability of the molecules, their conformational flexibility, and other valency-related properties, such as the maximum bond order (P_{AB}) for a given pair of atomic species A and B in the molecule [57].
6. **Thermodynamic Descriptors:** These descriptors are calculated on the basis of the total partition function (Q) of the molecule and its electronic, translational, rotational, and vibrational components. Examples include molecular vibrational enthalpy, translational enthalpy, vibrational entropy, rotational entropy, internal entropy, translational entropy, and vibrational heat capacity.
7. **Constructed Descriptors:** The descriptors generated by CODESSA do not provide the best modeling approach because functional group descriptors are neglected entirely. However, functional groups have been shown to play an important role in estimating properties [58, 59]. The concept of group contributions is based on the premise that each functional group in the molecule provides either a positive or negative increment to the molecular properties. Specifically, addition of functional groups is likely to alter the properties by increasing the polarizability and possibly the dipole moment of the molecule; thus, these functional groups redistribute electrons, increase or decrease internal strains, and also change the

molecular symmetry and rotational entropy [58-60]. Forty-eight functional group descriptors, which are presented in Table 11-4, were constructed for each molecule. If a functional group descriptor was present in the molecular structure, then that descriptor was assigned a numerical value corresponding to the frequency of the functional group appearance. If a functional group was absent from the molecular structure, then the descriptor was given a zero value. The functional group descriptors were then summed for each datum using either a linear combination or a power law combination of the functional group descriptors to develop a single descriptor, FGorg or FG, respectively. Regressions were done for each case study by minimizing the sum of squared errors between the infinite-dilution activity coefficient value and FGorg or FG. The values of FGorg and FG were then added to the descriptor pool after the regression. Other constructed descriptors consist of mathematical transformations of original descriptors, which included exponential changes and log values of the existing descriptors, and these are shown in Table 11-5.

11.3 QSPR Model Development and Results

Development of a QSPR model for each case study consists of strategies to (a) reduce the number of molecular descriptors and (b) generate a suitable model. The QSPR models can be classified as one of three types, depending on the combination of descriptor selection (linear or nonlinear) and model type (linear or nonlinear). Type I models are completely linear, Type II models are a hybrid using nonlinear descriptor reduction and linear model development, and Type III models are completely nonlinear

Type I Analysis

The Type I analysis employs CODESSA to generate a linear model between the desired property (γ^∞) and selected descriptors, using multiple linear regression. To insure the data were more than adequately described by the model, a greater than normal number of parameters (25 were chosen arbitrarily) were used in a Type I analysis of each case study. The results of these analyses were employed to determine outliers in the data. If there was a deviation greater than two standard deviations, the datum was determined to be an outlier and was eliminated from the case studies. During further QSPR development, no additional data reductions were made. An example of this is shown for the case study, CS3, in Figure 11-1, where the error lines correspond to two-standard-deviation differences, and information concerning the number of outliers for all case studies is provided in Table 11-6. While example figures are provided here, supporting figures for all case studies may be found elsewhere [61].

After elimination of outliers from the data set, the descriptor set was reduced to approximately 200 of the most significant descriptors for each case study by elimination of non-orthogonal descriptors using pairwise correlations employing the following strategy [48] (adapted from CODESSA documentation):

1. All orthogonal pairs of descriptors i and j are found in a given data set.
2. The property analyzed is treated by using the two-parameter regression with the pairs of descriptors, obtained in step 1.
3. For each descriptor pair, obtained in the previous step, a non-collinear descriptor, k is added, and the respective three-parameter regression treatment is performed.

4. For each descriptor set, chosen in the previous step, an additional non-collinear descriptor scale is added, and the respective (n+1) parameter regression treatment is performed.
5. The final product of the above steps is a linear relationship between molecular structure and the property of interest containing n parameters.

CODESSA maximizes R^2 by varying the descriptors in a descriptor set, where the number of descriptors has been specified. CODESSA was then used with the reduced data set and the final descriptor sets using 14, 12, 10, 8, 6, and 4 descriptors (in order to generate R^2 plots) for the determination of the optimum combination of R^2 value and number of descriptors, as used by the Jurs group at Penn State University. Although the R^2 value will continue to increase with an increasing number of descriptors, these plots provide a visual aid to subjectively determine the point at which the R^2 value ceases to significantly increase. Tabular results for all case studies are presented in Table 11-7. The optimum number of descriptors, which are synonymous with model parameters, is shown in an R^2 plot for CS3 in Figure 11-2. From this plot, the optimum number of descriptors used to construct a linear model for CS3 would be ten. While there is still an increase in the R^2 value with more than ten descriptors, significant change in R^2 has ceased by this point, and generally, model construction with the fewest possible descriptors permits the contributions of the individual descriptors to be observed more clearly [39]. Using these 10 parameters, calculated values of γ_i^∞ were generated resulting in a R^2 value of 0.9336, which is similar to the value found when using 25 parameters prior to elimination of outliers. The Type I results for the corrected data in CS3 are presented graphically in Figure 11-3.

Type II Analysis

Type II analysis involves the addition of linear and nonlinear descriptors, descriptor reduction using a genetic algorithm, and linear analysis with CODESSA. Prior to commencement of the Type II analyses, the data set of each case study was randomly divided into a training set, prediction set, and cross validation set composed of 70, 20 and 10%, respectively, of the total number of data in each case study. The prediction set was employed to test the viability of *a priori* predictive capability of the model, and the cross validation set was used as a measure of training in the Type III analyses.

The added descriptors included melting point, boiling point, octanol-water partition coefficient, functional group parameters based on molecular structure, and various mathematical transformations of such descriptors as the molecular weight, gravitational index, and molecular volume. Using the set of approximately 200 descriptors for each case study from the Type I analysis and the additional descriptors, a genetic algorithm in NeuralPower [62] was employed to reduce the descriptor set to 50 descriptors. Descriptor reduction was accomplished in a stepwise fashion where the set is reduced by approximately 25% each time over the course of five iterations of the genetic algorithm. Similar to the Type I analyses, CODESSA was used with the final descriptor set at various specifications of descriptors to generate R^2 plots for the determination of the optimum combination of R^2 value and number of descriptors. The results, %AAD and RMSE (both in $\ln\gamma^\infty$), for the Type II analyses are tabulated in Table 11-8, and plots for the various case studies are presented in Figures 11-4 - 11-8.

Type III Analysis

The twenty most significant descriptors from the Type II analyses were used as a descriptor set for the Type III analyses, which are non linear models employing neural networks, and these descriptors are presented in Table 11-9. A back propagation neural network was used in NeuralPower [62]. The initial weights for the network, type of transfer function, and network architecture were determined through trial and error.

Once a transfer function and architecture were selected, ten replicate analyses using randomized initial weights were performed. During these analyses, the root mean square errors (RMSE) of the training set and cross validation set were monitored as training cycles accumulated. Typically the RMSE of the training set decreases until insignificant changes in the RMSE are realized as training of the neural network progresses; however, allowing training to continue to this point often results in an over-trained network, which generally results in poor predictive capability. By monitoring the RMSE of the cross validation set as the network is trained, the point at which a minimum in the cross validation set RMSE is attained may be identified, which should correspond to the best predictive capability of that network. A contour plot can be constructed using the cross validation RMSE of the replicate analyses, which is utilized in determining the region of least RMSE of the cross validation set. The identified region will be the replicate analysis used for the Type III model.

When a replicate in a contour plot contains an extended “valley” of relatively unchanging RMSE values, such as shown by replicate number six in Figure 11-9, training is halted at a point in which the lowest RMSE value is obtained while using the fewest possible number of training cycles, which reduces computational burden. Calculation of

the %AAD (in $\ln\gamma^\infty$) for the training, prediction, and cross validation sets of replicate six at 42,600 training cycles results in 13.4, 12.9, and 12.4 %AAD, respectively and at 80,000 training cycles resulted in 13.1, 15.8, and 13.8, respectively. As shown by the change in the training set %AAD, these numbers illustrate that improvement in the training set correlation with additional training comes at the expense of a decrease in the predictive capability shown by the increase in %AAD for the prediction and cross validation sets. In this case for CS3, replicate number six with 42,000 training cycles would be selected. After selection of a particular replicate and number of training cycles, the results are obtained for the Type III models. The results for the Type III analyses are provided in Table 11-8, the descriptor sets used with the case studies are shown in, and plots of each case study are presented as Figures 11-10 - 11-14.

11.4 *Discussion*

QSPR Model Development

A critical review of the database was not attempted other than the deletion of outliers prior to the Type I analyses. In some cases there are many values reported which differ significantly for a particular molecule at a given temperature. Examples of this include chloroform ($\ln\gamma^\infty = 6.35-6.91$), 2-butanone ($\ln\gamma^\infty = 3.23-4.19$), and carbon tetrachloride ($\ln\gamma^\infty = 7.96-9.41$) all at 298 K. The effect of many data points may skew training to that particular structure, and the range of the values of a particular molecular structure can be large in magnitude, which results in larger errors. The absolute average deviation in γ^∞ for the prediction set of CS1 is 119%, but with the deletion from the average of just three values, which differ greatly from other reported values for the same

compound, the error is reduced to 79%. While this is an extreme example of the possible magnitude of error, other large errors are included in the error estimates. Since the majority of the data in the database are collected at ambient temperature, the lack of extended temperature data results in network training skewed to ambient temperature data, and there is not a descriptor retained which accounts for temperature dependence. The temperature of the infinite-dilution activity coefficient measurement was used as a descriptor, but this temperature descriptor was eliminated during the Type I analyses due to its insignificance.

Both Type II and III models showed substantial improvement over the Type I models. Non-linear models, Type III, for the case studies investigated showed better performance in predicting infinite-dilution activity coefficients when compared to the linear Type II models, which did employ a descriptor set reduced in a non-linear fashion. On average, the Type III models reduced the absolute average deviation of the predicted set by approximately 30% from the Type II results. Hybrid models, which involve the use of descriptors obtained from linear methods to develop non-linear models, are increasingly being employed due to the decrease in the amount of computational time required when using only non-linear methods.

For the Type III analyses, CS3 showed the lowest predictive error, which was the error only associated with the predictive set, of 12.9% (%AAD in $\ln \gamma^\infty$) and was followed closely by CS1 at 16.4% and CS4 at 19.5%. As mentioned previously, the predictive set of CS1 included some very large errors, which may give an artificially inflated predictive error. The other two case studies, CS2 and CS5, had predictive errors of 25.9 and 24.8%, respectively. Since the error increases with a decrease in the case

study size, the increase in the predictive errors of the case studies may reflect the number of structures involved in the training sets. Attempting to index matched data in CS4 and CS5 and the use of water molecular structure in CS5 resulted in larger errors than the combined data used in CS3. The hydrocarbon-in-water data used in CS1 showed much better fitting (7.6 %AAD) and predictive capability than the water-in-hydrocarbon data used in CS2 (28.5 %AAD for fit), which may be due to both the large number of data points per compound and the number of structures used in training.

The Type III model developed for the most generalized case study for predicting γ^∞ of hydrocarbon-water systems, CS3, provided satisfactory prediction of γ^∞ data (12.9 %AAD in $\ln \gamma^\infty$ and R^2 of 0.992) considering the database employed. The descriptors currently given in the literature and used in software packages do not adequately describe the molecular structure relationship with γ^∞ , but the addition of constructed descriptors improved the model predictions. However, predictions at extended temperatures are still poor. As stated previously, there is not a descriptor accounting for temperature dependence, which results in an average predictive value for a given structure. A possible solution is the provision of a theoretical backbone, or framework, to the model, which accounts for temperature dependence in the data.

While detailed interpretations of the molecular structure – infinite-dilution activity coefficient relationship are beyond the scope of this work, the resulting descriptor set obtained for the Type III analysis did provide insight into the relationship. Since the case studies were comprised of different structures and in some cases different numbers of the same structures, there were differences in the inclusion of particular descriptors in the final descriptor set for each case study. The final descriptor sets included the most

physically meaningful descriptors reflective of the intermolecular interactions, which lead to an infinite-dilution activity coefficient value. Molecular properties, which influence these values, include molecular structure descriptors, polarity descriptors, and descriptors concerned with hydrogen bonding.

Among the final descriptor sets were user-added descriptors including various functional group parameters (FG13, FG14, FG17, FG21, FG24, and FG32) and constructed descriptors based on the functional group parameters of each molecule (FG and FGorg). The appearance of these descriptors emphasizes the importance of physically describing the molecule. Other user-added descriptors such as boiling point (BP), octanol-water partition coefficient (logP), melting point (MP), and (gravitational index)^{0.33} have been shown to be highly correlated [63, 64] to aqueous solubility and, thus, infinite-dilution activity coefficient value.

Additional molecular shape/size information was provided by topological descriptors (Kier & Hall indexes, Randic indexes, and complementary information content) and geometrical descriptors (total molecular surface area (TMSA) and gravitational index). Generally, the remainder of the descriptors was from either the electrostatic or quantum-chemical class. These descriptors are important in the description of the electronic nature of the structure and the hydrogen-bonding capability. Examples of these descriptors include two methods of describing hydrogen acceptors (HA dependent HDCA-1 and HA dependent HDCA-2/SQRT(TMSA)) by either hydrogen donor charged surface area (HDCA) or HDCA divided by the square root of the TMSA, other quantum-chemical descriptors concerned with molecular energies (LUMO

energy, LUMO+1 energy, and min net atomic charge), and the surface weighted charged partial surface area.

BG-QSPR Model

Previous work [65] has shown that an integrated approach employing both theoretical and QSPR models is very capable of predicting saturated vapor pressures of pure fluids over a temperature range. The integrated approach employs a theoretical framework to develop the model that adequately describes the physical behavior of the fluid and QSPR to generalize the parameters in the theoretical model.

A preliminary study was done to develop an improved, integrated QSPR model based on the Bader-Gasem equation of state (BG EOS) [19]. Due to the lack of available extended temperature data, application of this model has previously been limited to a small database, but those initial results of 3.5 %AAD show marked improvement compared to models developed without a theoretical framework.

Using the current database, molecular structures for which data existed at extended temperature ranges were selected for validating the BG EOS as a theoretical backbone for a neural network model. In cases of multiple single temperature data points, averaging was used to generate a single datum, and simple plots of inverse temperature value versus the natural log activity coefficient value, which should be linear at infinite-dilution, were used to provide data consistency over a given temperature range. These plots should be linear according to the definition of an activity coefficient of species i in a solution given by the following equation, which is in the form of an equation for a straight line:

$$\frac{1}{T} = \left(\frac{R}{G_i^E} \right) \gamma_i \quad (11-2)$$

where T is the temperature, R is the universal gas constant, and G^E is the partial excess Gibbs energy. The final set of data for the BG EOS case study included 332 data points, which were randomly assigned to cross validation (10%), prediction (20%), and training (70%) sets, of a diverse collection of 79 molecular structures as shown in Table 11-10.

The two adjustable parameters, C_{12} and D_{12} , in the BG EOS were obtained by regression. These values were then employed in a Type III analysis using the same descriptors and methodology as described previously. As shown in the upper portion of Table 11-11, for C_{12} , the %AAD for the training, prediction, and cross validation sets were 1.21, 1.98, and 0.41, respectively and for D_{12} were 0.17, 0.10, and 0.15, respectively. Two structures, n-methylpyrrolidone and n,n-dimethylformamide, which account for a large portion of the prediction set error, were deleted from the database and not included in the final results. With increasing temperature, these structures have natural log activity coefficient values ranging from negative to positive values, and this behavior is unique to the database. The neural network model did not manage this behavior very well, and almost all the positive, higher temperature, data points were contained in the prediction and cross validation sets, which may have resulted in poor training. For an analogous reason, data from 1,3-butanediol and 1,2-butanediol were deleted due to inadequate training since all of the high temperature data appeared only in the prediction and cross validation sets. Due to the limited amount of higher temperature data available for the various structures in the database, re-randomization of the data into training, prediction, and cross validation sets was not attempted since this problem was expected to reoccur with other structures.

These generated values for C_{12} and D_{12} were used to calculate infinite-dilution activity coefficients with the BG EOS, which were then compared to the infinite-dilution activity coefficient values from the regressed parameters. As shown in Table 11-11, the %AAD in $\ln \gamma^\infty$ for the training, prediction, and cross validation sets was 3.23, 3.35, and 4.26, respectively. Figures 11-15 - 11-17 present the graphical results of the BG EOS case study for C_{12} , D_{12} , and the infinite-dilution activity coefficient, respectively. Details involving these case studies are available from the Thermodynamics Group of the School of Chemical Engineering at Oklahoma State University [61, 66].

Model Comparisons

Model comparisons are presented in Table 11-12. One class of models, which includes UNIFAC and ASOG, involves the calculation of mixture properties using molecular functional groups and their interactions with other functional groups. Zhang et al. [67] developed a UNIFAC modification and compared their predictive results for approximately 400 data points to those from ASOG, UNIFAC, and other UNIFAC modifications. The absolute average deviation in γ^∞ was 12.7% for the Zhang et al. modification compared to 56.1, 65.3, 45.4, 57.0, 55.6, and 53.1% for modified UNIFAC (Dortmund), modified UNIFAC (Lyngby), modified UNIFAC (Hooper), UNIFAC-LLE, UNIFAC, and ASOG, respectively. Another group contribution model, Group Contribution Solvation model (GCS), resulted in a predictive absolute error difference of 0.5 in $\ln \gamma^\infty$ for approximately 50 molecules consisting of alkanes, alkanols, and methylketones at a single temperature of 298 K [68]. A QSPR study of 325 molecules restricted to 298 K [69] resulted in an absolute difference of 0.52 and 0.02 in $\ln \gamma^\infty$ for the training and test set, respectively.

Another class of models, which includes the MOSCED and SPACE equations, is based on regular solution theory and involves the determination of mixture properties from molecular interactions. While an evaluation of MOSCED to aqueous systems is unavailable, Howell et al. [21] reported that for 78 varied solute compound classes in alkanes, alcohols, and alkyl nitriles MOSCED and a modified MOSCED provided overall prediction errors in γ^∞ of 13% and 15%, respectively. Earlier work [70] with the MOSCED model found similar errors when compared to UNIFAC. SPACE, which was developed from MOSCED, was applied to aqueous systems [71], and an error of 70% was found in fitting γ^∞ data at 298K.

In a related study involving prediction of hydrocarbon and water mutual solubilities over an extended temperature range, Klamt [72] used an *a priori* prediction method, conductor-like screening model – real solvents (COSMO-RS). While quantified results were not provided, the model was capable of qualitatively reproducing the hydrocarbon and water mutual solubility trends of 6-10 carbon *n*-alkanes, 1-alkenes, alkylbenzenes, and alkylcyclohexanes. Larger deviations were shown at temperatures below 298 K for hydrocarbon solubility in water and above 473 K for water solubility in water.

The predictive results, with the exception of CS1, for the Type III analyses from this work are better than or similar to the various UNIFAC modifications, with the exception of the Zhang et al. modification, and ASOG results. There are very large errors associated with the directly measured small chain alkanes (C₅-C₈) present in the predictive set of CS1, which contributes to the large overall error. The Zhang et al. modification demonstrated lower predictive errors, but the database used in their work

was limited to data at 298 K and they refined multiple data reported for a single compound. For all the case studies, the absolute average deviation for fitting the γ^∞ data was much smaller than the value found using the SPACE equation.

With use of the BG EOS approach, the absolute average deviations for predictions were 3.35% and 12.4% in $\ln\gamma^\infty$ and γ^∞ , respectively, which demonstrates considerably better predictive capability than any of the models previously discussed with the exception of Zhang et al. modification. A minor improvement is realized when compared to the Zhang et al. modification; however, a significant distinction between the models is the lack of temperature dependence in the Zhang et al. modification, which uses data collected at a single temperature. The BG-QSPR model, which accounts for temperature dependence and is not based on single temperature data, provides improved predictive capability when compared to the single temperature model constructed by Zhang et al.

The integrated BG-QSPR approach, with a predictive absolute difference of 0.11 in $\ln\gamma^\infty$, performs better than the GCS [68] and QSPR [69] work described previously. These results indicate that an integrated approach utilizing a theoretical framework with a QSPR model is an effective method for the prediction of infinite-dilution activity coefficients and, coupled with earlier work [65], provides additional evidence for the validity of an integrated approach.

11.5 References

1. Malmary, G.; Veziar, A.; Robert, A.; Mourgues, J.; Conte, T.; Molinier, J., *J. Chem. Technol. Biotechnol.* **1994**, 60, 67-71.
2. Skrylev, L. D.; Nevinskii, A. G.; Purich, A. N., *Zhurnal Prikladnoi Khimii (Sankt-Peterburg, Russian Federation)* **1984**, 57, 2026-30.
3. Yaws, C. L.; Yang, H. C.; Hopper, J. R.; Hansen, K. C., *Chem. Eng. (N.Y.)* **1990**, 97, 115-16, 8.
4. Yaws, C. L.; Yang, H. C.; Hopper, J. R.; Hansen, K. C., *Chem. Eng. (N.Y.)* **1990**, 97, 177-8, 80, 82.
5. Lazaridis, T.; Paulaitis, M. E., *AIChE J.* **1993**, 39, 1051-60.
6. Li, J.; Dallas, A. J.; Eikens, D. I.; Carr, P. W.; Bergmann, D. L.; Hait, M. J.; Eckert, C. A., *Anal. Chem.* **1993**, 65, 3212-18.
7. Bergmann, D. L.; Eckert, C. A., *Fluid Phase Equilib.* **1991**, 63, 141-50.
8. Belter, P. A.; Cussler, E. L.; Hue, W., *Bioseparations*, John Wiley & Sons, Inc.: New York, 1988.
9. Bader, M. S. H., Ph. D. Dissertation, Oklahoma State University: Stillwater, OK, 1993.
10. Bader, M. S. H.; Gasem, K. A. M., *Chem. Eng. Commun.* **1996**, 140, 41-72.
11. Tiegs, D.; Gmehling, J.; Medina, A.; Soares, M.; Bastos, J.; Alessi, P.; Kikic, I., *Chemistry Data Series, IX, Part 1*, D. Behrens and R. Eckermann; Schoon & Wetzell GmbH, Frankfurt, Germany, 1986.
12. Scatchard, G., *Chem. Rev.* **1931**, 8, 321-33.
13. Hildebrand, J. H.; Wood, S. E., *J. Chem. Phys.* **1933**, 1, 817-22.
14. Weimer, R. F.; Prausnitz, J. M., *Hydrocarbon Processing and Petroleum Refiner* **1965**, 44, 237-42.
15. Hansen, C. M., *J. Paint. Technol.* **1967**, 39, 505-10.
16. Helpinstill, J. G.; Van Winkle, M., *Ind. Eng. Chem. Process Des. Dev.* **1968**, 7, 213-20.
17. Karger, B.; Snyder, L. R.; Eon, C., *J. Chromatogr.* **1976**, 125, 71-88.
18. Tijssen, R.; Billiet, H. A. H.; Schoenmakers, P. J., *J. Chromatogr.* **1976**, 122, 185-203.
19. Bader, M. S. H.; Gasem, K. A. M., *Can. J. Chem. Eng.* **1998**, 76, 94-103.
20. Thomas, E. R.; Eckert, C. A., *Ind. Eng. Chem. Process Des. Dev.* **1984**, 23, 194-209.

21. Howell, W. J.; Karachewski, A. M.; Stephenson, K. M.; Eckert, C. A.; Park, J. H.; Carr, P. W.; Rutan, S. C., *Fluid Phase Equilib.* **1989**, 52, 151-60.
22. Hait, M. J.; Liotta, C. L.; Eckert, C. A.; Bergmann, D. L.; Karachewski, A. M.; Dallas, A. J.; Eikens, D. I.; Li, J. J.; Carr, P. W., *Ind. Eng. Chem. Res.* **1993**, 32, 2905-14.
23. Derr, E. L.; Deal, C. H., *Institute of Chemical Engineering Symposium Series* **1969**, 32, 40.
24. Fredenslund, A.; Jones, R. L.; Prausnitz, J. M., *AIChE J.* **1975**, 21, 1086-99.
25. Gmehling, J.; Li, J.; Schiller, M., *Ind. Eng. Chem. Res.* **1993**, 32, 178-93.
26. Pierotti, G. J.; Deal, C. H.; Derr, E. L., *J. Ind. Eng. Chem.* **1959**, 51, 95-102.
27. Tsonopoulos, C.; Prausnitz, M., *Ind. Eng. Chem. Fund.* **1971**, 10, 593-600.
28. Medir, M.; Giralt, F., *AIChE J.* **1982**, 28, 341-3.
29. Dutt, N. V. K.; Prasad, D. H. L., *Fluid Phase Equilib.* **1989**, 45, 1-5.
30. Yalkowsky, S. H.; Valvani, S. C., *J. Chem. Eng. Data* **1979**, 24, 127-9.
31. Mackay, D.; Shiu, W. Y., *J. Chem. Eng. Data* **1977**, 22, 399-402.
32. Taft, R. W.; Abboud, J. L. M.; Kamlet, M. J.; Abraham, M. H., *J. Solution Chem.* **1985**, 14, 153-86.
33. Sherman, S. R.; Trampe, D. B.; Bush, D. M.; Schiller, M.; Eckert, C. A.; Dallas, A. J.; Li, J.; Carr, P. W., *Ind. Eng. Chem. Res.* **1996**, 35, 1044-58.
34. Shing, K. S., *Chem. Phys. Lett.* **1985**, 119, 149-51.
35. Economou, I. G., *Fluid Phase Equilib.* **2001**, 183-184, 259-69.
36. Nelson, T. M.; Jurs, P. C., *J. Chem. Inf. Comput. Sci.* **1994**, 34, 601-9.
37. Mitchell, B. E.; Jurs, P. C., *J. Chem. Inf. Comput. Sci.* **1998**, 38, 200-9.
38. He, J.; Zhong, C., *Fluid Phase Equilib.* **2003**, 205, 303-16.
39. Huibers, P. D. T.; Katritzky, A. R., *J. Chem. Inform. Comput. Sci.* **1998**, 38, 283-92.
40. Martin, M. G.; Siepmann, J. I.; Schure, M. R., *Unified Chromatography, ACS Symposium Series* **2000**, 748,
41. Kubinyi, H., *Drug Discov. Today* **1997**, 2, 538-46.
42. Kojima, K.; Zhang, S.; Hiaki, T., *Fluid Phase Equilib.* **1997**, 131, 145-79.
43. *Chem3D Ultra* **2000**, Cambridge, MA, CambridgeSoft.com.
44. *Chem3D Pro* **2000**, Cambridge, MA, CambridgeSoft.com.
45. Stewart, J. P. P., *MOPAC Program Package* **1990**,
46. *AMPAC* **2000**, Shawnee Mission, KS, Semichem.

47. Katritzky, A. R.; Karelson, M.; Lobanov, V. S.; Dennington, R.; Keith, T., CODESSA **1999**, Shawnee, KS, Semichem, Inc.
48. Katritzky, A. R.; Lobanov, V. S.; Karelson, M., CODESSA User's Manual, (1994)
49. Kier, L. B.; Hall, L. H., Molecular Connectivity in Structure-Activity Analysis, John Wiley & Sons: New York, 1986.
50. Balaban, A. T., J. Chem. Inf. Comput. Sci. **1997**, 37, 645-50.
51. Karelson, M., Molecular Descriptors in QSAR/QSPR, John Wiley & Sons: New York, 2000.
52. Todeschini, R.; Consonni, V., Handbook of Molecular Descriptors, Wiley-VCH: Weinheim, 2000.
53. Doichinova, I. A.; Natcheva, R. N.; Mihailova, D. N., Eur. J. Med. Chem. **1994**, 29, 133.
54. Bodor, N.; Gabanyi, Z.; Wong, C.-K., J. Am. Chem. Soc. **1989**, 111, 3783.
55. Buydens, L.; Massart, D.; Geerlings, P., Anal. Chem. **1983**, 55, 738.
56. Klopman, G.; Iroff, L. D., J. Comput. Chem. **1981**, 2, 157.
57. Karelson, M.; Lobanov, V. S.; Katritzky, A. R., Chem. Rev. (Washington, D.C.) **1996**, 96, 1027-43.
58. Joback, K. G.; Reid, R. C., Chem. Eng. Communications **1987**, 57, 233.
59. Constantinou, L.; Gani, R., AIChE J. **1994**, 40, 1697.
60. Karelson, M.; Maran, U.; Wang, Y.; Katritzky, A. R., Czech. Chem. Commun. **1999**, 64, 1551-71.
61. Neely, B. J., Ph.D. Dissertation, Oklahoma State University: Stillwater, 2006.
62. NeuralPower **2003**, CPC-X Software.
63. Yalkowsky, S. H.; Banerjee, S., Aqueous Solubility. Methods of Estimation for Organic Compounds, Dekker: New York, 1992.
64. Katritzky, A. R.; Karelson, M.; Lobanov, V. S., Pure Appl. Chem. **1997**, 69, 245.
65. Godavarthy, S. S.; Robinson Jr., R. L.; Gasem, K. A. M., Fluid Phase Equilib. **2006**, 246, 39-51.
66. Godavarthy, S. S.; Robinson Jr., R. L.; Gasem, K. A. M., 2005 Spring AIChE National Meeting, Atlanta, GA, 2005.
67. Zhang, S.; Masaru, H.; Kojima, K., Fluid Phase Equilib. **2002**, 198, 15-27.
68. Nanu, D. E.; Mennucci, B.; de Loos, T. W., Fluid Phase Equilib. **2004**, 221, 127-37.
69. Giralt, F.; Espinosa, G.; Arenas, A.; Ferre-Gine, J.; Amat, L.; Girones, X.; Carbo-Dorca, R.; Cohen, Y., AIChE J. **2004**, 50, 1315-43.

70. Park, J. H.; Carr, P. W., *Anal. Chem.* **1987**, 2596-602.
71. Trampe, D. M., Ph.D. Dissertation, University of Illinois at Urbana: 1993.
72. Klamt, A., *Fluid Phase Equilib.* **2003**, 206, 223-35.

Table 11-1. Numerical Analysis of the Database

Type	Data	Error Data
Hydrocarbon-in-water		
Direct Measurement	776	438
Indirect Measurement	388	0
Water-in-hydrocarbon	236	66

Table 11-2. Database Hydrocarbon Structures

Hydrocarbon-in-water		Water-in-hydrocarbon
Direct	Indirect	
Alkanes	Aliphatic Alkanes	Aliphatic Alkanes
Alkenes	Cyclic Alkanes	Aromatic Hydrocarbons
Aromatic Hydrocarbons	Aliphatic Alkenes	Halogenated Hydrocarbons
Halogenated Hydrocarbons	Cyclic Alkenes	Alcohols
Alcohols	Alkynes	Ketones
Phenol and Derivatives	Monocyclic Aromatics	Acids
Aldehydes	Polycyclic Aromatics	Aldehydes
Ketones	Halogenated Hydrocarbons	Ethers
Acids	Alcohols	Esters
Esters	Phenol Derivatives	Compounds with Nitrogen
Ethers	Ketones	
Amines and Amides	Acids	
Nitriles	Esters	
Nitro Compounds	Ethers	
Compounds with Sulfur	Aldehydes	
	Amines and Amides	
	Nitro Compounds	
	Compounds with Sulfur	

Table 11-3. Summary of Case Studies

Case Study	Description	Number of Values
CS1-A	hydrocarbon-in-water data with error points – direct methods	
Data Values		776
Error Values		438
Total		1214
CS1-B	hydrocarbon-in-water data – indirect methods	
Data Values		388
Total		388
CS2	water-in-hydrocarbon data with error estimates	
Data Values		236
Error Values		66
Total		302
CS3	hydrocarbon-in-water data	
DIRECT		776
INDIRECT		388
WATER		236
Total		1400
CS4	matched data – hydrocarbon structure	
DIRECT		410
INDIRECT		30
WATER		154
Total		594
CS5	matched data – water structure	
DIRECT		410
INDIRECT		30
WATER		154
Total		594

Table 11-4. Functional Group Constructed Descriptors

Descriptor	Functional Group		Descriptor	Functional Group
FG1	CH3		FG25	NH2
FG2	CH2		FG26	NH
FG3	CH		FG27	N
FG4	C		FG28	-N=
FG5	=CH2		FG29	-S-
FG6	=CH		FG30	-CHO
FG7	=C		FG31	COOH
FG8	=C=		FG32	COO
FG9	#CH		FG33	=O
FG10	#C		FG34	-O-®
FG11	CH2®		FG35	O=C®
FG12	CH®		FG36	NH®
FG13	C®		FG37	-N=®
FG14	=CH®		FG38	N®
FG15	=C®		FG39	-S-®
FG16	F		FG40	C#N
FG17	Cl		FG41	NO2
FG18	Br		FG42	SH
FG19	I		FG43	Ortho
FG20	OH		FG44	Meta
FG21	OH (phenol)		FG45	Para
FG22	-O-		FG46	cis
FG23	O=C		FG47	trans
FG24	S=C		FG48	P=O

Where ® indicates a ring structure and # indicates a triple bond

Table 11-5. Constructed Descriptors

Descriptor Name (this work)	Descriptor	Transformation
Con1	Molecular volume	(Molecular volume) ²
Con2	Gravitational index	(Gravitational index)0.33
Con3	Gravitational index	(Gravitational index)0.5
Con4	Molecular weight	log(Molecular weight)
Con5	Relative molecular weight	log(Relative molecular weight)

Table 11-6. Summary of Type I Results

Results	Case Study					
	CS1-A	CS1-B	CS2	CS3	CS4	CS5
R ² with all Data at 25 Parameters	0.980	0.960	0.980	0.935	0.904	0.906
Numbers of Outliers	64	21	17	72	40	46
% of Outliers	5.3	5.4	5.6	5.1	6.7	7.7

Table 11-7. Summary of Type I Results after Outlier Elimination

Results	Case Study					
	CS1-A	CS1-B	CS2	CS3	CS4	CS5
R ² with Reduced Data Set and 25 Parameters	0.988	0.977	0.988	0.969	0.956	0.951
R ² at 14 Parameters	0.979	0.962	0.975	0.949	0.951	0.936
R ² at 12 Parameters	0.976	0.953	0.970	0.937	0.948	0.932
R ² at 10 Parameters	0.971	0.947	0.954	0.934	0.942	0.923
R ² at 8 Parameters	0.962	0.924	0.936	0.920	0.932	0.906
R ² at 6 Parameters	0.943	0.900	0.896	0.898	0.910	0.872
R ² at 4 Parameters	0.888	0.862	0.827	0.787	0.822	0.805

Table 11-8. Summary of Type I, Type II, and Type III Results

	CS1		CS2		CS3		CS4		CS5	
TYPE I										
Descriptors			12		10		10		12	
R ²			0.970		0.934		0.942		0.932	
TYPE II										
Descriptors	6		5		3		3		3	
R ²	0.955		0.903		0.956		0.960		0.893	
Results in γ^∞										
	%AAD	RMSE	%AAD	RMSE	%AAD	RMSE	%AAD	RMSE	%AAD	RMSE
Training Set	22.2	1.04	170.5	0.78	34.1	0.89	31.1	0.55	43.6	0.94
Prediction Set	21.2	0.98	116.0	0.95	17.4	0.68	21.3	0.54	41.9	1.08
Cross Validation Set	17.4	0.87	112.0	0.93	19.1	0.78	20.6	0.61	36.1	1.34
TYPE III										
R ²	0.991		0.965		0.992		0.984		0.949	
Results in γ^∞										
	%AAD	RMSE	%AAD	RMSE	%AAD	RMSE	%AAD	RMSE	%AAD	RMSE
Training Set	30.2	0.45	28.6	0.35	30.4	0.39	23.3	0.39	52.8	0.65
Prediction Set	119.1	0.94	33.4	0.72	35.1	0.52	48.3	0.51	42.8	1.06
Cross Validation Set	71.0	0.82	37.1	0.88	31.7	0.57	33.1	0.44	44.2	1.29
Results in $\ln \gamma^\infty$										
	%AAD		%AAD		%AAD		%AAD		%AAD	
Training Set	7.6		28.5		13.4		14.1		34.3	
Prediction Set	16.4		25.9		12.9		19.5		24.8	
Cross Validation Set	13.8		36.6		12.4		17.6		34.1	

Table 11-9. Descriptors Used in the Type III Analyses

Type III Descriptors				
CS1	CS2	CS3	CS4	CS5
FNSA-3 Fractional PNSA (PNSA-3/TMSA) [Quantum-Chemical PC]	Average Complementary Information content (order 0)	HA dependent HDCA-2/SQRT (TMSA) [Quantum-Chemical PC]	DPSA-3 Difference in CPSAs (PPSA3-PNSA3) [Quantum-Chemical PC]	HA dependent HDCA-2/TMSA [Quantum-Chemical PC]
FG13	BP	FGorg	BP	Number of O atoms
exch. eng. + e-e rep. for a C-O bond	DPSA-3 Difference in CPSAs (PPSA3-PNSA3) [Zefirov's PC]	FG21	count of H-donors sites [Quantum-Chemical PC]	HA dependent HDCA-1 [Zefirov's PC]
FG17	FG17	FG24	FG	FG
BP	Final heat of formation / # of atoms	(Gravitational Index) ^{0.33}	Average Information content (order 0)	1X GAMMA polarizability (DIP)
HA dependent HDCA-2/TMSA [Zefirov's PC]	HA dependent HDCA-1/TMSA [Zefirov's PC]	HA dependent HDCA-1 [Quantum-Chemical PC]	FG14	Gravitation index (all bonds)
HACA-2/TMSA [Zefirov's PC]	HACA-1 [Quantum-Chemical PC]	Complementary Information content (order 0)	FG32	Topographic electronic index (all bonds) [Zefirov's PC]
Min partial charge for a C atom [Zefirov's PC]	PPSA-3 Atomic charge weighted PPSA [Quantum-Chemical PC]	WNSA-1 Weighted PNSA (PNSA1*TMSA/1000) [Quan. Chem. PC]	FPSA-3 Fractional PPSA (PPSA-3/TMSA) [Zefirov's PC]	WNSA-1 Weighted PNSA (PNSA1*TMSA/1000) [Quan. Chem. PC]
Max n-n repulsion for a C-H bond	HACA-2/TMSA [Zefirov's PC]	Kier&Hall index (order 1)	Tot molecular 1-center E-N attraction	Max resonance energy for a C-O bond
Min atomic state energy for a C atom	logP	LUMO energy	HACA-2 [Zefirov's PC]	Max SIGMA-PI bond order
Min e-e repulsion for a O atom	Max resonance energy for a C-Cl bond	LUMO+1 energy	HACA-2/TMSA [Zefirov's PC]	Max total interaction for a C-Cl bond
Min nucleoph. react. index for a F atom	Min e-n attraction for a O atom	Max total interaction for a H-O bond	Max SIGMA-PI bond order	Min nucleoph. react. index for a Cl atom
logP	Min net atomic charge for a O atom	Min net atomic charge	Internal entropy (300K)	Min total interaction for a C-C bond
Min resonance energy for a Br-C bond	Min partial charge for a O atom [Zefirov's PC]	Min n-n repulsion for a C-O bond	Kier&Hall index (order 2)	min(#HA, #HD) [Zefirov's PC]
MP	Polarity parameter (Qmax-Qmin)	No. of occupied electronic levels	logP	Kier&Hall index (order 0)
Randic index (order 1)	HACA-2/SQRT(TMSA) [Zefirov's PC]	Randic index (order 0)	LUMO+1 energy	HA dependent HDCA-1 [Quantum-Chemical PC]
PNSA-1 Partial negative surface area [Zefirov's PC]	RNCS Relative neg. charged SA (SAMNEG*RNCG) [Quan.-Chem. PC]	TMSA Total molecular surface area [Zefirov's PC]	Image of the Onsager-Kirkwood solvation energy	RPCG Relative positive charge (QMPOS/QTPLUS) [Zefirov's PC]
Relative number of O atoms	Topographic electronic index (all bonds) [Zefirov's PC]	Tot molecular 2-center resonance energy	Molecular volume	Tot hybridization comp. of the molecular dipole
Vib enthalpy (300K)	Tot heat capacity (300K) / # of atoms	Vib heat capacity (300K)	Tot entropy (300K)	HACA-2/TMSA [Zefirov's PC]
TMSA Total molecular surface area [Zefirov's PC]	Translational entropy (300K) / # of atoms	Kier&Hall index (order 0)	HA dependent HDCA-2 [Quantum-Chemical PC]	WNSA-3 Weighted PNSA (PNSA3*TMSA/1000) [Zefirov's PC]

Table 11-10. Molecules in the BG EOS Case Study

1,1,1,2-Tetrachloroethane	Acetone	Iso-Butyl alcohol
1,1,1-Trichloroethane	Acetonitrile	m-Cresol
1,1,2,2-Tetrabromoethane	Acrylonitrile	m-Dichlorobenzene
1,1,2,2-Tetrachloroethane	Aniline	Methanol
1,1,2-Trichloroethane	Benzene	Methyl Acetate
1,1-Dichloroethene	Bromoform	Methyl Formate
1,2,3-Trimethylbenzene	Butyl Acetate	m-Xylene
1,2,4-Trimethylbenzene	Butyraldehyde	Nitromethane
1,2-Dibromoethane	Carbon Tetrachloride	o-Cresol
1,2-Dichloroethane	Chlorobenzene	o-Xylene
1,2-Dichloropropane	Chloroform	p-Cresol
1,3,5-Trimethylbenzene	Cis 1,2-Dichloroethene	p-Dichlorobenzene
1,4-Butanediol	Dibromomethane	Pentane
1-Butanol	Dichloromethane	Phenol
1-Chloropropane	Dimethylsulfoxide	Piperidine
1-Propanol	Ethanol	Propionaldehyde
2,3-butanediol	Ethyl Acetate	Propyl Acetate
2,4-Pentanedione	Ethyl Bromide	Propyl Formate
2-Butanol	Ethyl Ether	Propylamine
2-Butanone	Ethyl Formate	p-xylene
2-Butoxy ethanol	Ethyl Iodide	Pyridine
2-Heptanone	Ethylamine	Tert-Butanol
2-Pentanol	Ethylbenzene	Tetrahydrofuran
2-Propanol	Formic Acid	Toluene
3-Pentanone	Heptane	Trans 1,2-Dichloroethene
Acetaldehyde	Hexane	Trichloroethene
Acetic Acid		

Table 11-11. Summary of Bader-Gasem EOS Case Study

Bader-Gasem EOS			
	QSPR Model Results for C_{12}		QSPR Model Results for D_{12}
	%AAD		%AAD
Training Set	1.21		0.17
Prediction Set	1.98		0.10
Cross Validation Set	0.41		0.15
Comparison of BG-QSPR results to experimental values in $\ln \gamma^\infty$			
	C_{12} from QSPR and D_{12} from regression	D_{12} from QSPR and C_{12} from regression	C_{12} and D_{12} from QSPR
	%AAD	%AAD	%AAD
Training Set	1.15	3.25	3.23
Prediction Set	2.49	3.50	3.35
Cross Validation Set	1.86	3.46	4.26
Comparison of BG-QSPR results to experimental values in γ^∞			
	C_{12} from QSPR and D_{12} from regression	D_{12} from QSPR and C_{12} from regression	C_{12} and D_{12} from QSPR
	%AAD	%AAD	%AAD
Training Set	7.90	11.7	15.2
Prediction Set	7.60	11.9	12.4
Cross Validation Set	5.97	13.4	14.4

Table 11-12. Model Comparisons

Literature Model			This Work	
%AAD in γ^∞				
modified UNIFAC (Zhang)	12.7		QSPR	
modified UNIFAC (Dortmund)	56.1		CS1	119.1
modified UNIFAC (Lyngby)	65.3		CS2	33.4
modified UNIFAC (Hooper)	45.4		CS3	35.1
UNIFAC-LLE	57.0		CS4	48.3
UNIFAC	55.6		CS5	42.8
ASOG	53.1		BG-QSPR	12.4
MOSCED	13			
modified MOSCED	15			
SPACE	70			
Absolute Difference in $\ln \gamma^\infty$				
GCS	0.5		BG-QSPR	0.11
QSPR (Giralt)	0.52			

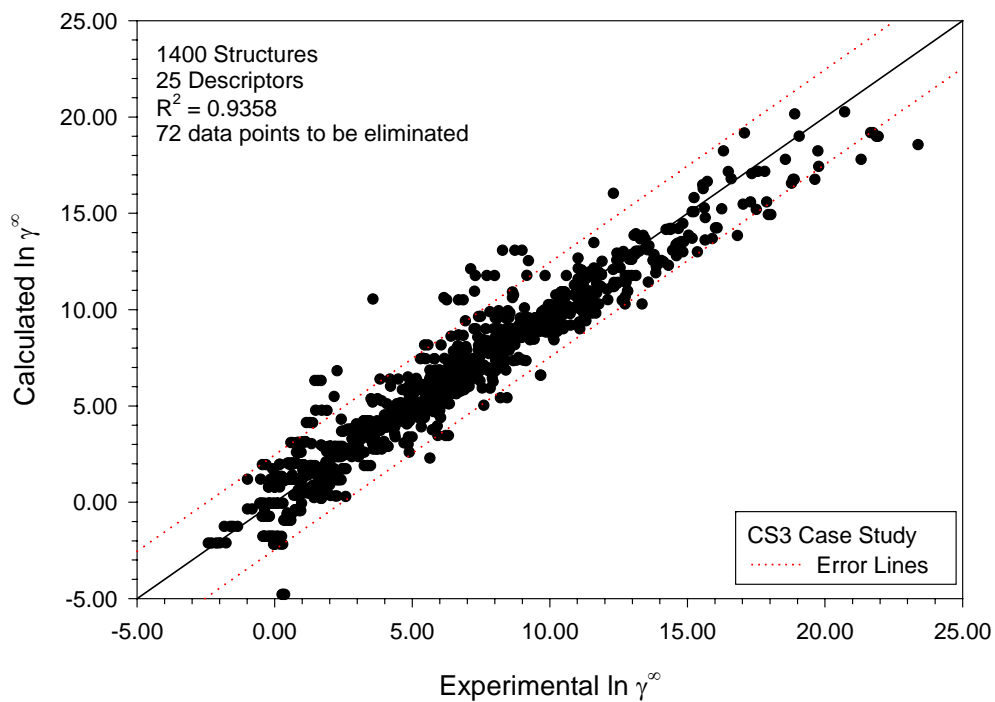


Figure 11-1. Infinite-Dilution Activity Coefficients of CS3 Case Study Showing Outliers

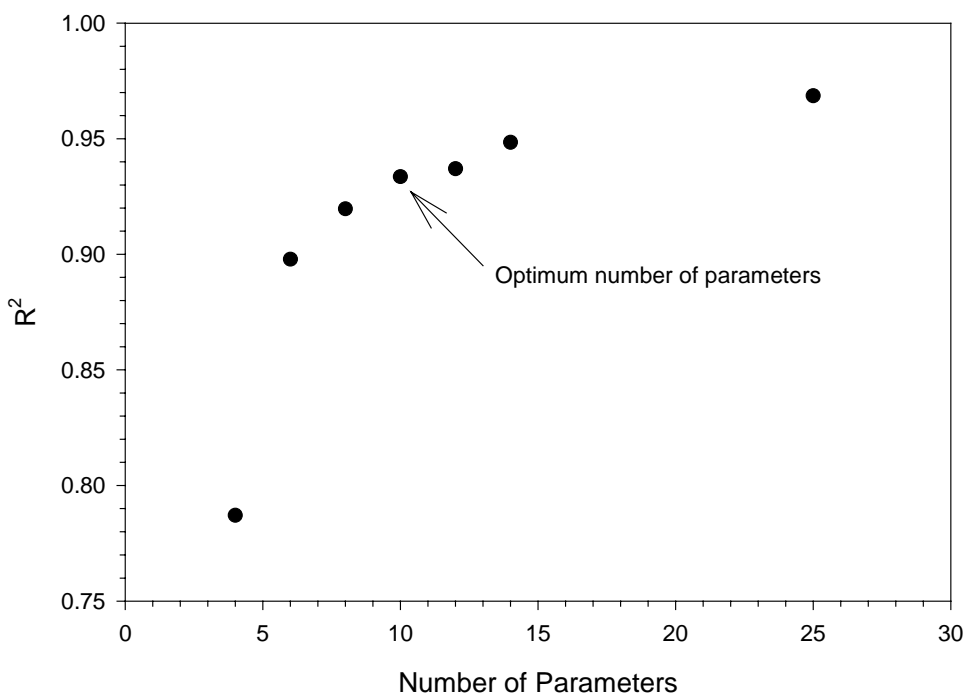


Figure 11-2. Optimum Number of Descriptors for the CS3 Case Study

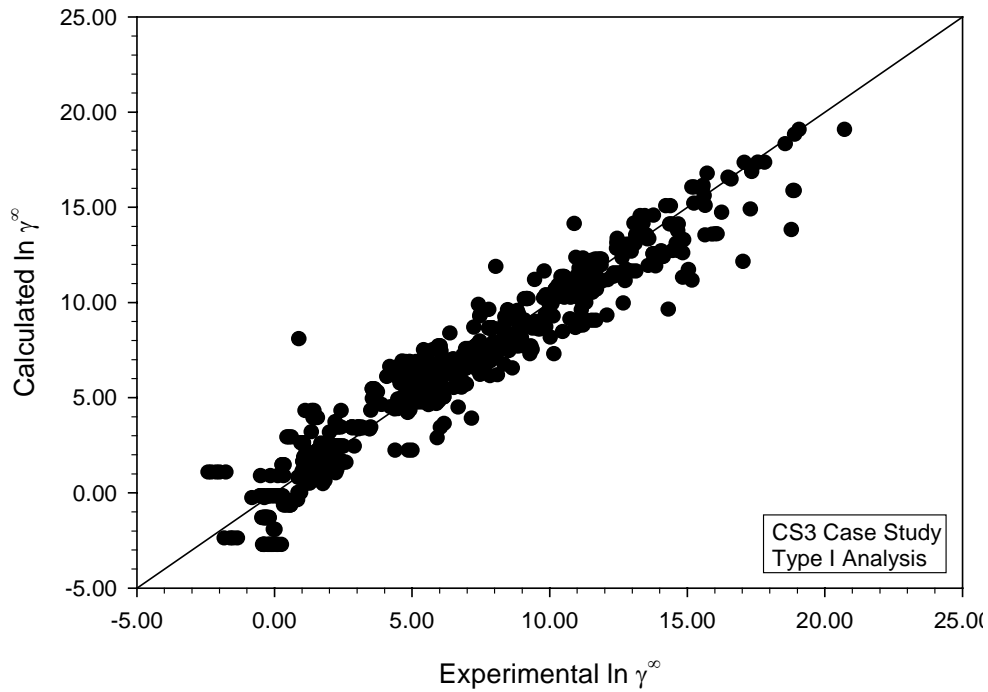


Figure 11-3. Infinite-Dilution Activity Coefficients of CS3 Case Study (Type I) using Ten Descriptors

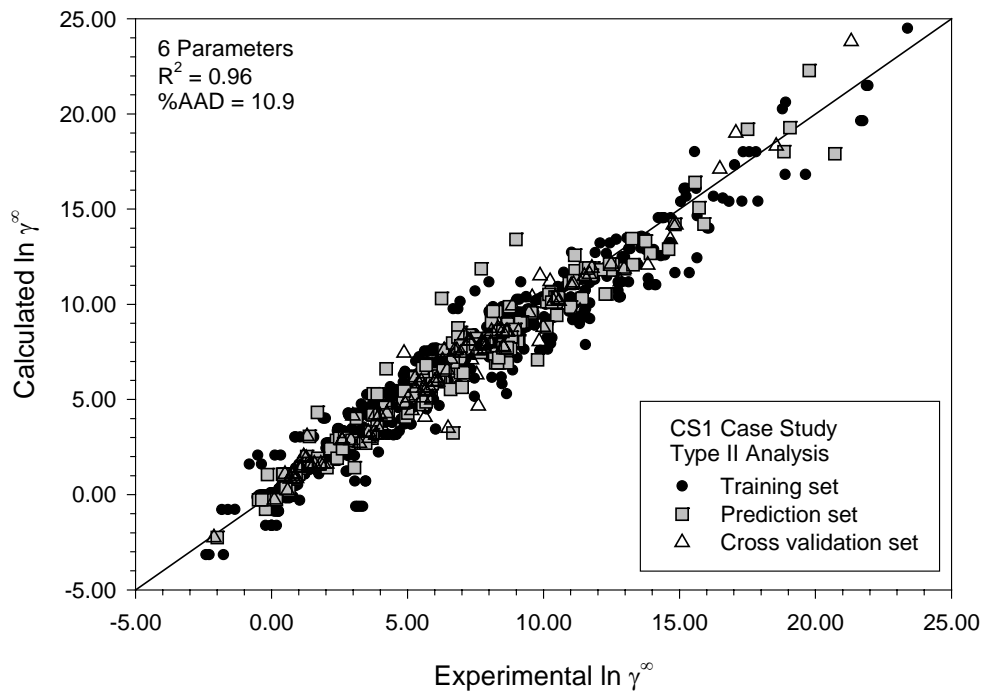


Figure 11-4. Infinite-Dilution Activity Coefficients of CS1 Case Study (Type II)

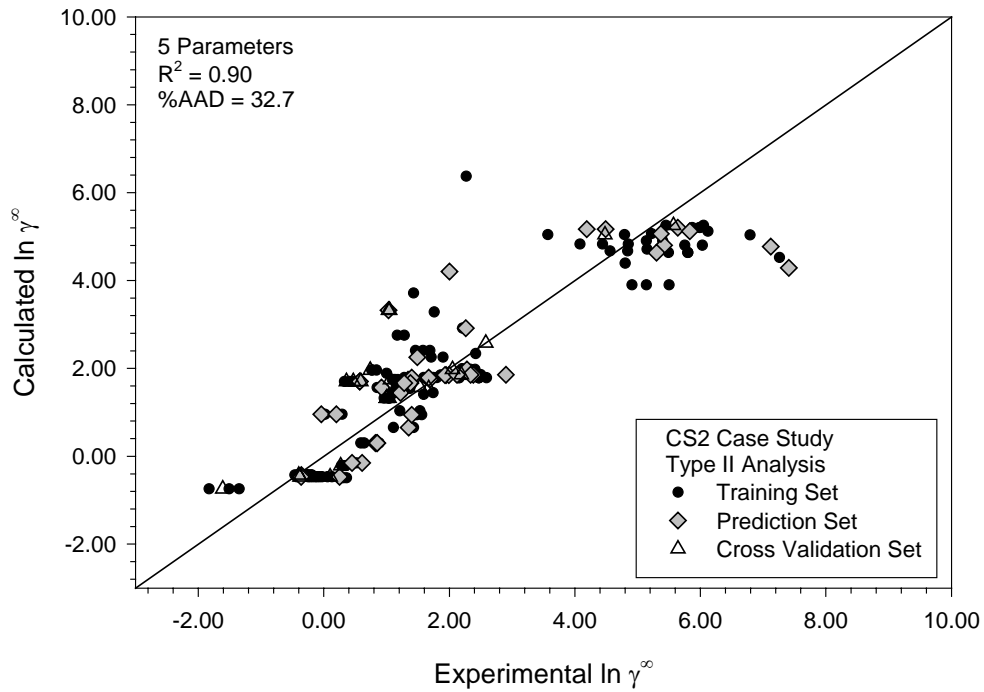


Figure 11-5. Infinite-Dilution Activity Coefficients of CS2 Case Study (Type II)

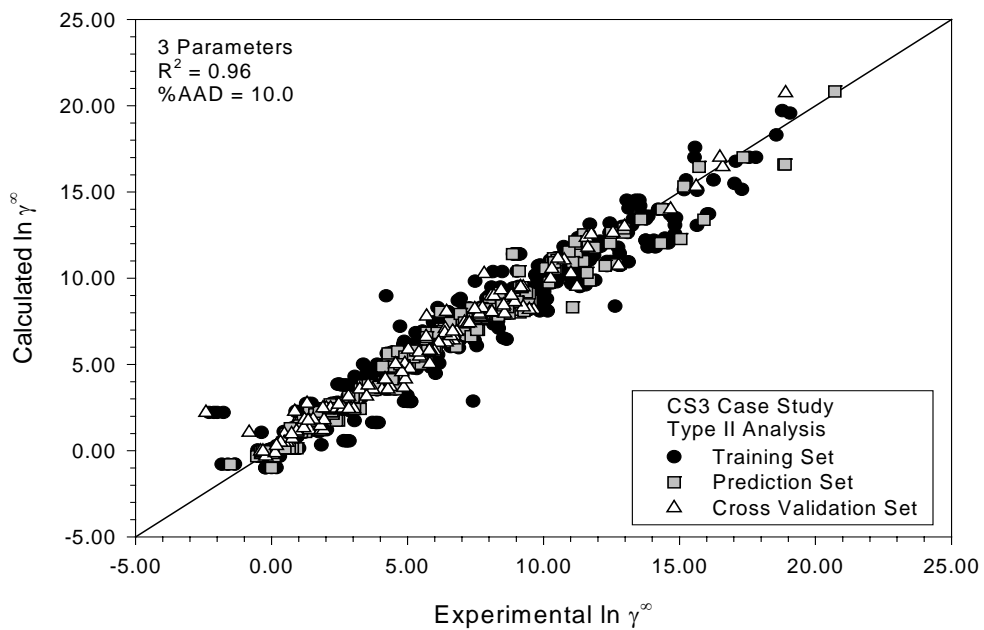


Figure 11-6. Infinite-Dilution Activity Coefficients of CS3 Case Study (Type II)

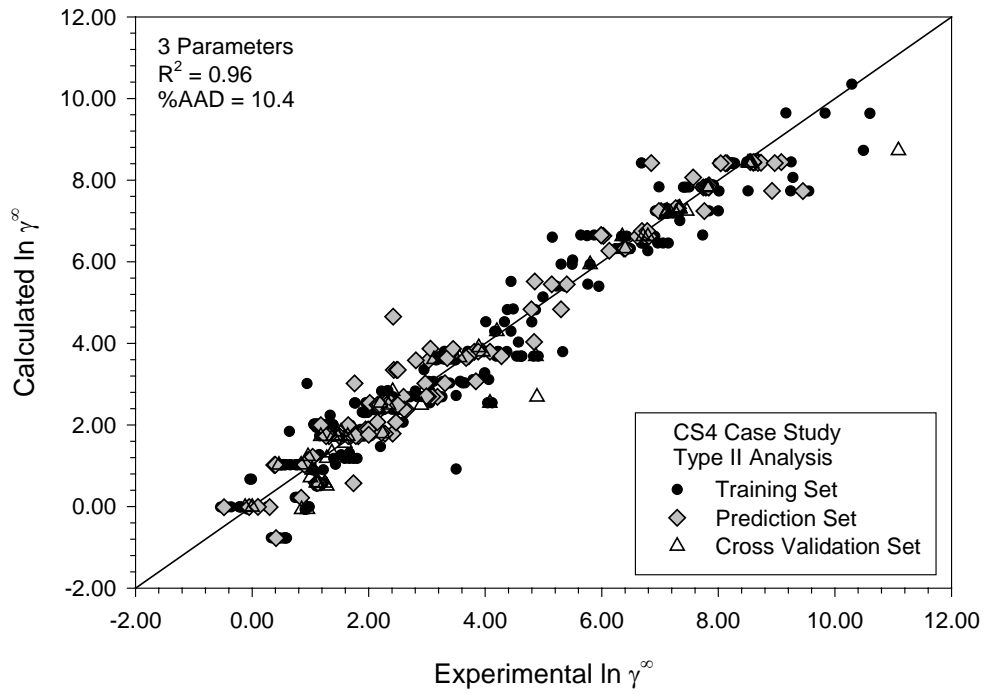


Figure 11-7. Infinite-Dilution Activity Coefficients of CS4 Case Study (Type II)

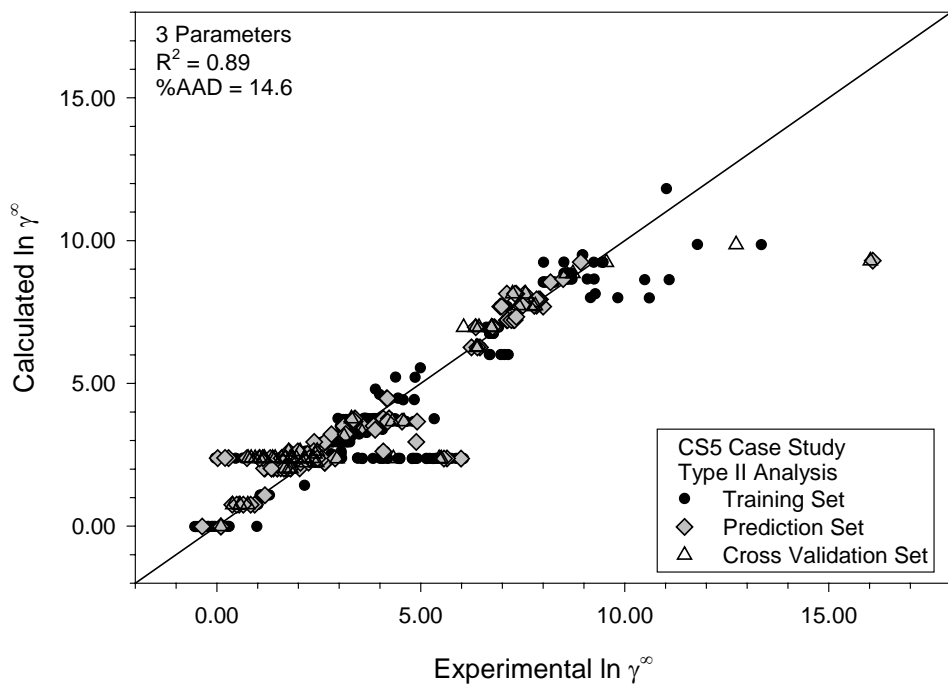


Figure 11-8. Infinite-Dilution Activity Coefficients of CS5 Case Study (Type II)

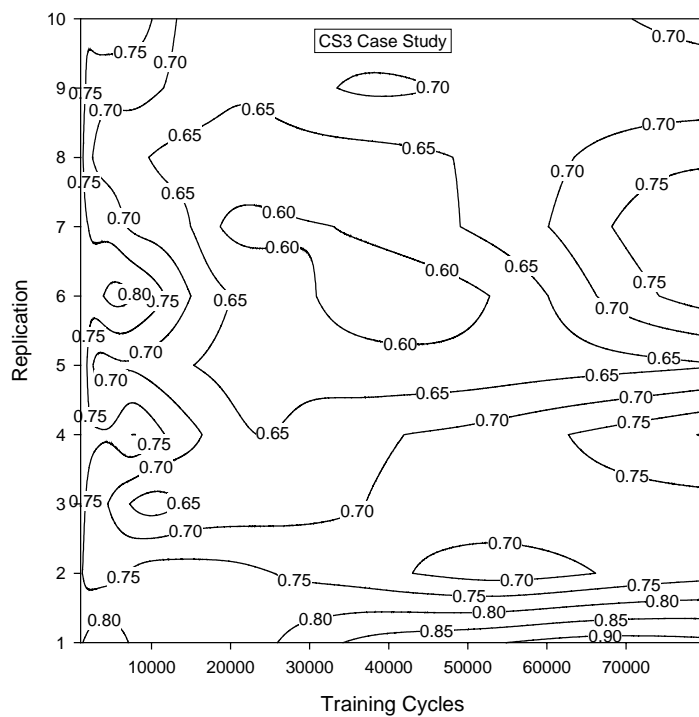


Figure 11-9. Contour Plot of Cross Validation RMSE of CS3

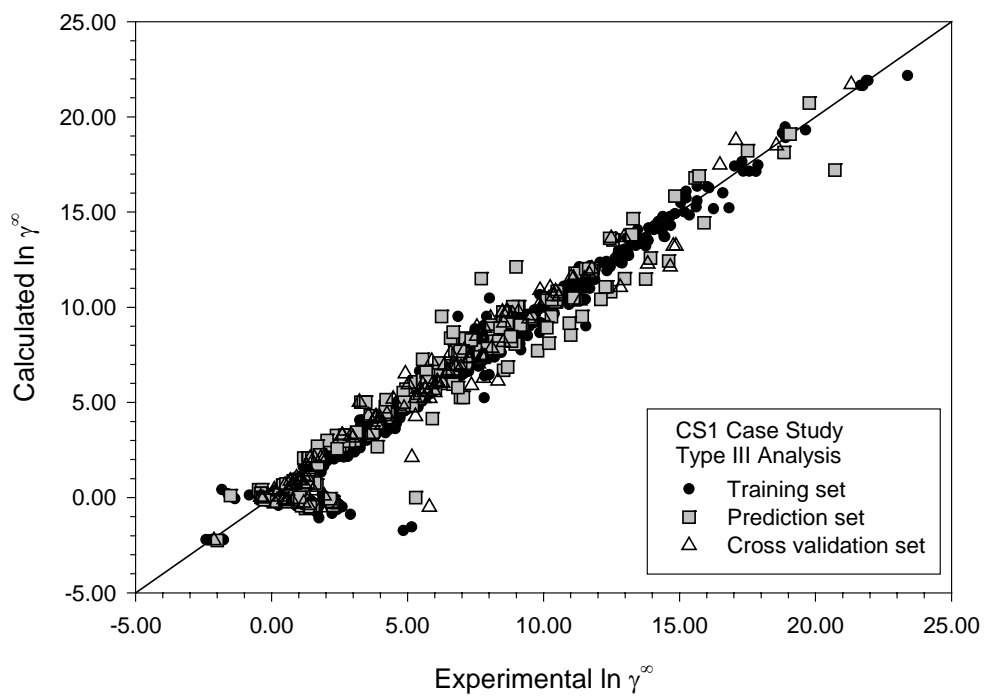


Figure 11-10. Infinite-Dilution Activity Coefficients of CS1 Case Study (Type III)

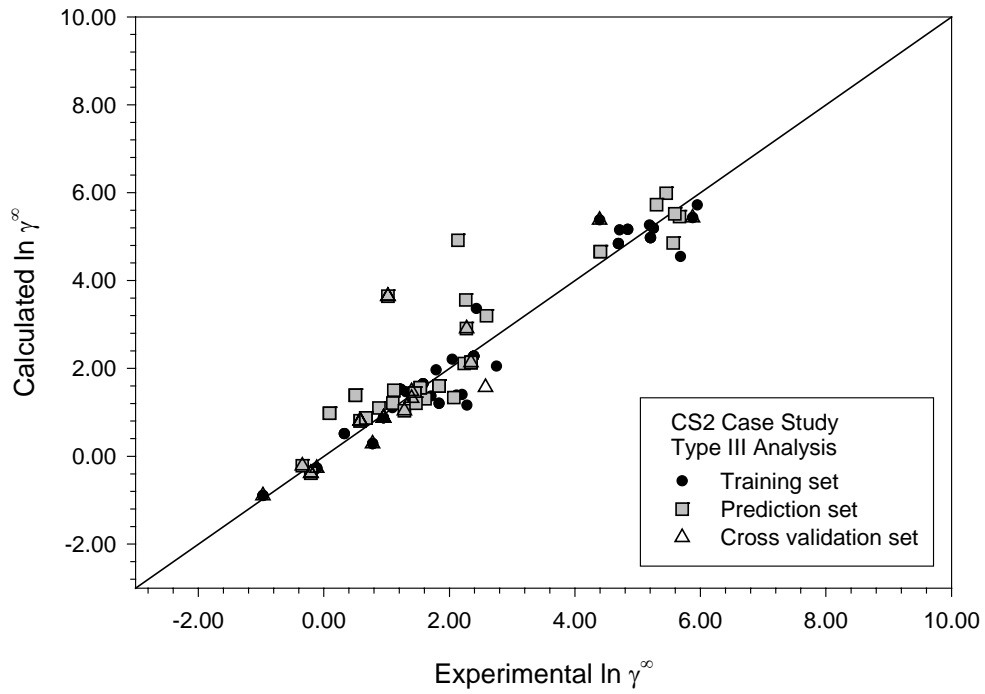


Figure 11-11. Infinite-Dilution Activity Coefficients of CS2 Case Study (Type III)

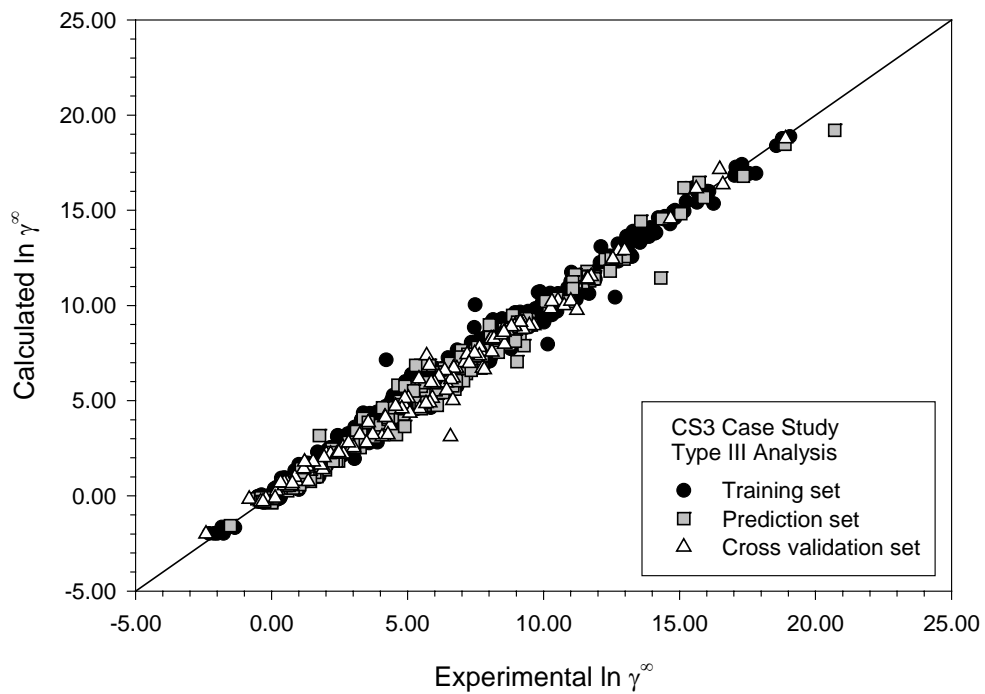


Figure 11-12. Infinite-Dilution Activity Coefficients of CS3 Case Study (Type III)

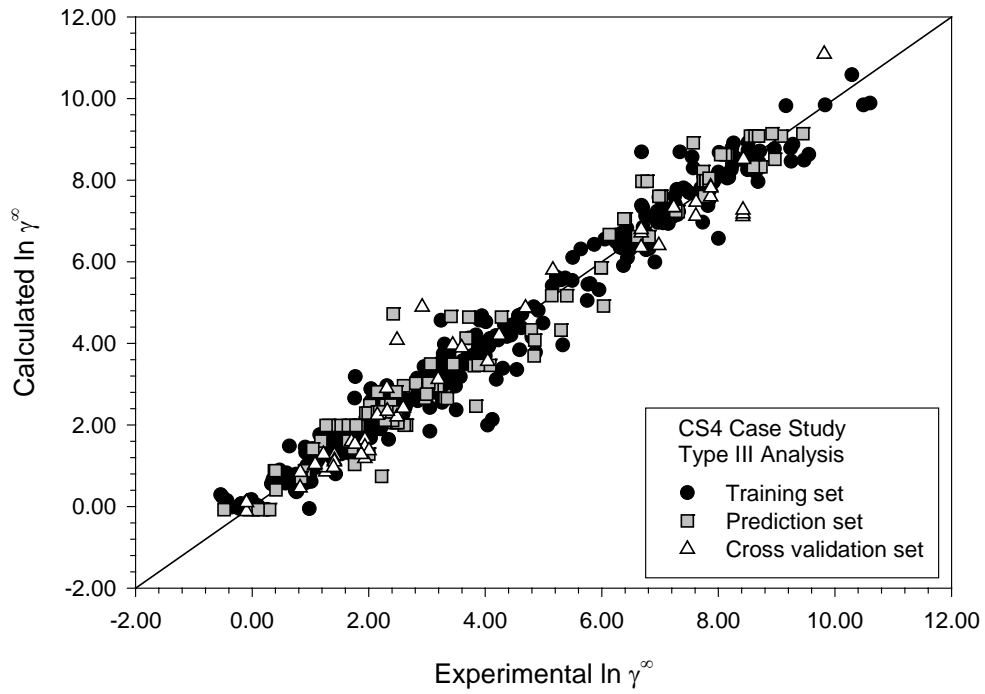


Figure 11-13. Infinite-Dilution Activity Coefficients of CS4 Case Study (Type III)

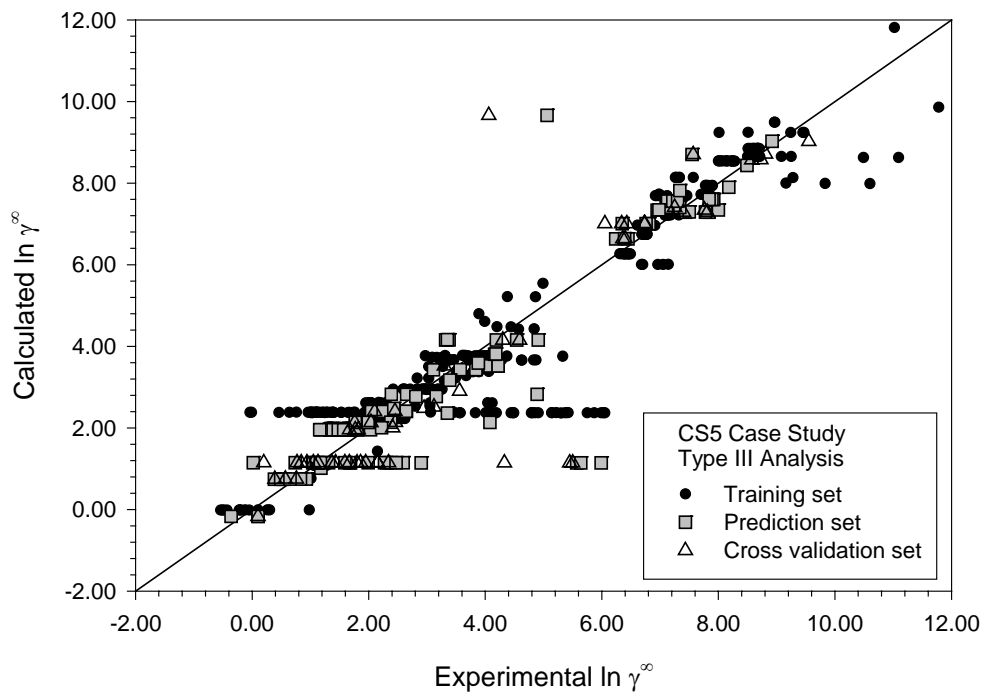


Figure 11-14. Infinite-Dilution Activity Coefficients of CS5 Case Study (Type III)

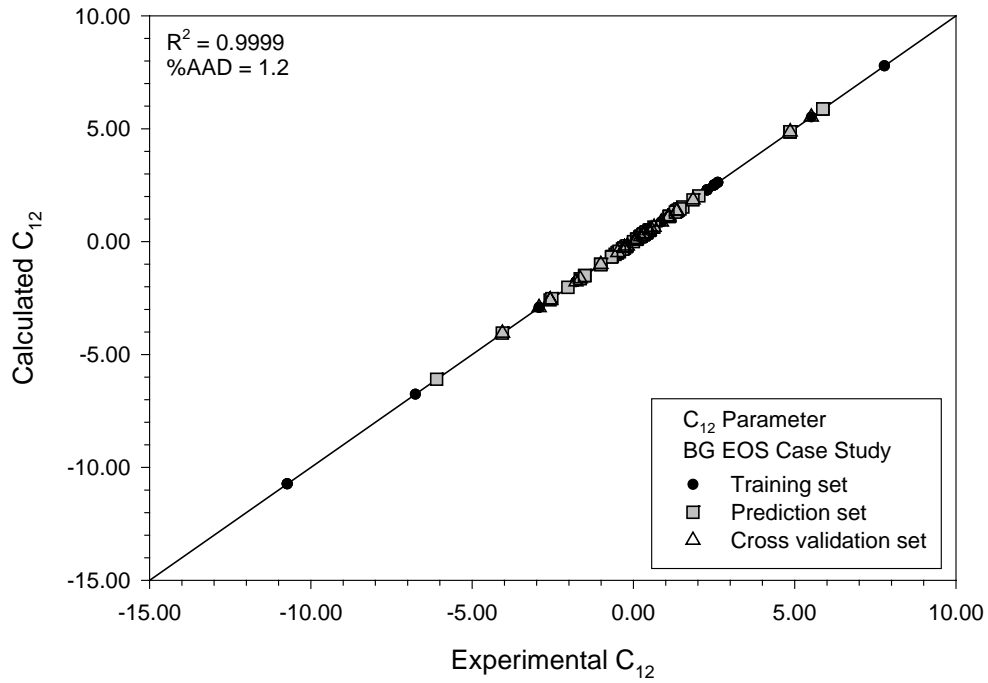


Figure 11-15. C_{12} Parameter of BG EOS Case Study

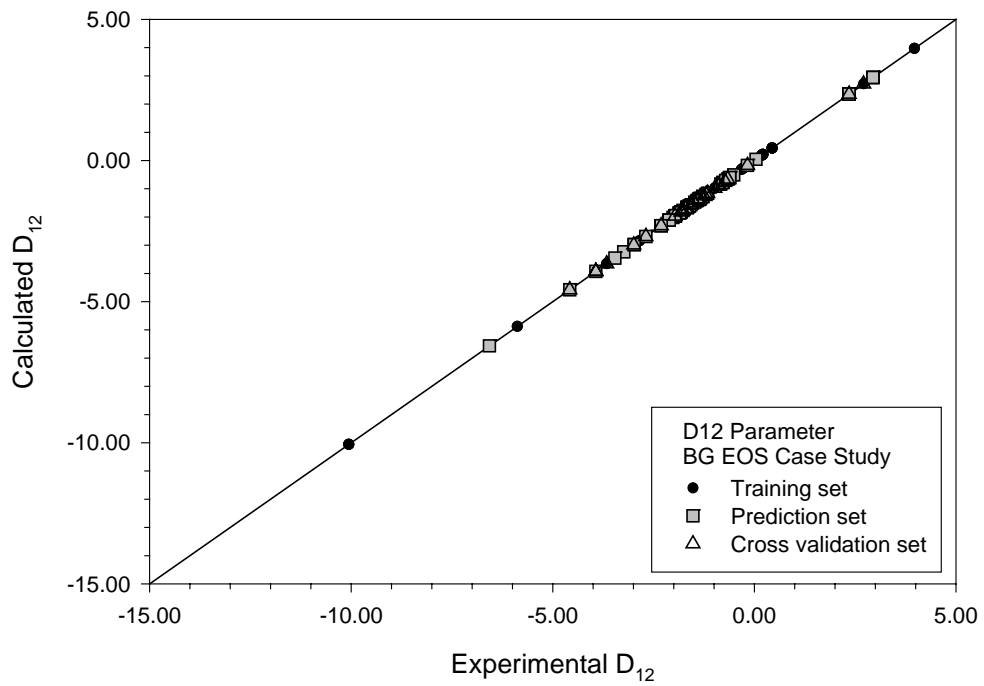


Figure 11-16. D_{12} Parameter of BG EOS Case Study

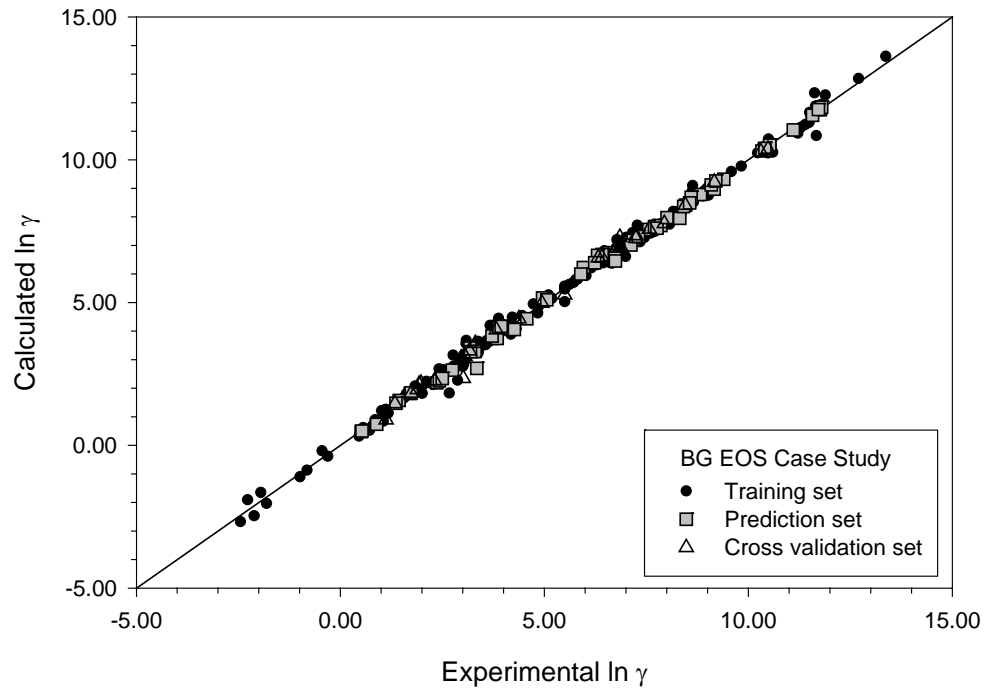


Figure 11-17. Infinite-Dilution Activity Coefficients of BG EOS Case Study

Chapter 12. Conclusions and Recommendations

12.1 *Conclusions*

The objectives of the modeling portion of this work were to (a) develop a quantitative structure-property relationship (QSPR) for prediction of γ_i^∞ values of hydrocarbon-water systems, (b) evaluate the efficacy of QSPR models using multiple linear regression analyses and back propagation neural networks, (c) develop a theory based QSPR model, and (d) evaluate the ability of the model to predict aqueous and hydrocarbon solubilities at multiple temperatures. Based on the present work, the following conclusions may be made:

1. Both Type II (nonlinear descriptor reduction and linear modeling) and Type III (nonlinear descriptor reduction and nonlinear modeling) models demonstrated substantially improved predictions when compared to the Type I (linear descriptor reduction and linear modeling) models.

2. Non-linear models for the case studies investigated showed either better or similar performance in predicting infinite-dilution activity coefficients when compared to the linear Type II models, which did employ a descriptor set reduced in a non-linear fashion.

3. The resulting descriptor set obtained for the Type III analysis provides insight into the relationship between structural molecular features and physical properties of an organic molecule.

4. Among the final descriptor set were user added descriptors including various functional group parameters (FG13, FG14, FG17, FG21, FG24, and FG32), constructed descriptors, which are based on the functional group parameters of each molecule (FG and FGorg), boiling point (BP), octanol-water partition coefficient (logP), melting point (MP), and a mathematical transformation of the gravitational index.

5. A theoretical backbone (using the Bader-Gasem equation-of-state), which accounts for temperature dependence in the data, was combined with QSPR to provide a model which provided significantly improved predictions relative to other available models.

12.2 Recommendations

Hybrid models, which involve the use of descriptors obtained from linear methods to develop non-linear models, are increasingly being employed due to the decrease in computational time required when using only non-linear methods. The Type III models developed in this study were generally satisfactory. For example, the most general model developed for predicting γ^∞ of hydrocarbon-water systems, CS3 provided predictions with 12.9 %AAD ($\ln \gamma^\infty$) and a R^2 of 0.992.

While the prediction of this model compared favorably to the majority of other predictive models found in literature, the model did not account for temperature dependence. An attempt to include a temperature descriptor failed when the descriptor was deemed insignificant in every case study during the Type I analyses. The addition of a theoretical framework, which accounts for temperature dependence, is a viable solution to this problem. Addition of the Bader-Gasem equation-of-state to account for

temperature dependence provided predictions with 3.35 %AAD ($\ln \gamma^\infty$). There are, however, other aspects of this study that, if addressed, may improve the overall predictive and the general nature of the models.

1. The descriptors currently given in the literature and used in software packages do not adequately describe the relationship between molecular structure relationship and γ^∞ , but the use of additional constructed or mathematically transformed descriptors may improve model predictions. Careful analysis of the spectrum of available models may provide useful suggestions or insight in the development of such descriptors.

2. In order to avoid skewing the training during modeling development, a critical review of the database may be necessary to both reduce the number of data points for particular molecules and establish a more accurate value when a wide range in data values exists at a given temperature.

3. The lack of infinite-dilution activity coefficients over wide temperature ranges and of water-in-hydrocarbon data may be remedied by the addition of either new experimental data or “pseudo-data” generated through computational methods such as molecular simulation.

4. To further validate the BG-QSPR model, an additional database for secondary predictions should be assembled.

Literature Cited

1. Tsonopoulos, C.; Wilson, G. M., *AIChE J.* **1983**, 29, 990-9.
2. Anderson, F. E.; Prausnitz, J. M., *Fluid Phase Equilib.* **1986**, 32, 63-76.
3. Kudchadker, A. P.; McKetta, J. J., *Hydrocarbon Process. Petrol Refiner* **1962**, 41, 191-2.
4. Krasnoshchekova, R. Y.; Gubergrits, M. Y., *Neftekhimiya* **1973**, 13, 885-8.
5. Schwarz, F. P., *Anal. Chem.* **1980**, 52, 10-5.
6. Sanemasa, I.; Araki, M.; Deguchi, T.; Nagai, H., *Chem. Lett.* **1981**, 225-8.
7. Chandler, K.; Eason, B.; Liotta, C. L.; Eckert, C. A., *Ind. Eng. Chem. Res.* **1998**, 37, 3515-8.
8. Jou, F.-Y.; Mather, A. E., *J. Chem. Eng. Data* **2003**, 48, 750-2.
9. Marche, C.; Delepine, H.; Ferronato, C.; Jose, J., *J. Chem. Eng. Data* **2003**, 48, 398-401.
10. Miller, D. J.; Hawthorne, S. B., *J. Chem. Eng. Data* **2000**, 45, 78-81.
11. Neely, B. J.; Ratzlaff, D. W.; Wagner, J.; Robinson, J., R.L.; Gasem, K. A. M., *Integrated Petroleum Environmental Consortium*, Albuquerque, NM, 2000.
12. Guseva, A. N.; Parnov, E. I., *Vestn. Mosk. Khim.* **1963**, 18, 76-9.
13. Englin, B. A.; Plate, A. F.; Tugolukov, V. M.; Pyranishnikova, M. A., *Chem. and Tech. Fuel and Oil* **1965**, 10, 722-6.
14. Bittrich, H. J.; Gedan, H.; Feix, G., *Z. Phys. Chem. Leipzig* **1979**, 260, 1009-13.
15. Hefter, G. T., 37, D. G. Shaw; Pergamon Press, Oxford, UK, 1989.
16. Chen, H. P.; Wagner, J., *J. Chem. Eng. Data* **1994**, 39, 470-4.
17. Umamo, S.; Hayano, I., *Kogyo Kagaku Zasshi* **1957**, 60, 1436-7.
18. Krasnoshchekova, R. Y.; Gubergrits, M. Y., *Vodnye. Resursy.* **1975**, 2, 170-3.
19. Pierotti, R. A.; Liabastre, A. A., Presented to U. S. Nat. Tech. Inform. Serv., 1972.
20. Chen, H. P.; Wagner, J., *J. Chem. Eng. Data* **1994**, 39, 475-9.
21. Tarassenkow, D. N.; Poloshinzewa, E. N., *Chem. Ber.* **1932**, 65, 184-6.
22. Tarassenkow, D. N.; Poloshinzewa, E. N., *Zh. Obshch. Khim.* **1931**, 1, 71-9.
23. Roddy, J. W.; Coleman, C. F., *Talanta* **1968**, 15, 1281-6.
24. Marche, C.; Ferronato, C.; De Hemptinne, J.-C.; Jose, J., *J. Chem. Eng. Data* **2006**, 51, 355-9.
25. McAuliffe, C., *J. Phys. Chem.* **1966**, 70, 1267-75.
26. Price, L. C., *Am. Assoc. Petrol. Geol. Bull.* **1976**, 60, 213-44.
27. Rudakov, E. S.; Lutsyk, A. I., *Zh. Fiz. Khim.* **1979**, 53, 1298-300.
28. Polak, J.; Lu, B. C. Y., *Can. J. Chem.* **1973**, 51, 4018-23.
29. McAuliffe, C., *Nature* **1963**, 200, 1092-3.
30. Karlsson, R. J., *Chem. Eng. Data* **1973**, 18, 290-2.
31. Franks, F., *Nature* **1966**, 210, 87-8.
32. Franks, F.; Gent, M.; Johnson, H. H., *J. Chem. Soc.* **1963**, 2716-23.

33. Goldman, S., *Can. J. Chem.* **1974**, 52, 1668-80.
34. Guerrant, R. P., M.S. Thesis, Pennsylvania State University: University Park, PA, 1964.
35. Ng, H.-J.; Chen, C.-J., Mutual Solubility in Water-Hydrocarbon Systems, Presented to DB Robinson Research Ltd., Edmonton, Alberta, Canada, 1995.
36. Schatzberg, P., *J. Phys. Chem.* **1963**, 67, 776-9.
37. Wang, Q.; Chao, K. C., *Fluid Phase Equilib.* **1990**, 59, 207-15.
38. Chen, H. P., Ph.D. Dissertation, Oklahoma State University: Stillwater, OK, 1994.
39. Chen, H. P.; Wagner, J., *J. Chem. Eng. Data* **1994**, 39, 679-84.
40. Bennett, B. D., M.S. Thesis, Oklahoma State University: Stillwater, OK, 1997.
41. Ratzlaff, D. W., M.S. Thesis, Oklahoma State University: Stillwater, OK, 1999.
42. Stevenson, R. L.; LaBracio, D. S.; Beaton, T. A.; Thies, M. C., *Fluid Phase Equilib.* **1994**, 93, 317-36.
43. Bradley, R. S.; Dew, M. J.; Munro, D. C., *High Temp. – High Press.* **1973**, 5, 169-76.
44. Arnold, D. S.; Plank, C. A.; Erickson, E. E.; Pike, F. P., *Chem. & Eng. Data Ser.* **1958**, 3, 253-6.
45. Chandler, J. P.; Jackson, L. W., Computer Program **1981**, Stillwater, OK, Department of Computing and Information Sciences, Oklahoma State University.
46. Malmory, G.; Vezier, A.; Robert, A.; Mourgues, J.; Conte, T.; Molinier, J., *J. Chem. Technol. Biotechnol.* **1994**, 60, 67-71.
47. Skrylev, L. D.; Nevinskii, A. G.; Purich, A. N., *Zhurnal Prikladnoi Khimii (Sankt-Peterburg, Russian Federation)* **1984**, 57, 2026-30.
48. Yaws, C. L.; Yang, H. C.; Hopper, J. R.; Hansen, K. C., *Chem. Eng. (N.Y.)* **1990**, 97, 115-16, 8.
49. Yaws, C. L.; Yang, H. C.; Hopper, J. R.; Hansen, K. C., *Chem. Eng. (N.Y.)* **1990**, 97, 177-8, 80, 82.
50. Lazaridis, T.; Paulaitis, M. E., *AIChE J.* **1993**, 39, 1051-60.
51. Li, J.; Dallas, A. J.; Eikens, D. I.; Carr, P. W.; Bergmann, D. L.; Hait, M. J.; Eckert, C. A., *Anal. Chem.* **1993**, 65, 3212-18.
52. Bergmann, D. L.; Eckert, C. A., *Fluid Phase Equilib.* **1991**, 63, 141-50.
53. Belter, P. A.; Cussler, E. L.; Hue, W., *Bioseparations*, John Wiley & Sons, Inc.: New York, 1988.
54. Bader, M. S. H., Ph. D. Dissertation, Oklahoma State University: Stillwater, OK, 1993.
55. Bader, M. S. H.; Gasem, K. A. M., *Chem. Eng. Commun.* **1996**, 140, 41-72.
56. Tiegs, D.; Gmehling, J.; Medina, A.; Soares, M.; Bastos, J.; Alessi, P.; Kikic, I., *Chemistry Data Series, IX, Part 1*, D. Behrens and R. Eckermann; Schoon & Wetzel GmbH, Frankfurt, Germany, 1986.
57. Scatchard, G., *Chem. Rev.* **1931**, 8, 321-33.
58. Hildebrand, J. H.; Wood, S. E., *J. Chem. Phys.* **1933**, 1, 817-22.
59. Weimer, R. F.; Prausnitz, J. M., *Hydrocarbon Processing and Petroleum Refiner* **1965**, 44, 237-42.
60. Hansen, C. M., *J. Paint. Technol.* **1967**, 39, 505-10.
61. Helpinstill, J. G.; Van Winkle, M., *Ind. Eng. Chem. Process Des. Dev.* **1968**, 7, 213-20.

62. Karger, B.; Snyder, L. R.; Eon, C., *J. Chromatogr.* **1976**, 125, 71-88.
63. Tijssen, R.; Billiet, H. A. H.; Schoenmakers, P. J., *J. Chromatogr.* **1976**, 122, 185-203.
64. Bader, M. S. H.; Gasem, K. A. M., *Can. J. Chem. Eng.* **1998**, 76, 94-103.
65. Thomas, E. R.; Eckert, C. A., *Ind. Eng. Chem. Process Des. Dev.* **1984**, 23, 194-209.
66. Howell, W. J.; Karachewski, A. M.; Stephenson, K. M.; Eckert, C. A.; Park, J. H.; Carr, P. W.; Rutan, S. C., *Fluid Phase Equilib.* **1989**, 52, 151-60.
67. Hait, M. J.; Liotta, C. L.; Eckert, C. A.; Bergmann, D. L.; Karachewski, A. M.; Dallas, A. J.; Eikens, D. I.; Li, J. J.; Carr, P. W., *Ind. Eng. Chem. Res.* **1993**, 32, 2905-14.
68. Derr, E. L.; Deal, C. H., *Institute of Chemical Engineering Symposium Series* **1969**, 32, 40.
69. Fredenslund, A.; Jones, R. L.; Prausnitz, J. M., *AIChE J.* **1975**, 21, 1086-99.
70. Gmehling, J.; Li, J.; Schiller, M., *Ind. Eng. Chem. Res.* **1993**, 32, 178-93.
71. Pierotti, G. J.; Deal, C. H.; Derr, E. L., *J. Ind. Eng. Chem.* **1959**, 51, 95-102.
72. Tsonopoulos, C.; Prausnitz, M., *Ind. Eng. Chem. Fund.* **1971**, 10, 593-600.
73. Medir, M.; Giralt, F., *AIChE J.* **1982**, 28, 341-3.
74. Dutt, N. V. K.; Prasad, D. H. L., *Fluid Phase Equilib.* **1989**, 45, 1-5.
75. Yalkowsky, S. H.; Valvani, S. C., *J. Chem. Eng. Data* **1979**, 24, 127-9.
76. Mackay, D.; Shiu, W. Y., *J. Chem. Eng. Data* **1977**, 22, 399-402.
77. Taft, R. W.; Abboud, J. L. M.; Kamlet, M. J.; Abraham, M. H., *J. Solution Chem.* **1985**, 14, 153-86.
78. Sherman, S. R.; Trampe, D. B.; Bush, D. M.; Schiller, M.; Eckert, C. A.; Dallas, A. J.; Li, J.; Carr, P. W., *Ind. Eng. Chem. Res.* **1996**, 35, 1044-58.
79. Shing, K. S., *Chem. Phys. Lett.* **1985**, 119, 149-51.
80. Economou, I. G., *Fluid Phase Equilib.* **2001**, 183-184, 259-69.
81. Nelson, T. M.; Jurs, P. C., *J. Chem. Inf. Comput. Sci.* **1994**, 34, 601-9.
82. Mitchell, B. E.; Jurs, P. C., *J. Chem. Inf. Comput. Sci.* **1998**, 38, 200-9.
83. He, J.; Zhong, C., *Fluid Phase Equilib.* **2003**, 205, 303-16.
84. Huibers, P. D. T.; Katritzky, A. R., *J. Chem. Inform. Comput. Sci.* **1998**, 38, 283-92.
85. Gao, H.; Shanmugasundaram, V.; Lee, P., *Pharm. Res.* **2002**, 19, 497-503.
86. Katritzky, A. R.; Karelson, M.; Lobanov, V. S., *Pure Appl. Chem.* **1997**, 69, 245.
87. Martin, M. G.; Siepmann, J. I.; Schure, M. R., *Unified Chromatography, ACS Symposium Series* **2000**, 748,
88. Kubinyi, H., *Drug Discov. Today* **1997**, 2, 538-46.
89. Malanowski, S.; Anderko, A., *Modeling Phase Equilibria Thermodynamic Background and Practical Tools*, John Wiley & Sons, Inc.: New York, 1992.
90. Peng, D.-Y.; Robinson, D. B., *Ind. Eng. Chem. Fund.* **1976**, 15, 59-64.
91. Kamlet, M. J.; Abboud, J. L. M.; Abraham, M. H.; Taft, R. W., *J. Org. Chem.* **1983**, 48, 2877-87.
92. Li, J.; Zhang, Y.; Dallas, A. J.; Carr, P. W., *J. Chromatogr.* **1991**, 550, 101-34.
93. Li, J.; Zhang, Y.; Ouyang, H.; Carr, P. W., *J. Am. Chem. Soc.* **1992**, 114, 9813-28.
94. Li, J.; Zhang, Y.; Carr, P. W., *Anal. Chem.* **1992**, 64, 210-18.

95. Mills, E. J., *Philos. Mag.* **1884**, 17, 173-87.
96. Meyer, H., *Arch. Exp. Pathol. Physiol.* **1899**, 42, 109-18.
97. Overton, E., *Zeitschr. Physik. Chem.* **1897**, 22, 189-209.
98. Fischer, E., *Ber. Dtsch. Chem. Ges.* **1894**, 27, 2985-93.
99. Hammett, L. P., *Chem. Rev.* **1935**, 17, 125-36.
100. Taft, R. W., *Steric Effects in Organic Chemistry*, M. S. Newman; John Wiley & Sons, New York, 1956.
101. Hansch, C.; Fujita, T., *J. Am. Chem. Soc.* **1964**, 86, 1616-26.
102. Free, S. M.; Wilson, J. W., *J. Med. Chem.* **1964**, 7, 395-9.
103. Vandeginste, B. G. M.; Massart, D. L.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J., *Handbook of Chemometrics and Qualimetrics: Part B*, Elsevier: New York, 1998.
104. Todeschini, R.; Consonni, V., *Handbook of Molecular Descriptors*, Wiley-VCH: Weinheim, 2000.
105. Karelson, M., *Molecular Descriptors in QSAR/QSPR*, John Wiley & Sons: New York, 2000.
106. Amzel, L. M., *Curr. Opin. Biotechnol.* **1998**, 9, 366.
107. Golender, V.; Vesterman, B.; Vorpapel, E., *Apex-3D System for Drug Design* **1998**,
108. Kubinyi, H., *Drug Discov. Today* **1997**, 2, 457.
109. Kubinyi, H., *Drug Discov. Today* **1997**, 2, 538.
110. Luo, Z.; Wang, R.; Lai, L., *J. Chem. Inf. Comput. Sci.* **1996**, 36, 1187.
111. Martin, Y. C., *Perspect. Drug Discov. Des.* **1997**, 7/8, 159.
112. Katritzky, A. R.; Lobanov, V. S.; Karelson, M., *J. Chem. Inf. Comp. Sci.* **1998**, 38, 28.
113. Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M., *J. Phys. Chem.* **1996**, 100, 10400-7.
114. Gallagher, D. A., *J. Chem. Inf. Comput. Sci.* **1998**, 38, 321.
115. Katritzky, A. R.; Mu, L.; Karelson, M., *J. Chem. Inf. Comput. Sci.* **1998**, 38, 293.
116. Katritzky, A. R.; Mu, L.; Karelson, M., *J. Chem. Inf. Comput. Sci.* **1996**, 36, 1162.
117. Katritzky, A. R.; Sild, S.; Karelson, M., *J. Chem. Inf. Comput. Sci.* **1998**, 38, 1171.
118. Rothman, J. H.; Kroemer, R. T., *J. Mol. Model* **1997**, 3, 261.
119. Clark, D. E.; Frenkel, D.; Levy, S. A.; Li, J.; Murray, C. W.; Robson, B.; Waszkowycz, B.; Westhead, D. R., *J. Comput.-Aided Mol. Des.* **1995**, 9, 13.
120. Weininger, D., *J. Chem. Inf. Comput. Sci.* **1988**, 28, 31.
121. Stewart, J. P. P., *MOPAC Program Package* **1990**,
122. Katritzky, A. R.; Karelson, M.; Lobanov, V. S.; Dennington, R.; Keith, T., *CODESSA* **1999**, Shawnee, KS, Semichem, Inc.
123. Merrifield, R. E.; Simmons, H. E., *Topological Methods in Chemistry*, J. Wiley & Sons: New York, 1989.
124. Wiener, H., *J. Am. Chem. Soc.* **1947**, 69, 17.
125. Kier, L. B., *Quant. Struct.-Act. Relat.* **1985**, 4, 109.
126. Klein, D. J.; Randic, M., *J. Math. Chem.* **1993**, 12, 85.
127. Basak, S. C.; Hariss, D. K.; Magnuson, V. R., *J. Pharm. Sci.* **1984**, 73, 429.
128. Lee, B.; Richards, F. M., *J. Mol. Biol.* **1971**, 55, 379.

129. Connolly, M. L., *J. Appl. Crystallogr.* **1983**, 16, 548.
130. Richards, F. M., *Annu. Rev. Biophys. Bioeng.* **1977**, 6, 151.
131. Connolly, M. L., *J. Am. Chem. Soc.* **1985**, 107, 1118.
132. Gasteiger, J.; Marsili, M., *Tetrahedron Lett.* **1978**, 3181.
133. Zefirov, N. S.; Kirpichenok, M. A.; Izmailov, F. F.; Trofimov, M. I., *Dokl. Akad. Nauk. SSSR* **1987**, 296, 883.
134. Price, S. L.; Stone, A. J., *J. Chem. Phys.* **1987**, 86, 2859.
135. Warshel, A., *Acc. Chem. Res.* **1981**, 14, 284.
136. Katritzky, A. R.; Lobanov, V. S.; Karelson, M., *CODESSA User's Manual*, (1994)
137. Stanton, D. T.; Jurs, P. C., *Anal. Chem.* **1990**, 62, 2323-9.
138. Stanton, D. T.; Egolf, L. M.; Jurs, P. C.; Hicks, M. G., *J. Chem Inf. Comput. Sci.* **1992**, 32, 306.
139. Sannigrahi, A. B., *Adv. Quant. Chem.* **1992**, 23, 301.
140. Cartier, A.; Rivail, J. L., *Chemometrics Intell. Lab. Syst.* **1987**, 1, 335.
141. Atkins, P. W., *Physical Chemistry, Second*; W. H. Freeman and Co.: San Francisco, 1982.
142. Akhiezer, A. I.; Peltminskii, S. V., *Methods of Statistical Physics*, Pergamon Press: Oxford, 1981.
143. McQuarrie, D. A., *Statistical Thermodynamics*, Harper & Row Publ.: New York, 1973.
144. Pierotti, R. A., *Chem. Rev.* **1976**, 76, 717.
145. Famini, G. R., *Using Theoretical Descriptors in Quantitative Structure-Activity Relationships*, Presented to U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, 1989.
146. Kurtz, H. A.; Stewart, J. P. P.; Dieter, K. M., *J. Comput. Chem.* **1990**, 11, 82.
147. Malanowski, E. R.; Howery, D. G., *Factor Analysis in Chemistry*, Wiley Interscience: New York, 1980.
148. Wold, S.; Geladi, P.; Esbensen, K.; Ohman, J., *J. Chemometrics* **1987**, 1, 41.
149. Zupan, J.; Gasteiger, J., *Neural Networks for Chemists*, VCH Publishers: Weinheim, 1993.
150. Rogers, D.; Hopfinger, A. J., *J. Chem. Inf. Comput. Sci.* **1994**, 34, 854.
151. Hartigan, J. A., *Clustering Algorithms*, John Wiley & Sons: New York, 1975.
152. Kohonen, T., *Self-Organising Maps*, Springer-Verlag: Heidelberg, 1995.
153. Brown, R. D.; Martin, Y. C., *J. Chem. Inf. Comput. Sci.* **1996**, 36, 572-84.
154. Wu, W.; Mallet, Y.; Walczak, B.; Penninckx, W.; Massart, D. L.; Heuerding, S.; Erni, F., *Anal. Chim. Acta* **1996**, 329, 257-65.
155. Lucic, B.; Trinajstic, N., *Neural Comp.* **1999**, 11, 1035.
156. Wessel, M. D.; Jurs, P. C., *J. Chem. Inf. Comput. Sci.* **1995**, 35, 841-50.
157. Ulmer, C. W.; Smith, D. A.; Sumpter, B. G.; Noid, D. I., *Comput. Theor. Polym.* **1998**, 8, 311-21.
158. Ramon y Cajal, S., *Histologie Du Systeme Nerveux de L'homme et Des Vertebres*, Maloine: Paris, 1911.
159. Shepherd, G. M.; Koch, C., *The Synaptic Organization of the Brain*, G. M. Shepherd; Oxford University Press, New York, 1990.

160. Faggin, F., International Joint Conference on Neural Networks, Seattle, WA, 1991.
161. McCulloch, W. S.; Pitts, W., Bull. Math. Biophys **1943**, 5, 115-33.
162. Hebb, D. O., The Organization of Behavior, John Wiley & Sons: New York, 1949.
163. von Neumann, J., Automata Studies, C. E. Shannon and J. McCarthy; Princeton University Press, Princeton, 1956.
164. Rosenblatt, F., Psychol. Rev. **1958**, 65, 386-408.
165. Widrow, B.; Hoff, M. E., Jr., IRE WESCON Convention Record. part 4, 1960.
166. Minsky, M. L.; Papert, S. A., Perceptrons, First , 1969; The MIT Press: Cambridge, MA, 1969.
167. Cowan, J. D., Ph.D. Dissertation, University of London: London, UK, 1967.
168. Sherrington, D.; Kirkpatrick, S., Phys. Rev. Lett. **1975**, 35,
169. Willshaw, D. J.; von der Malsburg, C., Proc. R. Soc. London, B **1976**, 194, 431.
170. von der Malsburg, C., Kybernetik **1973**, 14, 85.
171. Hopfield, J. J., Proceedings of the National Academy of Sciences, 1982.
172. Cragg, B. G.; Tamperley, H. N. V., Electroencephalogr. Clin. Neurophysiol. **1954**, 6, 85.
173. Cragg, B. G.; Tamperley, H. N. V., Brain **1955**, 78, 304.
174. Grossberg, S., J. Math. Anal. Appl. **1968**, 22, 490.
175. Grossberg, S., Proc. Natl. Acad. Sci. U. S. A. **1967**, 22, 490.
176. Amari, S. I., IEEE Trans. Syst. Man Cybern. **1972**, 2, 643.
177. Wilson, H. R.; Cowan, J. D., J. Biophys. **1972**, 12, 1.
178. Little, W. A.; Shaw, G. L., Behav. Biol. **1975**, 14, 115.
179. Anderson, J. A.; Silverstein, J. W.; Ritz, S. A.; Jones, R. S., Psychol. Rev. **1977**, 84, 413-51.
180. Rumelhart, D. E.; Hinton, G. E.; Williams, R. J., Nature (London) **1986**, 323, 533.
181. Rumelhart, D. E.; McClelland, J. L., 1, D. E. Rumelhart, J. L. McClelland and P. R. Group; MIT Press, Cambridge, MA, 1986.
182. Parker, D. B., Learning-logic: Casting the Cortex of the Human Brain in Silicon, Presented to Center for Computational Research in Economics and Management Science, MIT, Cambridge, MA, 1985.
183. LeCun, Y., Cognitiva **1985**, 85, 599.
184. Kirkpatrick, S.; Gelatt, J., C. D.; Vecchi, M. P., Science **1983**, 220, 671.
185. Ackley, S. C.; Hinton, G. E.; Sejnowski, T. J., Cognit. Sci. **1985**, 9, 147.
186. Linkser, R., Computer **1988**, 21, 105.
187. Broomhead, D. S.; Lowe, D., Complex Systems **1988**, 2, 321.
188. Aoyama, T.; Suzuki, Y.; Ichikawa, H., J. Med. Chem. **1990**, 33, 2583-90.
189. Aoyama, T.; Suzuki, Y.; Ichikawa, H., J. Med. Chem. **1990**, 33, 905-8.
190. Aoyama, T.; Ichikawa, H., Chem. Pharm. Bull. **1991**, 39, 358-66.
191. Aoyama, T.; Ichikawa, H., Chem. Pharm. Bull. **1991**, 39, 372-8.
192. Chastrette, M.; De Saint Laumer, J. Y., Eur. J. Med. Chem. **1991**, 26, 829-33.
193. Hammerstrom, D., Digital VLSI for Neural Networks. The Handbook of Brain Theory and Neural Networks, MIT Press: Cambridge, MA, 1995.
194. Andrea, T. A.; Kalayeh, H., J. Med. Chem. **1991**, 34, 2824-36.
195. Andrea, T. A.; Kalayeh, H., Pharmacochem. Libr. **1991**, 16, 209-12.

196. Ichikawa, H.; Aoyama, T., SAR QSAR Environ. Res. **1993**, 1, 115-30.
197. Hirst, J. D.; King, R. D.; Sternberg, M. J. E., J. Comput. Aided Mol. Des. **1994**, 8, 421-32.
198. Hirst, J. D.; King, R. D.; Sternberg, M. J. E., J. Comput. Aided Mol. Des. **1994**, 8, 405-20.
199. So, S.-S.; Richards, W. G., J. Med. Chem. **1992**, 35, 3201-7.
200. Tetko, I. V.; Tanchuk, V. Y.; Chentsova, N. P.; Antonenko, S. V.; Poda, G. I.; Kukhar, V. P.; Luik, A. I., J. Med. Chem. **1994**, 37, 2520-6.
201. Wiese, M.; Schaper, K. J., Med. Pharm. Chemie **1993**, 1, 137-52.
202. Calleja, M. C.; Geladi, P.; Persoone, G., SAR QSAR Environ. Res. **1994**, 2, 193-234.
203. Devillers, J.; Bintein, S.; Domine, D.; Karcher, W., SAR QSAR Environ. Res. **1995**,
204. Devillers, J.; Cambon, B., Polycyclic Aromat. Compd. **1993**, 3, 257-65.
205. Villemin, D.; Cherqaoui, D.; Mesbah, A., J. Chem. Inf. Comput. Sci. **1994**, 34, 1288-93.
206. Brinn, M.; Walsh, P. T.; Payne, M. P.; Bott, B., SAR QSAR Environ. Res. **1993**, 1, 169-210.
207. Brinn, M. W.; Payne, M. P.; T., W. P., Chem. Eng. Res. Des **1993**, 71, 337-9.
208. Ghoshal, N.; Mukhopadhyay, S. N.; Ghoshal, T. K.; Achari, B., Indian J. Chem. **1993**, 32B, 1045-50.
209. Villemin, D.; Cherqaoui, D.; Cense, J. M., J. Chim. Phys. Phys.-Chim. Biol. **1993**, 90, 1505-19.
210. Grunenber, J.; Herges, R., J. Chem. Inf. Comput. Sci. **1995**, 35, 905-11.
211. Devillers, J.; Domine, D.; Karcher, W., Polycyclic Aromat. Compd. **1996**, 11, 211-7.
212. Briendl, A.; Bernd, B.; Clark, T.; Glen, R. C., J. Mol. Model. **1997**, 3, 142.
213. Bodor, N.; Huang, M. J., J. Pharm. Sci. **1992**, 81, 954-60.
214. Bodor, N.; Harget, A.; Huang, M. J., J. Am. Chem. Soc. **1991**, 113, 9480-3.
215. Katritzky, A. R.; Wang, Y.; Sild, S.; Tamm, T.; Karelson, M., J. Chem. Inf. Comput. Sci. **1998**, 38, 720-5.
216. Chow, H.; Chen, H.; Ng, T.; Myrdal, P.; Yalkowsky, S. H., J. Chem. Inf. Comput. Sci. **1995**, 35, 723-8.
217. Gakh, A. A.; Gakh, E. G.; Sumpter, B. G.; Noid, D. W., J. Chem. Inf. Comput. Sci. **1994**, 34, 832-9.
218. Cherqaoui, D.; Villemin, D.; Kvasnicka, V., Chemometr. Intell. Lab. Syst. **1994**, 24, 117-28.
219. Cherqaoui, D.; Villemin, D., J. Chem. Soc., Faraday Trans. **1994**, 90, 97-102.
220. Cherqaoui, D.; Villemin, D.; Mesbah, A.; Cense, J. M.; Kvasnicka, V., J. Chem. Soc., Faraday Trans. **1994**, 90, 2015-9.
221. Cherqaoui, D.; Villemin, D.; Mesbah, A.; Cense, J. M.; Kvasnicka, V., AIP Conf. Proc. (Comp. Chem.), 1995.
222. Egolf, L. M.; Jurs, P. C., J. Chem. Inf. Comput. Sci. **1993**, 33, 616-25.
223. Egolf, L. M.; Wessel, M. D.; Jurs, P. C., J. Chem. Inf. Comput. Sci. **1994**, 34, 947-56.

224. Balaban, A. T.; Basak, S. C.; Colburn, T.; Grunwald, G. D., *J. Chem. Inf. Comput. Sci.* **1994**, 34, 1118-21.
225. Kan, D.; Ma, D., *Beijing Huagong Xueyuan Xuebao* **1993**, 20, 106-17.
226. Zhang, X.; Zhao, L.; Zhang, G., *Huagong Xuebao* **1995**, 46, 66-74.
227. Bishop, C. M., *Neural Networks for Pattern Recognition*, Oxford University Press: Oxford, 1995.
228. Fausett, L., *Fundamentals of Neural Networks*, Prentice Hall: Englewood Cliffs, NJ, 1994.
229. Hagen, M. T.; Demuth, H. B.; Beale, M., *Neural Network Design*, Thomson Learning, Inc.: Boston, MA, 1996.
230. Haykin, S., *Neural Networks: A Comprehensive Foundation, Second*; Prentice Hall: Upper Sadle River, 1999.
231. Hertz, J.; Krogh, A.; Palmer, R., *Introduction to the Theory of Neural Computation*, Addison-Wesley: Redwood City, California, 1991.
232. Patterson, D., *Artificial Neural Networks*, Prentice Hall: Singapore, 1996.
233. Foresee, F. D., Ph.D. Dissertation, Oklahoma State University: Stillwater, OK, 1996.
234. Sarle, W. S., *Neural Network FAQ, Part 1 of 7: Introduction*, <ftp://ftp.sas.com/pub/neural/FAQ.html>, 1997.
235. Deco, G.; Obradovic, D., *An Information-Theoretic Approach to Neural Computing*, Springer-Verlag: New York, 1996.
236. Hecht-Nielsen, R., *Neurocomputing*, Addison-Wesley: Reading, MA, 1990.
237. Kosko, B., *Neural Networks and Fuzzy Systems*, Prentice-Hall: Englewood Cliffs, N.J., 1992.
238. Kohonen, T., *Self-Organizing Maps, First*; Springer-Verlag: Berlin, 1995.
239. Minsky, M. L.; Papert, S. A., *Perceptrons, Second*; The MIT Press: Cambridge, MA, 1988.
240. Pao, Y. H., *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley Publishing Company: Reading, MA, 1989.
241. Reed, R. D.; Marks, R. J., II, *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*, The MIT Press: Cambridge, MA, 1999.
242. Rumelhart, D. E.; Hinton, G. E.; Williams, R. J., *Learning Internal Representations by Error Propagation*, The MIT Press: Cambridge, MA, 1986.
243. Fahlman, S. E.; Lebiere, C., *The Cascade-Correlation Learning Architecture*, Morgan Kaufmann Publishers: Los Altos, CA, 1990.
244. Fahlman, S. E., *Proceedings of the 1988 Connectionist Models Summer School*, 1989.
245. Riedmiller, M.; Braun, H., *Proceedings of the IEEE International Conference on Neural Networks*, San Francisco, 1993.
246. Moody, J.; Darken, C. J., *Neural Comput.* **1989**, 1, 281-94.
247. Orr, M. J. L., *Introduction to radial basis function networks* **1996**, <http://www.anc.ed.ac.uk/~mjo/papers/intro.ps>,
248. Chen, S.; Cowan, C. F. N.; Grant, P. M., *IEEE Trans. Neural Network.* **1991**, 2, 302-9.
249. Albus, J. S., *Trans. ASME J. Dyn. Syst. Meas. Contr.* **1975**, Sep-75, 220-27.

250. Brown, M.; Harris, C., Neurofuzzy Adaptive Modeling and Control, Prentice Hall: New York, 1994.
251. Kohonen, T., Neural Network. **1988**, 1 suppl 1, 303.
252. Hand, D. J., Kernel Discriminant Analysis, John Wiley & Sons: New York, 1982.
253. Masters, T., Practical Neural Network Recipes in C++, Academic Press: San Diego, 1993.
254. Specht, D. F., Neural Network. **1990**, 3, 110-8.
255. Specht, D. F., IEEE Trans. Neural Network. **1991**, Nov-91, 568-76.
256. Nadaraya, E. A., Theory Probab. Applic. **1964**, 10, 186-90.
257. Watson, G. S., Sankhya **1964**, 26, 359-72.
258. Medsker, L. R.; Jain, L. C., Recurrent Neural Networks: Design and Applications, CRC Press, Boca Raton, FL, 2000.
259. Werbos, P. J., Proc. IEEE **1990**, 78, 1550-60.
260. Elman, J. L., Cognit. Sci. **1990**, 14, 179-211.
261. Wan, E. A., Proceedings of the 1990 Connectionist Models Summer School, San Mateo, CA, 1990.
262. Jordan, M. I., Proceedings of the Eighth Annual conference of the Cognitive Science Society, 1986.
263. Williams, R. J.; Zipser, D., Neural Comput. **1989**, 1, 270-80.
264. Pineda, F. J., Neural Comput. **1989**, 1, 161-72.
265. Lang, K. J.; Waibel, A. H.; Hinton, G., Neural Network. **1990**, 3, 23-44.
266. Carpenter, G. A.; Grossberg, S.; Rosen, D. B., Neural Network. **1991**, 4, 493-504.
267. Carpenter, G. A.; Grossberg, S.; Markuzon, N.; Reynolds, J. H.; Rosen, D. B., IEEE Trans. Neural Network. **1992**, 3, 698-713.
268. Kasuba, T., AI Expert **1993**, 8, 18-25.
269. Williamson, J. R., Gaussian ARTMAP: A Neural Network for Fast Incremental Learning of Noisy Multidimensional Maps, Presented to Center of Adaptive Systems and Department of Cognitive and Neural Systems, Boston University, Boston, 1995.
270. Hecht-Nielsen, R., Appl. Opt. **1987**, 26, 4979-84.
271. Hecht-Nielsen, R., Neural Network. **1988**, 1, 131-9.
272. Fukushima, K., Neural Network. **1988**, 1, 119-30.
273. Grossberg, S., Biol. Cybern. **1976**, 23, 121-34.
274. Kohonen, T., Self-Organization and Associative Memory, Springer: Berlin, 1984.
275. Desieno, D., Proc. Int. Conf. on Neural Networks, 1988.
276. Bishop, C. M.; Svensén, M.; Williams, C. K. I., GTM: A Principled Alternative to the Self-Organizing Map, The MIT Press: Cambridge, MA, 1997.
277. Mulier, F.; Cherkassky, V., Neural Comput. **1995**, 7, 1165-77.
278. Carpenter, G. A.; Grossberg, S., Comput. Vis. Graph.Image Process. **1987**, 37, 54-115.
279. Moore, B., Proceedings of the 1988 Connectionist Models Summer School, San Mateo, CA, 1988.
280. Carpenter, G. A.; Grossberg, S., Appl. Opt. **1987**, 26, 4919-30.
281. Carpenter, G. A.; Grossberg, S., Neural Network. **1990**, 3, 129-52.
282. Carpenter, G. A.; Grossberg, S.; Rosen, D. B., Neural Network. **1991**, 4, 759-71.

283. Diamantaras, K. I.; Kung, S. Y., *Principal Component Neural Networks: Theory and Applications*, John Wiley & Sons: New York, 1996.
284. Oja, E., *Int. J Neural Syst.* **1989**, 1, 61-8.
285. Sanger, T. D., *Neural Network.* **1989**, 2, 459-73.
286. Cichocki, A.; Unbehauen, R., *Neural Networks for Optimization and Signal Processing*, John Wiley & Sons: New York, 1993.

Appendix A. Standard Operating Procedure

The collection of operating procedures, sampling methods, and analytical techniques, which accompany the experimental apparatus used in this study to obtain liquid-liquid equilibrium data, are described in this appendix. Figure A-1 presents a schematic of the experimental apparatus.

A.1 Initialization

Pressurization of the Backpressure Regulator

Before sampling at elevated temperatures and pressures, the hydrocarbon-water system must be raised to a pressure greater than the mixture vapor pressure. By application of a nitrogen blanket to the system, controlled by a backpressure regulator (BPR), a sufficiently higher pressure is obtained. In order to control the pressure, the BPR must be “loaded” to the desired system pressure using the following procedure:

1. Allow nitrogen gas flow to the BPR by turning the Whitey three-way valve (V9).
2. Turn the setscrew on the BPR labeled, “load,” counterclockwise. This allows nitrogen to fill the diaphragm of the BPR.
3. The desired BPR pressure is reached by increasing the pressure from the nitrogen source. The pressure gauge (P4) will reflect changes in pressure.
4. When the desired experimental pressure is reached, turn the setscrew on the BPR labeled, “load,” clockwise until closed.

Isolation of the Apparatus

Before pressurization of the system, the apparatus must be isolated from the atmosphere by the following procedure:

5. Prevent nitrogen ventilation to the atmosphere by closure of the needle valves (V6 and V7) and rotation of the three-way valve (V4).
6. Direct the organic phase sample flow to the trash collection cell (C2) by manipulation of the three-way valve (V2).
7. Connect the water phase sample line to the three-way valve (V8), via a 0.318 cm-o.d. tubing sleeve.
8. After tightening the tube fittings on the connection, turn V8 such that the sample tubing is opened to the system pressure. This acts as a shutoff valve for the water phase sample.

Pressurization of the System

The system may now be pressurized. The apparatus is designed so the BPR is set at the desired pressure, and the nitrogen source regulator is set at a slightly higher pressure. This effectively controls the system pressure by allowing any excess nitrogen to vent to the atmosphere at the BPR. The following steps should be taken to pressurize the system:

9. Allow nitrogen flow to the system by turning V9.
10. Adjustment of the system pressure may be accomplished by turning the regulator on the nitrogen source. The pressure readout (P1) will reflect the change in pressure.

11. If pressure at the BPR has been set too high, the pressure may require a reset to a lower system pressure in order for the pressure to be accurately controlled.
12. Pressure can be lowered in the BPR by turning the setscrew on the BPR labeled, “vent,” counterclockwise. The pressure gauge (P4) will reflect the change in pressure.
13. When the desired system pressure has been reached, close the vent by turning the setscrew clockwise.

Preparation for Sample Collection

With the adjustments to the system pressure finalized, the apparatus is readied for sample collection by the following procedures:

14. Allowing three hours for thermal equilibration, the oven temperature is set to the desired temperature.
15. After temperature stabilization, the duplex pump (DP1) is engaged and allowed to flush the system with at least one system volume (120 cm^3) of the hydrocarbon-water mixture. The pump should never be started against pressures in excess of 6.9 MPa (1000 psi) and should never be operated without fluids present.
16. Vent the 0.159 cm-o.d. stainless steel tubing (water phase sample) by opening V8 to the atmosphere. When venting, a waste bottle is utilized to capture the small amount of water phase sample that may exit from V8.
17. The water phase sample line from V8 is removed, which allows the water phase to exit from the bottom of the phase separation cell (labeled as such in Figure A-1) and through the water phase sample tubing.

18. Adjustment of the metering valve (MV1) inside the oven controls the amount of water phase exiting the phase separation cell. While the rate at which the water phase exits the phase separation cell is not critical, the hydrocarbon-water interface must be kept at the level of the inlet, which corresponds to the center of the phase separation cell. By maintaining the interface at this level, the water phase flow rate will be approximately one half the total flow rate. The MV1 should never be used as a shutoff valve as this will cause damage to the stem or packing and render the valve inoperable.

Preparation of the Sample Bottles

After flushing the system with the hydrocarbon-water mixture and controlling the hydrocarbon-water interface near the inlet of the phase separation cell, the sample bottles are prepared according to the following procedures:

19. Empty sample bottles, along with a cap and Teflon liner, are numbered and weighed.
20. Add the extractant (decane or 2,2,4-trimethylpentane) or cosolvent (ethanol) to the empty sample bottles.
21. The amount of extractant or cosolvent added is determined by weighing the sample bottles.
22. Sample bottles for the aqueous phase sample collection, which contain the extractant, are placed in an ice bath. The ice bath aids in the prevention of vaporization of the volatile hydrocarbon.
23. A sample bottle containing the cosolvent is placed in the sampling cell (C1) with the organic phase sample tubing inserted in the bottle.

24. The cell is closed with the connected sample bottle.
25. Pressurization of C1 is accomplished by directing the flow of nitrogen to C1 by closing the needle valve (V5) and turning the three-way valve (V3). The pressure differential between the system pressure and C1 will result in the flow of nitrogen from both the system and the source, which causes a decrease in the system pressure. Manipulation of V3 should occur slowly such that the flow of nitrogen into C1 is not greater than the flow of nitrogen from the nitrogen source; thus ensuring the system pressure does not decrease below the mixture vapor pressure.

A.2 Sampling

Sample Collection

After placing the organic phase sample bottle in C1 and pressurizing C1 to the elevated system pressure, the following steps are used to collect the sample:

26. The organic phase sample is directed to the sample bottle in C1 through manipulation of V2.
27. A sample bottle placed in an ice bath is capped with a cap and Teflon liner, which the water phase sample tube passes through into the bottle. The bottle is capped to prevent contact with the atmosphere.
28. Filling sample bottles to the neck reduces headspace and restricts vapor phase mass transfer to a minimum. The liquid level in the water phase bottle can be determined visually, but since the organic phase sample is collected in a sightless cell, the liquid level in the organic phase bottle is determined from the organic phase flow rate.

29. When sampling of the water phase is completed, the sample tubing is removed and the cap is replaced with the original cap and liner from the initial weighing.
30. When sampling of the organic phase is completed, V2 is used to direct the organic phase sample flow to C2.
31. C1 is closed off from the flow of nitrogen by manipulation of V3.
32. V5 is opened allowing the slow depressurization of C1. Vaporization of the organic phase sample can occur if venting is rapid.
33. Once C1 has been completely vented, open C1.
34. Remove the organic phase sample bottle and cap with the original cap and liner from the initial weighing.

A.3 Shut Down and Preventive Maintenance

Apparatus Shut Down

After completion of sampling, the apparatus is shut down in a state where sampling can be resumed with minimal preparatory time.

35. The water phase sample line is connected to V8.
36. Pressurization of the water phase sample line is completed by opening V8 to the system pressure, which prevents liquid from leaking out of the apparatus from the water phase sample line.
37. V2 is used to direct the flow of the organic phase to either C1 at system pressure or C2 at system pressure.
38. Turn off DP1.

Preventive Maintenance

The apparatus requires little preventive maintenance, however, C2 does require periodic emptying by the following procedures:

39. Ensure C1 is closed and direct the organic phase sample flow to C1 by the use of V2.
40. Use the three-way valve (V4) to isolate C2 from the flow of nitrogen.
41. Vent C2 by opening the needle valve (V6).
42. After venting is complete, collect the waste contents of C2 by opening the needle valve (V7).
43. After emptying C2, close V7 and V6.
44. Pressurization of C2 is accomplished by directing the flow of nitrogen to C2 using V4. The pressure differential between the system pressure and C2 will result in the flow of nitrogen from both the system and the source, which causes a decrease in the system pressure. Manipulation of V4 should occur slowly such that the flow of nitrogen into C1 is not greater than the flow of nitrogen from the nitrogen source; thus ensuring the system pressure does not decrease below the mixture vapor pressure.

A.4 Analysis

Preparation for Analysis

After sample collection, the following procedures are followed in preparation for sample analysis:

45. The organic phase and water phase sample bottles are weighed to determine the sample weight.
46. The sample bottles are vigorously shaken resulting in the homogenization of the hydrocarbon/water/ethanol mixture of the organic phase sample and the extraction of the hydrocarbon from the water phase sample.
47. Refrigeration of the water phase for several hours allows the less dense extractant to separate from the water, and the chilled environment aids in the prevention of vaporization of the volatile hydrocarbon.

Sample Analysis

The samples are analyzed using the following procedures:

48. The organic phase may be analyzed immediately after collection.
49. 0.003 cm^3 ($3 \text{ }\mu\text{L}$) of the homogenized organic phase is injected into the gas chromatograph (GC) for analysis. Since the calibration involves the area ratio of water/ethanol versus the weight ratio of water/ethanol, the peaks of interest in the GC analysis of the organic phase are water and ethanol.
50. After the water phase sample has separated into two phases (extractant phase and water phase), 0.003 cm^3 ($3 \text{ }\mu\text{L}$) of the extractant phase is injected into the GC for analysis. Since the calibration is the weight ratio of hydrocarbon/extractant (decane or 2,2,4-trimethylpentane) versus the area ratio of hydrocarbon/extractant, the peaks of interest in the GC analysis of the water phase are the hydrocarbon and extractant.
51. The syringe is rinsed with the solution to be analyzed, which prevents cross-contamination between sample bottles.

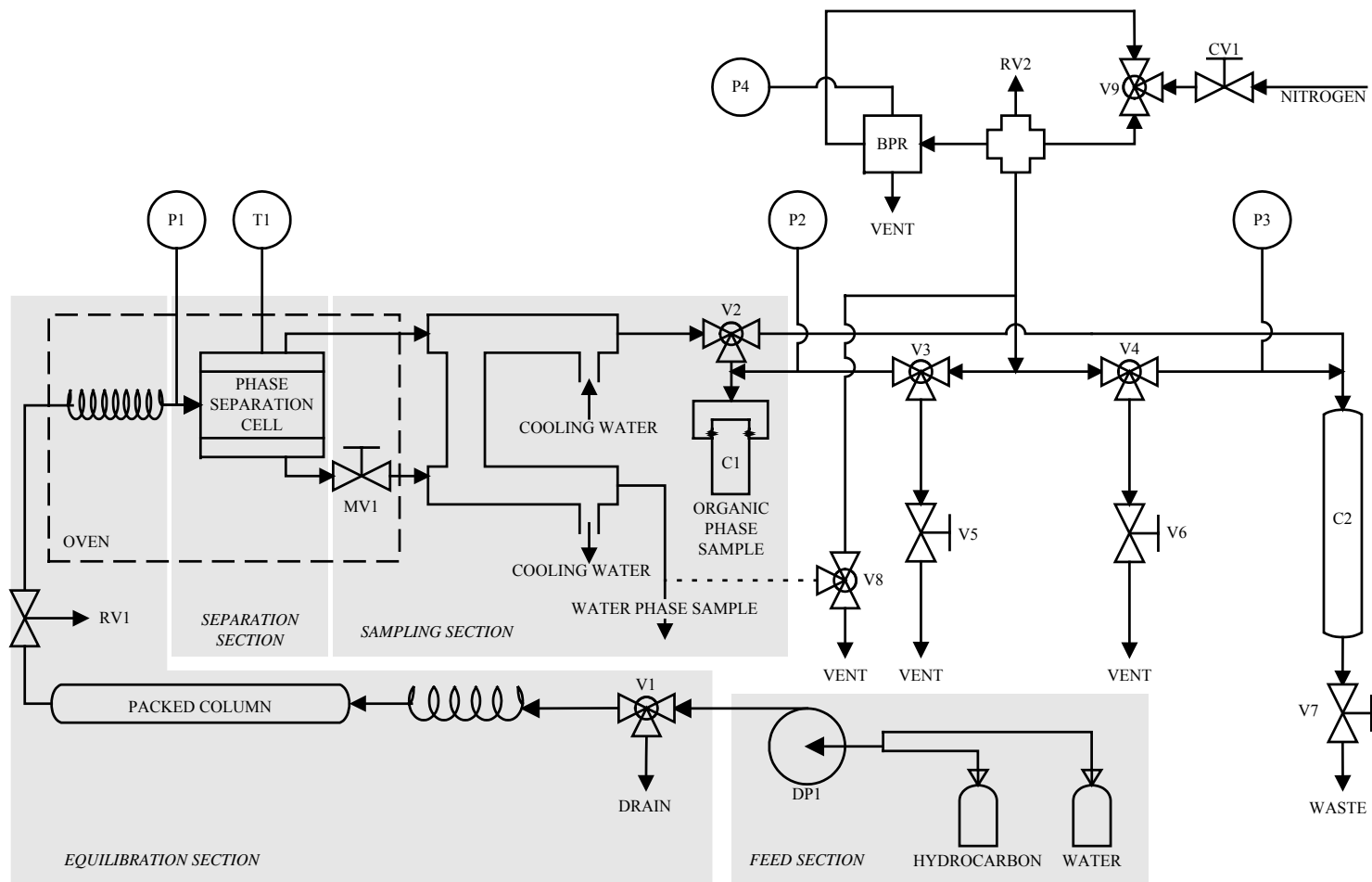


Figure A-1. Schematic Diagram of the Experimental Apparatus

Appendix B. Calibration Technique and Data

The gas chromatograph (GC) is calibrated by one of two techniques. For the aqueous phase (hydrocarbon in water) a serial dilution technique was utilized, and for the organic phase (water in hydrocarbon) individual external standards were employed.

B.1 Aqueous Phase

The serial dilution technique involves the dilution of a fixed amount of solute with increasing proportions of solvent through a series of dilutions. The weights of the solute and solvent are recorded to determine the solute-to-solvent weight ratio of each dilution. Using the GC, the mixtures are analyzed to obtain the corresponding solute-to-solvent area ratio. A calibration curve is produced, which yields the weight ratio as a function of the area ratio.

Procedure

Dilutions of the hydrocarbon of interest (solute) in either decane or 2,2,4-trimethylpentane (solvent) are prepared to calibrate the GC for the aqueous phase samples. The following calibration procedure was employed:

1. An empty 16 cm³ vial is weighed.
2. The solvent is added to the vial
3. The vial is weighed to obtain the weight of the solvent.

4. If the dilution is the first in the series, then pure solute is added to the vial, otherwise a portion of the previous dilution, the diluent, is added to the vial. The vials are filled to the neck to reduce headspace evaporation effects.
5. The vial is weighed to obtain the weight of the solute or the weight of the diluent.
6. The same Hamilton 10 *ml* syringe is used to make each serial dilution. In order to eliminate cross-contamination the syringe is rinsed with acetone between dilutions, allowed to air dry, and flushed with the diluent to be transferred before making the next dilution.
7. Immediately after preparation, the solution is analyzed by gas chromatography using a Hamilton 10 μ l syringe to inject 3 μ l samples. During analysis of the dilution, the vial is chilled in an ice bath to reduce evaporation.
8. After completion of the analysis, the next dilution is made in a similar fashion repeating steps 1-6.

Material Balance

A material balance is used to determine the mass of the solute and the mass of the solvent in each calibration mixture. The weight ratio is the weight of the solute in the mixture, A_i , divided by the weight of the solvent in the mixture, B_i :

$$WR_i = \frac{A_i}{B_i} \quad (\text{B-1})$$

If the mixture is first in the serial dilution, then the composition results from the addition of pure solute, A_i , and pure solvent, B_i ; otherwise, the mixture composition reflects the addition of solvent and diluent, D_i , which is a fraction, x , of the previous mixture

consisting of solute from the previous dilution, A_{i-1} , and solvent from the previous dilution, B_{i-1} , expressed as:

$$D_i = x(A_{i-1} + B_{i-1}) \quad (\text{B-2})$$

As the series of dilutions continues, the amount of solute in each dilution decreases; and Equation (B-1), may be written as:

$$WR_i = \frac{A_i + x(A_{i-1})}{B_i + x(B_{i-1})} \quad (\text{B-3})$$

Expressing Equation (B-2) in terms of A_{i-1} gives:

$$\frac{D_i}{x(A_{i-1})} = 1 + \frac{x(B_{i-1})}{x(A_{i-1})} \quad (\text{B-4})$$

but, substitution with Equation (B-1) results in:

$$\frac{D_i}{x(A_{i-1})} = 1 + \frac{1}{WR_{i-1}} \quad (\text{B-5})$$

and with rearrangement,

$$\frac{D_i}{x(A_{i-1})} = \frac{1 + WR_{i-1}}{WR_{i-1}} \quad (\text{B-6})$$

Solving for A_{i-1} , Equation (B-6) becomes:

$$x(A_{i-1}) = D_i \left(\frac{WR_{i-1}}{1 + WR_{i-1}} \right) \quad (\text{B-7})$$

Expressing Equation (B-2) in terms of B_{i-1} gives:

$$\frac{D_i}{x(B_{i-1})} = \frac{x(A_{i-1})}{x(B_{i-1})} + 1 \quad (\text{B-8})$$

but, substitution with Equation (B-1) results in:

$$\frac{D_i}{x(B_{i-1})} = WR_{i-1} + 1 \quad (\text{B-9})$$

Solving for B_{i-1} , Equation (B-9) becomes:

$$x(B_{i-1}) = D_i \left(\frac{1}{1 + WR_{i-1}} \right) \quad (\text{B-10})$$

Combination of Equations (B-3), (B-7), and (B-10) results in a general equation for the calibration weight ratio as given by:

$$WR_i = \frac{A_i + D_i \left(\frac{WR_{i-1}}{1 + WR_{i-1}} \right)}{B_i + D_i \left(\frac{1}{1 + WR_{i-1}} \right)} \quad (\text{B-11})$$

Rearrangement of Equation (B-11) gives:

$$WR_i = \frac{\left\{ \frac{(A_i)(1 + WR_{i-1}) + (D_i)(WR_{i-1})}{1 + WR_{i-1}} \right\}}{\left\{ \frac{(B_i)(1 + WR_{i-1}) + D_i}{1 + WR_{i-1}} \right\}} \quad (\text{B-12})$$

or

$$WR_i = \frac{(A_i)(1 + WR_{i-1}) + (D_i)(WR_{i-1})}{(B_i)(1 + WR_{i-1}) + D_i} \quad (\text{B-13})$$

where, WR_i is the solute-to-solvent weight ratio, A_i is the weight of the pure solute, B_i is the weight of the pure solvent, WR_{i-1} is the solute-to-solvent weight ratio of the previous dilution, and D_i is the weight of the diluent from the previous dilution added to the i^{th} dilution. Following this section, a sample calculation of the calibration is provided.

The calibration data appear at the end of the appendix in Tables B-1 - B-3. The uncertainty in the weight ratio was determined by propagated error analysis, as described in Appendix C. The uncertainty was used to weight each data point in the nonlinear weighted-least-squares regression of the calibration data.

Sample Calculation

Toluene-Water

1st Mixture

$$A_1 = 0.9578 \text{ grams} \quad B_1 = 9.0486 \text{ grams} \quad WR_0 = 0.0000 \quad D_1 = 0.0000$$

$$WR_1 = \frac{A_1}{B_1} = \frac{0.9578}{9.0486} = 0.1059$$

2nd Mixture

$$A_2 = 0.0000 \text{ grams} \quad B_2 = 5.4519 \text{ grams} \quad WR_1 = 0.1059 \quad D_2 = 4.8252$$

$$WR_2 = \frac{(0.0000)(1 + 0.1059) + (4.8252)(0.1059)}{(5.4519)(1 + 0.1059) + 4.8252} = \frac{0.5110}{10.8545} = 0.0471$$

3rd Mixture

$$A_3 = 0.0000 \text{ grams} \quad B_3 = 6.0905 \text{ grams} \quad WR_2 = 0.0471 \quad D_3 = 4.0457$$

$$WR_3 = \frac{(0.0000)(1 + 0.0471) + (4.0457)(0.0471)}{(6.0905)(1 + 0.0471) + 4.0457} = \frac{0.1906}{10.4231} = 0.0183$$

4th Mixture

$$A_4 = 0.0000 \text{ grams} \quad B_4 = 5.0320 \text{ grams} \quad WR_3 = 0.0183 \quad D_4 = 2.3742$$

$$WR_4 = \frac{(0.0000)(1 + 0.0183) + (2.3742)(0.0183)}{(5.0320)(1 + 0.0183) + 2.3742} = \frac{0.0434}{7.4983} = 0.0058$$

5th Mixture

$$A_5 = 0.0000 \text{ grams} \quad B_5 = 5.7385 \text{ grams} \quad WR_4 = 0.0058 \quad D_5 = 3.1046$$

$$WR_5 = \frac{(0.0000)(1 + 0.0058) + (3.1046)(0.0058)}{(5.7385)(1 + 0.0058) + 3.1046} = \frac{0.0180}{8.8764} = 0.0020$$

B.2 Organic Phase

The employment of external standards involves creating a series of standards with known chemical amounts covering the expected range of the quantity of interest. For the organic phase analysis a desired solvent (ethanol) to solute (water) weight ratio (WR) was used with a constant solvent to sample (hydrocarbon + water) ratio (SSR) to calculate the amounts of ethanol, hydrocarbon, and water required for the calibration standard. The weights of the solute and solvent are recorded to determine the solute-to-solvent weight ratio of each standard. Using the GC, the mixtures are analyzed to obtain the corresponding solute-to-solvent area ratio. A calibration curve is produced, which yields the weight ratio as a function of the area ratio.

Procedure

External standards containing the hydrocarbon of interest, ethanol (solvent), and water (solute) are prepared to calibrate the GC for the organic phase samples. The following calibration procedure was employed:

1. An empty 1 oz bottle is weighed.
2. Ethanol (solvent) is added to the empty bottle.
3. The vial is weighed to obtain the weight of the solvent.
4. Using the SSR and the weight of the solvent, the amount of the hydrocarbon of interest required is calculated and added to the bottle. This assumes that the weight of water has a negligible contribution to the total sample weight. During production of the standards, the bottles are filled nearly to the neck for the reduction of headspace evaporation effects.
5. The vial is weighed to obtain the weight of the hydrocarbon.
6. Using the WR and the weight of the solvent, the amount of solute (water) required for the calibration standard is calculated and added to the bottle.
7. The bottle is shaken vigorously for a period of one minute to insure homogeneity.
8. Immediately after preparation, the solution is analyzed by gas chromatography using a Hamilton 10 μ l syringe to inject 3 μ l samples. During analysis of the dilution, the vial is chilled in an ice bath to reduce evaporation.
9. After completion of the analysis, the next standard is made in a similar fashion repeating steps 1-8.

Material Balance

A material balance is used to determine the mass of the solute and the mass of the solvent in each calibration standard. The weight ratio is the weight of the solute in the mixture, A_i , divided by the weight of the solvent in the mixture, B_i :

$$WR_i = \frac{A_i}{B_i} \quad (\text{B-14})$$

A sample calculation of the weight ratio calibration is provided following this section. The calibration data appear at the end of this appendix in Tables B-1 - B-3. The uncertainty in the weight ratio was determined by propagated error analysis, as described in Appendix C. The uncertainty was used to weight each data point in the nonlinear weighted-least-squares regression of the calibration data.

Sample Calculation

Toluene-Water

1st Standard

$$A_1 = 0.0128 \text{ grams} \quad B_1 = 5.5368 \text{ grams}$$

$$WR_1 = \frac{A_1}{B_1} = \frac{0.0128}{5.5368} = 0.0023$$

2nd Standard

$$A_1 = 0.0210 \text{ grams} \quad B_1 = 5.3286 \text{ grams}$$

$$WR_1 = \frac{A_1}{B_1} = \frac{0.0210}{5.3286} = 0.0039$$

3rd Standard

$$A_1 = 0.3450 \text{ grams} \quad B_1 = 5.0548 \text{ grams}$$

$$WR_1 = \frac{A_1}{B_1} = \frac{0.3450}{5.0548} = 0.0683$$

4th Standard

$$A_1 = 1.4165 \text{ grams} \quad B_1 = 7.4979 \text{ grams}$$

$$WR_1 = \frac{A_1}{B_1} = \frac{1.4165}{7.4979} = 0.1889$$

B.3 Calibration Results

Results from the calibrations of the systems of interest are given in the following tables:

Table B-1. Calibration Data for the Benzene-Water System

Benzene - Water Calibration		
Aqueous Phase		
Area Ratio	Weight Ratio	Uncertainty in the Weight Ratio
0.1228	0.1019	0.00025
0.0201	0.0093	0.00016
0.0068	0.0031	0.00003
0.0016	0.0007	0.00003
0.0103	0.0047	0.00020
Organic Phase		
Area Ratio	Weight Ratio	Uncertainty in the Weight Ratio
0.0109	0.0060	0.00056
0.1506	0.1007	0.00219
0.0231	0.0134	0.00177
0.0033	0.0012	0.00011

Table B-2. Calibration Data for the Toluene-Water System

Toluene-Water Calibration		
Aqueous Phase		
Area Ratio	Weight Ratio	Uncertainty in the Weight Ratio
0.4817	0.1059	0.00580
0.0856	0.0183	0.00291
0.0288	0.0058	0.00030
0.0099	0.0020	0.00005
Organic Phase		
Area Ratio	Weight Ratio	Uncertainty in the Weight Ratio
0.0049	0.0023	0.00020
0.0078	0.0039	0.00036
0.0683	0.0683	0.00137
0.1892	0.1889	0.02474

Table B-3. Calibration Data for the 3-Methylpentane-Water System

3-Methylpentane - Water Calibration		
Aqueous Phase		
Area Ratio	Weight Ratio	Uncertainty in the Weight Ratio
0.0905	0.0580	0.00092
0.0282	0.0099	0.00082
0.0177	0.0056	0.00027
0.0035	0.0010	0.00008
0.0405	0.0182	0.00084
Organic Phase – High Area Ratio		
Area Ratio	Weight Ratio	Uncertainty in the Weight Ratio
0.0170	0.0008	0.00006
0.0259	0.0013	0.00006
0.0347	0.0021	0.00019
0.3629	0.0350	0.00070
0.7953	0.0831	0.00528
Organic Phase – Low Area Ratio		
Area Ratio	Weight Ratio	Uncertainty in the Weight Ratio
0.0173	0.0008	0.00007
0.0148	0.0004	0.00009
0.0118	0.0001	0.00004

Appendix C. Propagated Calibration Error

C.1 Introduction

The calibration of the gas chromatograph involves known solute-to-solvent weight ratios, which are measured and given as a function of their respective solute-to-solvent gas chromatograph area ratios. A calculated weight ratio is a function of the weight measurements and the chromatographic area ratio.

C.2 Uncertainty

The weight ratio, as a function of the weight measurements, is given in the simplified form of the governing material balance equation as:

$$WR_i = \frac{A_i}{B_i} \quad (C-1)$$

where WR_i is the solute-to-solvent weight ratio, A_i is the weight of the pure solute, and B_i is the weight of the pure solvent. When utilizing serial dilutions, the simplified equation is expanded to reflect the addition of previous mixtures as shown:

$$WR_i = \frac{(A_i)(1 + WR_{i-1}) + (D_i)(WR_{i-1})}{(B_i)(1 + WR_{i-1}) + D_i} \quad (C-2)$$

where WR_{i-1} is the solute-to-solvent weight ratio of the previous dilution, and D_i is the weight of the diluent from the previous dilution added to the i^{th} dilution. The dependence of the weight ratio on the area ratio is described by the calibrating equation:

$$WR_i = \alpha AR_i^\beta \quad (C-3)$$

where AR_i is the solute-to-solvent area ratio and α and β are regressed parameters.

The propagated uncertainty in the weight ratio from Equation (C-3) is expressed in terms of variances as shown in the following equation:

$$\begin{aligned} \sigma_{WR_i}^2 = & \left(\frac{\partial WR_i}{\partial A_i} \right)^2 \sigma_{A_i}^2 + \left(\frac{\partial WR_i}{\partial B_i} \right)^2 \sigma_{B_i}^2 + \left(\frac{\partial WR_i}{\partial WR_{i-1}} \right)^2 \sigma_{WR_{i-1}}^2 \\ & + \left(\frac{\partial WR_i}{\partial D_i} \right)^2 \sigma_{D_i}^2 + \left(\frac{\partial WR_i}{\partial AR_i} \right)^2 \sigma_{AR_i}^2 \end{aligned} \quad (C-4)$$

This propagated uncertainty in the weight ratio may then be expressed in terms of fractional uncertainty:

$$\begin{aligned} \left(\frac{\sigma_{WR_i}}{WR_i} \right)^2 = & \left(\frac{\partial WR_i}{\partial A_i} \right)^2 \left(\frac{\sigma_{A_i}}{WR_i} \right)^2 + \left(\frac{\partial WR_i}{\partial B_i} \right)^2 \left(\frac{\sigma_{B_i}}{WR_i} \right)^2 + \left(\frac{\partial WR_i}{\partial WR_{i-1}} \right)^2 \left(\frac{\sigma_{WR_{i-1}}}{WR_i} \right)^2 \\ & + \left(\frac{\partial WR_i}{\partial D_i} \right)^2 \left(\frac{\sigma_{D_i}}{WR_i} \right)^2 + \left(\frac{\partial WR_i}{\partial AR_i} \right)^2 \left(\frac{\sigma_{AR_i}}{WR_i} \right)^2 \end{aligned} \quad (C-5)$$

If the weight ratios were exact, there would still exist error in the value calculated from the calibration relation due to the uncertainty in the area ratio obtained from the gas chromatograph. The final term of Equation (C-5) accounts for this, and since this term is independent of any previous measurements, the uncertainty associated with the area ratio is not propagated.

C.3 Uncertainty Associated with the Weight of Pure Solute

The partial derivative of WR_i with respect to A_i is given as follows:

$$\left(\frac{\partial WR_i}{\partial A_i}\right)^2 = \left(\frac{1 + WR_{i-1}}{(B_i)(1 + WR_{i-1}) + D_i}\right)^2 \quad (C-6)$$

This equation is divided by WR_i^2 from Equation (C-2) to find the fractional uncertainty associated with the weight of pure solute added to the i^{th} dilution. The first term in Equation (C-5) then becomes:

$$\left(\frac{\partial WR_i}{\partial A_i}\right)^2 \left(\frac{\sigma_{A_i}}{WR_i}\right)^2 = \left\{ \frac{1 + WR_{i-1}}{(A_i)(1 + WR_{i-1}) + (D_i)(WR_{i-1})} \right\}^2 \sigma_{A_i}^2 \quad (C-7)$$

C.4 Uncertainty Associated with the Weight of Pure Solvent

The partial derivative of WR_i with respect to B_i is as follows:

$$\left(\frac{\partial WR_i}{\partial B_i}\right)^2 = \left\{ -\frac{[(A_i)(1 + WR_{i-1}) + (D_i)(WR_{i-1})](1 + WR_{i-1})}{[(B_i)(1 + WR_{i-1}) + D_i]^2} \right\}^2 \quad (C-8)$$

Dividing this equation by WR_i^2 provides the fractional uncertainty associated with the weight of pure solvent added to the i^{th} dilution, and the second term of Equation (C-5) then is shown as:

$$\left(\frac{\partial WR_i}{\partial B_i}\right)^2 \left(\frac{\sigma_{B_i}}{WR_i}\right)^2 = \left\{ \frac{1 + WR_{i-1}}{(B_i)(1 + WR_{i-1}) + D_i} \right\}^2 \sigma_{B_i}^2 \quad (C-9)$$

C.5 Uncertainty Associated with the Weight Ratio of the Diluent

The partial derivative of WR_i with respect to WR_{i-1} is:

$$\left(\frac{\partial WR_i}{\partial WR_{i-1}}\right)^2 = \left\{ \frac{(A_i + D_i)[(B_i)(1 + WR_{i-1}) + D_i] - [(A_i)(1 + WR_{i-1}) + (D_i)(WR_{i-1})](B_i)}{[(B_i)(1 + WR_{i-1}) + D_i]^2} \right\}^2 \quad (C-10)$$

The fractional uncertainty associated with the weight ratio of the $i-1^{\text{th}}$ dilution, the third term in Equation (C-5), is obtained by division of this equation by WR_i^2 as follows:

$$\left(\frac{\partial WR_i}{\partial WR_{i-1}}\right)^2 \left(\frac{\sigma_{WR_{i-1}}}{WR_i}\right)^2 = \left\{ \frac{A_i + D_i}{(A_i)(1 + WR_{i-1}) + (D_i)(WR_{i-1})} - \frac{B_i}{(B_i)(1 + WR_{i-1}) + D_i} \right\}^2 \sigma_{WR_{i-1}}^2 \quad (C-11)$$

C.6 Uncertainty Associated with the Weight of Diluent

The partial derivative of WR_i with respect to D_i is:

$$\left(\frac{\partial WR_i}{\partial D_i}\right)^2 = \left\{ \frac{(WR_{i-1})[(B_i)(1 + WR_{i-1}) + D_i] - [(A_i)(1 + WR_{i-1}) + (D_i)(WR_{i-1})](B_i)}{[(B_i)(1 + WR_{i-1}) + D_i]^2} \right\}^2 \quad (C-12)$$

The fractional uncertainty associated with the weight of the $i-1^{\text{th}}$ dilution added to the i^{th} dilution, the fourth term of Equation (C-5), is determined by dividing this equation by WR_i^2 as follows:

$$\left(\frac{\partial WR_i}{\partial D_i}\right)^2 \left(\frac{\sigma_{D_i}}{WR_i}\right)^2 = \left\{ \frac{WR_{i-1}}{(A_i)(1 + WR_{i-1}) + (D_i)(WR_{i-1})} - \frac{1}{(B_i)(1 + WR_{i-1}) + D_i} \right\}^2 \sigma_{D_i}^2 \quad (C-13)$$

C.7 Uncertainty Associated with the Gas Chromatograph Area Ratio

The partial derivative of the calibration equation, Equation (C-3), is given by taking the partial derivative of WR_i with respect to AR_i , as follows:

$$\frac{\partial WR_i}{\partial AR_i} = \alpha \beta AR_i^{(\beta-1)} \quad (C-14)$$

This expression is divided by WR_i to determine the fractional uncertainty associated with the gas chromatograph area ratio, the final term in Equation (C-5), as follows:

$$\left(\frac{\partial WR_i}{\partial AR_i} \right)^2 \left(\frac{\sigma_{AR_i}}{WR_i} \right)^2 = \left\{ \frac{\alpha \beta AR_i^{(\beta-1)}}{WR_i} \right\}^2 \sigma_{AR_i}^2 \quad (C-15)$$

C.8 Total Fractional Uncertainty

Combining Equations (C-7), (C-9), (C-11), (C-13), and (C-15) gives the total propagated fractional uncertainty in the weight ratio as shown by:

$$\begin{aligned} \left(\frac{\sigma_{WR_i}}{WR_i} \right)^2 &= \left\{ \frac{1 + WR_{i-1}}{(A_i)(1 + WR_{i-1}) + (D_i)(WR_{i-1})} \right\}^2 \sigma_{A_i}^2 \\ &+ \left\{ \frac{1 + WR_{i-1}}{(B_i)(1 + WR_{i-1}) + D_i} \right\}^2 \sigma_{B_i}^2 \\ &+ \left\{ \frac{A_i + D_i}{(A_i)(1 + WR_{i-1}) + (D_i)(WR_{i-1})} - \frac{B_i}{(B_i)(1 + WR_{i-1}) + D_i} \right\}^2 \sigma_{WR_{i-1}}^2 \\ &+ \left\{ \frac{WR_{i-1}}{(A_i)(1 + WR_{i-1}) + (D_i)(WR_{i-1})} - \frac{1}{(B_i)(1 + WR_{i-1}) + D_i} \right\}^2 \sigma_{D_i}^2 \\ &+ \left\{ \frac{\alpha \beta AR_i^{(\beta-1)}}{WR_i} \right\}^2 \sigma_{AR_i}^2 \end{aligned} \quad (C-16)$$

This uncertainty expression was used to provide weights for the weighted least squares regression of the calibration data for the calibration relation, Equation (C-3). Values of α and β were first estimated from an unweighted least squares regression and subsequently used in Equation (C-16). An iterative procedure was then employed to

determine final values of the calibration constants and the weighting of the calibration data. The variance in the weight measurements, σ_{A_i} , σ_{B_i} , and σ_{D_i} , was determined by repeated measurements and the variance in the area ratio, σ_{AR_i} , value was determined from multiple analyses. Since the uncertainty associated with the area ratio, the last term in Equation (C-16), is independent of any previous measurements, this uncertainty is not propagated. Therefore, this term was not included when determining the variance in the weight ratio of the $i-1^{\text{th}}$ dilution, $\sigma_{WR_{i-1}}$.

The total uncertainty is largely dependent upon the uncertainty associated with the gas chromatograph area ratio, the last term in Equation (C-16). For the initial calibration dilutions and the external standard calibration mixtures, all other terms are negligible. For the final dilutions, or the mixtures lowest in concentration, the uncertainty associated with the gas chromatograph area ratio is approximately one-half the total uncertainty.

C.9 Sample Calculation (Toluene-Water)

Sample calculations are provided for the uncertainty of the first two solutions associated with the aqueous phase serial dilution calibration of the toluene-water system. If the calibration of interest involves external standards, then all calculations of uncertainty would be similar to those of the first example, Solution #1.

Solution #1

$$A_1 = 0.9578 \text{ grams} \quad B_1 = 9.0486 \text{ grams} \quad WR_0 = 0.0000 \quad D_1 = 0.0000 \text{ grams}$$

$$\sigma_{A_1} = 0.00021 \quad \sigma_{B_1} = 0.00021 \quad \sigma_{WR_0} = 0.0000 \quad \sigma_{D_1} = 0.0000$$

$$WR_1 = 0.1059 \quad AR_1 = 0.4817 \quad \sigma_{AR_1} = 2.58E - 02$$

$$\alpha = 0.2214 \quad \beta = 1.0198$$

$$\left(\frac{\sigma_{WR_1}}{WR_1} \right)^2 = \left\{ \frac{1}{(0.9578)(1)} \right\}^2 (0.00021)^2$$

$$+ \left\{ \frac{1}{(9.0486)(1)} \right\}^2 (0.00021)^2$$

$$+ \left\{ \frac{0.9578}{(0.9578)(1)} - \frac{9.0486}{(9.0486)(1)} \right\}^2 (0.0000)^2$$

$$+ \left\{ \frac{0.0000}{(0.9578)(1)} - \frac{1}{(9.0486)(1)} \right\}^2 (0.0000)^2$$

$$+ \left\{ \frac{(0.2214)(1.0198)(0.4817)^{(1.0198-1)}}{0.1059} \right\}^2 (2.58E - 02)^2$$

$$\left(\frac{\sigma_{WR_1}}{WR_1} \right)^2 = 4.8072E-08 + 5.3861E-10 + 0.0000 + 0.0000 + 2.9395E-03$$

$$= 2.9395E-03$$

or,

$$\sigma_{WR_1} = (2.9395E - 03)(WR_1)$$

$$= (2.9395E - 03)(0.1059)$$

$$= 5.7416E-03$$

This is the total uncertainty for the first dilution in the calibration, which is used to weight the first point in the weighted least squares regression of the calibration curve.

Solution #2 (First Dilution)

$$A_2 = 0.0000 \text{ grams} \quad B_2 = 6.0905 \text{ grams} \quad WR_1 = 0.1059 \quad D_2 = 4.0457 \text{ grams}$$

$$\sigma_{A_2} = 0.00021 \quad \sigma_{B_2} = 0.00021 \quad \sigma_{WR_1} = 2.357E - 05 \quad \sigma_{D_2} = 0.00021$$

$$WR_2 = 0.0183 \quad AR_2 = 0.0856 \quad \sigma_{AR_1} = 1.38E - 02$$

$$\alpha = 0.2214 \quad \beta = 1.0198$$

$$\begin{aligned} \left(\frac{\sigma_{WR_2}}{WR_2} \right)^2 &= \left\{ \frac{1 + 0.1059}{(4.0457)(0.1059)} \right\}^2 (0.00021)^2 \\ &+ \left\{ \frac{1 + 0.1059}{(6.0905)(1 + 0.1059) + 4.0457} \right\}^2 (0.00021)^2 \\ &+ \left\{ \frac{4.0457}{(4.0457)(0.1059)} - \frac{6.0905}{(6.0905)(1 + 0.1059) + 4.0457} \right\}^2 (2.357E - 05)^2 \\ &+ \left\{ \frac{0.1059}{(4.0457)(0.1059)} - \frac{1}{(6.0905)(1 + 0.1059) + 4.0457} \right\}^2 (0.00021)^2 \\ &+ \left\{ \frac{(0.2214)(1.0198)(0.0856)^{(1.0198-1)}}{0.0183} \right\}^2 (1.38E - 02)^2 \end{aligned}$$

$$\left(\frac{\sigma_{WR_2}}{WR_2} \right)^2 = 2.938E-07 + 4.640E-10 + 4.379E-08 + 1.052E-09 + 2.630E-02$$

$$= 2.630E-02$$

or,

$$\sigma_{WR_2} = (1.622E - 01)(WR_2)$$

$$= (1.622E - 01)(0.0183)$$

$$= 2.968E-03$$

This is the total uncertainty for the second dilution in the calibration. This value is used to weight the second point in the weighted least squares regression of the calibration curve.

Appendix D. Solubility Calculations and Data

Mutual solubilities are expressed as the mole fraction of solute in the sample. Since the weight of the solute in the sample is unknown, the solute-to-solvent weight ratio (WR), the solvent-to-sample weight ratio (SSR), and the molecular weights (MW) of the two species of interest, are used to calculate the solute mole fraction. The resulting expression for the mole fraction of a binary system is as follows:

$$x_1 = \frac{\frac{(WR)(SSR)}{MW_1}}{\frac{(WR)(SSR)}{MW_1} + \frac{1}{MW_1}} \quad (D-1)$$

where the subscript indicates either of the two species of interest.

The solute-to-solvent weight ratio is given in the calibration equation as a function of the solute-to-solvent area ratio (AR), which is determined by gas chromatography. The solvent-to-sample weight ratio is a ratio of the weight of the solvent added (ethanol in the organic phase and decane or 2,2,4-trimethylpentane in the water phase) to the weight of the sample collected.

Each solubility measurement reported is an average of nine to fifteen measurements. At each temperature studied, three samples were collected of each phase, and each phase sample was then analyzed a minimum of three times. An example calculation is shown below.

D.1 Sample Calculation (Toluene-Water)

Temperature = 350.9 K Pressure = 0.358 Mpa

Aqueous Phase				
Sample #	SSR	AR	WR	Mole Fraction Toluene
16	0.1421	0.02935	0.00606	1.671E-04
16	0.1421	0.02972	0.00614	1.693E-04
16	0.1421	0.02959	0.00611	1.685E-04
16	0.1421	0.02963	0.00612	1.687E-04
16	0.1421	0.02930	0.00605	1.668E-04
17	0.1617	0.02840	0.00586	1.838E-04
17	0.1617	0.02854	0.00589	1.847E-04
17	0.1617	0.02818	0.00581	1.824E-04
17	0.1617	0.02846	0.00587	1.842E-04
18	0.1502	0.03074	0.00635	1.850E-04
18	0.1502	0.03033	0.00626	1.825E-04
18	0.1502	0.03056	0.00631	1.839E-04
18	0.1502	0.03032	0.00626	1.825E-04
Mole Fraction of Toluene =				1.776E-04

Organic Phase				
Sample #	SSR	AR	WR	Mole Fraction Water
13	0.36223	0.00926	0.00528	8.840E-03
13	0.36223	0.00931	0.00533	8.907E-03
13	0.36223	0.00992	0.0058	9.618E-03
13	0.36223	0.01007	0.00591	9.793E-03
14	0.30871	0.01264	0.00942	1.291E-02
14	0.30871	0.01278	0.00956	1.309E-02
14	0.30871	0.01319	0.00997	1.360E-02
14	0.30871	0.01331	0.01009	1.375E-02
16	0.37777	0.01178	0.00699	1.184E-02
16	0.37777	0.01143	0.00671	1.142E-02
16	0.37777	0.01121	0.00654	1.115E-02
16	0.37777	0.01217	0.0073	1.232E-02
16	0.37777	0.01230	0.00741	1.248E-02
Mole Fraction of Water =				1.152E-02

Appendix E. Organic Phase Sample Analysis Correction

E.1 Introduction

Ethanol, which is employed as a cosolvent for the homogenization of the organic phase samples, is hygroscopic and contains a small amount of water (typically less than 0.015% by GC analysis). While the ethanol is stored over molecular sieves and contact with the atmosphere is kept to a minimum, a correction for the water introduced to the sample by the ethanol addition is necessary.

The total weight of an ethanol aliquot, W_e , is the sum of the weight of the water fraction of the ethanol, $W_{w,e}$, and the weight of the ethanol fraction, $W_{e,e}$:

$$W_e = W_{w,e} + W_{e,e} \quad (\text{E-1})$$

The total weight of the organic phase sample, W_s , from the experiment is the sum of the weight of the water in the sample, $W_{w,s}$, and the weight of the hydrocarbon in the sample, $W_{h,s}$:

$$W_s = W_{w,s} + W_{h,s} \quad (\text{E-2})$$

Prior to defining the mass balance, several terms are defined, as follows:

$ER_{w-e} = W_{w,e}/W_{e,e}$; weight ratio of the water fraction of the ethanol, $W_{w,e}$, to the ethanol fraction of the ethanol, $W_{e,e}$

$ER_w = W_{w,e}/W_e$; mass fraction of water in the ethanol

$WR_{s-s} = W_s/W_e$; weight ratio of the sample, W_s , to the solvent, W_e

$WR_{w-h} = W_{w,s}/W_{h,s}$; weight ratio of the water in the sample, $W_{w,s}$ to the hydrocarbon in the sample, $W_{h,s}$

$MF_w = W_{w,s}/W_s$; mass fraction of water in the sample

E.2 Material Balance

The total weight of the ethanol in a given analysis mixture, consisting of an organic phase sample mixed with ethanol, is given as:

$$W_e = W_{w,e} + \frac{W_{w,e}}{ER_{w-e}} \quad (E-3)$$

$$= W_{w,e} \left(\frac{1 + ER_{w-e}}{ER_{w-e}} \right) \quad (E-4)$$

$$= \frac{W_{w,e}}{ER_w} \quad (E-5)$$

$$= W_{e,e} ER_{w-e} + W_{e,e} \quad (E-6)$$

$$= W_{e,e} (1 + ER_{w-e}) \quad (E-7)$$

The total weight of the sample is expressed as:

$$W_s = W_{w,s} + \frac{W_{w,s}}{WR_{w-h}} \quad (E-8)$$

$$= W_{w,s} \left(\frac{1 + WR_{w-h}}{WR_{w-h}} \right) \quad (E-9)$$

$$= \frac{W_{w,s}}{MF_w} \quad (E-10)$$

$$= W_{h,s} WR_{w-h} + W_{h,s} \quad (E-11)$$

$$= W_{h,s}(1 + WR_{w-h}) \quad (\text{E-12})$$

The equation used in the calibration of the GC is:

$$WR_{w-e} = \alpha AR_{w-e}^\beta \quad (\text{E-13})$$

where, WR_{w-e} is the weight ratio of the total amount of water to the total amount of ethanol, AR_{w-e} is the area ratio of the total amount of water to the total amount of ethanol, and α and β are the calibration parameters.

E.3 Organic Phase Sample Calculations

After GC analysis, the results from an organic phase sample would then be calculated in the following manner:

$$WR_{w-e} = \alpha AR_s^\beta \quad (\text{E-14})$$

where, AR_s is the area ratio of the total amount of water to the total amount of ethanol in the organic phase sample. By mass balance, Equation (E-14) is expressed as:

$$\frac{MF_w(W_s) + W_e(ER_w)}{W_e(1 - ER_w)} = \frac{WR_{s-s}(MF_w)}{1 - ER_w} + \frac{ER_w}{1 - ER_w} \quad (\text{E-15})$$

$$= \frac{WR_{s-s}(MF_w) + ER_w}{1 - ER_w} \quad (\text{E-16})$$

$$= \alpha AR_s^2 + \beta AR_s + \gamma \quad (\text{E-17})$$

rearranging,

$$MF_w = \frac{(\alpha AR_s^2 + \beta AR_s + \gamma)(1 - ER_w) - ER_w}{WR_{s-s}} \quad (\text{E-18})$$

While the calibration parameters in Equation (E-18), α , β , and γ are regressed from calibration data, the variables, AR_s and WR_{s-s} , are measured. The mass fraction of

water in the ethanol, ER_w , is unknown. The value of ER_w is accounted for by relating the weight ratio of the solvent to the sample of the calibration (SSR_c) and sample (SSR_s) mixtures, as shown by:

$$SSR_{c,s} = \frac{1}{WR_{s-s}} = \frac{W_e}{W_s} \quad (E-19)$$

where the subscript on SSR indicates either a calibration or sample mixture.

E.4 Calibration Calculations

Analyses of distilled water, ethanol, and hydrocarbon (of interest) external standard mixtures provide the data necessary for calibration. Equation (E-18) is applied to the calibration in the following form:

$$\frac{WR_{s-s}MF_w + ER_w}{1 - ER_w} = \alpha AR_s^\beta \quad (E-20)$$

but, for the calibration

$$WR_{s-s} = \frac{W_w}{W_e} \quad (E-21)$$

and

$$MF_w = 1 \quad (E-22)$$

which results in the following expression:

$$\frac{WR_{s-s} + ER_w}{1 - ER_w} = \alpha AR_c^\beta \quad (E-23)$$

where, AR_c is the area ratio of the total amount of water to the total amount of ethanol in the calibration mixture.

E.5 Water Correction Calculations

The mass fraction of water in the ethanol, ER_w , is unknown and is accounted for by relating the weight ratios, SSR_c and SSR_s , of the calibration and sample mixtures. The calibration standards are composed of distilled water, ethanol, and the hydrocarbon of interest. These standards are prepared gravimetrically, and with due attention, a constant SSR_c for each standard in the calibration set is maintained. Ideally, just enough sample, distilled water, and hydrocarbon, is collected in the sample bottle, which contains ethanol, to replicate the value of SSR_c ; thus, equating SSR_c and SSR_s . Since the SSR 's of the calibration and sample mixtures are equal, the mass fractions of the water in the ethanol, ER_w , of each mixture will also have equal values. With equal values of ER_w , a correction for water present in the ethanol is unnecessary since analysis of the calibration and sample mixtures are based on the same proportion of ethanol to sample.

However, in practice exact duplication of SSR_s is difficult, which results in the use of a correction factor to adjust the mass fraction of water in the ethanol, ER_w . The correction factor, CF , is described as:

$$CF = \left(1 - \frac{SSR_c}{SSR_s} \right) \quad (E-24)$$

Application of Equation (E-23) based on the weight ratio of the sample along with the inclusion of CF from Equation (E-24) to the mass fraction of water in the ethanol results in the following expression:

$$\frac{ER_w}{1 - ER_w} = \alpha AR_s^\beta \left(1 - \frac{SSR_c}{SSR_s} \right) \quad (E-25)$$

As applied in this case, W_w , the total weight of water from an additional source, is zero.

Solving for ER_w , Equation (E-25) becomes:

$$ER_w = \frac{\alpha AR_s^\beta \left(1 - \frac{SSR_c}{SSR_s}\right)}{1 + \alpha AR_s^\beta \left(1 - \frac{SSR_c}{SSR_s}\right)} \quad (E-26)$$

In order to finalize the correction for differences in SSR_c and SSR_s , which reflects the introduction of water to the sample from the ethanol, Equations (E-20) and (E-26) are combined to provide the corrected mass fraction of water in the sample, as given by:

$$MF_w = \frac{\alpha AR_s^\beta - \alpha AR_s^\beta \left(1 - \frac{SSR_c}{SSR_s}\right)}{WR_{s-s} \left(1 + \alpha AR_s^\beta \left(1 - \frac{SSR_c}{SSR_s}\right)\right)} \quad (E-27)$$

And solving for the weight ratio of water in the sample to ethanol, WR_s , gives,

$$WR_s = \frac{\alpha AR_s^\beta - \alpha AR_s^\beta \left(1 - \frac{SSR_c}{SSR_s}\right)}{1 + \alpha AR_s^\beta \left(1 - \frac{SSR_c}{SSR_s}\right)} \quad (E-28)$$

Appendix F. Experimental Error Analysis

F.1 Introduction

The mole fraction, x , of a component in a binary mixture is expressed as:

$$x_1 = \frac{n_1}{n_1 + n_2} \quad (\text{F-1})$$

where n is the number of moles and the subscript indicates either component 1 or 2. In Equation (F-1), component 1 is considered to be the solute. Replacing n_1 and n_2 in terms of the weight ratio of solute to solvent, WR , the weight ratio of solvent to sample, SSR , and the molecular weights of both component 1 and 2, MW_1 and MW_2 , respectively, expresses x_1 as the following:

$$x_1 = \frac{\left[\frac{WR (SSR)}{MW_1} \right]}{\left[\frac{WR (SSR)}{MW_1} + \frac{1}{MW_2} \right]} \quad (\text{F-2})$$

Expected uncertainties associated with mole fractions calculated from solubility data are estimated by error propagation. In general, where R is a function of the measured variables x_1, x_2, \dots, x_n , the expected variance, σ_R^2 , is calculated as [32]:

$$\sigma_R^2 = \sum_{i=1}^n \left[\left(\frac{\partial R}{\partial x_i} \right)^2 \sigma_{x_i}^2 \right] \quad (\text{F-3})$$

Since liquid-liquid mutual solubilities at equilibrium increase with temperature, the mole fraction, x_1 , is also a function of temperature, T , as well as WR and SSR . Equation (F-3) can be rewritten in the terms of variances, which account for uncertainty in the mole fraction, as the following:

$$\sigma_{x_1}^2 = \left(\frac{\partial x_1}{\partial SSR} \right)^2 \sigma_{SSR}^2 + \left(\frac{\partial x_1}{\partial WR} \right)^2 \sigma_{WR}^2 + \left(\frac{\partial x_1}{\partial T} \right)^2 \sigma_T^2 \quad (F-4)$$

however, WR is also a function of the gas chromatograph area ratio of solute-to-solvent.

Using Equation (F-3), the variance in WR can be expressed as:

$$\sigma_{WR}^2 = \left(\frac{\partial WR}{\partial AR} \right)^2 \sigma_{AR}^2 \quad (F-5)$$

Combination of Equations (F-4) and (F-5) results in the following expression:

$$\sigma_{x_1}^2 = \left(\frac{\partial x_1}{\partial SSR} \right)^2 \sigma_{SSR}^2 + \left(\frac{\partial x_1}{\partial WR} \right)^2 \left(\frac{\partial WR}{\partial AR} \right)^2 \sigma_{AR}^2 + \left(\frac{\partial x_1}{\partial T} \right)^2 \sigma_T^2 \quad (F-6)$$

Equation (F-6) is then expressed in terms of fractional uncertainty as:

$$\left(\frac{\sigma_{x_1}}{x_1} \right)^2 = \left(\frac{\partial x_1}{\partial SSR} \right)^2 \left(\frac{\sigma_{SSR}}{x_1} \right)^2 + \left(\frac{\partial x_1}{\partial WR} \right)^2 \left(\frac{\partial WR}{\partial AR} \right)^2 \left(\frac{\sigma_{AR}}{x_1} \right)^2 + \left(\frac{\partial x_1}{\partial T} \right)^2 \left(\frac{\sigma_T}{x_1} \right)^2 \quad (F-7)$$

F.2 Uncertainty Associated with the Solvent-to-Sample Weight Ratio

The derivation of the uncertainty associated with SSR begins by taking the partial derivative of x_1 , from Equation (F-2), with respect to SSR as follows:

$$\frac{\partial x_1}{\partial SSR} = \frac{\left(\frac{WR}{MW_1} \right) \left(\frac{(WR)(SSR)}{MW_1} + \frac{1}{MW_2} \right) - \left(\frac{(WR)(SSR)}{MW_1} \right) \left(\frac{WR}{MW_1} \right)}{\left[\frac{(WR)(SSR)}{MW_1} + \frac{1}{MW_2} \right]^2} \quad (F-8)$$

Combining terms, Equation (F-8) becomes:

$$\frac{\partial x_1}{\partial SSR} = \frac{\frac{(WR)^2(SSR)}{(MW_1)^2} + \frac{WR}{(MW_1)(MW_2)} - \frac{(WR)^2(SSR)}{(MW_1)^2}}{\left[\frac{(WR)(SSR)}{MW_1} + \frac{1}{MW_2} \right]^2} \quad (F-9)$$

Reducing the equation results in the following expression:

$$\left(\frac{\partial x_1}{\partial SSR} \right)^2 = \left\{ \frac{\frac{WR}{(MW_1)(MW_2)}}{\left[\frac{(WR)(SSR)}{MW_1} + \frac{1}{MW_2} \right]^2} \right\}^2 \quad (F-10)$$

Division of Equation (F-10) by x_1^2 obtains the fractional uncertainty associated with the solvent-to-sample weight ratio, as follows:

$$\left(\frac{\partial x_1}{\partial SSR} \right)^2 \left(\frac{1}{x_1} \right)^2 = \left\{ \frac{1}{\left[\frac{(WR)(SSR)}{MW_1} + \frac{1}{MW_2} \right] (MW_2)(SSR)} \right\}^2 \quad (F-11)$$

but, from Equation (F-2),

$$\left[\frac{(WR)(SSR)}{MW_1} + \frac{1}{MW_2} \right] = \frac{(WR)(SSR)}{(MW_1)(x_1)} \quad (F-12)$$

Substitution of Equation (F-12) into Equation (F-11) results in the following:

$$\left(\frac{\partial x_1}{\partial SSR} \right)^2 \left(\frac{1}{x_1} \right)^2 = \left\{ \frac{1}{\left[\frac{(WR)(SSR)}{(MW_1)(x_1)} \right] (MW_2)(SSR)} \right\}^2 \quad (F-13)$$

The fractional uncertainty associated with the solvent-to-sample weight ratio then becomes:

$$\left(\frac{\partial x_1}{\partial SSR}\right)^2 \left(\frac{\sigma_{SSR}}{x_1}\right)^2 = \left\{ \frac{(MW_1)(x_1)}{(MW_2)(WR)(SSR)^2} \right\}^2 \sigma_{SSR}^2 \quad (\text{F-14})$$

The solvent-to-sample ratio, SSR, is calculated as follows:

$$SSR = \left(\frac{m_{sol}}{m_{sam}} \right) \quad (\text{F-15})$$

where m_{sol} is the average solvent mass and m_{sam} is the average sample mass. Variance in SSR is found by application of error propagation to Equation (F-15). The propagated uncertainty in SSR, as expressed in terms of variances, is a function of solvent mass and sample mass:

$$\sigma_{SSR}^2 = \left(\frac{\partial SSR}{\partial m_{sol}} \right)^2 \sigma_{m_{sol}}^2 + \left(\frac{\partial SSR}{\partial m_{sam}} \right)^2 \sigma_{m_{sam}}^2 \quad (\text{F-16})$$

The partial differentials from Equation (F-16) are:

$$\frac{\partial SSR}{\partial m_{sol}} = \left(\frac{1}{m_{sam}} \right) \quad (\text{F-17})$$

and

$$\frac{\partial SSR}{\partial m_{sam}} = \left(-\frac{m_{sol}}{m_{sam}^2} \right) \quad (\text{F-18})$$

Combination of Equations (F-16), (F-17), and (F-18) provides the variance associated with the solvent-to-sample weight ratio, as shown by:

$$\sigma_{SSR}^2 = \left(\frac{1}{m_{sam}} \right)^2 \sigma_{m_{sol}}^2 + \left(\frac{m_{sol}}{m_{sam}^2} \right)^2 \sigma_{m_{sam}}^2 \quad (\text{F-19})$$

The propagated fractional uncertainty associated with the solvent-to-sample weight ratio, which is the first term in Equation (F-7), is found by combining Equations (F-14) and (F-19), as follows:

$$\left(\frac{\partial x_1}{\partial SSR}\right)^2 \left(\frac{\sigma_{SSR}}{x_1}\right)^2 = \left\{ \frac{(MW_1)(x_1)}{(MW_2)(WR)(SSR)^2} \right\}^2 \left\{ \left(\frac{1}{m_{sam}}\right)^2 \sigma_{m_{sol}}^2 + \left(\frac{m_{sol}}{m_{sam}^2}\right)^2 \sigma_{m_{sam}}^2 \right\} \quad (F-20)$$

F.3 Uncertainty Associated with the Weight Ratio

Derivation of the uncertainty associated with WR begins with taking the partial derivative of x_1 , from Equation (F-2), with respect to WR, as given below:

$$\frac{\partial x_1}{\partial WR} = \frac{\left(\frac{SSR}{MW_1}\right) \left(\frac{(WR)(SSR)}{MW_1} + \frac{1}{MW_2}\right) - \left(\frac{(WR)(SSR)}{MW_1}\right) \left(\frac{SSR}{MW_1}\right)}{\left[\frac{(WR)(SSR)}{MW_1} + \frac{1}{MW_2}\right]^2} \quad (F-21)$$

Combining terms, Equation (F-21) becomes:

$$\frac{\partial x_1}{\partial WR} = \frac{\frac{(WR)(SSR)^2}{(MW_1)^2} + \frac{SSR}{(MW_1)(MW_2)} - \frac{(WR)(SSR)^2}{(MW_1)^2}}{\left[\frac{(WR)(SSR)}{MW_1} + \frac{1}{MW_2}\right]^2} \quad (F-22)$$

Reducing the equation results in the following expression::

$$\left(\frac{\partial x_1}{\partial WR}\right)^2 = \left\{ \frac{\frac{SSR}{(MW_1)(MW_2)}}{\left[\frac{(WR)(SSR)}{MW_1} + \frac{1}{MW_2}\right]^2} \right\}^2 \quad (F-23)$$

Division of Equation (F-23) by x_1^2 obtains the fractional uncertainty associated with the weight ratio, as shown by the following expression:

$$\left(\frac{\partial x_1}{\partial WR}\right)^2 \left(\frac{1}{x_1}\right)^2 = \left\{ \frac{1}{\left[\frac{(WR)(SSR)}{MW_1} + \frac{1}{MW_2} \right] (MW_2)(WR)} \right\}^2 \quad (\text{F-24})$$

but, from Equation (F-2),

$$\left[\frac{(WR)(SSR)}{MW_1} + \frac{1}{MW_2} \right] = \frac{(WR)(SSR)}{(MW_1)(x_1)} \quad (\text{F-25})$$

Substitution of Equation (F-25) into Equation (F-24) results in the following:

$$\left(\frac{\partial x_1}{\partial WR}\right)^2 \left(\frac{1}{x_1}\right)^2 = \left\{ \frac{1}{\left[\frac{(WR)(SSR)}{(MW_1)(x_1)} \right] (MW_2)(WR)} \right\}^2 \quad (\text{F-26})$$

The fractional uncertainty associated with the weight ratio then becomes:

$$\left(\frac{\partial x_1}{\partial WR}\right)^2 \left(\frac{\sigma_{WR}}{x_1}\right)^2 = \left\{ \frac{(MW_1)(x_1)}{(MW_2)(WR)^2 (SSR)} \right\}^2 \sigma_{WR}^2 \quad (\text{F-27})$$

Since the weight ratio, WR, is a function of the area ratio, AR, the variance in the weight ratio is a function of the variance in the gas chromatograph area ratio, as given by:

$$\sigma_{WR}^2 = \left(\frac{\partial WR}{\partial AR}\right)^2 \sigma_{AR}^2 \quad (\text{F-28})$$

The partial derivative of the weight ratio with respect to the area ratio can be estimated from the slope of the calibration curve, which is expressed as:

$$WR = \alpha AR^\beta \quad (\text{F-29})$$

where α and β are regressed calibration parameters. The slope of the calibration curve is determined as the following:

$$\frac{\partial WR}{\partial AR} = \alpha \beta AR^{(\beta-1)} \quad (\text{F-30})$$

The second term in Equation (F-7), the propagated fractional uncertainty associated with the weight ratio, is found by combination of Equations (F-27), (F-28), and (F-30), as follows:

$$\left(\frac{\partial x_1}{\partial WR} \right)^2 \left(\frac{\partial WR}{\partial AR} \right)^2 \left(\frac{\sigma_{AR}}{x_1} \right)^2 = \left\{ \frac{(MW_1)(x_1)}{(MW_2)(WR)^2(SSR)} \right\}^2 \{ \alpha \beta AR^{(\beta-1)} \}^2 \sigma_{AR}^2 \quad (\text{F-31})$$

The variance in the area ratio was determined from repeated measurements.

F.4 Uncertainty Associated with the Temperature

Two sources account for the uncertainty associated with the temperature; namely thermometer imprecision and thermal fluctuation in the oven. Deviation in the mole fraction with respect to temperature is determined by use of the slope of the solubility curve. For each phase, aqueous and organic, a generally accepted correlation is used for the solubility curve. The variance in the temperature was estimated to be $\pm 0.3\text{K}$.

Aqueous Phase

Using non-linear regression, the solubility data from the aqueous phase are fit to the following correlation for a hydrocarbon mole fraction:

$$\ln(x) = A + \frac{B}{T_r} + \frac{C}{T_r^2} \quad (\text{F-32})$$

where A, B, and C are regressed parameters and T_r , the reduced temperature, is found by,

$$T_r = \frac{T}{T_c} \quad (\text{F-33})$$

where T is the temperature (absolute) of the system and T_c is the critical temperature of the hydrocarbon of interest. The slope of the solubility curve is found by taking the partial derivative of x from Equation (F-32) with respect to T, as shown by:

$$\frac{\partial x_1}{\partial T} = \left[\frac{-BT_c}{T^2} - \frac{2CT_c^2}{T^3} \right] \exp \left[A + \frac{BT_c}{T} + \frac{CT_c^2}{T^2} \right] \quad (\text{F-34})$$

Division by x_1^2 expresses Equation (F-34) in terms of fractional uncertainty:

$$\left(\frac{\partial x_1}{\partial T} \right)^2 \left(\frac{1}{x_1} \right)^2 = \left\{ \left[\frac{-BT_c}{T^2} - \frac{2CT_c^2}{T^3} \right] \exp \left[A + \frac{BT_c}{T} + \frac{CT_c^2}{T^2} \right] \right\}^2 \left(\frac{1}{x_1} \right)^2 \quad (\text{F-35})$$

The final term in variance form, which is the fractional uncertainty associated with the temperature, for Equation (F-7) is written as:

$$\left(\frac{\partial x_1}{\partial T} \right)^2 \left(\frac{\sigma_T}{x_1} \right)^2 = \left\{ \left[\frac{-BT_c}{T^2} - \frac{2CT_c^2}{T^3} \right] \exp \left[A + \frac{BT_c}{T} + \frac{CT_c^2}{T^2} \right] \right\}^2 \left(\frac{\sigma_T}{x_1} \right)^2 \quad (\text{F-36})$$

Hydrocarbon Phase

Using non-linear regression, the solubility data from the hydrocarbon phase are fit to the following correlation for a hydrocarbon mole fraction:

$$\ln(x) = A + B \ln T_r \quad (\text{F-37})$$

where A and B are regressed parameters and T_r, the reduced temperature, is found by,

$$T_r = \frac{T}{T_c} \quad (\text{F-38})$$

T is the temperature (absolute) of the system and T_c is the critical temperature of the hydrocarbon of interest. The slope of the solubility curve is found by taking the partial derivative of x from Equation (F-37) with respect to T, as shown by:

$$\frac{\partial x_1}{\partial T} = \frac{B}{T} \exp \left[A + B \ln \left(\frac{T}{T_c} \right) \right] \quad (\text{F-39})$$

Division by x_1^2 expresses Equation (F-39) in terms of fractional uncertainty:

$$\left(\frac{\partial x_1}{\partial T} \right) \left(\frac{1}{x_1} \right) = \left\{ \frac{B}{T} \exp \left[A + B \ln \left(\frac{T}{T_c} \right) \right] \right\}^2 \left(\frac{1}{x_1} \right)^2 \quad (\text{F-40})$$

The final term in variance form, which is the fractional uncertainty associated with the temperature, for Equation (F-7) is written as:

$$\left(\frac{\partial x_1}{\partial T} \right) \left(\frac{\sigma_T}{x_1} \right) = \left\{ \frac{B}{T} \exp \left[A + B \ln \left(\frac{T}{T_c} \right) \right] \right\}^2 \left(\frac{\sigma_T}{x_1} \right)^2 \quad (\text{F-41})$$

F.5 Total Fractional Uncertainty

The total fractional uncertainty is found by substitution of Equations (F-20), (F-31), and either (F-36) or (F-41), when dealing with the aqueous or organic phase, respectively, into Equation (F-7).

Total uncertainty of the aqueous phase is calculated using the following equation:

$$\begin{aligned} \left(\frac{\sigma_{x_1}}{x_1} \right)^2 = & \left\{ \frac{(MW_1)(x_1)}{(MW_2)(WR)(SSR)^2} \right\}^2 \left\{ \left(\frac{1}{m_{sam}} \right)^2 \sigma_{m_{sol}}^2 + \left(\frac{m_{sol}}{m_{sam}^2} \right)^2 \sigma_{m_{sam}}^2 \right\} \\ & + \left\{ \frac{(MW_1)(x_1)}{(MW_2)(WR)^2(SSR)} \right\}^2 \{ \alpha \beta AR^{(\beta-1)} \}^2 \sigma_{AR}^2 \\ & + \left\{ \left[\frac{-BT_c}{T^2} - \frac{2CT_c^2}{T^3} \right] \exp \left[A + \frac{BT_c}{T} + \frac{CT_c^2}{T^2} \right] \right\}^2 \left(\frac{\sigma_T}{x_1} \right)^2 \end{aligned} \quad (F-42)$$

while the total uncertainty of the organic phase is calculated by:

$$\begin{aligned} \left(\frac{\sigma_{x_1}}{x_1} \right)^2 = & \left\{ \frac{(MW_1)(x_1)}{(MW_2)(WR)(SSR)^2} \right\}^2 \left\{ \left(\frac{1}{m_{sam}} \right)^2 \sigma_{m_{sol}}^2 + \left(\frac{m_{sol}}{m_{sam}^2} \right)^2 \sigma_{m_{sam}}^2 \right\} \\ & + \left\{ \frac{(MW_1)(x_1)}{(MW_2)(WR)^2(SSR)} \right\}^2 \{ \alpha \beta AR^{(\beta-1)} \}^2 \sigma_{AR}^2 \\ & + \left\{ \frac{B}{T} \exp \left[A + B \ln \left(\frac{T}{T_c} \right) \right] \right\}^2 \left(\frac{\sigma_T}{x_1} \right)^2 \end{aligned} \quad (F-43)$$

The first term represents the uncertainty associated with the solvent-to-sample weight ratio, the second term represents the uncertainty associated with the solute-to-solvent weight ratio, and the final term represents the uncertainty associated with the temperature.

Repeated measurements provided estimates of uncertainty required for the calculation of the total uncertainty associated with the measured mole fractions. The uncertainty in the solvent mass, $\sigma_{m_{sol}}$, and sample mass, $\sigma_{m_{sam}}$, is taken as 0.0002, which is the standard deviation of twenty measurements of an empty, capped vial. The

uncertainty in the temperature, σ_T , is 0.3 K, which is the standard deviation of ten ice point measurements of distilled water. The uncertainty in the area ratio, σ_{AR} , is the standard deviation in the GC analyses. With the exception of σ_{AR} , all uncertainties were considered to remain constant through the course of the study. The uncertainties in the area ratio are presented at the end of the appendix in Table F-1.

F.6 Sample Calculations Toluene - Water

Toluene Solubility in Water

Expressing uncertainty estimates as standard deviations, a sample calculation for the solubility of toluene in water at 324.0 K is provided, as follows:

$$\begin{aligned} \left(\frac{\sigma_{x_1}}{x_1}\right)^2 = & \left\{ \frac{(MW_1)(x_1)}{(MW_2)(WR)(SSR^2)} \right\}^2 \left\{ \left(\frac{1}{m_{sam}}\right)^2 \sigma_{m_{sol}}^2 + \left(\frac{m_{sol}}{m_{sam}^2}\right)^2 \sigma_{m_{sam}}^2 \right\} \\ & + \left\{ \frac{(MW_1)(x_1)}{(MW_2)(WR)^2(SSR)} \right\}^2 \left\{ \alpha \beta AR^{(\beta-1)} \right\}^2 \sigma_{AR}^2 \\ & + \left\{ \left[\frac{-BT_c}{T^2} - \frac{2CT_c^2}{T^3} \right] \exp \left[A + \frac{BT_c}{T} + \frac{CT_c^2}{T^2} \right] \right\}^2 \left(\frac{\sigma_T}{x_1}\right)^2 \end{aligned}$$

$$\sigma_{m_{sol}} = 0.00021 \text{ grams} \quad \sigma_{m_{sam}} = 0.00021 \text{ grams}$$

$$\sigma_{AR} = 1.361E - 04 \quad \sigma_T = 0.3 \text{ K}$$

$$\alpha = 0.2214 \quad \beta = 1.020$$

$$m_{sol} = 3.4535 \text{ grams} \quad m_{sam} = 22.5932 \text{ grams}$$

$$MW_1 = 92.834 \text{ grams/mole} \quad MW_2 = 18.015 \text{ grams/mole}$$

$$x_1 = 1.307E - 04 \quad T = 324.0 \text{ K} \quad T_c = 591.8 \text{ K}$$

$$WR = 0.00441 \quad SSR = 0.1529 \quad AR = 0.02147$$

$$A = 10.71 \quad B = -19.81 \quad C = 4.951$$

According to Equation (F-42), the total fractional propagated uncertainty in the solubility measurement is calculated as:

$$\begin{aligned}
\left(\frac{\sigma_{x_1}}{x_1}\right)^2 &= \left\{ \frac{(92.834)(1.307E-04)}{(18.015)(0.00441)(0.1529)^2} \right\}^2 \left\{ \left(\frac{1}{22.59}\right)^2 (0.0002)^2 + \left(\frac{3.4535}{(22.59)^2}\right)^2 (0.0002)^2 \right\} \\
&+ \left\{ \frac{(92.834)(1.307E-04)}{(18.015)(0.00441)^2(0.1529)} \right\}^2 \left\{ (0.2214)(1.020)(0.02147)^{(1.020-1)} \right\}^2 (1.361E-04)^2 \\
&+ \left\{ \left[\frac{(19.81)(591.8)}{(324.0)^2} - \frac{(2)(4.95)(591.8)^2}{(324.0)^3} \right] \exp \left[10.71 + \frac{(-19.81)(591.8)}{324.0} + \frac{(4.95)(591.8)^2}{(324.0)^2} \right] \right\}^2 \\
&\quad \times \left(\frac{0.3}{1.307E-04} \right)^2 \\
&= (3.773E-09) + (4.156E-05) + (8.308E-06) \\
&= 4.987E-05
\end{aligned}$$

or,

$$\begin{aligned}
\sigma_{x_1} &= (0.00706)(x_1) \\
&= (0.00706)(1.307E-04) \\
&= 9.230E-07
\end{aligned}$$

Division of the uncertainty by the mole fraction of toluene determines the percentage of uncertainty in the mole fraction of toluene at 324.0 K, as follows:

$$\% \text{ uncertainty} = \left(\frac{9.230E-07}{1.307E-04} \right) (100\%) = 0.71\%$$

Division of each individual fractional uncertainty by the total fractional uncertainty, as shown below, determines the percentage of uncertainty associated with each term:

$$\% \text{ uncertainty associated with SSR} = \left(\frac{3.773E-09}{4.987E-05} \right) (100\%) < 1\%$$

$$\% \text{ uncertainty associated with W} = \left(\frac{4.156E-05}{4.987E-05} \right) (100\%) = 83\%$$

$$\% \text{ uncertainty associated with } T = \left(\frac{8.308E - 06}{4.987E - 05} \right) (100\%) = 17\%$$

The uncertainty associated with the solvent-to-sample weight ratio is negligible. For the hydrocarbon solubility in water measurements, the uncertainty associated with the weight ratio is typically 90% of the total uncertainty, while the uncertainty associated with the temperature accounts for the balance of the total uncertainty.

Water Solubility in Toluene

Expressing uncertainty estimates as standard deviations, a sample calculation for the solubility of water in toluene at 324.3 K is provided, as follows:

$$\begin{aligned} \left(\frac{\sigma_{x_1}}{x_1}\right)^2 = & \left\{ \frac{(MW_1)(x_1)}{(MW_2)(WR)(SSR)^2} \right\}^2 \left\{ \left(\frac{1}{m_{sam}}\right)^2 \sigma_{m_{sol}}^2 + \left(\frac{m_{sol}}{m_{sam}^2}\right)^2 \sigma_{m_{sam}}^2 \right\} \\ & + \left\{ \frac{(MW_1)(x_1)}{(MW_2)(WR)^2(SSR)} \right\}^2 \left\{ \alpha \beta AR^{(\beta-1)} \right\}^2 \sigma_{AR}^2 \\ & + \left\{ \frac{B}{T} \exp \left[A + B \ln \left(\frac{T}{T_c} \right) \right] \right\}^2 \left(\frac{\sigma_T}{x_1} \right)^2 \end{aligned}$$

$$\sigma_{m_{sol}} = 0.00021 \text{ grams} \quad \sigma_{m_{sam}} = 0.00021 \text{ grams}$$

$$\sigma_{AR} = 2.427E-04 \quad \sigma_T = 0.3 \text{ K}$$

$$\alpha = 1.572 \quad \beta = 1.227$$

$$A = 1.122 \quad B = 9.035$$

$$m_{sol} = 6.1963 \text{ grams} \quad m_{sam} = 13.6404 \text{ grams}$$

$$MW_1 = 18.015 \text{ grams/mole} \quad MW_2 = 92.834 \text{ grams/mole}$$

$$x_1 = 6.747E-03 \quad T = 324.3 \text{ K} \quad T_c = 647.1 \text{ K}$$

$$WR = 0.00290 \quad SSR = 0.4543 \quad AR = 0.00742$$

According to Equation (F-43), the total fractional propagated uncertainty in the solubility measurement is calculated as:

$$\begin{aligned}
\left(\frac{\sigma_{x_1}}{x_1}\right)^2 &= \left\{ \frac{(18.015)(6.747E-03)}{(92.834)(0.0029)(0.4543)^2} \right\}^2 \left\{ \left(\frac{1}{13.64}\right)^2 (0.0002)^2 + \left(\frac{6.1963}{(13.64)^2}\right)^2 (0.0002)^2 \right\} \\
&+ \left\{ \frac{(18.015)(6.747E-03)}{(92.834)(0.0029)^2(0.4543)} \right\}^2 \left\{ (1.572)(1.227)(0.00742)^{(1.227-1)} \right\}^2 (2.427E-04)^2 \\
&+ \left\{ \left(\frac{9.035}{324.3}\right) \exp\left[1.122 + (9.035)\ln\left(\frac{324.3}{647.1}\right)\right] \right\}^2 \left(\frac{0.3}{6.747E-03}\right)^2 \\
&= (1.241E-09) + (2.558E-03) + (5.484E-05) \\
&= 2.613E-03
\end{aligned}$$

or,

$$\begin{aligned}
\sigma_{x_1} &= (0.0511)(x_1) \\
&= (0.0511)(6.747E-03) \\
&= 3.448E-04
\end{aligned}$$

Division of the uncertainty by the mole fraction of toluene determines the percentage of uncertainty in the mole fraction of toluene at 324.3 K, as follows:

$$\% \text{ uncertainty} = \left(\frac{3.448E-04}{6.747E-03}\right)(100\%) = 5.1\%$$

Division of each individual fractional uncertainty by the total fractional uncertainty, as shown below, determines the percentage of uncertainty associated with each term:

$$\% \text{ uncertainty associated with SSR} = \left(\frac{1.241E-09}{2.612E-03}\right)(100\%) < 1\%$$

$$\% \text{ uncertainty associated with W} = \left(\frac{2.558E-03}{2.612E-03}\right)(100\%) = 98\%$$

$$\% \text{ uncertainty associated with T} = \left(\frac{5.403E-05}{2.612E-03}\right)(100\%) = 2\%$$

The uncertainty associated with the solvent-to-sample weight ratio is negligible. For the water solubility in hydrocarbon measurements, the uncertainty associated with the weight ratio is typically 97% of the total uncertainty, while the uncertainty associated with the temperature accounts for the balance of the total uncertainty.

Table F-1. Uncertainty Estimates for the Area Ratio

Benzene-Water System			
Benzene Solubility		Water Solubility	
Temperature (K)	Area Ratio Uncertainty (10^5)	Temperature (K)	Area Ratio Uncertainty (10^5)
299.1	11.73	299.1	7.354
324.3	3.859	324.3	15.46
350.2	8.550	350.2	33.78
376.2	13.72	376.2	42.73
400.3	6.913	400.3	53.74
431.4	109.2	431.4	176.1
461.8	244.2	461.8	315.0
490.8	101.4	490.8	377.4


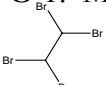



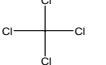
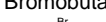


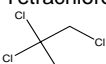
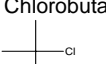
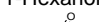

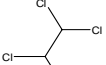

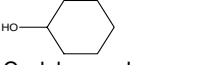

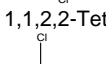
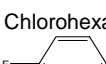


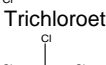
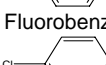
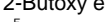
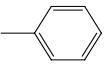
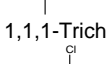
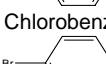
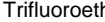
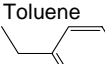
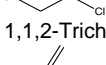
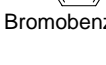

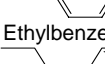
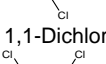
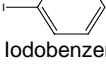
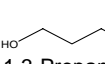
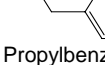
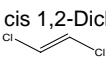
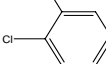
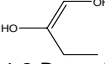
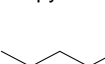
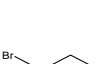
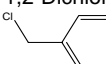
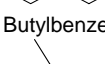
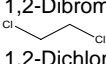
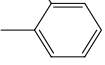
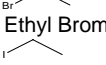

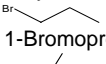
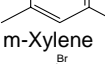
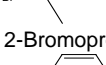
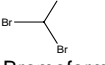
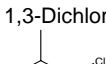
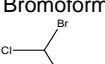
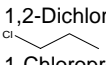
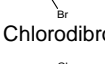
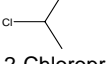
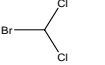

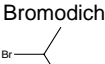

Toluene-Water System			
Toluene Solubility		Water Solubility	
Temperature (K)	Area Ratio Uncertainty (10^5)	Temperature (K)	Area Ratio Uncertainty (10^5)
297.8	60.58	298.5	8.284
324.0	13.51	324.3	16.99
350.6	18.04	350.9	40.68
376.1	15.69	376.6	78.20
401.6	111.7	401.5	128.0
431.8	334.5	431.7	142.5
460.8	84.68	461.8	205.8
490.4	771.1	491.4	198.3

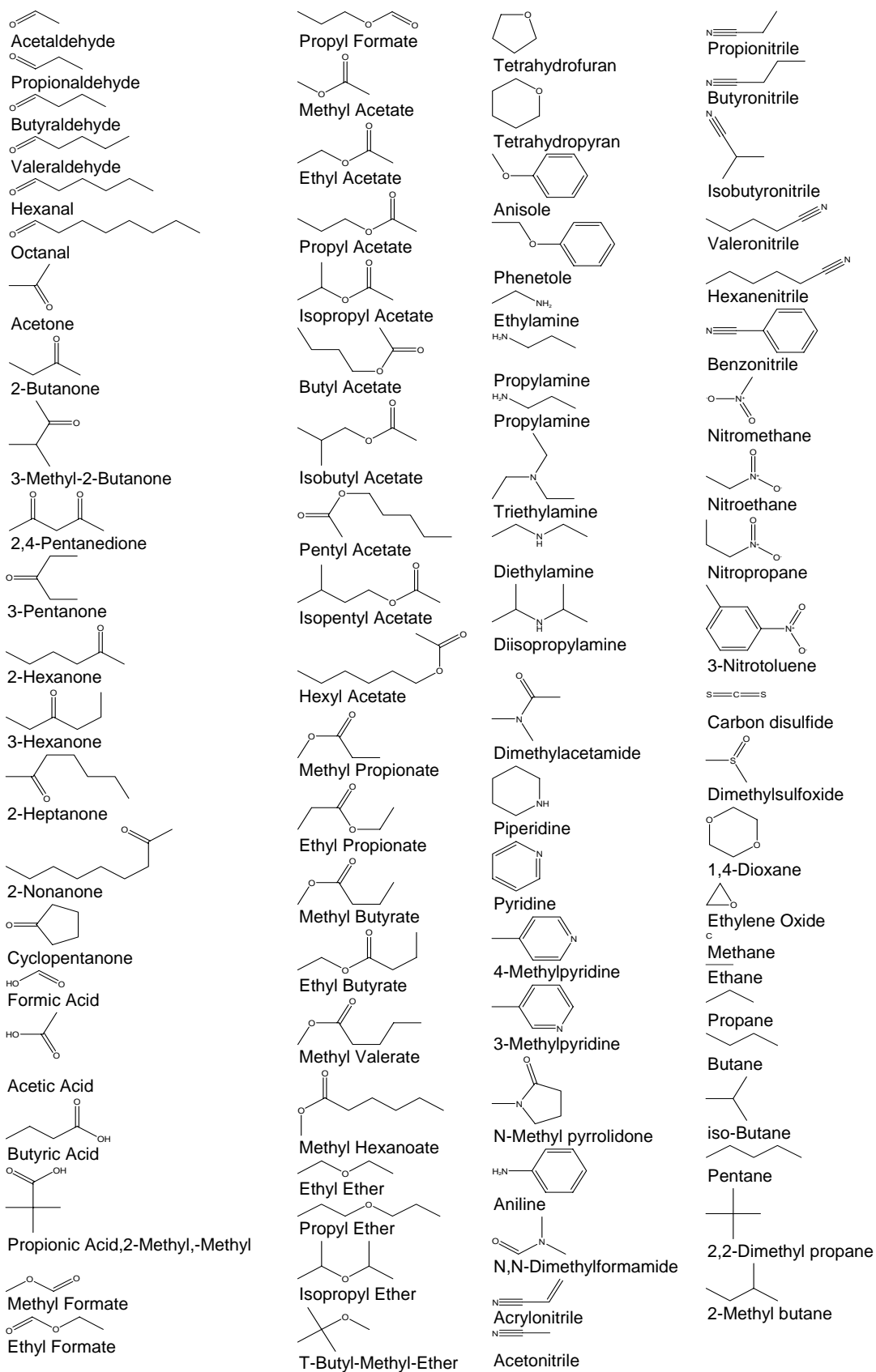
3-Methylpentane-Water System			
3-Methylpentane Solubility		Water Solubility	
Temperature (K)	Area Ratio Uncertainty (10^5)	Temperature (K)	Area Ratio Uncertainty (10^5)
298.3	0.5510	299.1	12.86
324.5	0.9267	324.3	26.93
351.2	1.414	350.9	30.52
377.2	0.7014	376.4	137.5
401.2	3.011	400.3	39.36
432.2	7.495	432.4	173.4
462.1	10.89	491.5	1157
491.9	26.72	295.5	11.44

Appendix G. Database Information

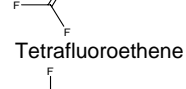
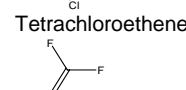
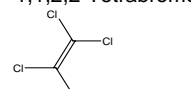
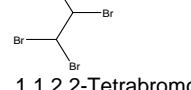
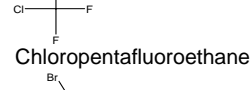
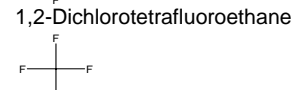
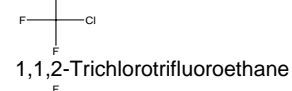
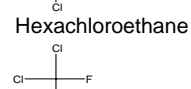
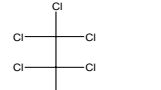
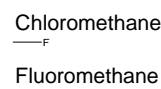
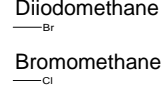
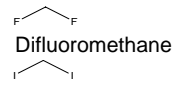
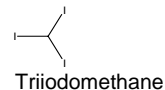
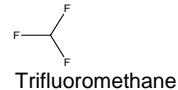
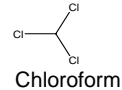
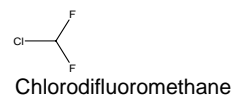
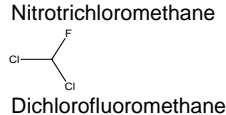
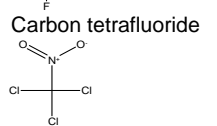
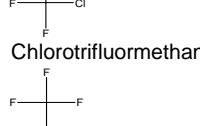
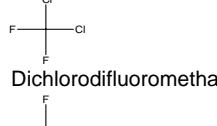
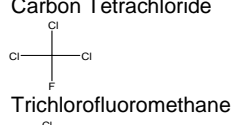
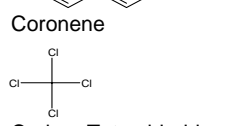
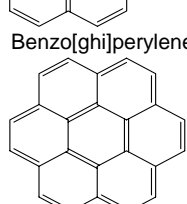
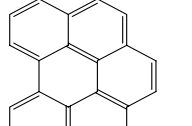
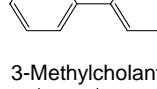
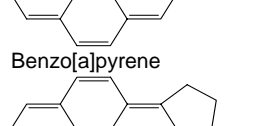
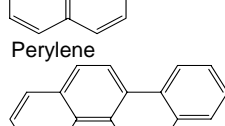
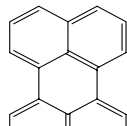
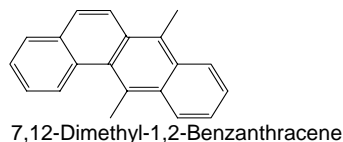
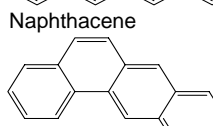
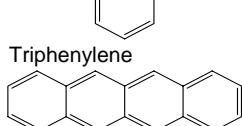
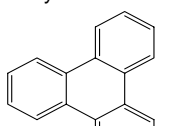
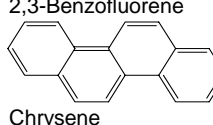
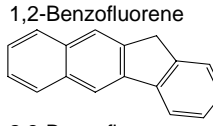
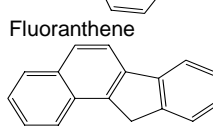
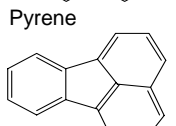
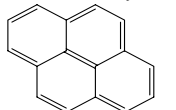
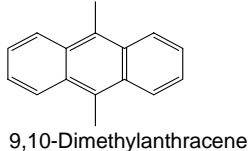
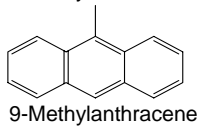
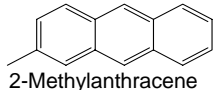
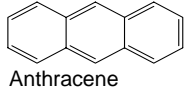
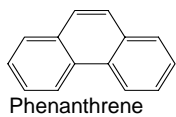
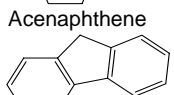
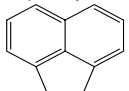
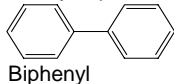
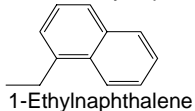
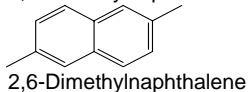
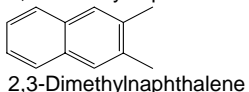
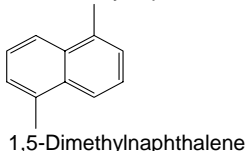
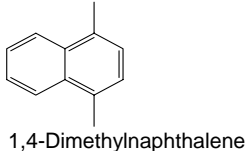
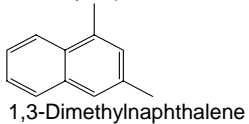
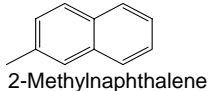
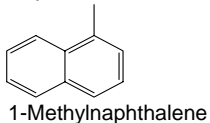
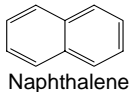
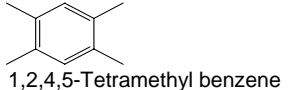
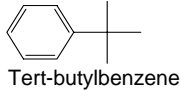
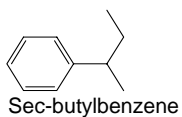
This appendix contains illustrations of the molecular structures and the infinite-dilution activity coefficient values for the molecules found in the database. The structure graphics generally follow the same order as the tabulated values. For more detailed information concerning the source of the experimental values please refer to the original publication of the database, “Measuring methods of infinite-dilution activity coefficients and a database for systems including water,” by Kojima, Zhang, and Hiaki in Fluid Phase Equilibria (1997, volume 131, pages 145-179).

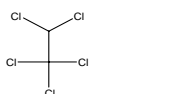
Table G-1. Molecular Structures Found in the Database

			
Pentane	1,1,2,2-Tetrabromoethane	Bromobutane	3-Methyl-Butanol
			
Hexane	Tetrachloromethane	Chlorobutane	2-Hexanone
			
Heptane	1,1,1,2-Tetrachloroethane	Chloropentane	1-Heptanol
			
Cyclohexane	1,1,2,2-Tetrachloroethane	Fluorobenzene	Trifluoroethanol
			
Cyclohexene	Trichloroethene	Bromobenzene	1,2-Propanediol
			
Benzene	1,1,1-Trichloroethane	1,2-Dichlorobenzene	1,2-Butandiol
			
Toluene	1,1,2-Trichloroethane	Methanol	1,4-Butanediol
			
Ethylbenzene	1,1-Dichloroethene	2-Propen-1-ol	Phenol
			
Propylbenzene	cis 1,2-Dichloroethene	1-Butanol	m-Cresol
			
Butylbenzene	trans 1,2-Dichloroethene	Iso-Butyl alcohol	Formaldehyde
			
o-Xylene	1,2-Dibromoethane	1-Pentanol	
			
m-Xylene	1,2-Dichloroethane		
			
Bromoform	Ethyl Bromide		
			
Chlorodibromomethane	Ethyl Iodide		
			
Bromodichloromethane	1-Bromopropane		
			
Dibromomethane	2-Bromopropane		
			
Carbon Tetrachloride	1,3-Dichloropropylene		
			
Chloroform	1,2-Dichloropropane		
			
Dichloromethane	1-Chloropropane		
			
Methyl iodide	2-Chloropropane		

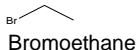








Pentachloroethane



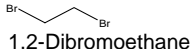
Bromoethane



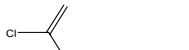
Chloroethene



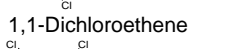
1-Bromo-2-Chloroethane



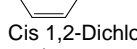
1,2-Dibromoethane



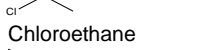
1,1-Dichloroethene



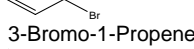
Cis 1,2-Dichloroethene



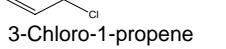
Chloroethane



3-Bromo-1-Propene



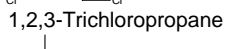
3-Chloro-1-propene



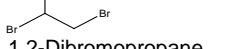
1,2,3-Trichloropropane



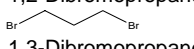
1,2-Dibromopropane



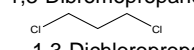
1,3-Dibromopropane



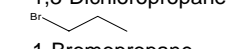
1,3-Dichloropropane



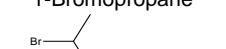
1-Bromopropane



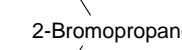
2-Bromopropane



2-Iodopropane



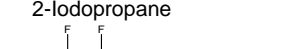
Octafluorocyclobutane



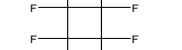
2-Chlorobutane



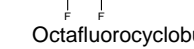
1-Chloro-2-Methylpropane



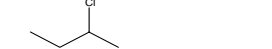
1-Bromo-2-Methylpropane



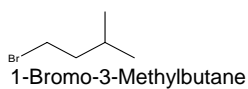
1-Bromobutane



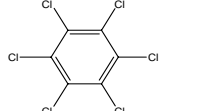
1-Bromopentane



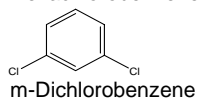
2-Chloro-2-Methylbutane



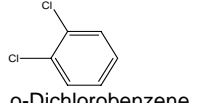
1-Bromo-3-Methylbutane



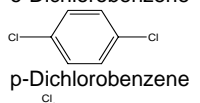
Hexachlorobenzene



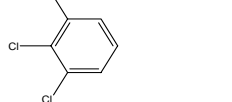
m-Dichlorobenzene



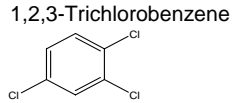
o-Dichlorobenzene



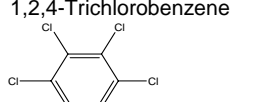
p-Dichlorobenzene



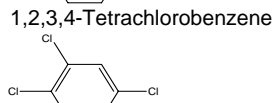
1,2,3-Trichlorobenzene



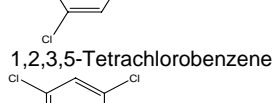
1,2,4-Trichlorobenzene



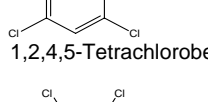
1,2,3,4-Tetrachlorobenzene



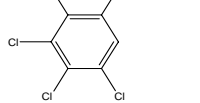
1,2,3,5-Tetrachlorobenzene



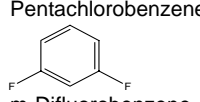
1,2,4,5-Tetrachlorobenzene



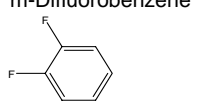
Pentachlorobenzene



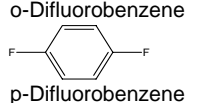
m-Difluorobenzene



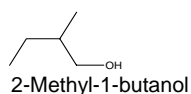
o-Difluorobenzene



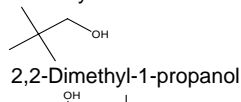
p-Difluorobenzene



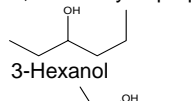
1-Pentanol



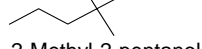
2-Methyl-1-butanol



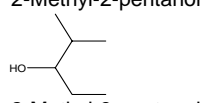
2,2-Dimethyl-1-propanol



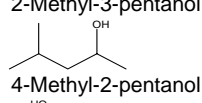
3-Hexanol



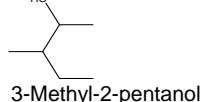
2-Methyl-2-pentanol



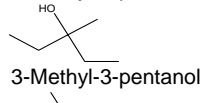
2-Methyl-3-pentanol



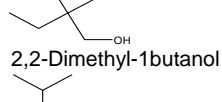
4-Methyl-2-pentanol



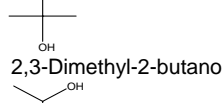
3-Methyl-2-pentanol



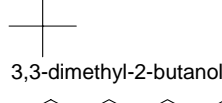
3-Methyl-3-pentanol



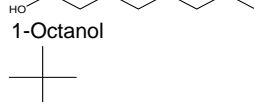
2,2-Dimethyl-1butanol



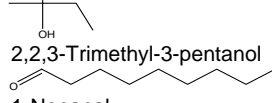
2,3-Dimethyl-2-butanol



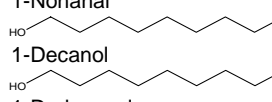
3,3-dimethyl-2-butanol



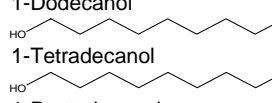
1-Octanol



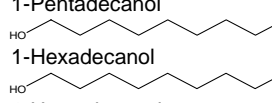
2,2,3-Trimethyl-3-pentanol



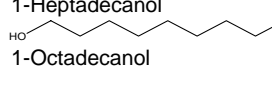
1-Nonanal



1-Decanol



1-Dodecanol



1-Tetradecanol



1-Pentadecanol

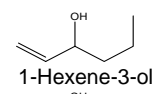


1-Hexadecanol

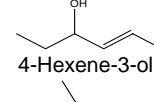


1-Heptadecanol

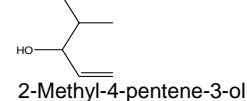
1-Octadecanol



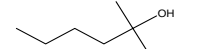
1-Hexene-3-ol



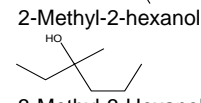
4-Hexene-3-ol



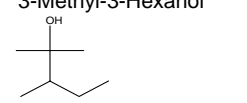
2-Methyl-4-pentene-3-ol



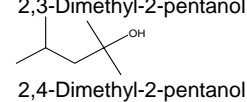
2-Methyl-2-hexanol



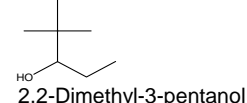
3-Methyl-3-Hexanol



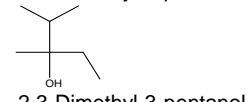
2,3-Dimethyl-2-pentanol



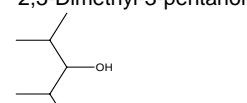
2,4-Dimethyl-2-pentanol



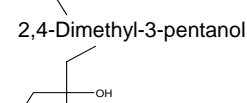
2,2-Dimethyl-3-pentanol



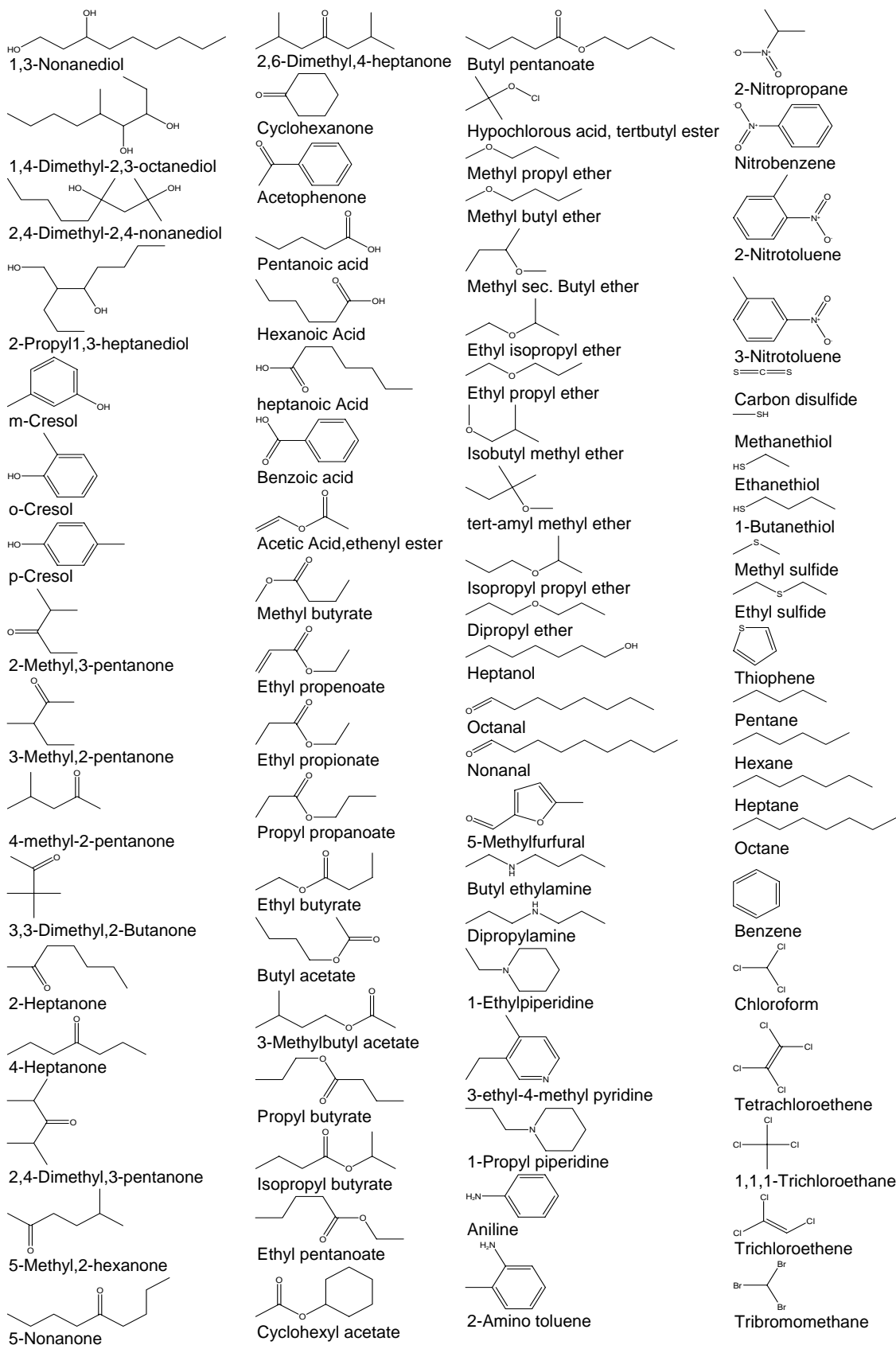
2,3-Dimethyl-3-pentanol

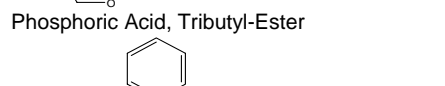
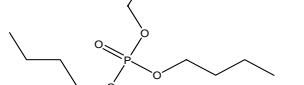
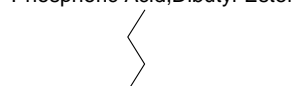
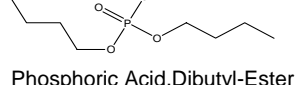
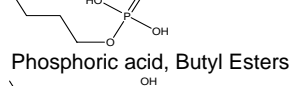
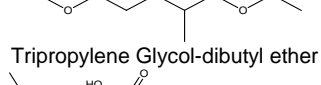
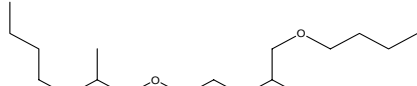
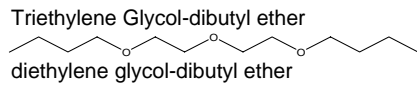
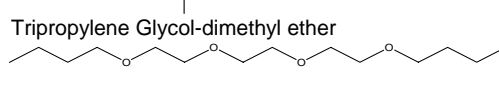
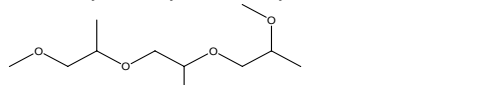
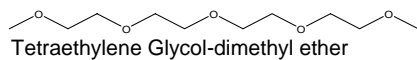
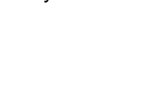
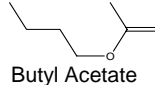
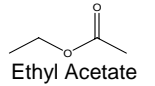
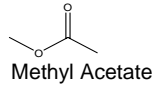
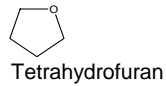
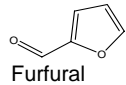
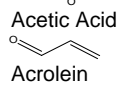
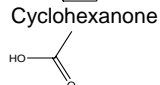
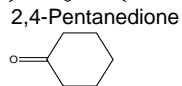
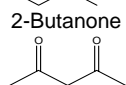
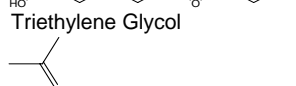
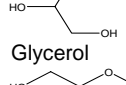
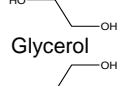
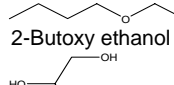
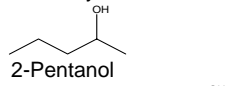
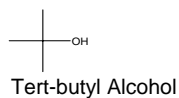
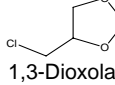
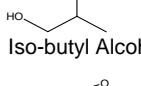
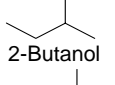
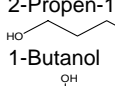
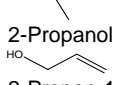
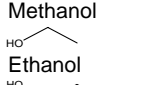
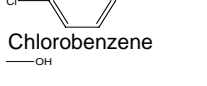
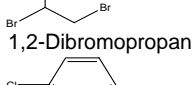
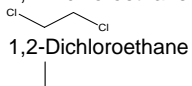
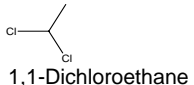
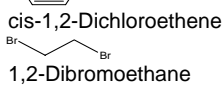
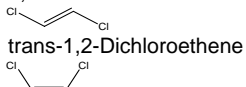
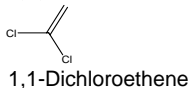
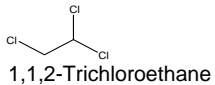
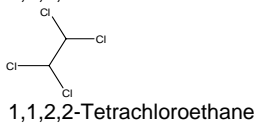
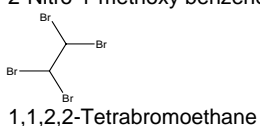
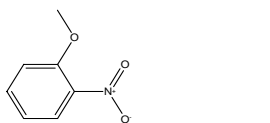


2,4-Dimethyl-3-pentanol



3-Ethyl-3-Pentanol







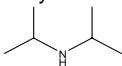
Nitromethane



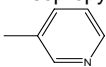
Acetonitrile



Acrylonitrile



Diisopropylamine



3-Methylpyridine



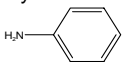
Dimethylformamide



Piperidine



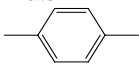
Pyridine



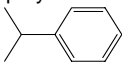
Aniline



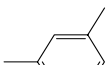
Water



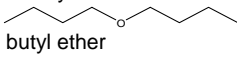
p-xylene



cumene



mesitylene



butyl ether

Table G-2. Organics in Water (Direct Measurements)

Solute	T/K	Exp γ	In γ	Error γ	Solute	T/K	Exp γ	In γ	Error γ
Alkanes									
Pentane	293.15	945	6.85		Toluene	293.15	4500	8.41	
	303.15	796	6.68			296.15	10400	9.25	
	313.15	517	6.25			298.15	9186	9.13	
Hexane	293.15	2940	7.99		298.15	9490	9.16		
	298.15	40000	10.60		298.15	9170	9.12		
	303.15	2225	7.71		313.15	3249	8.09	38	
	313.15	1465	7.29		313.15	3211	8.07		
	333.15	18500	9.83		313.15	3287	8.10		
Heptane	373.15	9500	9.16		Ethylbenzene	298.15	32670	10.39	
	293.15	8050	8.99		298.15	35500	10.48		
	303.15	6160	8.73		298.15	33400	10.42		
	313.15	3925	8.28		Propylbenzene	298.15	135800	11.82	
Cyclohexane	313.15	1026.5	6.93	30	298.15	127000	11.75		
	313.15	996.5	6.90		298.15	133000	11.80		
	313.15	1056.5	6.96		Butylbenzene	298.15	566260	13.25	
Alkenes									
Cyclohexene	296.15	20400	9.92		298.15	502000	13.13		
					298.15	533000	13.19		
Aromatic Hydrocarbons					o-Xylene	298.15	30540	10.33	
Benzene	293.15	2500	7.82		298.15	33300	10.41		
	293.15	2505	7.83		298.15	32500	10.39		
	293.15	2510	7.83		313.15	5859	8.68	84	
	296.15	2350	7.76		313.15	5775	8.66		
	296.15	2360	7.77	52	313.15	5943	8.69		
	296.15	2308	7.74		m-Xylene	298.15	33214	10.41	
	296.15	2412	7.79		298.15	37400	10.53		
	298.15	2495	7.82		298.15	39000	10.57		
	298.15	2289	7.74		p-xylene	298.15	33257	10.41	
	298.15	2400	7.78		298.15	34100	10.44		
	298.15	2320	7.75		298.15	37900	10.54		
	298.15	2200	7.70		313.15	5634	8.64	68	
	298.15	1700	7.44		313.15	5566	8.62		
298.15	2475	7.81		313.15	5702	8.65			
298.15	2420	7.79		Cumene	298.15	101540	11.53		
298.15	2530	7.84		298.15	102000	11.53			
303.15	2402	7.78		298.15	99100	11.50			
303.15	2422	7.79		mesitylene	298.15	118810	11.69		
303.15	2570	7.85		298.15	130000	11.78			
313.05	2535	7.84		298.15	128000	11.76			
313.15	1635	7.40	49	Halogenated Hydrocarbons					
313.15	1586	7.37		Bromoform	293.15	3530	8.17	260	
313.15	1684	7.43		293.15	3270	8.09			
323.15	2465	7.81		293.15	3790	8.24			
333.15	1800	7.50		293.15	4340	8.38	510		
373.15	1080	6.98		293.15	3830	8.25			

Solute	T/K	Exp γ	In γ	Error γ	Solute	T/K	Exp γ	In γ	Error γ
Bromoform	293.15	4850	8.49		Carbon Tetrachloride	308.15	787	6.67	87
	298.15	3380	8.13			308.15	700	6.55	
	298.15	9220	9.13			308.15	874	6.77	
	303.15	4068	8.31	480		308.15	801	6.69	35
	303.15	3588	8.19			308.15	766	6.64	
	303.15	4548	8.42			308.15	836	6.73	
	308.15	3080	8.03	425		313.15	768	6.64	76
	308.15	2655	7.88			313.15	692	6.54	
	308.15	3505	8.16			313.15	844	6.74	
	308.15	5820	8.67			323.15	740	6.61	32
	313.15	3800	8.24	440		323.15	708	6.56	
	313.15	3360	8.12			323.15	772	6.65	
	313.15	4240	8.35			285.65	7100	8.87	
	323.15	4050	8.31	220		293.15	2870	7.96	
	323.15	3830	8.25			293.15	6300	8.75	
323.15	4270	8.36		293.15	4500	8.41			
323.15	3330	8.11		293.15	12200	9.41	250		
Chlorodibromomethane	293.15	1990	7.60	250	293.15	11950	9.39		
	293.15	1740	7.46		293.15	12450	9.43		
	293.15	2240	7.71		293.15	9190	9.13	930	
	303.15	1928	7.56	240	293.15	8260	9.02		
	303.15	1688	7.43		293.15	10120	9.22		
	303.15	2168	7.68		298.15	15400	9.64		
	313.15	1896	7.55	240	298.15	10300	9.24		
	313.15	1656	7.41		298.15	3450	8.15		
	313.15	2136	7.67		303.15	13100	9.48	100	
	313.15	2136	7.67		303.15	13000	9.47		
Bromodichloromethane	293.15	1025	6.93	130	303.15	13200	9.49		
	293.15	895	6.80		303.15	9799	9.19	1000	
	293.15	1155	7.05		303.15	8799	9.08		
	303.15	1050	6.96	130	303.15	10799	9.29		
	303.15	920	6.82		303.15	2150	7.67		
	303.15	1180	7.07		308.15	8999	9.10	1000	
	313.15	1070	6.98	130	308.15	7999	8.99		
	313.15	940	6.85		308.15	9999	9.21		
	313.15	1200	7.09		313.15	8706	9.07	860	
Dibromomethane	293.15	869	6.77	45	313.15	7846	8.97		
	293.15	824	6.71		313.15	9566	9.17		
	293.15	914	6.82		313.15	1399.4	7.24	45	
	293.15	846	6.74	85	313.15	1354.4	7.21		
	293.15	761	6.63		313.15	1444.4	7.28		
	293.15	931	6.84		313.15	1490	7.31		
	298.15	850	6.75		313.15	13100	9.48	300	
	303.15	803	6.69	81	313.15	12800	9.46		
	303.15	722	6.58		313.15	13400	9.50		
	303.15	884	6.78						

Solute	T/K	Exp γ	In γ	Error γ	Solute	T/K	Exp γ	In γ	Error γ	
Chloroform	293.15	571	6.35			303.15	201	5.30		
	293.15	821	6.71			308.15	242	5.49	27	
	293.15	1000	6.91			308.15	215	5.37		
	293.15	818	6.71	20		308.15	269	5.59		
	293.15	798	6.68			313.15	242	5.49	24	
	293.15	838	6.73			313.15	218	5.38		
	298.15	903	6.81			313.15	266	5.58		
	298.15	857	6.75			313.15	226	5.42	7	
	298.15	623	6.43			313.15	219	5.39		
	303.15	835	6.73	21		313.15	233	5.45		
	303.15	814	6.70			313.15	153	5.03		
	303.15	856	6.75			Methyl iodide	293.15	870	6.77	
	303.15	568	6.34			1,1,2,2-Tetrabromoethane	298.15	625000	13.35	
	308.15	847	6.74	30			308.15	337000	12.73	
	308.15	817	6.71				323.15	130000	11.78	
308.15	877	6.78			Tetrachloromethane	298.15	35970	10.49		
308.15	892	6.79				298.15	65500	11.09		
313.15	742	6.61	21		1,1,1,2-Tetrachloroethane	293.15	9280	9.14	730	
313.15	721	6.58				293.15	8550	9.05		
313.15	763	6.64				293.15	10010	9.21		
313.15	425	6.05				293.15	7730	8.95	760	
323.15	862	6.76	27			293.15	6970	8.85		
323.15	835	6.73				293.15	8490	9.05		
323.15	889	6.79				298.15	8910	9.09		
323.15	910	6.81				303.15	7282	8.89	720	
Dichloromethane	283.15	235	5.46	5		303.15	6562	8.79		
	283.15	230	5.44			303.15	8002	8.99		
	283.15	240	5.48			303.15	8530	9.05	635	
	293.15	370	5.91			303.15	7895	8.97		
	293.15	209	5.34			303.15	9165	9.12		
	293.15	251	5.53	14		308.15	7216	8.88	790	
	293.15	237	5.47			308.15	6426	8.77		
	293.15	265	5.58			308.15	8006	8.99		
	293.15	245	5.50	25		313.15	7057	8.86	690	
	293.15	220	5.39			313.15	6367	8.76		
	293.15	270	5.60			313.15	7747	8.96		
	298.15	253	5.53			313.15	8830	9.09	645	
	298.15	259	5.56			313.15	8185	9.01		
	298.15	185	5.22			313.15	9475	9.16		
	303.15	250	5.52	8		1,1,2,2-Tetrachloroethane	293.15	3850	8.26	660
303.15	242	5.49				293.15	3190	8.07		
303.15	258	5.55				293.15	4510	8.41		
303.15	238	5.47	24			29315	3758	8.23	480	
303.15	214	5.37				29315	3278	8.09		
303.15	262	5.57				29315	4238	8.35		

Solute	T/K	Exp γ	ln γ	Error γ	Solute	T/K	Exp γ	ln γ	Error γ
1,1,2,2-Tetrachloroethane	293.15	3960	8.28			323.15	5893	8.68	
	298.15	3460	8.15		1,1,1-Trichloroethane	293.15	5880	8.68	75
	298.15	3790	8.24			293.15	5805	8.67	
	298.15	3360	8.12			293.15	5955	8.69	
	303.15	3726	8.22	410		293.15	5245	8.57	690
	303.15	3316	8.11			293.15	4555	8.42	
	303.15	4136	8.33			293.15	5935	8.69	
	303.15	2970	8.00	625		293.15	5660	8.64	
	303.15	2345	7.76			298.15	5903	8.68	
	303.15	3595	8.19			298.15	6030	8.70	
	308.15	3197	8.07	440		303.15	5480	8.61	60
	308.15	2757	7.92			303.15	5420	8.60	
	308.15	3637	8.20			303.15	5540	8.62	
	308.15	3460	8.15			303.15	5324	8.58	700
	313.15	3100	8.04	390		303.15	4624	8.44	
	313.15	2710	7.90			303.15	6024	8.70	
	313.15	3490	8.16			308.15	5097	8.54	700
	313.15	3570	8.18	245		308.15	4397	8.39	
	313.15	3325	8.11			308.15	5797	8.67	
	313.15	3815	8.25			308.15	6210	8.73	
	323.15	3010	8.01			313.15	4986	8.51	650
Trichloroethene	293.15	5410	8.60	160		313.15	4336	8.37	
	293.15	5250	8.57			313.15	5636	8.64	
	293.15	5570	8.63			313.15	5410	8.60	80
	293.15	4922	8.50	500		313.15	5330	8.58	
	293.15	4422	8.39			313.15	5490	8.61	
	293.15	5422	8.60			323.15	5850	8.67	
	293.15	5450	8.60		1,1,2-Trichlorethane	293.15	1500	7.31	
	296.15	4880	8.49			293.15	1520	7.33	10
	298.15	8750	9.08			293.15	1510	7.32	
	298.15	10400	9.25			293.15	1530	7.33	
	303.15	5034	8.52	520		293.15	1540	7.34	150
	303.15	4514	8.41			293.15	1390	7.24	
	303.15	5554	8.62			293.15	1690	7.43	
	303.15	5180	8.55	195		303.15	1472	7.29	140
	303.15	4985	8.51			303.15	1332	7.19	
	303.15	5375	8.59			303.15	1612	7.39	
	303.15	6061	8.71			308.15	1410	7.25	125
	313.15	4973	8.51	510		308.15	1285	7.16	
	313.15	4463	8.40			308.15	1535	7.34	
	313.15	5483	8.61			308.15	1520	7.33	
	313.15	5580	8.63	290		313.15	1424	7.26	140
	313.15	5290	8.57			313.15	1284	7.16	
	313.15	5870	8.68			313.15	1564	7.36	
	313.15	5943	8.69			323.15	1220	7.11	20

Solute	T/K	Exp γ	In γ	Error γ	Solute	T/K	Exp γ	In γ	Error γ	
1,1,2-Trichlorethane	323.15	1200	7.09			303.15	1375	7.23		
	323.15	1240	7.12			303.15	1435	7.27		
	323.15	1430	7.27			313.15	1243	7.13	130	
1,1-Dichloroethene	293.15	1894	7.55	270		313.15	1113	7.01		
	293.15	1624	7.39			313.15	1373	7.22		
	293.15	2164	7.68			313.15	1370	7.22	75	
	303.15	1930	7.57	270		313.15	1295	7.17		
	303.15	1660	7.41			313.15	1445	7.28		
	303.15	2200	7.70			313.15	1477	7.30		
	313.15	1936	7.57	270		323.15	1509	7.32		
	313.15	1666	7.42			1,2-Dibromoethane	298.15	2340	7.76	
	313.15	2206	7.70			308.15	2990	8.00		
	Cis 1,2-Dichloroethene	293.15	856	6.75	44		323.15	1740	7.46	
293.15		812	6.70			293.15	1080	6.98		
293.15		900	6.80			293.15	1100	7.00	15	
293.15		819	6.71	118		293.15	1085	6.99		
293.15		701	6.55			293.15	1115	7.02		
293.15		937	6.84			293.15	1046	6.95	110	
298.15		870	6.77			293.15	936	6.84		
303.15		803	6.69	116		293.15	1156	7.05		
303.15		687	6.53			303.15	1034	6.94	100	
303.15		919	6.82			303.15	934	6.84		
303.15		884	6.78	43		303.15	1134	7.03		
303.15		841	6.73			308.15	1120	7.02		
303.15		927	6.83			308.15	1240	7.12	20	
313.15		807	6.69	116		308.15	1220	7.11		
313.15		691	6.54			308.15	1260	7.14		
313.15		923	6.83			313.15	1017	6.92	100	
313.15		866	6.76	65		313.15	917	6.82		
313.15		801	6.69			313.15	1117	7.02		
313.15		931	6.84			318.15	1050	6.96	25	
Trans 1,2-Dichloroethene		293.15	1200	7.09	60		318.15	1025	6.93	
	293.15	1140	7.04			318.15	1075	6.98		
	293.15	1260	7.14			323.15	1080	6.98		
	293.15	1202	7.09	130		1,2-Dichloroethane	293.15	585	6.37	6
	293.15	1072	6.98			293.15	579	6.36		
	293.15	1332	7.19			293.15	591	6.38		
	293.15	1216	7.10			293.15	626	6.44		
	293.15	1220	7.11			293.15	660	6.49		
	298.15	1260	7.14			293.15	647	6.47	54	
	303.15	1202	7.09	130		293.15	593	6.39		
	303.15	1072	6.98			293.15	701	6.55		
	303.15	1332	7.19			298.15	641	6.46		
	303.15	1310	7.18	65		298.15	635	6.45		
	303.15	1245	7.13			298.15	511	6.24		

Solute	T/K	Exp γ	ln γ	Error γ		Solute	T/K	Exp γ	ln γ	Error γ
1,2-Dichloroethane	303.15	610	6.41	51			293.15	2329	7.75	
	303.15	559	6.33				298.15	2330	7.75	
	303.15	661	6.49				303.15	2006	7.60	230
	308.15	600	6.40	56			303.15	1776	7.48	
	308.15	544	6.30				303.15	2236	7.71	
	308.15	656	6.49				303.15	2310	7.75	30
	308.15	597	6.39	12			303.15	2280	7.73	
	308.15	585	6.37				303.15	2340	7.76	
	308.15	609	6.41				313.15	1925	7.56	220
	308.15	604	6.40				313.15	1705	7.44	
	313.15	587	6.38	49			313.15	2145	7.67	
	313.15	538	6.29				313.15	2090	7.64	30
	313.15	636	6.46				313.15	2060	7.63	
	313.15	579	6.36	12			313.15	2120	7.66	
	313.15	567	6.34			1-Chloropropane	293.15	3500	8.16	
	313.15	591	6.38				298.15	1747	7.47	
	323.15	552	6.31				298.15	1720	7.45	
	323.15	559	6.33	6			298.15	913	6.82	
	323.15	553	6.32			2-Chloropropane	298.15	1477	7.30	
	323.15	565	6.34				298.15	1320	7.19	
Ethyl Bromide	293.15	970	6.88			1-Iodopropane	298.15	8550	9.05	
	298.15	679	6.52				298.15	9730	9.18	
	298.15	633	6.45			Bromobutane	298.15	12240	9.41	
Ethyl Iodide	293.15	2200	7.70				298.15	12800	9.46	
	298.15	2192	7.69			2-Bromobutane	298.15	8315	9.03	
	298.15	2180	7.69				298.15	15900	9.67	
1-Bromopropane	298.15	2850	7.96			Chlorobutane	298.15	7609	8.94	
	298.15	2900	7.97				298.15	7640	8.94	
2-Bromopropane	298.15	2093	7.65			tert-Butyl chloride	293.15	5300	8.58	
	298.15	1950	7.58			Chloropentane	298.15	32070	10.38	
1,3-Dichloropropylene	293.15	1360	7.22	30			298.15	29500	10.29	
	293.15	1330	7.19			Chlorohexane	298.15	140800	11.86	
	293.15	1390	7.24			Fluorobenzene	298.15	4796	8.48	
	298.15	1400	7.24				298.15	3440	8.14	
	303.15	1430	7.27	25		Chlorobenzene	293.15	12960	9.47	
	303.15	1405	7.25				293.15	3000	8.01	
	303.15	1455	7.28				298.15	13990	9.55	
	313.15	1460	7.29	85			298.15	12700	9.45	
	313.15	1375	7.23				303.15	10280	9.24	
	313.15	1545	7.34				313.15	7450	8.92	
1,2-Dichloropropane	293.15	2340	7.76	30			323.15	4952	8.51	
	293.15	2310	7.75			Bromobenzene	298.15	22460	10.02	
	293.15	2370	7.77				298.15	20900	9.95	
	293.15	2089	7.64	240		Iodobenzene	298.15	54130	10.90	
	293.15	1849	7.52				298.15	53500	10.89	

Solute	T/K	Exp γ	In γ	Error γ		Solute	T/K	Exp γ	In γ	Error γ
1,2-Dichlorobenzene	298.15	68200	11.13				357.75	1.64	0.49	
	298.15	63900	11.07				357.75	1.92	0.65	
Benzyl Chloride	298.15	31960	10.37				373.15	2.47	0.90	
	298.15	35900	10.49				373.15	2.25	0.81	
	Alcohols						373.15	2.75	1.01	
Methanol	293.15	1.68	0.52	0.12			373.15	2.24	0.81	
	293.15	1.56	0.44			Ethanol	283.15	4.38	1.48	0.21
	293.15	1.8	0.59				283.15	4.17	1.43	
	293.15	2.69	0.99				283.15	4.59	1.52	
	297.45	2.12	0.75				293.15	4.81	1.57	0.19
	298.15	1.58	0.46				293.15	4.62	1.53	
	298.15	1.65	0.50	0.01			293.15	5	1.61	
	298.15	1.64	0.49				293.15	6.51	1.87	
	298.15	1.66	0.51				293.15	4.5	1.50	
	298.15	1.74	0.55				297.45	4.74	1.56	
	298.15	1.64	0.49				298.15	3.74	1.32	
	298.15	1.46	0.38				298.15	4.03	1.39	
	298.15	1.47	0.39				298.15	3.8	1.34	
	298.15	1.64	0.49				298.15	3.73	1.32	
	300.45	2.2	0.79				298.15	3.91	1.36	
	303.15	2.53	0.93				298.15	3.92	1.37	
	303.15	1.77	0.57	0.1			298.15	3.76	1.32	
	303.15	1.67	0.51				298.15	3.69	1.31	
	303.15	1.87	0.63				298.15	3.27	1.18	0.05
	313.15	2.3	0.83				298.15	3.22	1.17	
	313.15	1.92	0.65	0.08			298.15	3.32	1.20	
	313.15	1.84	0.61				298.15	3.55	1.27	
	313.15	2	0.69				298.15	3.83	1.34	
	317.85	1.46	0.38	0.16			298.15	3.88	1.36	
	317.85	1.3	0.26				298.15	4	1.39	
	317.85	1.62	0.48				303.15	6.15	1.82	
	323.15	1.93	0.66				303.15	4.12	1.42	
	328.15	2.31	0.84				313.15	5.17	1.64	0.16
	328.15	2.13	0.76				313.15	5.01	1.61	
	328.15	2.14	0.76				313.15	5.33	1.67	
	328.45	1.49	0.40	0.05			313.15	5.5	1.70	
	328.45	1.44	0.36				313.15	6.1	1.81	
	328.45	1.54	0.43				323.15	5.01	1.61	
	337.65	1.59	0.46	0.05			323.15	5.42	1.69	
	337.65	1.54	0.43				323.15	5.21	1.65	
	337.65	1.64	0.49				323.15	4.4	1.48	
	348.25	1.52	0.42	0.12			328.15	5.94	1.78	
	348.25	1.4	0.34				328.15	5.32	1.67	
	348.25	1.64	0.49				328.15	5.11	1.63	
	357.75	1.78	0.58	0.14			328.15	6.8	1.92	

Solute	T/K	Exp γ	In γ	Error γ		Solute	T/K	Exp γ	In γ	Error γ
Ethanol	333.15	5.59	1.72	0.12			328.15	14.11	2.65	
	333.15	5.47	1.70				328.15	12.25	2.51	
	333.15	5.71	1.74				337.55	9.5	2.25	0.4
	333.15	5.2	1.65				337.55	9.1	2.21	
	338.15	4.8	1.57				337.55	9.9	2.29	
	343.15	5.61	1.72				349.15	11	2.40	0.8
	343.15	4.3	1.46				349.15	10.2	2.32	
	348.15	7.6	2.03				349.15	11.8	2.47	
	353.15	3.2	1.16				353.15	13.62	2.61	
	363.15	5.9	1.77				357.75	11.6	2.45	0.5
	373.15	6.85	1.92				357.75	11.1	2.41	
	373.15	5.82	1.76				357.75	12.1	2.49	
	383.15	6.05	1.80				363.15	13.68	2.62	
2-Propan-1-ol	371.15	8.58	2.15				373.15	14	2.64	
1-Propanol	293.15	24	3.18				373.15	12.63	2.54	
	298.15	13.36	2.59			1-Butanol	293.15	41.4	3.72	
	298.15	14.17	2.65				298.15	50	3.91	
	298.15	13.5	2.60				298.15	53.33	3.98	
	298.15	17.2	2.84				298.15	52.24	3.96	
	298.15	10.9	2.39	0.2			298.15	53.7	3.98	
	298.15	10.7	2.37				298.15	50.5	3.92	
	298.15	11.1	2.41				298.15	45.1	3.81	1.4
	298.15	15	2.71				298.15	43.7	3.78	
	298.15	13.8	2.62				298.15	46.5	3.84	
	298.15	11.2	2.42				298.15	205.6	5.33	
	298.15	133.5	4.89				298.15	51.6	3.94	
	303.15	26	3.26				313.15	49.5	3.90	
	313.15	22	3.09				323.23	78.7	4.37	
	328.15	21.2	3.05				333.15	59.3	4.08	
	328.15	18.53	2.92				343.15	59.3	4.08	
	328.15	20.44	3.02				343.15	67.8	4.22	
	373.15	19.35	2.96				343.15	68.03	4.22	
2-Propanol	288.15	12.9	2.56	0.57			353.15	46.5	3.84	
	288.15	12.33	2.51				353.15	57.2	4.05	
	288.15	13.47	2.60				353.15	48.53	3.88	
	298.15	7.47	2.01				363.15	55.5	4.02	
	298.15	8.13	2.10				372.15	27.1	3.30	
	298.15	7.75	2.05				373.15	54	3.99	
	298.15	8.14	2.10				376.15	40.5	3.70	
	317.85	8.8	2.17	0.5			378.15	54.6	4.00	
	317.85	8.3	2.12			2-Butanol	293.15	20.8	3.03	
	317.85	9.3	2.23				298.15	26.2	3.27	
	328.05	9.6	2.26	0.5			298.15	22.4	3.11	
	328.05	9.1	2.21				323.18	35.5	3.57	
	328.05	10.1	2.31				333.55	46.49	3.84	

Solute	T/K	Exp γ	ln γ	Error γ		Solute	T/K	Exp γ	ln γ	Error γ
Iso-Butyl alcohol	293.15	44.4	3.79				317.85	1	0.00	0.1
	298.15	49	3.89				317.85	0.9	-0.11	
	298.15	48.4	3.88				317.85	1.1	0.10	
	323.15	58.1	4.06				328.15	1	0.00	0.1
Tert-Butanol	293.15	11.4	2.43				328.15	0.9	-0.11	
	298.15	11.9	2.48				328.15	1.1	0.10	
	298.15	12.2	2.50				338.05	1	0.00	0.1
	323.13	19.2	2.95				338.05	0.9	-0.11	
1-Pentanol	298.15	197	5.28				338.05	1.1	0.10	
	298.15	197.5	5.29				348.05	1.2	0.18	0.1
	298.15	192	5.26	8			348.05	1.1	0.10	
	298.15	184	5.21				348.05	1.3	0.26	
	298.15	200	5.30			1,2-Propanediol	296.75	1	0.00	0.1
	298.15	338.4	5.82				296.75	0.9	-0.11	
	298.15	198.1	5.29				296.75	1.1	0.10	
	298.15	225.4	5.42				308.25	1.1	0.10	0.1
	298.15	208.9	5.34				308.25	1	0.00	
2-Pentanol	298.15	97	4.57				308.25	1.2	0.18	
	363.15	126	4.84	6			318.75	1.2	0.18	0.1
	363.15	120	4.79				318.75	1.1	0.10	
	363.15	132	4.88				318.75	1.3	0.26	
3-Methyl-Butanol	298.15	208	5.34				328.45	1.2	0.18	0.2
1-Hexanol	298.15	799	6.68				328.45	1	0.00	
	298.15	791.8	6.67				328.45	1.4	0.34	
	298.15	738.9	6.61				337.85	1.3	0.26	0.2
	298.15	1012	6.92				337.85	1.1	0.10	
	298.15	645	6.47	32			337.85	1.5	0.41	
	298.15	613	6.42				348.25	1.3	0.26	0.2
	298.15	677	6.52				348.25	1.1	0.10	
2-Hexanonal	298.15	282	5.64				348.25	1.5	0.41	
Cyclohexanol	298.15	157	5.06			1,3-Propanediol	297.95	1.2	0.18	0.2
1-Heptanol	298.15	3270	8.09				297.95	1	0.00	
	298.15	4364	8.38				297.95	1.4	0.34	
2-Butoxy ethanol	278.15	19.5	2.97				307.85	1.2	0.18	0.2
	298.15	27.4	3.31				307.85	1	0.00	
	318.15	29.7	3.39				307.85	1.4	0.34	
	338.15	37	3.61				318.65	1.7	0.53	0.2
	358.15	38.5	3.65				318.65	1.5	0.41	
Trifluoroethanol	298.15	8.65	2.16				318.65	1.9	0.64	
1,2-Ethenediol	297.45	0.8	-0.22	0.1			328.65	1.8	0.59	0.2
	297.45	0.7	-0.36				328.65	1.6	0.47	
	297.45	0.9	-0.11				328.65	2	0.69	
	308.15	0.8	-0.22	0.1			338.35	1.9	0.64	0.2
	308.15	0.7	-0.36				338.35	1.7	0.53	
	308.15	0.9	-0.11				338.35	2.1	0.74	

Solute	T/K	Exp γ	In γ	Error γ		Solute	T/K	Exp γ	In γ	Error γ	
1,3-Propanediol	347.45	1.9	0.64	0.2			318.35	2.9	1.06	0.3	
	347.45	1.7	0.53				318.35	2.6	0.96		
	347.45	2.1	0.74				318.35	3.2	1.16		
1,2-Butendiol	299.15	2	0.69	0.2			326.75	3	1.10	0.3	
	299.15	1.8	0.59				326.75	2.7	0.99		
	299.15	2.2	0.79				326.75	3.3	1.19		
	308.35	2.1	0.74	0.2			337.55	3	1.10	0.3	
	308.35	1.9	0.64				337.55	2.7	0.99		
	308.35	2.3	0.83				337.55	3.3	1.19		
	318.05	2.2	0.79	0.2			349.85	3.1	1.13	0.4	
	318.05	2	0.69				349.85	2.7	0.99		
	318.05	2.4	0.88				349.85	3.5	1.25		
2,3-butanediol	327.25	2.3	0.83	0.3		2,3-butanediol	298.95	1.6	0.47	0.2	
	327.25	2	0.69				298.95	1.4	0.34		
	327.25	2.6	0.96				298.95	1.8	0.59		
	338.55	2.4	0.88	0.3			308.85	1.7	0.53	0.2	
	338.55	2.1	0.74				308.85	1.5	0.41		
	338.55	2.7	0.99				308.85	1.9	0.64		
	348.35	2.4	0.88	0.3			318.45	1.9	0.64	0.2	
	348.35	2.1	0.74				318.45	1.7	0.53		
	348.35	2.7	0.99				318.45	2.1	0.74		
	1,3-Butanediol	299.05	2.2	0.79	0.2			327.35	2.1	0.74	0.2
		299.05	2	0.69				327.35	1.9	0.64	
		299.05	2.4	0.88				327.35	2.3	0.83	
308.75		2.3	0.83	0.2			337.65	2.2	0.79	0.3	
308.75		2.1	0.74				337.65	1.9	0.64		
308.75		2.5	0.92				337.65	2.5	0.92		
318.25		2.4	0.88	0.3			347.35	1.9	0.64	0.2	
318.25		2.1	0.74				347.35	1.7	0.53		
318.25		2.7	0.99				347.35	2.1	0.74		
327.25		2.4	0.88	0.3			Phenol and Derivatives				
327.25		2.1	0.74			Phenol	298.15	551.2	6.31		
327.25		2.7	0.99				303.15	495.3	6.21		
1,4-Butanediol	338.05	2.5	0.92	0.3			308.15	375.3	5.93		
	338.05	2.2	0.79			o-Cresol	298.15	403.1	6.00		
	338.05	2.8	1.03				303.15	379.4	5.94		
	347.95	2.6	0.96	0.3			308.15	314.4	5.75		
	347.95	2.3	0.83			m-Cresol	298.15	574.3	6.35		
	347.95	2.9	1.06				303.15	554.3	6.32		
1,4-Butanediol	299.35	2.8	1.03	0.3			308.15	547.7	6.31		
	299.35	2.5	0.92			p-Cresol	298.15	738.2	6.60		
	299.35	3.1	1.13				303.15	611.6	6.42		
	309.05	2.8	1.03	0.3			308.15	510	6.23		
	309.05	2.5	0.92				Aldehydes				
309.05	3.1	1.13			Formaldehyde	298.15	2.8	1.03			

Solute	T/K	Exp γ	In γ	Error γ		Solute	T/K	Exp γ	In γ	Error γ
Acetaldehyde	293.15	3.59	1.28				318.15	8.99	2.20	
	298.15	3.94	1.37				373.15	11.66	2.46	
	298.15	29.4	3.38				373.15	10.81	2.38	
	303.15	4.36	1.47				373.15	8.76	2.17	
	303.15	26.04	3.26			2-Butanone	298.15	25.6	3.24	
	308.15	32.45	3.48				298.15	27.8	3.33	
	313.15	3.71	1.31				298.15	25.98	3.26	
Propionaldehyde	293.15	14.7	2.69				298.15	27.6	3.32	
	298.15	13.03	2.57				298.15	41.2	3.72	
	298.15	35.8	3.58				298.15	65.74	4.19	
	298.15	48.67	3.89				298.15	25.3	3.23	
	303.15	17.6	2.87				298.15	26.4	3.27	0.7
	303.15	40.1	3.69				298.15	25.7	3.25	
	308.15	41.79	3.73				298.15	27.1	3.30	
	313.15	15.4	2.73				303.15	29.5	3.38	
Butyraldehyde	293.15	63	4.14				303.15	72.32	4.28	
	298.15	48.6	3.88				308.15	73.86	4.30	
	298.15	29.14	3.37				343.15	26.7	3.28	
	298.15	69.2	4.24				343.15	29.5	3.38	2.5
	303.15	72.11	4.28				343.15	27	3.30	
	303.15	73	4.29				343.15	32	3.47	
	308.15	65.38	4.18				353.15	28.5	3.35	
	313.15	61	4.11				363.15	30.2	3.41	
	298.15	220.2	5.39				373.15	31.8	3.46	
	298.15	813.4	6.70				298.15	93.4	4.54	
Hexanal	298.15	824	39.02				298.15	98.9	4.59	
	298.15	8920	9.10				298.15	102	4.62	
	Ketones						298.15	102.5	4.63	
Acetone	288.15	5.85	1.77				298.15	135.1	4.91	
	288.15	5.83	1.76				303.15	125	4.83	
	298.15	7.01	1.95				308.15	129.7	4.87	
	298.15	7.56	2.02			3-Methyl-2-Butanone	298.15	84	4.43	
	298.15	7.31	1.99			2,4-Pentanedione	343.15	28.6	3.35	
	298.15	21.1	3.05				354.15	26.1	3.26	
	298.15	61.86	4.12				363.15	23.9	3.17	
	298.15	7.69	2.04	0.3			373.15	22	3.09	
	298.15	7.39	2.00			3-Pentanone	298.15	107.4	4.68	
	298.15	7.99	2.08				298.15	113	4.73	
	303.15	7.69	2.04				298.15	113	4.73	
	303.15	7.7	2.04				298.15	134.8	4.90	
	303.15	7.65	2.03				298.15	106	4.66	2
	303.15	7.42	2.00				298.15	104	4.64	
	303.15	59.11	4.08				298.15	108	4.68	
308.15	56.85	4.04				303.15	139.6	4.94		
318.15	9.85	2.29				308.15	180	5.19		

Solute	T/K	Exp γ	In γ	Error γ	Solute	T/K	Exp γ	In γ	Error γ
2-Hexanone	298.15	355.7	5.87			323.15	66.8	4.20	
	298.15	329.1	5.80		Propyl Acetate	298.15	242	5.49	
3-Hexanone	298.15	412	6.02			298.15	301.6	5.71	
2-Heptanone	298.15	1397	7.24			303.15	314.4	5.75	
	298.15	1055	6.96			308.15	326.8	5.79	
	298.15	882.2	6.78		Isopropyl Acetate	298.15	195.5	5.28	
	303.15	626	6.44			298.15	251.2	5.53	
	308.15	713	6.57			303.15	233.8	5.45	
2-Nonanone	298.15	16290	9.70			308.15	268	5.59	
	298.15	24800	10.12		Butyl Acetate	298.15	814	6.70	
Cyclopentanone	298.15	29.2	3.37			298.15	1058	6.96	
	Acids					303.15	1155	7.05	
Formic Acid	298.15	0.64	-0.45			308.15	1261	7.14	
	373.15	0.736	-0.31		Isobutyl Acetate	298.15	844.4	6.74	
Acetic Acid	298.5	0.92	-0.08		Pentyl Acetate	298.15	3233	8.08	
	298.15	2.9	1.06		Isopentyl Acetate	298.15	2977	8.00	
	339.58	3.03	1.11		Hexyl Acetate	298.15	12490	9.43	
	356.11	3.11	1.13		Methyl Propionate	293.15	87.1	4.47	
	366.66	3.25	1.18		Ethyl Propionate	298.15	256	5.55	
	373.15	3.65	1.29		Methyl Butyrate	293.15	331	5.80	
	373.15	2.92	1.07		Ethyl Butyrate	298.15	730	6.59	
Butyric Acid	298.15	52.9	3.97		Methyl Valerate	293.15	1259	7.14	
Propionic Acid,2-Methyl,-Methyl	293.15	309	5.73		Methyl Hexanoate	293.15	3981	8.29	
	Esters					Ethers			
Methyl Formate	293.15	16.1	2.78		Ethyl Ether	268.15	72.9	4.29	
	298.15	15.5	2.74			298.15	109.6	4.70	
	303.15	17.4	2.86			298.15	69.7	4.24	
	313.5	14.2	2.65		Propyl Ether	298.15	2315	7.75	
Ethyl Formate	293.15	46	3.83			298.15	2330	7.75	
	298.15	47.3	3.86			298.15	2210	7.70	
	303.15	51	3.93			298.15	1560	7.35	
	313.15	41	3.71		Isopropyl Ether	296.15	639	6.46	
Propyl Formate	293.15	150	5.01			298.15	628	6.44	
	298.15	169	5.13			298.15	2810	7.94	
	303.15	168	5.12			298.15	496.1	6.21	
	313.15	131	4.88			298.15	667	6.50	
Methyl Acetate	298.15	22.6	3.12		butyl ether	298.15	47180	10.76	
	313.15	27.2	3.30	1		298.15	28100	10.24	
	313.15	26.2	3.27			298.15	37500	10.53	
	313.15	28.2	3.34		T-Butyl-Methyl-Ether	298.15	112.5	4.72	
Ethyl Acetate	288.15	63.9	4.16	1.96	Tetrahydrofuran	293.15	16.6	2.81	0.2
	288.15	61.94	4.13			293.15	16.4	2.80	
	288.15	65.86	4.19			293.15	16.8	2.82	
	298.15	65.3	4.18			298.15	17	2.83	
	313.15	84.5	4.44			308.15	23.5	3.16	0.2

Solute	T/K	Exp γ	In γ	Error γ	Solute	T/K	Exp γ	In γ	Error γ
Tetrahydrofuran	308.15	23.3	3.15		Aniline	328.35	1.13	0.12	0.03
	308.15	23.7	3.17			328.35	1.1	0.10	
	313.15	20.8	3.03	1		328.35	1.16	0.15	
	313.15	19.8	2.99			337.85	1.39	0.33	0.2
	313.15	21.8	3.08			337.85	1.19	0.17	
	323.15	32.8	3.49	0.5		337.85	1.59	0.46	
Tetrahydropyran	323.15	32.3	3.48		N,N-Dimethylformamide	289.35	0.58	-0.54	0.03
	323.15	33.3	3.51			373.15	80	4.38	
	298.15	78.6	4.36						
Anisole	298.15	3650	8.20			289.35	0.55	-0.60	
	298.15	4610	8.44			289.35	0.61	-0.49	
Phenetole	298.15	15730	9.66			291.95	0.6	-0.51	0.1
	298.15	15800	9.67			291.95	0.5	-0.69	
Amines and Amides						291.95	0.7	-0.36	
Ethylamine	293.15	0.37	-0.99			297.95	0.65	-0.43	0.11
	303.15	0.69	-0.37			297.95	0.54	-0.62	
	313.15	0.44	-0.82			297.95	0.76	-0.27	
Propylamine	293.15	2.33	0.85			298.15	0.83	-0.19	
	303.15	2.85	1.05			298.15	0.62	-0.48	
	313.15	2.44	0.89			307.75	0.81	-0.21	0.08
Propylamine	293.15	4	1.39			307.75	0.73	-0.31	
	303.15	4.7	1.55			307.75	0.89	-0.12	
	313.15	3.7	1.31			308.25	0.7	-0.36	0.1
Triethylamine	298.15	67.5	4.21			308.25	0.6	-0.51	
Diethylamine	303.15	5.4	1.69			308.25	0.8	-0.22	
Diisopropylamine	293.15	33.24	3.50			317.85	0.95	-0.05	0.11
Dimethylacetamide	298.15	1.04	0.04			317.85	0.84	-0.17	
Piperidine	343.15	6.62	1.89			317.85	1.06	0.06	
	263.15	6.89	1.93			318.05	0.8	-0.22	0.1
	273.15	7.24	1.98			318.05	0.7	-0.36	
Pyridine	298.15	19.9	2.99			318.05	0.9	-0.11	
	343.15	24.6	3.20			323.15	0.89	-0.12	0.4
	363.15	20	3.00			323.15	0.49	-0.71	
	373.15	17	2.83			323.15	1.29	0.25	
4-Methylpyridine	298.5	42.3	3.74			328.25	1.11	0.10	0.04
3-Methylpyridine	298.15	49.1	3.89			328.25	1.07	0.07	
N-Methyl pyrrolidone	298.75	0.37	-0.99	0.014		328.25	1.15	0.14	
	298.75	0.356	-1.03			328.25	1.11	0.10	0.1
	298.75	0.384	-0.96			328.25	1.01	0.01	
	308.35	0.6	-0.51	0.09		328.25	1.21	0.19	
	308.35	0.51	-0.67			333.05	1.35	0.30	0.26
	308.35	0.69	-0.37			333.05	1.09	0.09	
	317.95	0.86	-0.15	0.15		333.05	1.61	0.48	
	317.95	0.71	-0.34			337.75	1.1	0.10	0.1
317.95	1.01	0.01			337.75	1	0.00		

Solute	T/K	Exp γ	In γ	Error γ		Solute	T/K	Exp γ	In γ	Error γ
N,N-Dimethylformamide	337.75	1.2	0.18				308.35	0.1	-2.30	0.002
	337.95	1.3	0.26	0.4			308.35	0.098	-2.32	
	337.95	0.9	-0.11				308.35	0.102	-2.28	
	337.95	1.7	0.53				317.95	0.12	-2.12	0.017
	343.05	2.67	0.98	0.15			317.95	0.103	-2.27	
	343.05	2.52	0.92				317.95	0.137	-1.99	
	343.05	2.82	1.04				328.35	0.135	-2.00	0.015
	347.35	1.1	0.10	0.1			328.35	0.12	-2.12	
	347.35	1	0.00				328.35	0.15	-1.90	
	347.35	1.2	0.18				337.85	0.17	-1.77	0.004
	Nitriles						337.85	0.166	-1.80	
Acrylonitrile	298.15	39.4	3.67				337.85	0.174	-1.75	
	373.15	35	3.56				Others			
Acetonitrile	298.15	11.1	2.41			1,4-Dioxane	298.15	5.42	1.69	
	298.15	10.09	2.31				298.15	5.45	1.70	
	298.15	16.37	2.80			Ethylene Oxide	298.15	6.23	1.83	
	298.15	9.24	2.22							
	364.15	21.1	3.05							
	373.15	10.3	2.33							
Propionitrile	298.15	35.29	3.56							
	298.15	37.56	3.63							
	298.15	34.9	3.55							
Butyronitrile	298.15	117.75	4.77							
	298.15	112.56	4.72							
	298.15	115.18	4.75							
Isobutyronitrile	298.15	117	4.76							
	298.15	134.69	4.90							
Valeronitrile	298.15	401.9	6.00							
	298.15	413	6.02							
	298.15	382.8	5.95							
Hexanenitrile	298.15	1401	7.24							
Benzonitrile	298.15	1741	7.46							
	Nito Compunds									
Nitromethane	298.15	31.6	3.45							
	298.15	31.68	3.46							
	323.15	21.4	3.06							
Nitroethane	298.15	88.6	4.48							
Nitropropane	298.15	299.2	5.70							
3-Nitrotoluene	298.15	7089	8.87							
	298.15	14500	9.58							
	Compunds with Sulfur									
Carbon disulfide	298.15	3300	8.10							
Dimethylsulfoxide	298.75	0.09	-2.41	0.02						
	298.75	0.07	-2.66							
	298.75	0.11	-2.21							

Table G-3. Organics in Water (Indirect Measurements)

Solute	T/K	Exp γ	In γ	Solute	T/K	Exp γ	In γ
Aliphatic Alkanes				Nonane	298.15	5.85E+07	17.88
Methane	298.15	137.4	4.92		298.15	3.24E+07	17.29
Ethane	298.15	717.1	6.58	4-Methyloctane	298.15	6.21E+07	17.94
Propane	298.15	4230	8.35	Decane	298.15	1.52E+08	18.84
Butane	298.15	2.19E+04	9.99		298.15	1.58E+08	18.88
iso-Butane	298.15	1.96E+04	9.88		298.15	3.34E+08	19.63
	298.15	1.87E+04	9.84	Dodecane	298.15	2.78E+09	21.74
Pentane	298.15	9.09E+04	11.42		298.15	2.56E+09	21.66
	298.15	1.04E+05	11.55	Cyclic Alkanes			
2,2-Dimethyl propane	298.15	1.20E+05	11.70	Cyclopentane	298.15	2.49E+04	10.12
	298.15	7.42E+04	11.21	Cyclohexane	298.15	8.13E+04	11.31
	298.15	7.11E+04	11.17	Methyl cyclopentane	298.15	1.09E+05	11.60
2-Methyl butane	298.15	8.26E+04	11.32	Cycloheptane	298.15	1.82E+05	12.11
	298.15	8.39E+04	11.34	1-Methyl cyclohexane	298.15	3.40E+05	12.74
Hexane	298.15	5.04E+05	13.13	Cyclooctane	298.15	7.89E+05	13.58
	298.15	3.40E+05	12.74	1,2-Dimethyl cyclohexane(cis)	298.15	1.04E+06	13.85
	298.15	4.35E+05	12.98	Pentylcyclopentane	298.15	6.78E+07	18.03
	298.15	3.89E+05	12.87	Aliphatic Alkenes			
2,2-Dimethyl butane	298.15	2.27E+05	12.33	Ethylene	298.15	1.99E+02	5.29
	298.15	2.60E+05	12.47		298.15	2.02E+02	5.31
2,3-Dimethyl butane	298.15	2.27E+05	12.33	Propene	298.15	1.04E+03	6.95
	298.15	2.13E+05	12.27		298.15	1.15E+03	7.05
2-Methyl pentane	298.15	3.48E+05	12.76	1-Butene	298.15	4.79E+03	8.47
	298.15	3.47E+05	12.76		298.15	5.07E+03	8.53
3-Methyl pentane	298.15	3.70E+05	12.82	2-Methylpropene	298.15	4.65E+03	8.44
	298.15	3.74E+05	12.83	1-Pentene	298.15	2.63E+04	10.18
Heptane	298.15	2.33E+06	14.66	2-Pentene	298.15	1.92E+04	9.86
	298.15	1.90E+06	14.46	3-Methyl-1-Butene	298.15	2.50E+04	10.13
2-Methylhexane	298.15	2.19E+06	14.60		298.15	29940	10.31
3-Methylhexane	298.15	1.12E+06	13.93	2-Methyl-2-Butene	298.15	17760	9.78
2,2-Dimethyl pentane	298.15	1.27E+06	14.05	1-Hexene	298.15	87700	11.38
2,3-Dimethyl pentane	298.15	1.06E+06	13.87		298.15	84300	11.34
2,4-Dimethyl pentane	298.15	1.32E+06	14.09		298.15	93500	11.45
	298.15	1.37E+06	14.13	2-Methyl-1-Pentene	298.15	59900	11.00
	298.15	1.01E+06	13.83	4-Methyl-1-Pentene	298.15	97400	11.49
3,3-Dimethyl pentane	298.15	9.43E+05	13.76	2,3-Dimethyl-1-butene	298.15	467	6.15
	298.15	9.37E+05	13.75	2-Heptene	298.15	363000	12.80
Octane	298.15	9.08E+06	16.02	1-Octene	298.15	2310000	14.65
	298.15	9.62E+06	16.08	1-Nonene	298.15	6290000	15.65
3-Methylheptane	298.15	8.01E+06	15.90	1,3-Butadiene	298.15	4090	8.32
2,2,4-Trimethyl pentane	298.15	2.86E+06	14.87	1,4-Pentadiene	298.15	6780	8.82
	298.15	2.60E+06	14.77	2-Methyl-1,3-butadiene	298.15	6210	8.73
2,3,4-Trimethyl pentane	298.15	2.76E+06	14.83		298.15	5900	8.68
	298.15	4.67E+06	15.36	1,5-Hexadiene	298.15	27000	10.20
2,3,5-Trimethyl pentane	298.15	6.20E+06	15.64	1,6-Heptadiene	298.15	121000	11.70

Solute	T/K	Exp γ	In γ	Solute	T/K	Exp γ	In γ
	Cyclic Alkenes				298.15	30700	10.33
Cyclopentene	298.15	7070	8.86		308.15	29850	10.30
Cyclohexene	298.15	28600	10.26		318.15	27510	10.22
Cycloheptene	298.15	80900	11.30	m-Xylene	288.15	37250	10.53
1-Methyl cyclohexene	298.15	103000	11.54		298.15	36440	10.50
4-Ethenyl cyclohexene	298.15	120000	11.70		298.15	33800	10.43
1,4-Cyclohexadiene	298.15	4760	8.47		308.15	35180	10.47
Cycloheptatriene	298.15	800	6.68		318.15	31800	10.37
	Alkynes			p-Xylene	288.15	37500	10.53
Propyne	298.15	104	4.64		295.15	36315.5	10.50
1-Butyne	298.15	583	6.37		298.15	36200	10.50
1-Pentyne	298.15	2410	7.79		308.15	34300	10.44
1-Hexyne	298.15	12700	9.45		318.15	33140	10.41
1-Heptyne	298.15	56800	10.95	1,2,3-Trimethylbenzene	288.15	111500	11.62
1-Octyne	298.15	255000	12.45		298.15	88700	11.39
1-Nonyne	298.15	958000	13.77		298.15	106500	11.58
1,6-Heptadiyne	298.15	3100	8.04		308.15	92500	11.43
1,8-Nonanediynes	298.15	53400	10.89		318.15	78360	11.27
	Monocyclic Aromatics			1,2,4-Trimethylbenzene	288.15	127600	11.76
Benzene	278.15	2684.7	7.90		298.15	115000	11.65
	288.15	2748.8	7.92		298.15	118100	11.68
	295.15	2670.4	7.89		308.15	107400	11.58
	298.15	2676.9	7.89		318.15	96340	11.48
	298.15	2480	7.82	1,3,5-Trimethylbenzene	288.15	144900	11.88
	298.15	2400	7.78		298.15	139000	11.84
	308.15	2543.2	7.84		298.15	133500	11.80
	318.15	2413.8	7.79		308.15	121400	11.71
Styrene	298.15	18000	9.80		318.15	113400	11.64
Indan	298.15	60200	11.01	o-Ethyl toluene	298.15	71940	11.18
	298.15	60000	11.00	p-Ethyl toluene	298.15	70400	11.16
m-Methyl styrene	298.15	73500	11.21	Propylbenzene	288.15	143100	11.87
p-Methyl styrene	298.15	73500	11.21		298.15	131000	11.78
Toluene	288.15	9948.2	9.21		308.15	121300	11.71
	295.15	9701.2	9.18	Propylbenzene	318.15	104200	11.55
	298.5	9701.8	9.18		288.15	112100	11.63
	298.15	9412	9.15		295.15	109097.8	11.60
	308.15	9392.8	9.15		298.15	116000	11.66
	318.15	8760.9	9.08		298.15	108600	11.60
Ethylbenzene	288.15	36760	10.51		308.15	97220	11.48
	298.15	34840	10.46		318.15	86230	11.36
	298.15	35700	10.48	Sec-butylbenzene	298.15	424000	12.96
	308.15	33470	10.42	Tert-butylbenzene	298.15	253000	12.44
	318.15	30060	10.31	1,2,4,5-Tetramethyl benzene	298.15	2140000	14.58
o-Xylene	288.15	35130	10.47		Polycyclic Aromatics		
	298.15	32970	10.40	Naphthalene	298.15	6.83E+04	11.13

Solute	T/K	Exp γ	ln γ	Solute	T/K	Exp γ	ln γ
Naphthalene	298.15	6.42E+04	11.07	Trichlorofluoromethane	298.15	7040	8.86
1-Methylnaphthalene	298.15	2.82E+05	12.55	Dichlorodifluoromethane	298.15	22400	10.02
2-Methylnaphthalene	298.15	2.65E+05	12.49	Chlorotrifluoromethane	298.15	64500	11.07
	298.15	2.51E+05	12.43	Carbon tetrafluoride	298.15	306000	12.63
1,3-Dimethylnaphthalene	298.15	1.09E+05	11.60	Nitrotrichloromethane	298.15	5630	8.64
1,4-Dimethylnaphthalene	298.15	7.63E+05	13.55	Dichlorofluoromethane	298.15	299	5.70
1,5-Dimethylnaphthalene	298.15	7.45E+05	13.52	Chlorodifluoromethane	298.15	1730	7.46
2,3-Dimethylnaphthalene	298.15	5.02E+05	13.13	Chloroform	293.15	810	6.70
2,6-Dimethylnaphthalene	298.15	4.69E+05	13.06	Trifluoromethane	298.15	4310	8.37
	298.15	6.53E+05	13.39	Triiodomethane	298.15	219000	12.30
1-Ethylnaphthalene	298.15	8.06E+05	13.60	Difluoromethane	298.15	658	6.49
Biphenyl	298.15	4.90E+05	13.10	Diiodomethane	298.15	11990	9.39
	298.15	4.33E+05	12.98	Bromomethane	298.15	389	5.96
Acenaphthene	298.15	3.98E+05	12.89	Chloromethane	298.15	474	6.16
	298.15	4.33E+05	12.98	Fluoromethane	298.15	787	6.67
	298.15	4.25E+05	12.96	Hexachloroethane	298.15	1640000	14.31
Fluorene	298.15	7.01E+05	13.46	1,1,2-Trichlorotrifluoroethane	298.15	61400	11.03
	298.15	5.93E+05	13.29	1,2-Dichlorotetrafluoroethane	298.15	69400	11.15
Phenanthrene	298.15	1.76E+06	14.38	Chloropentafluoroethane	298.15	148000	11.90
	298.15	1.50E+06	14.22	1,1,2,2-Tetrabromoethane	303.15	29500	10.29
Anthracene	298.15	1.84E+06	14.43	Tetrachloroethene	298.15	61360	11.02
	298.15	1.72E+06	14.36	Tetrafluoroethene	298.15	35100	10.47
	298.15	2.35E+06	14.67	Hexafluoroethane	298.15	971000	13.79
2-Methylantracene	298.15	4.17E+06	15.24	Pentachloroethane	293.15	23900	10.08
9-Methylantracene	298.15	1.14E+07	16.25	Bromoethane	298.15	670	6.51
9,10-Dimethylantracene	298.15	5.77E+06	15.57	Chloroethene	298.15	1290	7.16
Pyrene	298.15	4.17E+06	15.24	1-Bromo-2-Chloroethane	303.15	1160	7.06
	298.15	3.87E+06	15.17	1,2-Dibromoethane	298.15	2530	7.84
Fluoranthene	298.15	6.02E+06	15.61	1,1-Dichloroethene	293.15	10760	9.28
1,2-Benzofluorene	298.15	6.73E+06	15.72		298.15	25600	10.15
2,3-Benzofluorene	298.15	1.60E+07	16.59	Cis 1,2-Dichloroethene	298.15	1540	7.34
Chrysene	298.15	3.41E+07	17.34	Chloroethane	298.15	394	5.98
Triphenylene	298.15	5.65E+06	15.55	3-Bromo-1-Propene	298.15	1750	7.47
Naphthacene	298.15	1.43E+07	16.48	3-Chloro-1-propene	298.15	1060	6.97
1,2-Benzanthracene	298.15	4.24E+07	17.56	1,2,3-Trichloropropane	298.15	4310	8.37
	298.15	5.41E+07	17.81	1,2-Dibromopropane	298.15	7810	8.96
7,12-Dimethyl-1,2-Benzanthracene	298.15	2.60E+07	17.07		298.15	7840	8.97
Perylene	298.15	1.15E+08	18.56	1,3-Dibromopropane	303.15	6670	8.81
Benzo[a]pyrene	298.15	3.75E+08	19.74	1,3-Dichloropropane	298.15	2300	7.74
	298.15	1.21E+07	16.31	1-Bromopropane	303.15	2770	7.93
3-Methylcholanthrene	298.15	1.61E+08	18.90	2-Bromopropane	293.15	2130	7.66
Benzo[ghi]perylene	298.15	1.89E+08	19.06	2-Iodopropane	298.15	6760	8.82
Coronene	298.15	9.82E+08	20.71	Octafluorocyclobutane	298.15	222000	12.31
				Halogenated Hydrocarbons			
Carbon Tetrachloride	293.15	11000	9.31	2-Chlorobutane	298.15	5130	8.54
				1-Chloro-2-Methylpropane	298.15	5560	8.62

Solute	T/K	Exp γ	In γ		Solute	T/K	Exp γ	In γ
Butyl acetate	298.15	799.5	6.68					
3-Methylbutyl acetate	303.15	3146	8.05					
Propyl butyrate	290.15	4454	8.40					
Isopropyl butyrate	298.15	3077	8.03					
Ethyl pentanoate	298.15	2880	7.97					
Cyclohexyl acetate	298.15	2720	7.91					
Butyl pentanoate	298.15	19400	9.87					
Hypochlorous acid, tertbutyl ester	293.15	1880	7.54					
	Ethers					Nitro Compounds		
Methyl propyl ether	298.15	132	4.88		2-Nitropropane	298.15	292	5.68
Methyl butyl ether	298.15	546	6.30		Nitrobenzene	298.15	3530	8.17
Methyl sec. Butyl ether	298.15	302	5.71		2-Nitrotoluene	303.15	11680	9.37
Ethyl isopropyl ether	298.15	200	5.30		2-Nitro-1-methoxy benzene	303.15	5030	8.52
Ethyl propyl ether	298.15	257	5.55		3-Nitrotoluene	303.15	15300	9.64
Isobutyl methyl ether	298.15	442	6.09					
tert-amyl methyl ether	293.15	449	6.11					
Isopropyl propyl ether	298.15	1202	7.09		Carbon disulfide	298.15	1140	7.04
Dipropyl ether	298.15	1444	7.28		Methanethiol	298.15	110	4.70
	Aldehydes				Ethanethiol	298.15	231	5.44
Heptanol	303.15	4178	8.34		1-Butanethiol	298.15	8330	9.03
Octanal	303.15	19230	9.86					
Nonanal	303.15	75188	11.23		Methyl sulfide	298.15	174	5.16
5-Methylfurfural	298.15	128	4.85		Ethyl sulfide	298.15	1600	7.38
					Thiophene	298.15	1550	7.35

Table G-4. Water in Organic Measurements

Solute	T/K	Exp γ	ln γ	Error γ	Solute	T/K	Exp γ	ln γ	Error γ
Aliphatic Alkanes					1,2-Dichloroethane	293.15	1224.80		
Pentane	298.15	2273	7.73			308.15	764.33		
Hexane	298.15	1650	7.41			323.15	554.01		
Heptane	298.15	1422	7.26		1,2-Dibromopropane	298.15	1735.15		
Octane	298.15	1235	7.12		Chlorobenzene	293.15	3297.5	80	
Aromatic Hydrocarbons						293.15	3295.80		
Benzene	293.15	245.4	5.50			308.15	243.25	49	
	308.15	170	5.14			323.15	201.35	30	
	323.15	135.5	4.91		1,3-Dioxolan-2-one, 4-Chloromethyl	324.95	9.58	2.26	
Halogenated Hydrocarbons						329.95	9.12	2.21	
Chloroform	293.15	200	5.30		Alcohols				
	308.15	120	4.79		Methanol	307.75	1.60	47	0.2
	323.15	88	4.48			307.75	1.40	34	
1,1,2,2-Tetrabromoethane	298.15	426	6.05			307.75	1.80	59	
	308.15	263	5.57			317.85	1.80	59	0.2
	323.15	233	5.45			317.85	1.60	47	
1,1,2,2-Tetrachloroethane	293.15	89	4.49			317.85	20.69		
	308.15	66	4.19			323.15	1.50	41	
	323.15	46	3.83			327.85	1.70	53	0.25
Tetrachloroethene	298.15	35.6	3.57			327.85	1.45	37	
1,1,1-Trichloroethane	293.15	417	6.03			327.85	1.95	67	
	308.15	313	5.75			337.05	1.76	57	0.05
	323.15	227	5.42			337.65	1.44	36	
Trichloroethene	293.15	400.5	5.99			337.65	1.58	46	
	293.15	400	5.99			337.75	1.77	57	
	308.15	354.2	5.87			373.15	1.39	33	
	323.15	282.1	5.64		Ethanol	323.15	2.35	85	
Tribromomethane	298.15	455	6.12			323.15	2.49	91	
	308.15	340	5.83			351.45	2.58	95	
	323.15	256	5.55			351.45	2.52	92	
1,1,2-Trichloroethane	293.15	128	4.85		1-Propanol	370.35	3.39	1.22	
	308.15	85	4.44		2-Propanol	288.15	5.72	1.74	0.11
	323.15	59	4.08			288.15	5.61	1.72	
1,1-Dichloroethene	285.15	890.3	6.79			288.15	5.83	1.76	
	293.15	461.4	6.13			318.35	3.51	1.25	0.4
trans-1,2-Dichloroethene	285.15	385.6	5.95			318.35	3.11	1.13	
	293.15	215.5	5.37			318.35	3.91	1.36	
	308.15	185.3	5.22			328.15	3.39	1.22	0.07
cis-1,2-Dichloroethene	298.15	35	3.56			328.15	3.32	1.20	
1,2-Dibromoethane	298.15	317	5.76			328.15	3.46	1.24	
	308.15	222	5.40			337.95	31.10	0.16	
	323.15	171	5.14			337.95	2.84	1.04	
1,1-Dichloroethane	293.15	172	5.15			337.95	3.16	1.15	
	308.15	127	4.84			355.35	3.11	1.13	
	323.15	96	4.56			355.55	3.04	1.11	

Solute	T/K	Exp γ	In γ	Error γ	Solute	T/K	Exp γ	In γ	Error γ
2-Propanol	363.15	3.09	1.13			393.15	1.28	0.25	
	373.15	3.01	1.10		Triethylene Glycol	322.55	0.63	-0.46	
2-Propen-1-ol	371.15	2.72	1.00			323.15	0.668	-0.40	
1-Butanol	308.15	5.41	1.69			328.15	0.672	-0.40	
	323.23	5.2	1.65			331.45	0.66	-0.42	
	333.15	5.31	1.67			333.15	0.685	-0.38	
	343.15	3.27	1.18			338.15	0.698	-0.36	
	353.15	3.12	1.14			340.75	0.68	-0.39	
	363.15	3.07	1.12			343.15	0.714	-0.34	
	363.15	4	1.39			348.15	0.728	-0.32	
	372.15	2.97	1.09			348.95	0.716	-0.33	
	373.15	4	1.39			353.15	0.735	-0.31	
	376.15	3.98	1.38			358.15	0.748	-0.29	
	378.15	3.71	1.31			358.75	0.752	-0.29	
	383.15	2.9	1.06			363.15	0.758	-0.28	
2-Butanol	323.18	4.6	1.53			368.15	0.774	-0.26	
	353.15	3.16	1.15			373.15	0.779	-0.25	
Iso-butyl Alcohol	323.15	5.3	1.67			378.15	0.807	-0.21	
	379.15	3.94	1.37			383.15	0.828	-0.19	
Tert-butyl Alcohol	323.13	4.9	1.59			Ketones			
2-Pentanol	363.15	3.8	1.34	0.2	Acetone	307.85	6.02	1.80	0.15
	363.15	3.6	1.28			307.85	5.87	1.77	
	363.15	4	1.39			307.85	6.17	1.82	
2-Butoxy ethanol	278.15	2.77	1.02			318.05	5.68	1.74	0.1
	298.15	2.83	1.04			318.05	5.58	1.72	
	318.15	2.83	1.04			318.05	5.78	1.75	
	338.15	2.86	1.05			328.45	5.3	1.67	0.02
	358.15	2.72	1.00			328.45	5.28	1.66	
	383.15	2.72	1.00			328.45	5.32	1.67	
	383.15	2.61	0.96			329.25	4.06	1.40	
Glycerol	330.85	0.648	-0.43			329.25	4.92	1.59	
	333.15	0.78	-0.25			373.15	3.6	1.28	
	338.15	0.805	-0.22		2-Butanone	323.15	7.3	1.99	
	339.75	0.667	-0.40			333.15	7.2	1.97	
	343.15	0.836	-0.18			343.15	6.4	1.86	0.2
	348.15	0.866	-0.14			343.15	6.2	1.82	
	349.95	0.676	-0.39			343.15	6.6	1.89	
	353.15	0.899	-0.11			343.15	7	1.95	
	358.15	0.931	-0.07			347.15	7.32	1.99	
	359.25	0.685	-0.38			353.15	6.9	1.93	
	363.15	0.966	-0.03		2,4-Pentanedione	354.15	4.6	1.53	
	364.05	0.696	-0.36			363.15	3.36	1.21	
	373.15	1.06	0.06			373.15	1.89	0.64	
Glycerol	378.15	1.1	0.10		Cyclohexanone	363.15	9.03	2.20	
	383.15	1.14	0.13						

Solute	T/K	Exp γ	ln γ	Error γ		Solute	T/K	Exp γ	ln γ	Error γ
Nitromethane	314.25	13.5	2.60				353.65	1.08	0.08	
	322.45	11.8	2.47	0.5			362.95	0.99	-0.01	0.05
	322.45	11.3	2.42				362.95	0.94	-0.06	
	322.45	12.3	2.51				362.95	1.04	0.04	
	323.15	11.8	2.47				367.65	1.02	0.02	0.03
	323.15	11.8	2.47				367.65	0.99	-0.01	
	323.15	11.8	2.47				367.65	1.05	0.05	
	333.05	11.8	2.47	0.8			382.95	1.33	0.29	0.07
	333.05	11	2.40				382.95	1.26	0.23	
	333.05	12.6	2.53				382.95	1.4	0.34	
	343.35	8.6	2.15	0.4			397.75	1.22	0.20	0.15
	343.35	8.2	2.10				397.75	1.07	0.07	
	343.35	9	2.20				397.75	1.37	0.31	
	353.65	8.7	2.16	0.8		Piperidine	343.15	3	1.10	
	353.65	7.9	2.07				363.15	3.23	1.17	
	353.65	9.5	2.25				373.15	3.61	1.28	
Acetonitrile	364.15	4.18	1.43			Pyridine	343.15	2.87	1.05	
Acrylonitrile	352.15	7.39	2.00				363.15	2.8	1.03	
Diisopropylamine	293.15	33.24	3.50				373.15	2.8	1.03	
3-Methylpyridine	343.15	3	1.10			Aniline	323.15	2.55	0.94	
Dimethylformamide	353.65	0.96	-0.04	0.12			373.15	5.81	1.76	
	353.65	0.84	-0.17							

Appendix H. Type I Analysis

This appendix contains results from the Type I analyses, which are comprised of a general summary, a table of the optimum descriptor set used for each case study, plots showing data outliers, determination plots for optimum number of parameters, and plots of the calculated values of the infinite-dilution activity coefficients.

Table H-1. Summary of Type I Results

	CS1-A	CS1-B	CS2	CS3	CS4	CS5
Type I						
Descriptors	10	10	12	10	10	12
R ²	0.971	0.947	0.970	0.934	0.942	0.932

Table H-2. Optimum Descriptors Used in the Type I Analyses

Type I Descriptors					
CS1-A	CS1-B	CS2	CS3	CS4	CS5
HA dependent HDSA-2/TMSA [Zefirov's PC]	Kier shape index (order 1)	Max e-e repulsion for a O atom	Number of O atoms	Vib heat capacity (300K)	Kier shape index (order 1)
Max net atomic charge	Min partial charge for a O atom [Zefirov's PC]	count of H-acceptor sites [Zefirov's PC]	TMSA Total molecular surface area [Zefirov's PC]	Max partial charge for a N atom [Zefirov's PC]	Max SIGMA-PI bond order
temperature	temperature	Moment of inertia A	index	index2	index2
1X GAMMA polarizability (DIP)	min(#HA, #HD) [Quantum-Chemical PC]	Max net atomic charge for a N atom	temperature	temperature	temperature
WNSA-1 Weighted PNSA (PNSA1*TMSA/1000) [Zefirov's PC]	Min atomic orbital electronic population	FHBSA Fractional HBSA (HBSA/TMSA) [Quantum-Chemical PC]	Max net atomic charge for a N atom	Min partial charge for a O atom [Zefirov's PC]	count of H-donors sites [Quantum-Chemical PC]
Max partial charge for a N atom [Zefirov's PC]	Molecular volume / XYZ Box	FNSA-1 Fractional PNSA (PNSA-1/TMSA) [Quantum-Chemical PC]	RNCG Relative negative charge (QMNEG/QTMINUS) [Zefirov's PC]	HASA-2 [Zefirov's PC]	Min (>0.1) bond order of a C atom
Number of O atoms	Min n-n repulsion for a N-O bond	DPSA-1 Difference in CPSAs (PPSA1-PNSA1) [Zefirov's PC]	Min atomic state energy for a C atom	PPSA-2 Total charge weighted PPSA [Zefirov's PC]	WNSA-2 Weighted PNSA (PNSA2*TMSA/1000) [Quantum-Chemical PC]
WPSA-2 Weighted PPSA (PPSA2*TMSA/1000) [Quantum-Chemical PC]	Min atomic state energy for a C atom	PPSA-3 Atomic charge weighted PPSA [Zefirov's PC]	count of H-donors sites [Quantum-Chemical PC]	ZX Shadow / ZX Rectangle	Max partial charge for a N atom [Zefirov's PC]
Number of double bonds	Min nucleoph. react. index for a N atom	Min (>0.1) bond order of a O atom	LUMO+1 energy	Min e-n attraction for a C-H bond	Min n-n repulsion for a C-O bond
Min (>0.1) bond order of a O atom	PPSA-1 Partial positive surface area [Quantum-Chemical PC]	Max partial charge for a N atom [Zefirov's PC]	RNCG Relative negative charge (QMNEG/QTMINUS) [Quantum-Chemical PC]	Min electroph. react. index for a N atom	WNSA-2 Weighted PNSA (PNSA2*TMSA/1000) [Zefirov's PC]
		Min resonance energy for a N-O bond			HA dependent HDSA-1 [Zefirov's PC]
		Min e-e repulsion for a H atom			Number of O atoms

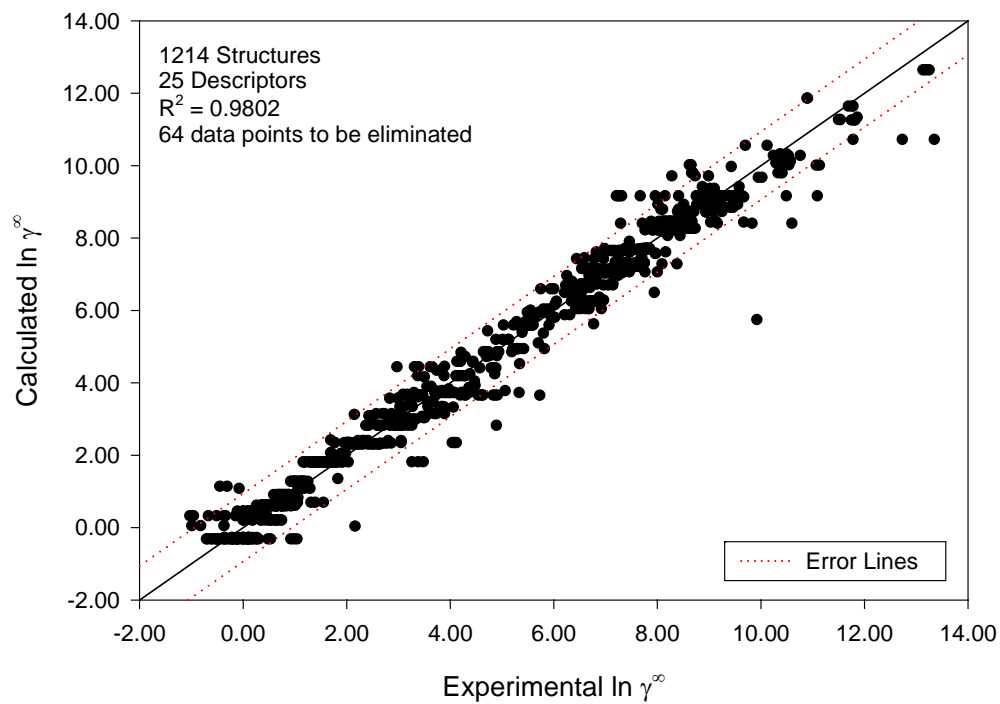


Figure H-1. Infinite-Dilution Activity Coefficients of CS1-A with Outliers

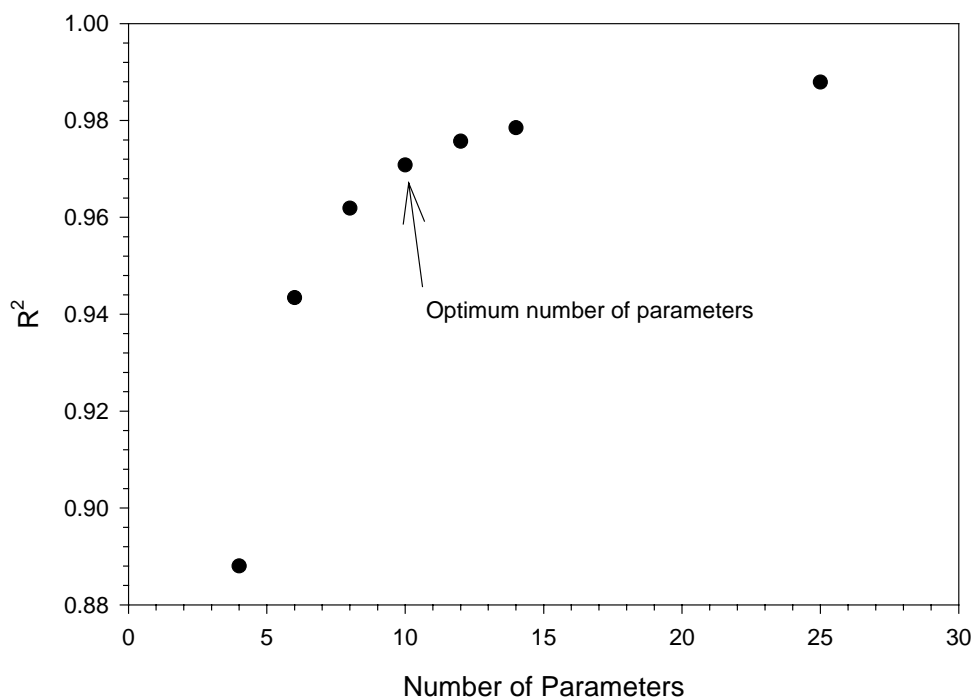


Figure H-2. R^2 Plot for Number of Parameters Determination (CS1-A Case Study)

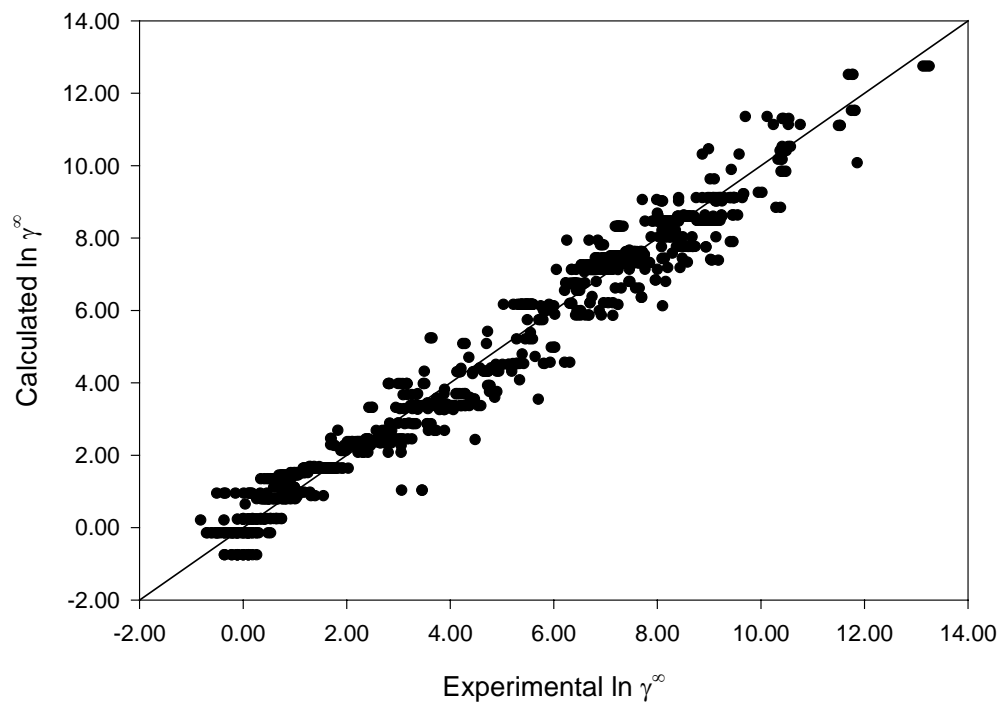


Figure H-3. Infinite-Dilution Activity Coefficients of CS1-A Case Study (Type I)

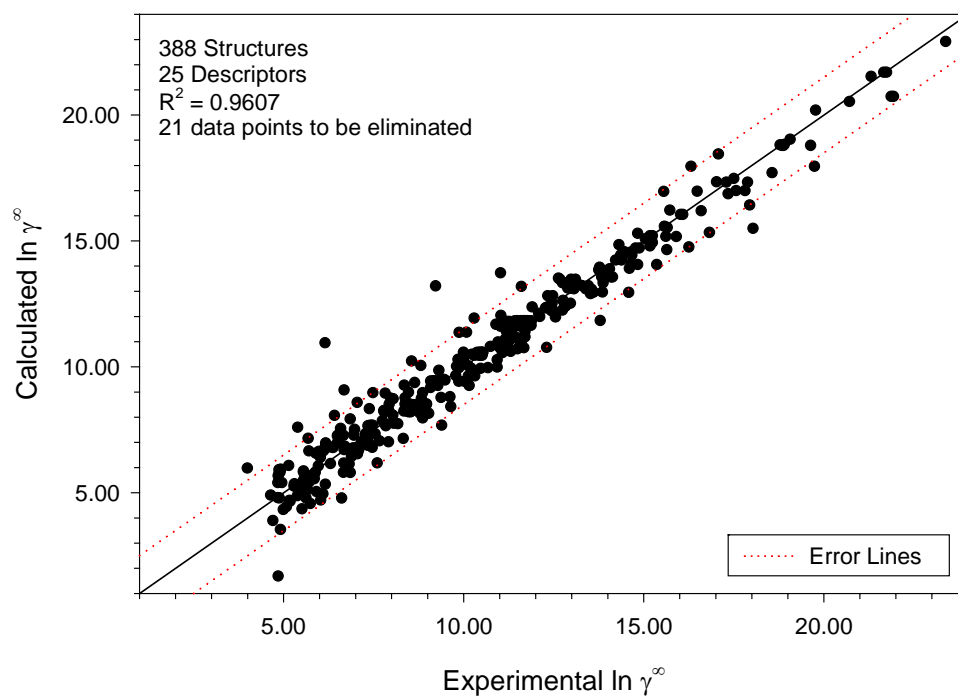


Figure H-4. Infinite-Dilution Activity Coefficients of CS1-B with Outliers

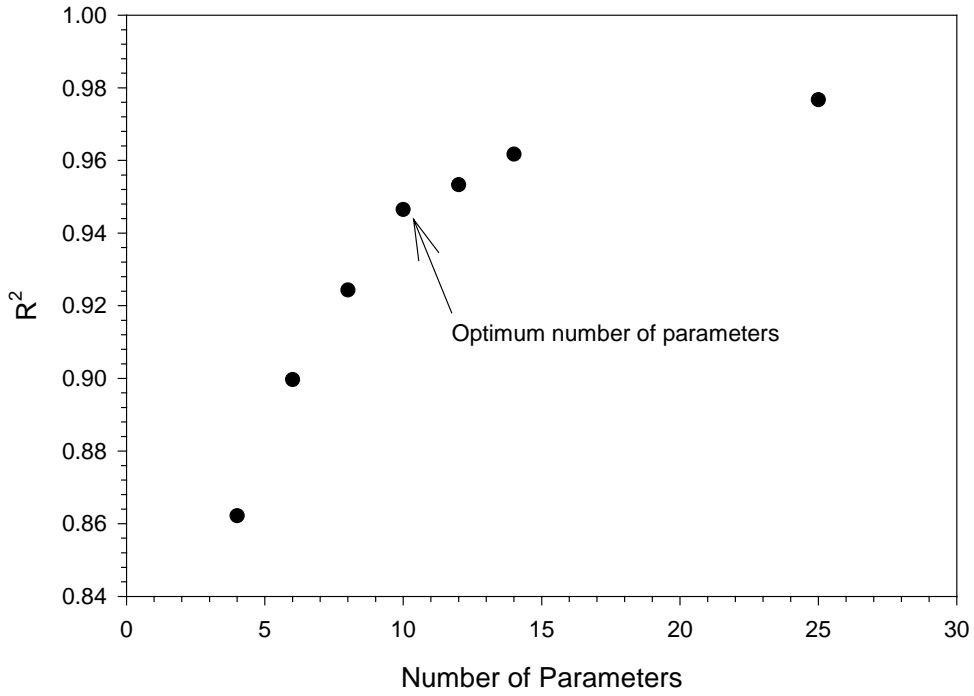


Figure H-5. R^2 Plot for Number of Parameters Determination (CS1-B Case Study)

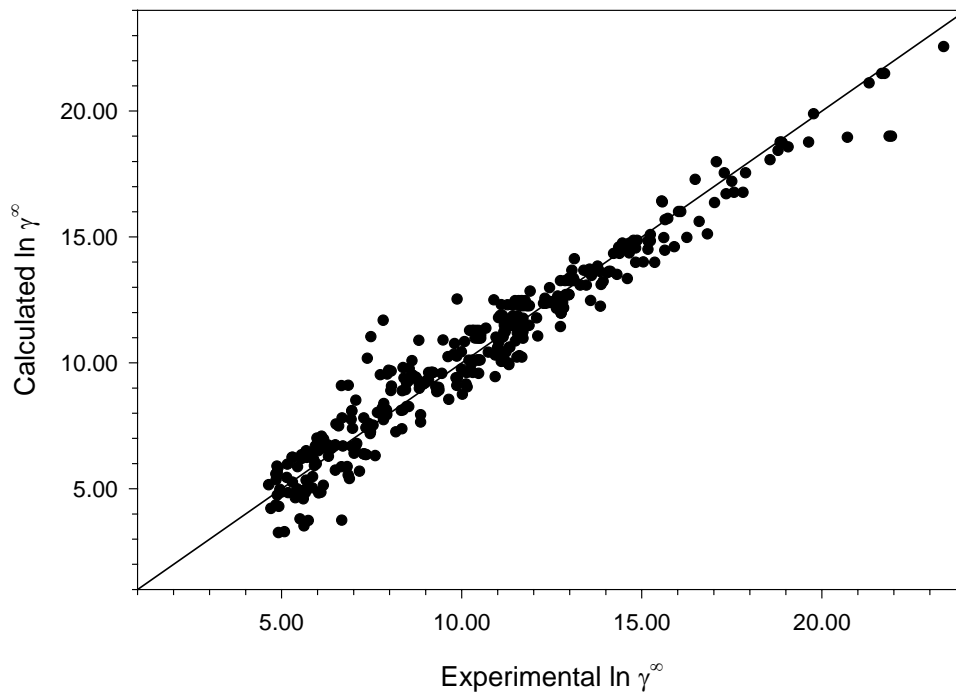


Figure H-6. Infinite-Dilution Activity Coefficients of CS1-B Case Study (Type I)

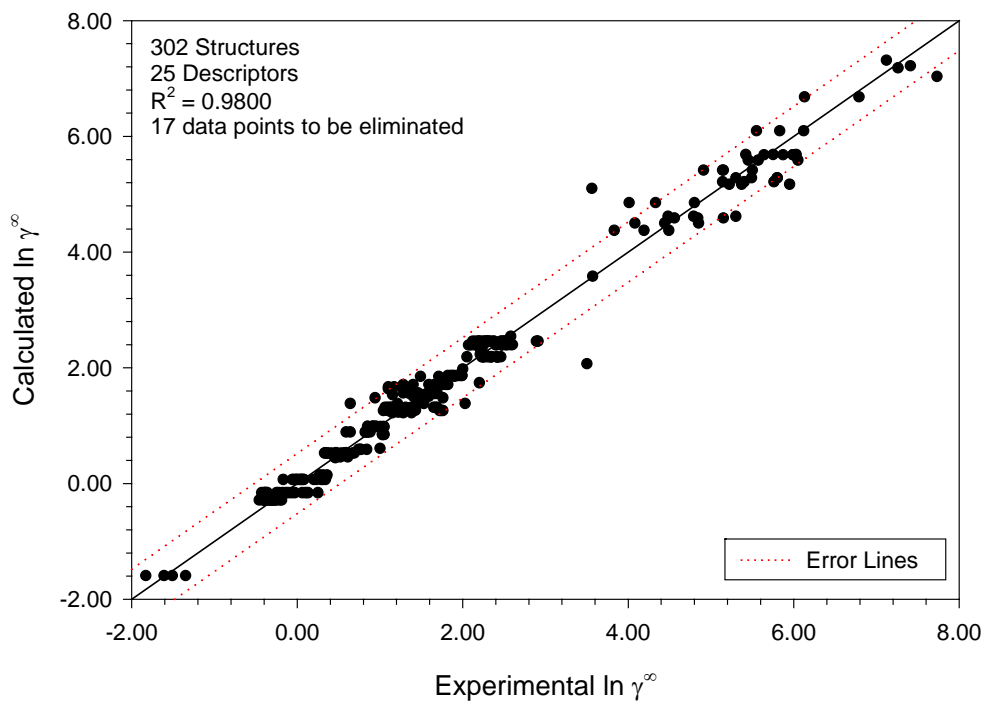


Figure H-7. Infinite-Dilution Activity Coefficients of CS2 with Outliers

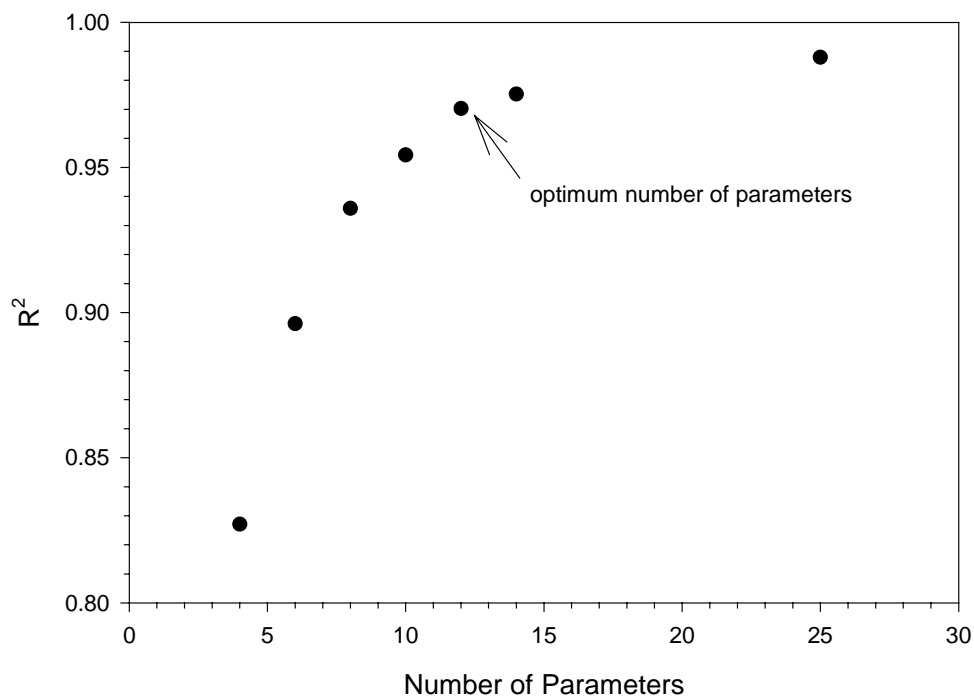


Figure H-8. R^2 Plot for Number of Parameters Determination (CS2Case Study)

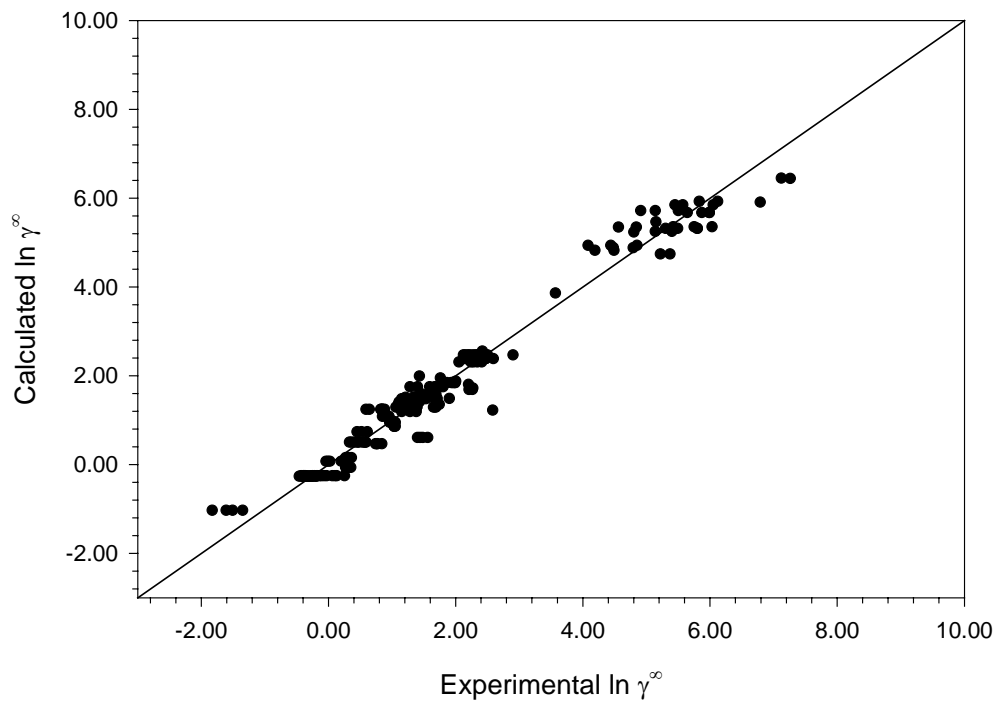


Figure H-9. Infinite-Dilution Activity Coefficients of CS2Case Study (Type I)

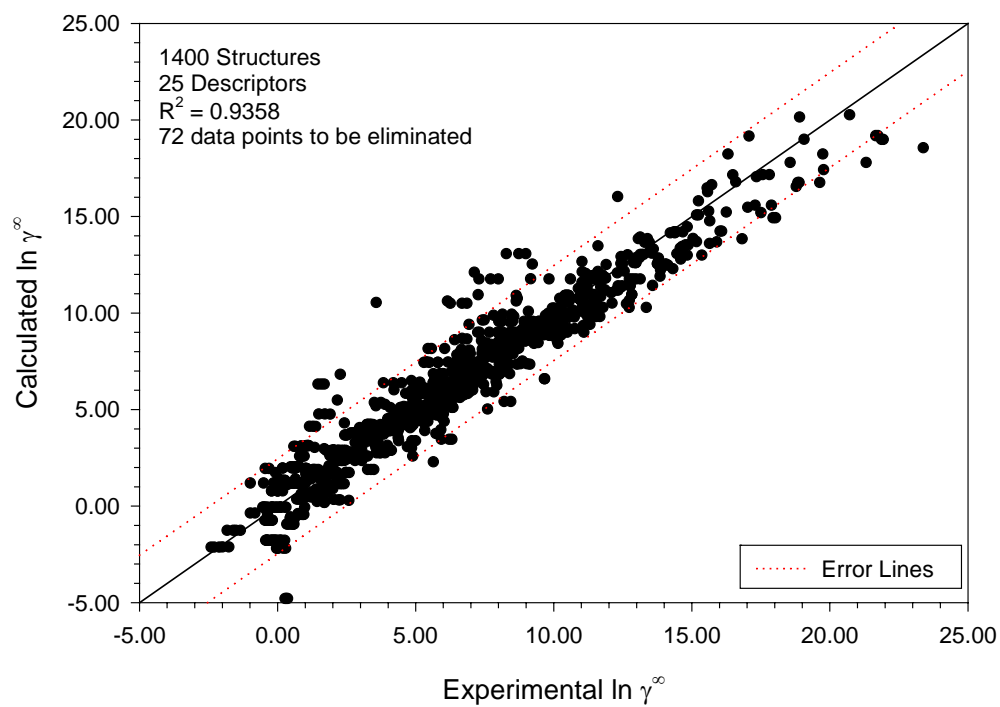


Figure H-10. Infinite-Dilution Activity Coefficients of CS3 with Outliers

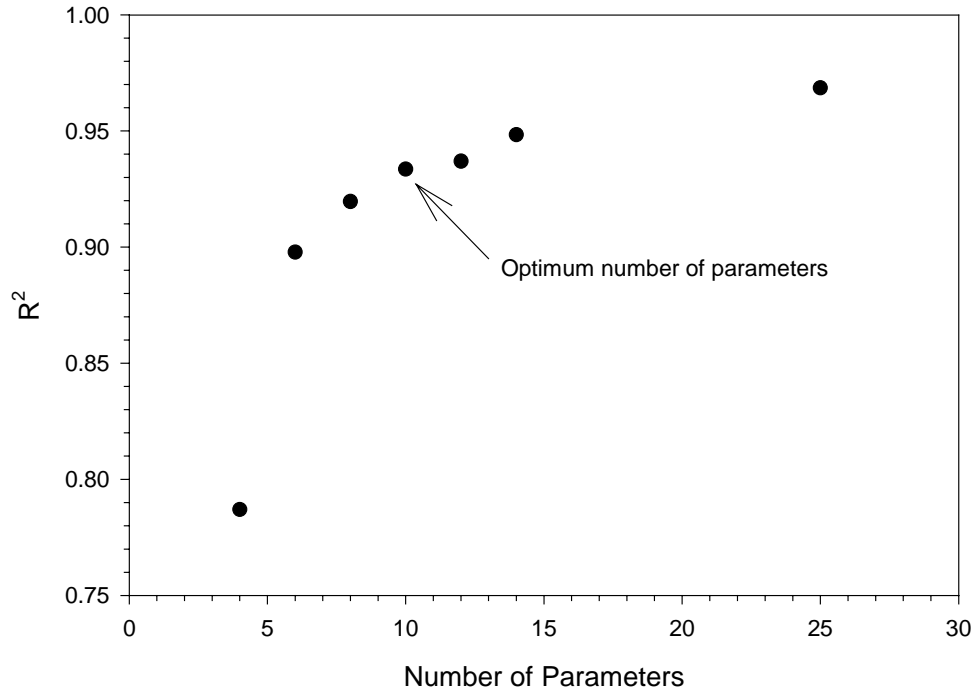


Figure H-11. R^2 Plot for Number of Parameters Determination (CS3 Case Study)

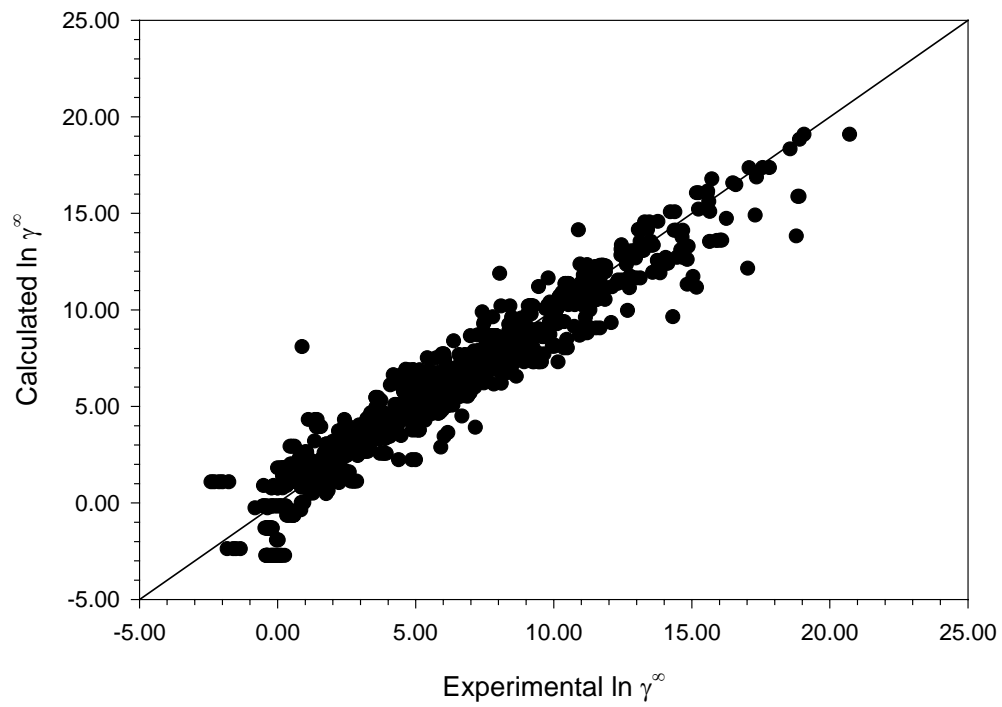


Figure H-12. Infinite-Dilution Activity Coefficients of CS3 Case Study (Type I)

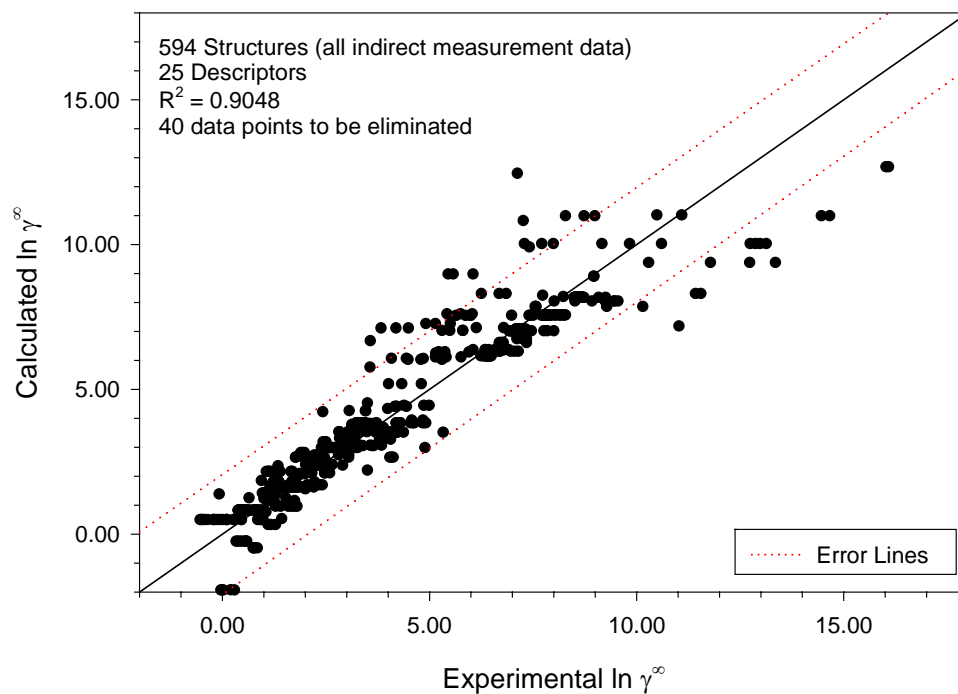


Figure H-13. Infinite-Dilution Activity Coefficients of CS4 with Outliers

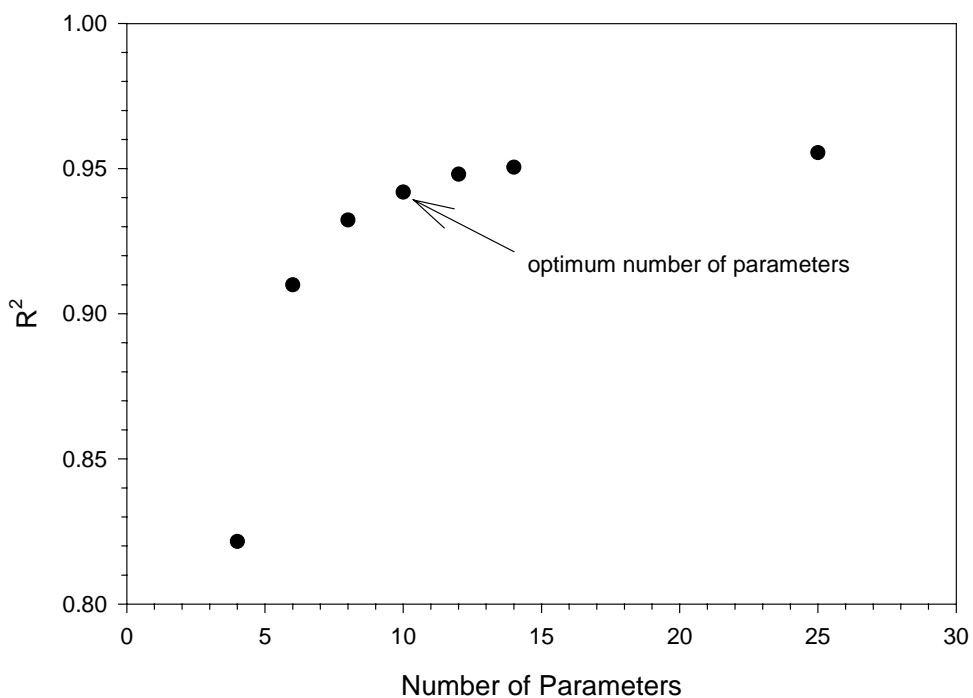


Figure H-14. R^2 Plot for Number of Parameters Determination (CS4 Case Study)

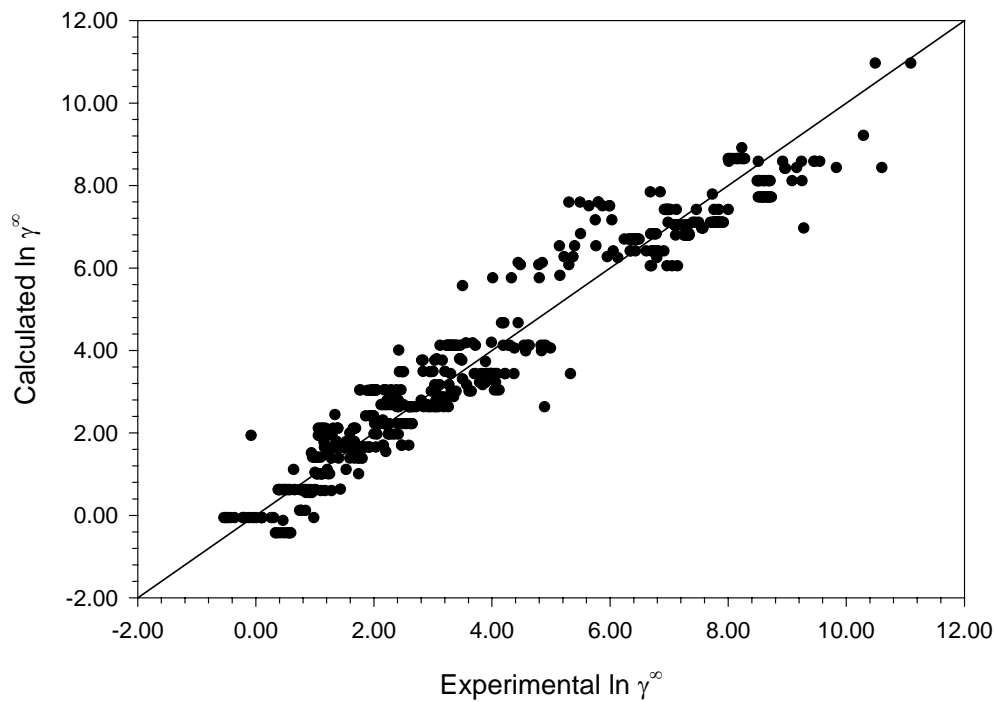


Figure H-15. Infinite-Dilution Activity Coefficients of CS4 Case Study (Type I)

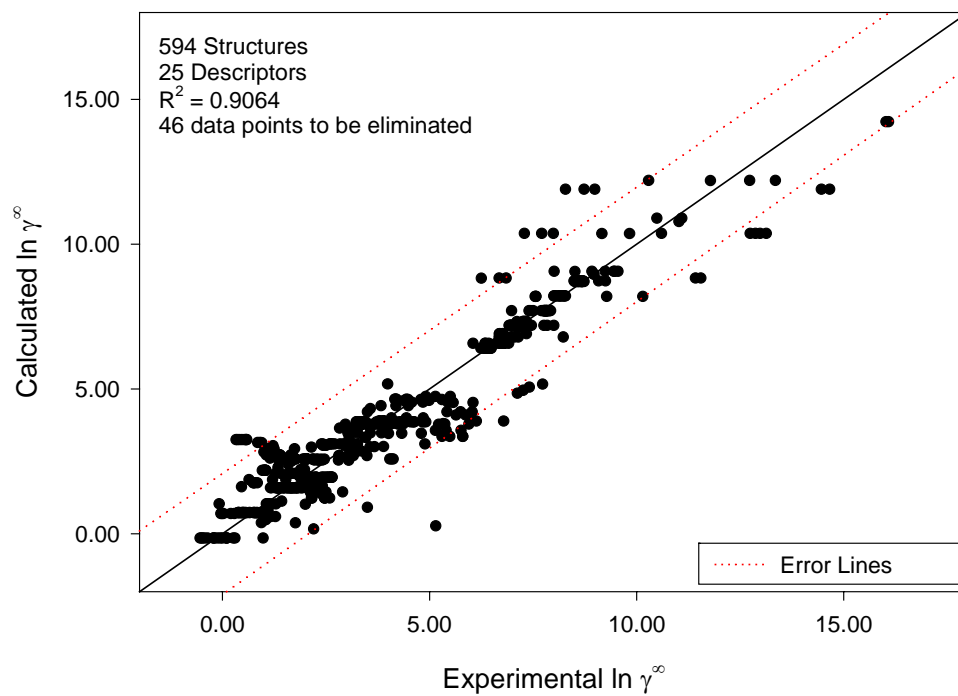


Figure H-16. Infinite-Dilution Activity Coefficients of CS5 with Outliers

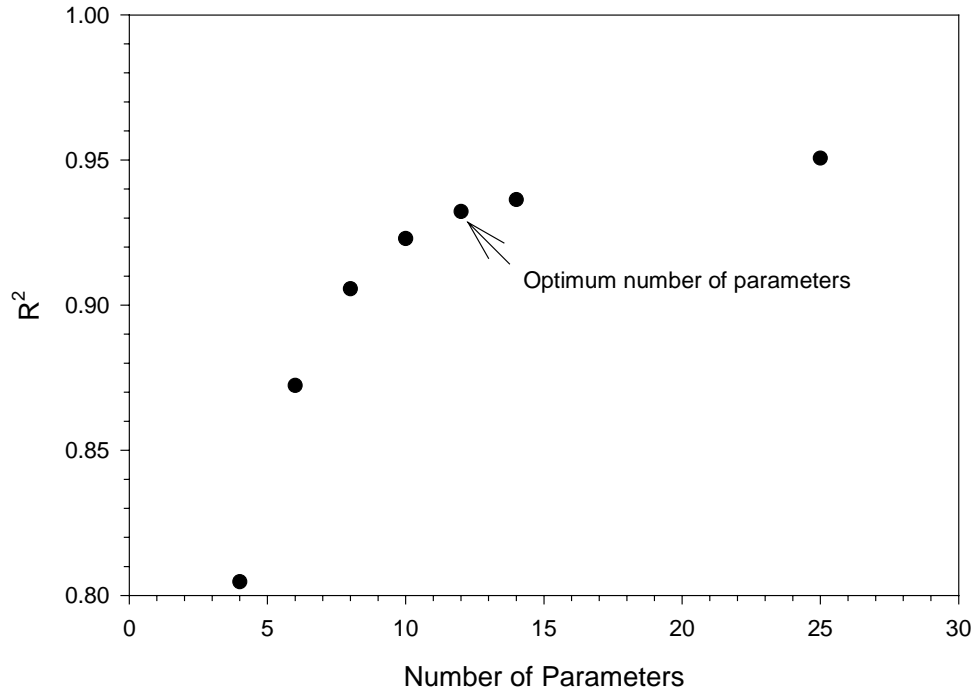
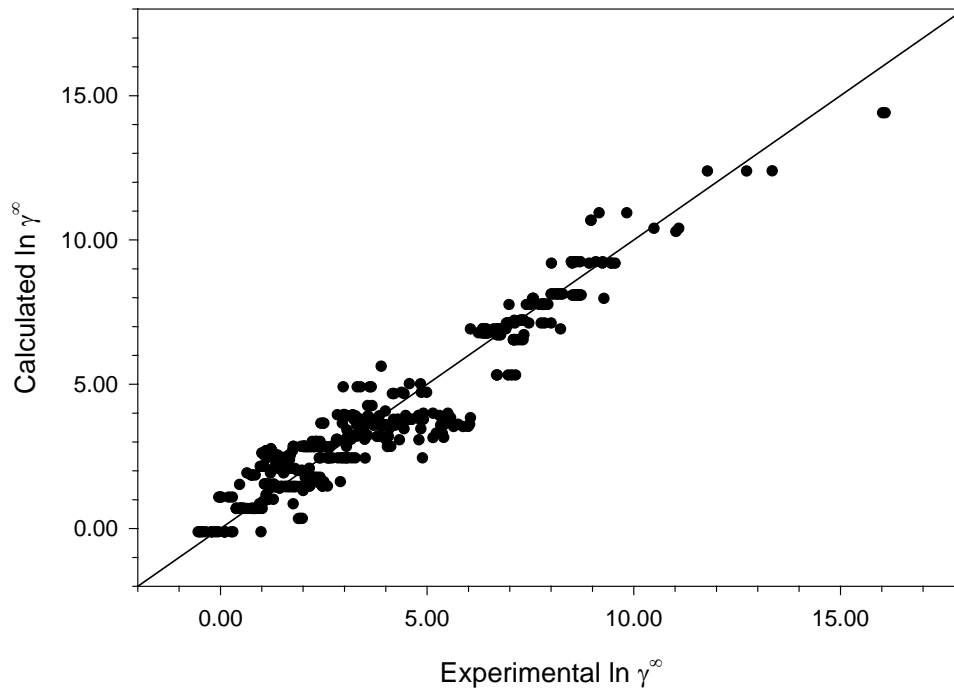


Figure H-17. R^2 Plot for Number of Parameters Determination (CS5 Case Study)



Appendix I. Type II Analysis

This appendix contains results from the Type II analyses, which are comprised of a general summary, a table of the optimum descriptor set used for each case study, determination plots for optimum number of parameters, and plots of both the calculated values of the infinite-dilution activity coefficients and the difference between the experimental and calculated values.

Table I-1. Summary of Type II Results

	CS1		CS2		CS3		CS4		CS5	
TYPE II										
Descriptors	6		5		3		3		3	
R ²	0.955		0.903		0.956		0.960		0.893	
Results in γ^∞										
	%AAD	RMSE	%AAD	RMSE	%AAD	RMSE	%AAD	RMSE	%AAD	RMSE
Training Set	22.2	1.04	170.5	0.78	34.1	0.89	31.1	0.55	43.6	0.94
Prediction Set	21.2	0.98	116.0	0.95	17.4	0.68	21.3	0.54	41.9	1.08
Cross Validation Set	17.4	0.87	112.0	0.93	19.1	0.78	20.6	0.61	36.1	1.34

Table I-2. Optimum Descriptors Used in the Type II Analyses

Type II Descriptors				
CS1	CS2	CS3	CS4	CS5
TMSA Total molecular surface area [Zefirov's PC]	Min e-n attraction for a O atom	Min n-n repulsion for a C-O bond	Max SIGMA-PI bond order	Kier&Hall index (order 0)
exch. eng. + e-e rep. for a C-O bond	DPSA-3 Difference in CPSAs (PPSA3-PNSA3) [Zefirov's PC]	Complementary Information content	Tot molecular 1-center E-N attraction	Max SIGMA-PI bond order
logP	HACA-1 [Quantum-Chemical PC]	FGorg (functional group)	FG (functional group)	FG (functional group)
Max n-n repulsion for a C-H bond	Tot heat capacity (300K) / # of atoms			
MP (melting point)	Polarity parameter (Qmax-Qmin)			
Relative number of O atoms				

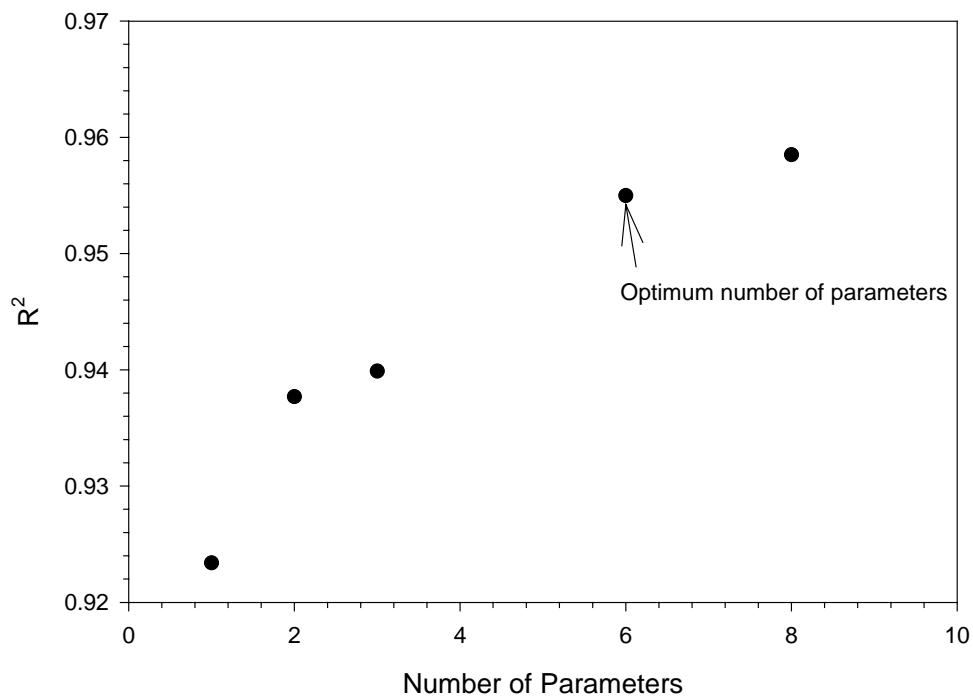


Figure I-1. R^2 Plot for Number of Parameters Determination (CS1 Case Study)

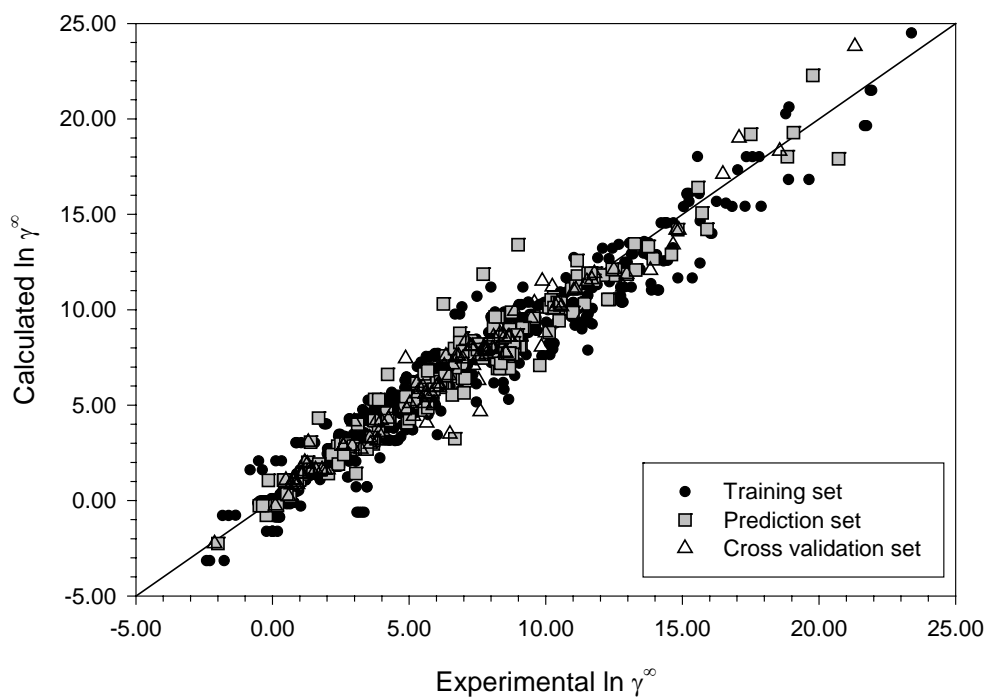


Figure I-2. Infinite-Dilution Activity Coefficients of CS1 Case Study (Type II)

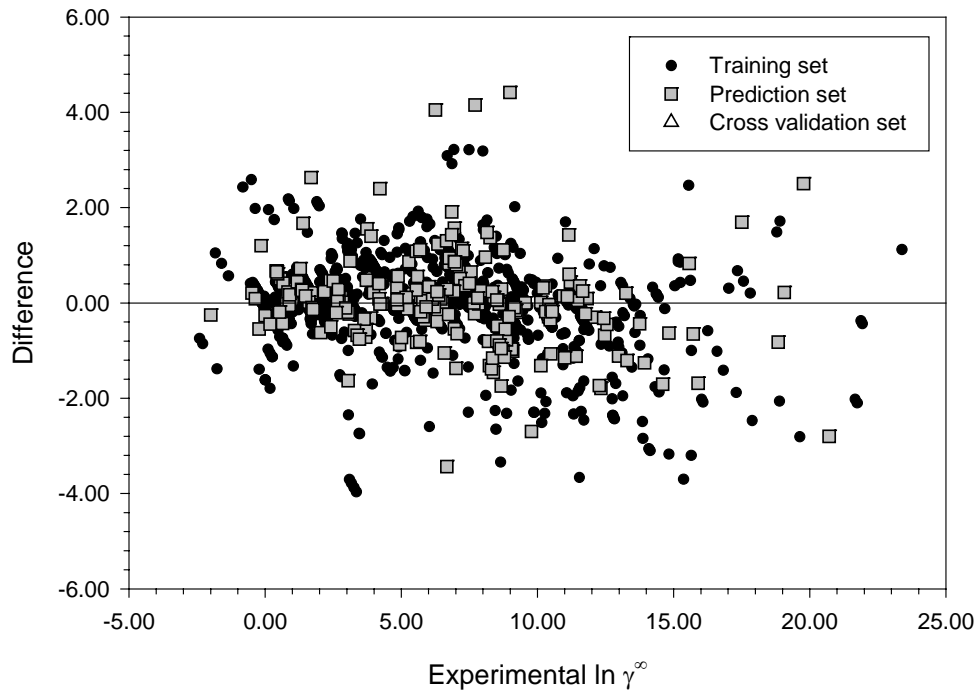


Figure I-3. Difference Plot of CS1 Case Study (Type II)

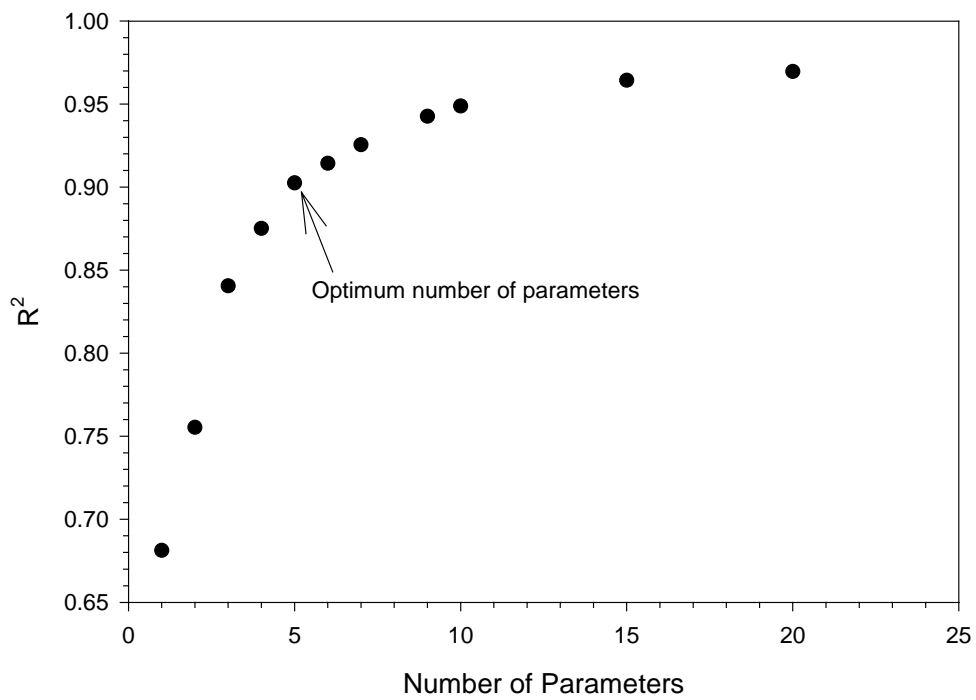


Figure I-4. R^2 Plot for Number of Parameters Determination (CS2 Case Study)

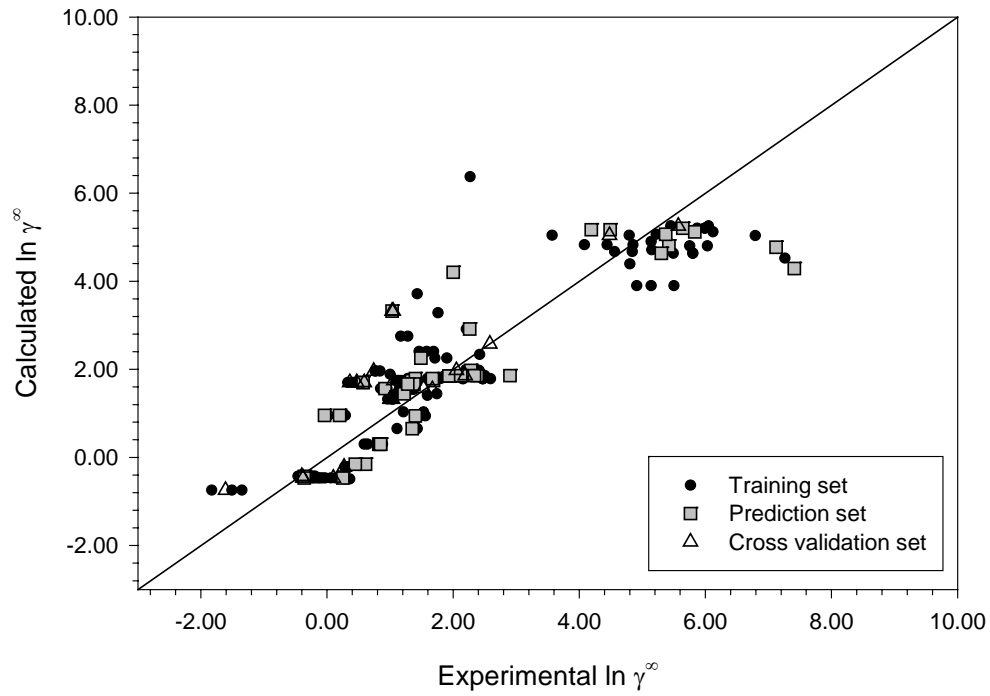


Figure I-5. Infinite-Dilution Activity Coefficients of CS2 Case Study (Type II)

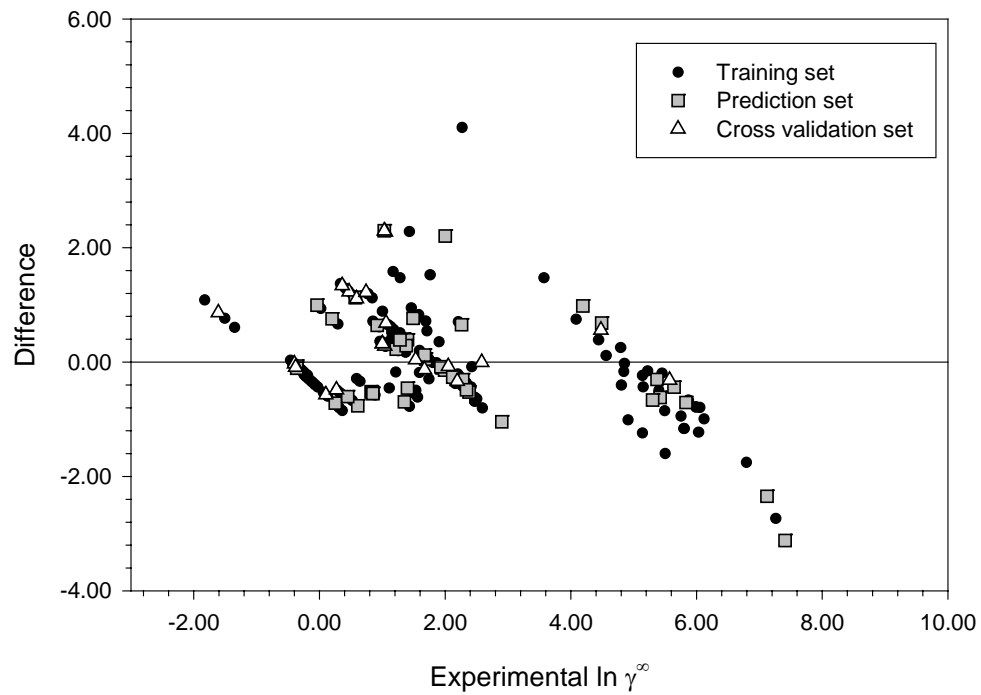


Figure I-6. Difference Plot of CS2 Case Study (Type II)

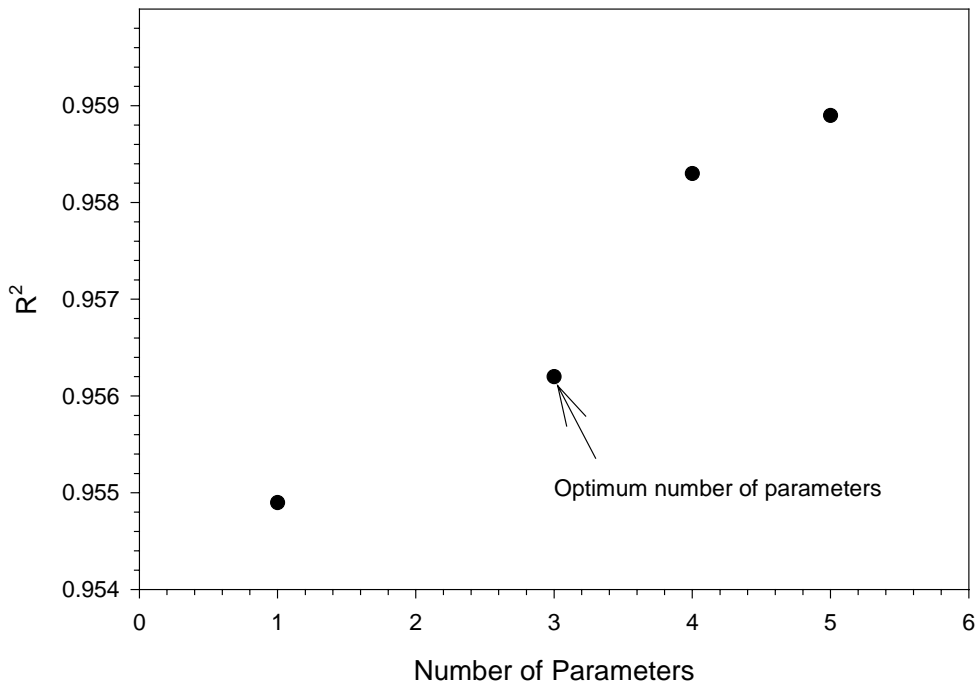


Figure I-7. R² Plot for Number of Parameters Determination (CS3 Case Study)

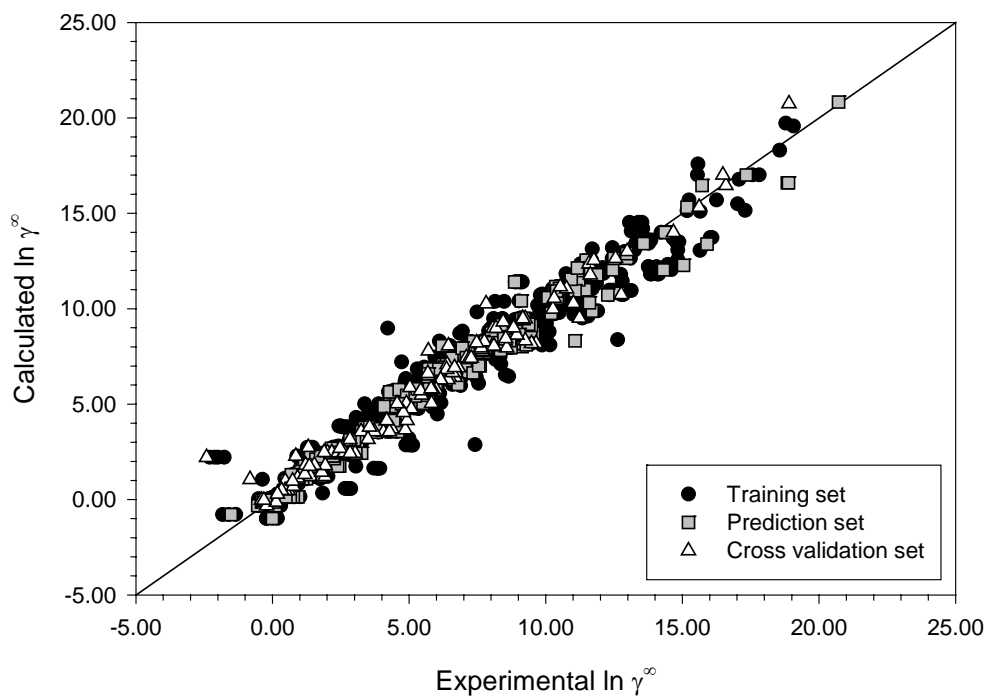


Figure I-8. Infinite-Dilution Activity Coefficients of CS3 Case Study (Type II)

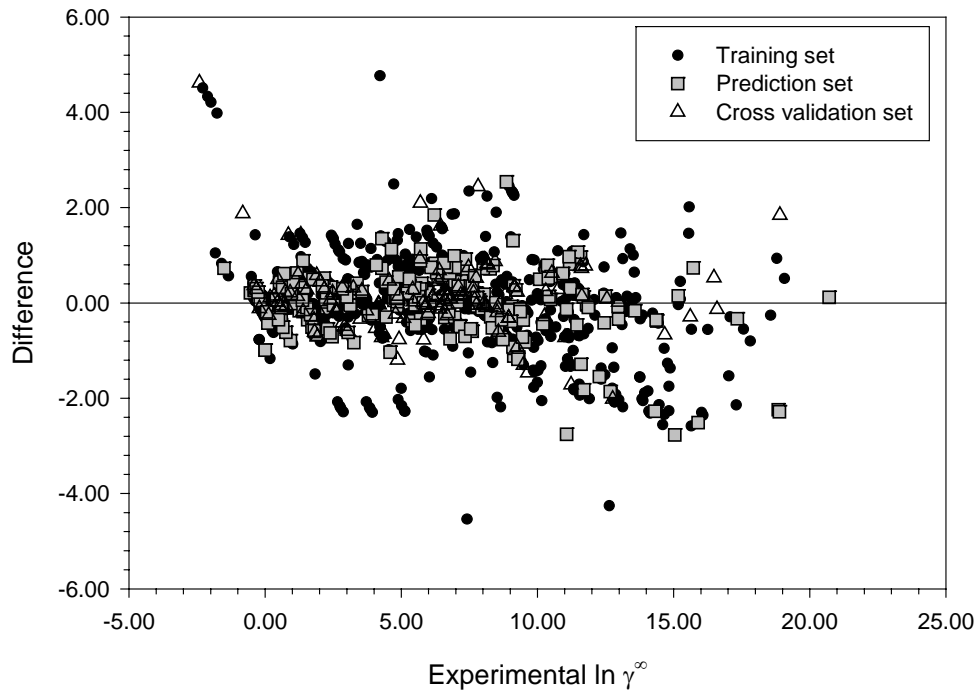


Figure I-9. Difference Plot of CS3 Case Study (Type II)

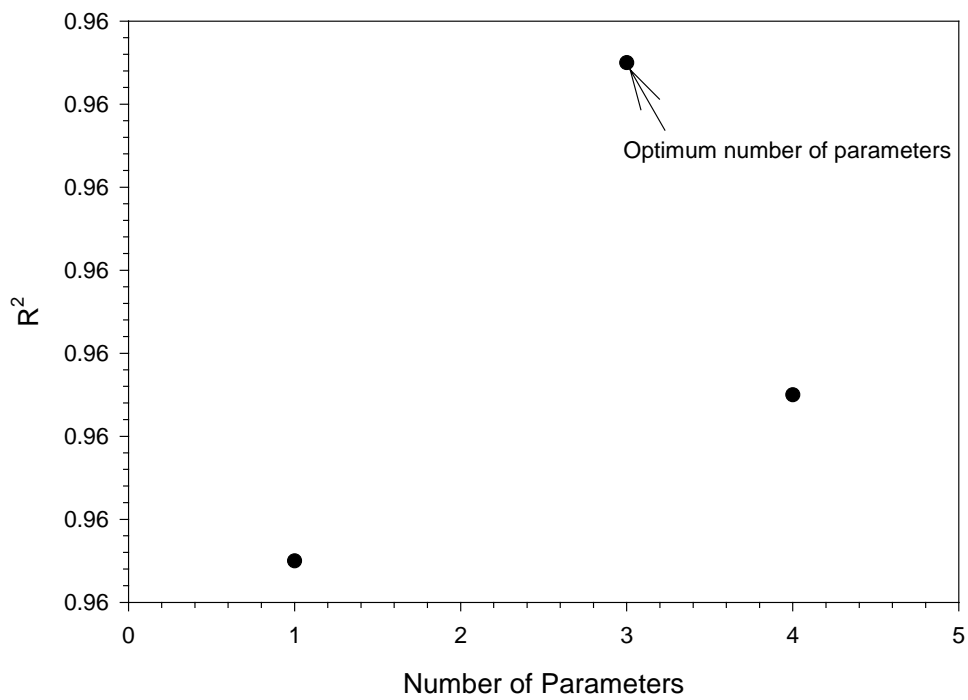


Figure I-10. R^2 Plot for Number of Parameters Determination (CS4 Case Study)

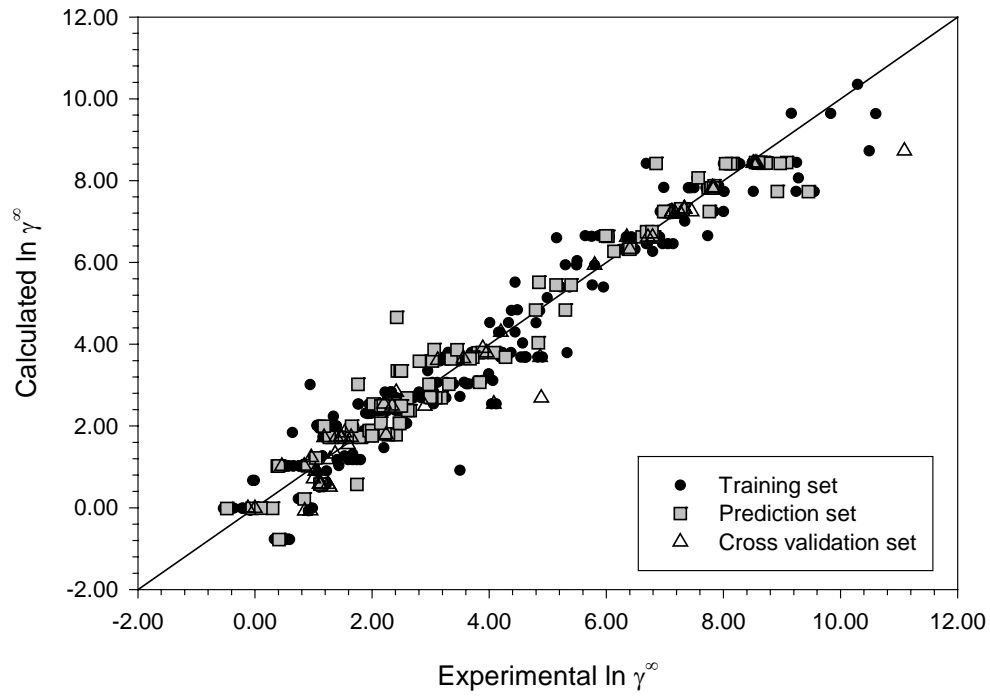


Figure I-11. Infinite-Dilution Activity Coefficients of CS4 Case Study (Type II)

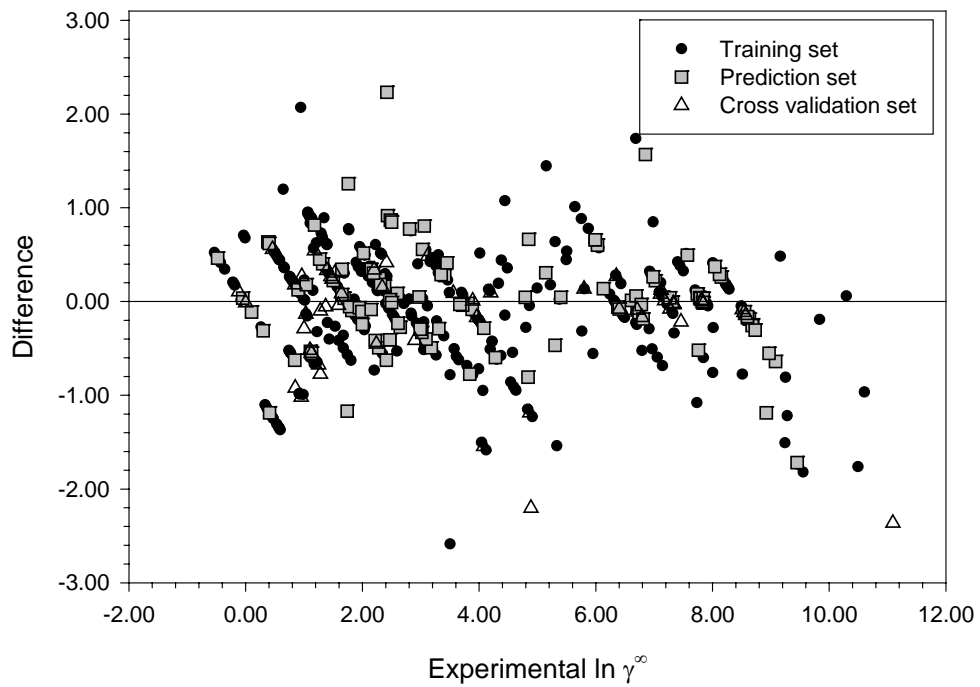


Figure I-12. Difference Plot of CS4 Case Study (Type II)

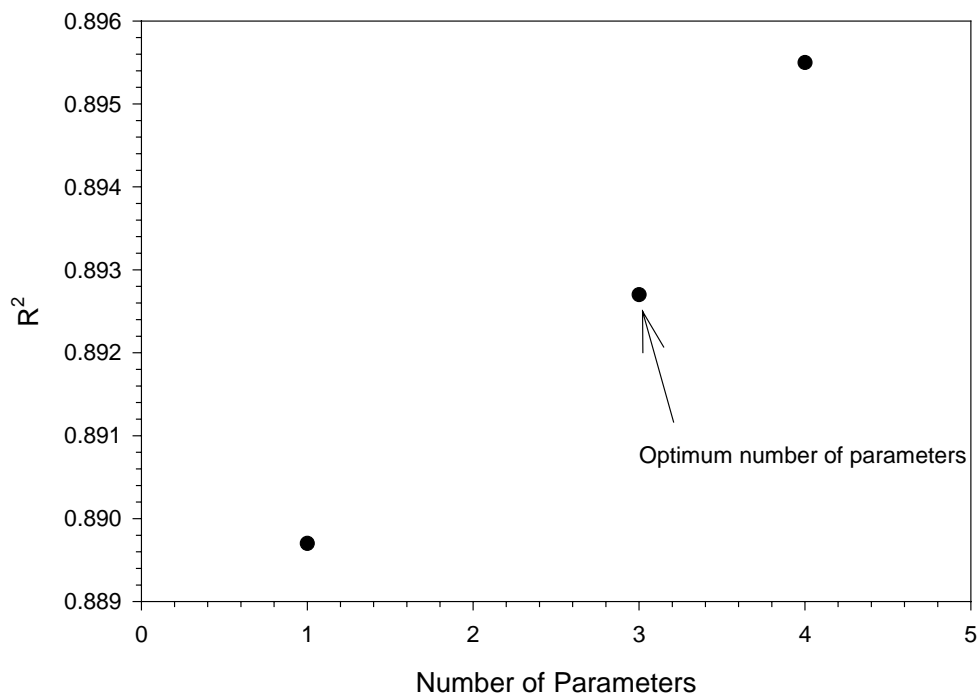


Figure I-13. R^2 Plot for Number of Parameters Determination (CS5 Case Study)

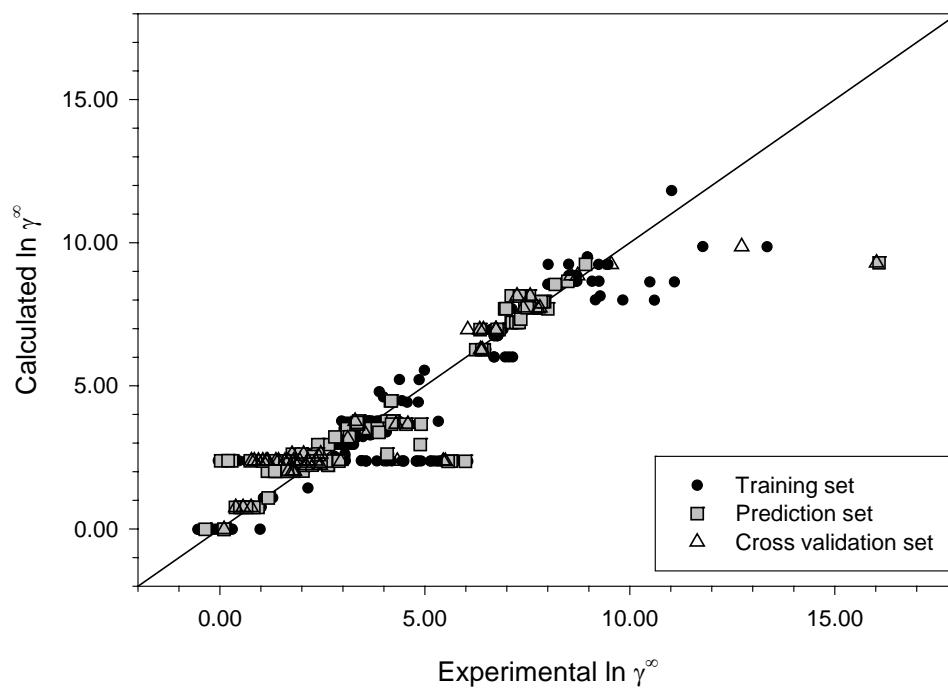


Figure I-14. Infinite-Dilution Activity Coefficients of CS5 Case Study (Type II)

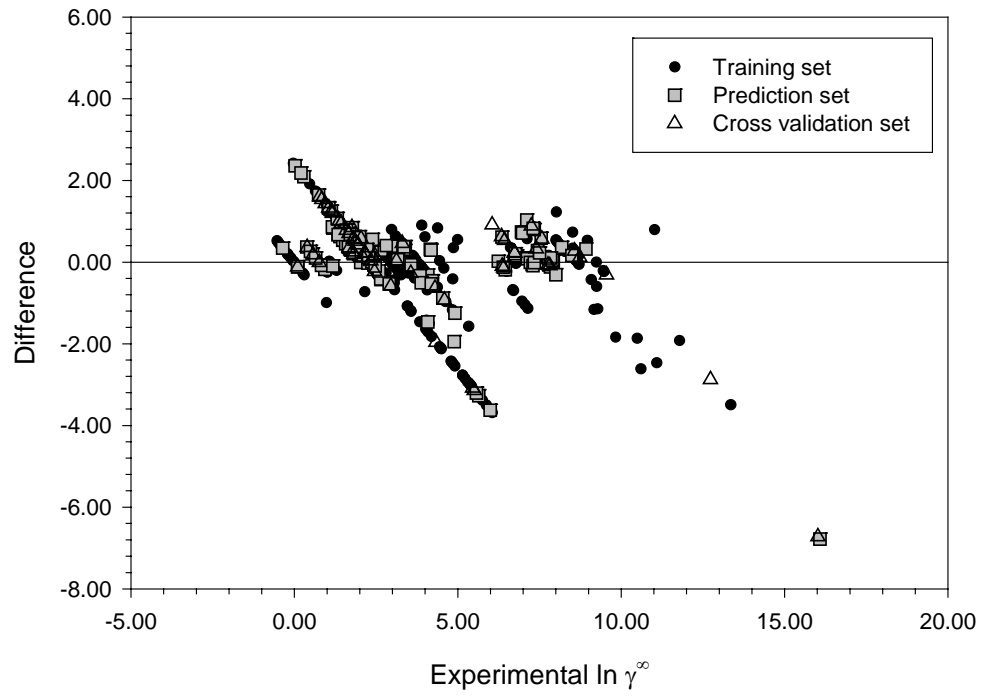


Figure I-15. Difference Plot of CS5 Case Study (Type II)

Appendix J. Type III Analysis

This appendix contains results from the Type III analyses, which are comprised of a general summary, a table of the final descriptor set used for each case study, contour plots used to determine the cessation point of training, and plots of both the calculated values of the infinite-dilution activity coefficients and the difference between the experimental and calculated values.

Table J-1. Summary of Type III Results

	CS1		CS2		CS3		CS4		CS5	
TYPE III										
R ²	0.991		0.965		0.992		0.984		0.949	
	Results in γ^∞									
	%AAD	RMSE	%AAD	RMSE	%AAD	RMSE	%AAD	RMSE	%AAD	RMSE
Training Set	30.2	0.45	28.6	0.35	30.4	0.39	23.3	0.39	52.8	0.65
Prediction Set	119.1	0.94	33.4	0.72	35.1	0.52	48.3	0.51	42.8	1.06
Cross Validation Set	71.0	0.82	37.1	0.88	31.7	0.57	33.1	0.44	44.2	1.29
	Results in $\ln \gamma^\infty$									
	%AAD		%AAD		%AAD		%AAD		%AAD	
Training Set	7.6		28.5		13.4		14.1		34.3	
Prediction Set	16.4		25.9		12.9		19.5		24.8	
Cross Validation Set	13.8		36.6		12.4		17.6		34.1	

Table J-2. Descriptors Used in the Type III Analyses

Type III Descriptors				
CS1	CS2	CS3	CS4	CS5
FNSA-3 Fractional PNSA (PNSA-3/TMSA) [Quantum-Chemical PC]	Average Complementary Information content (order 0)	HA dependent HDCA-2/SQRT (TMSA) [Quantum-Chemical PC]	DPSA-3 Difference in CPSAs (PPSA3-PNSA3) [Quantum-Chemical PC]	HA dependent HDCA-2/TMSA [Quantum-Chemical PC]
FG13	BP	FGorg	BP	Number of O atoms
exch. eng. + e-e rep. for a C-O bond	DPSA-3 Difference in CPSAs (PPSA3-PNSA3) [Zefirov's PC]	FG21	count of H-donors sites [Quantum-Chemical PC]	HA dependent HDSA-1 [Zefirov's PC]
FG17	FG17	FG24	FG	FG
BP	Final heat of formation / # of atoms	(Gravitational Index) ^{0.33}	Average Information content (order 0)	1X GAMMA polarizability (DIP)
HA dependent HDCA-2/TMSA [Zefirov's PC]	HA dependent HDCA-1/TMSA [Zefirov's PC]	HA dependent HDCA-1 [Quantum-Chemical PC]	FG14	Gravitation index (all bonds)
HACA-2/TMSA [Zefirov's PC]	HACA-1 [Quantum-Chemical PC]	Complementary Information content (order 0)	FG32	Topographic electronic index (all bonds) [Zefirov's PC]
Min partial charge for a C atom [Zefirov's PC]	PPSA-3 Atomic charge weighted PPSA [Quantum-Chemical PC]	WNSA-1 Weighted PNSA (PNSA1*TMSA/1000) [Quan. Chem. PC]	FPSA-3 Fractional PPSA (PPSA-3/TMSA) [Zefirov's PC]	WNSA-1 Weighted PNSA (PNSA1*TMSA/1000) [Quan. Chem. PC]
Max n-n repulsion for a C-H bond	HACA-2/TMSA [Zefirov's PC]	Kier&Hall index (order 1)	Tot molecular 1-center E-N attraction	Max resonance energy for a C-O bond
Min atomic state energy for a C atom	logP	LUMO energy	HACA-2 [Zefirov's PC]	Max SIGMA-PI bond order
Min e-e repulsion for a O atom	Max resonance energy for a C-Cl bond	LUMO+1 energy	HACA-2/TMSA [Zefirov's PC]	Max total interaction for a C-Cl bond
Min nucleoph. react. index for a F atom	Min e-n attraction for a O atom	Max total interaction for a H-O bond	Max SIGMA-PI bond order	Min nucleoph. react. index for a Cl atom
logP	Min net atomic charge for a O atom	Min net atomic charge	Internal entropy (300K)	Min total interaction for a C-C bond
Min resonance energy for a Br-C bond	Min partial charge for a O atom [Zefirov's PC]	Min n-n repulsion for a C-O bond	Kier&Hall index (order 2)	min(#HA, #HD) [Zefirov's PC]
MP	Polarity parameter (Qmax-Qmin)	No. of occupied electronic levels	logP	Kier&Hall index (order 0)
Randic index (order 1)	HACA-2/SQRT(TMSA) [Zefirov's PC]	Randic index (order 0)	LUMO+1 energy	HA dependent HDCA-1 [Quantum-Chemical PC]
PNSA-1 Partial negative surface area [Zefirov's PC]	RNCS Relative neg. charged SA (SAMNEG*RNCG) [Quan.-Chem. PC]	TMSA Total molecular surface area [Zefirov's PC]	Image of the Onsager-Kirkwood solvation energy	RPCG Relative positive charge (QMPOS/QTPLUS) [Zefirov's PC]
Relative number of O atoms	Topographic electronic index (all bonds) [Zefirov's PC]	Tot molecular 2-center resonance energy	Molecular volume	Tot hybridization comp. of the molecular dipole
Vib enthalpy (300K)	Tot heat capacity (300K) / # of atoms	Vib heat capacity (300K)	Tot entropy (300K)	HACA-2/TMSA [Zefirov's PC]
TMSA Total molecular surface area [Zefirov's PC]	Translational entropy (300K) / # of atoms	Kier&Hall index (order 0)	HA dependent HDCA-2 [Quantum-Chemical PC]	WNSA-3 Weighted PNSA (PNSA3*TMSA/1000) [Zefirov's PC]

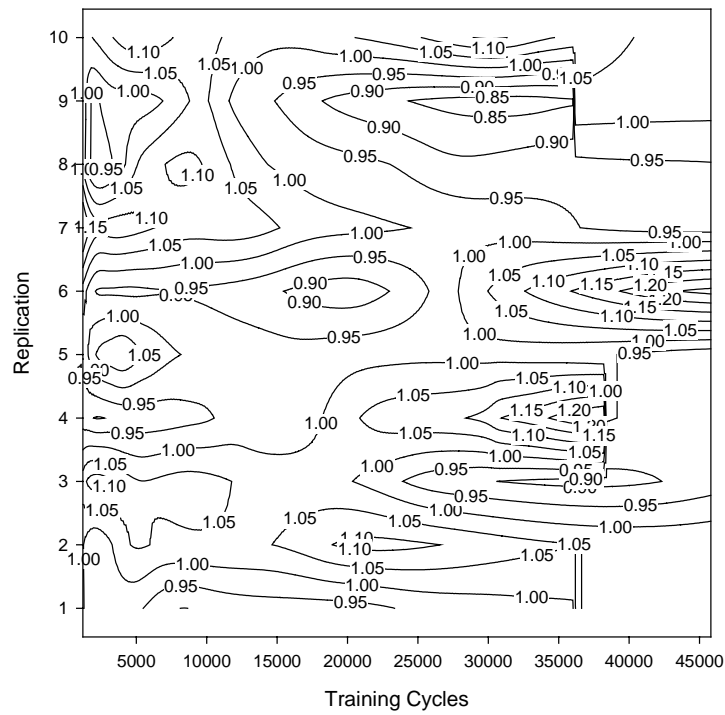


Figure J-1. Contour Plot of Cross Validation RMSE of CS1 Case Study

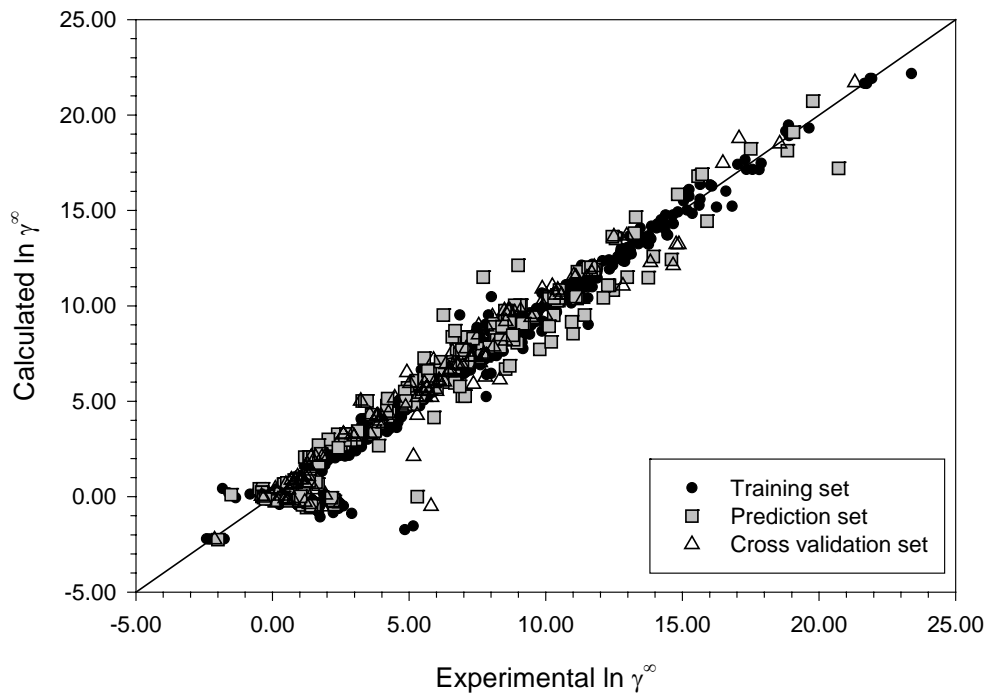


Figure J-2. Infinite-Dilution Activity Coefficients of CS1 Case Study (Type III)

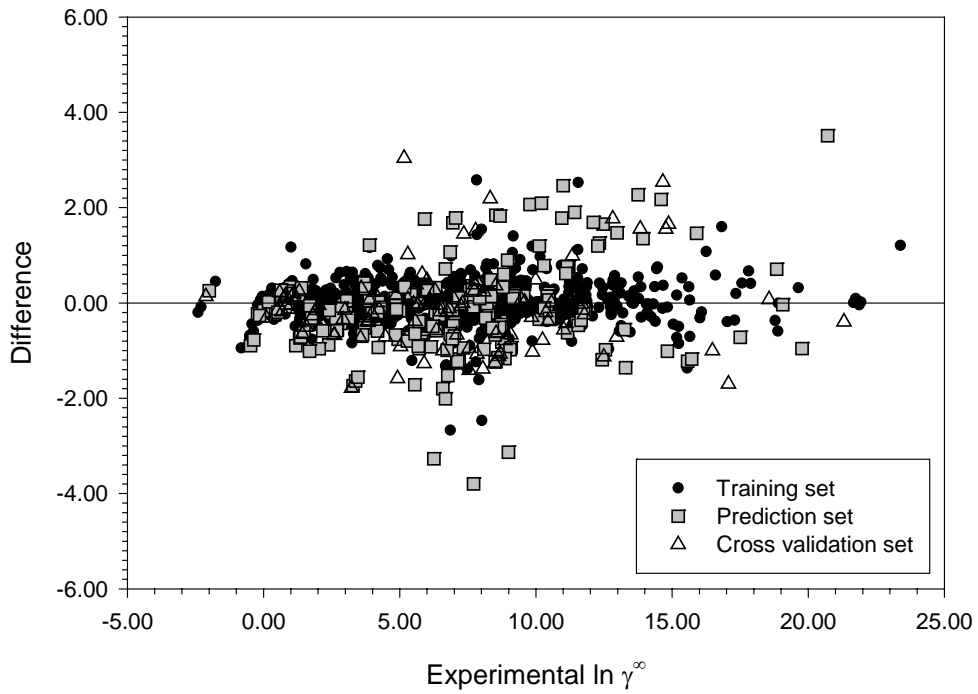


Figure J-3. Difference Plot of CS1 Case Study (Type III)

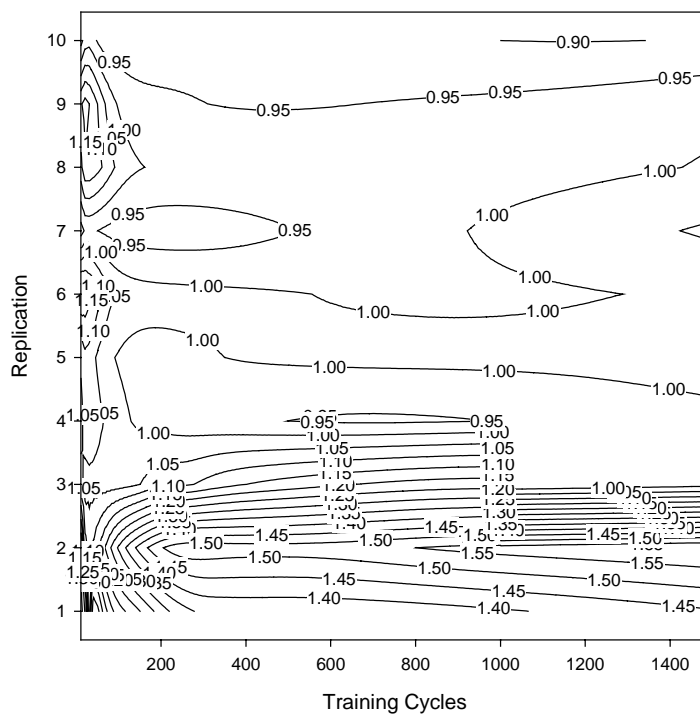


Figure J-4. Contour Plot of Cross Validation RMSE of CS2 Case Study

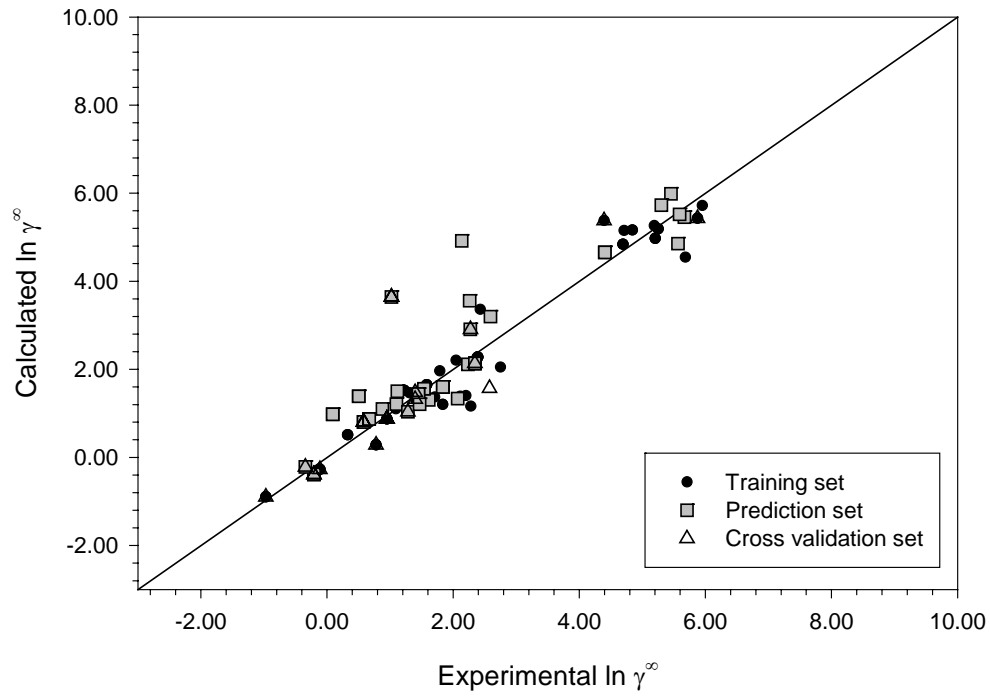


Figure J-5. Infinite-Dilution Activity Coefficients of CS2Case Study (Type III)

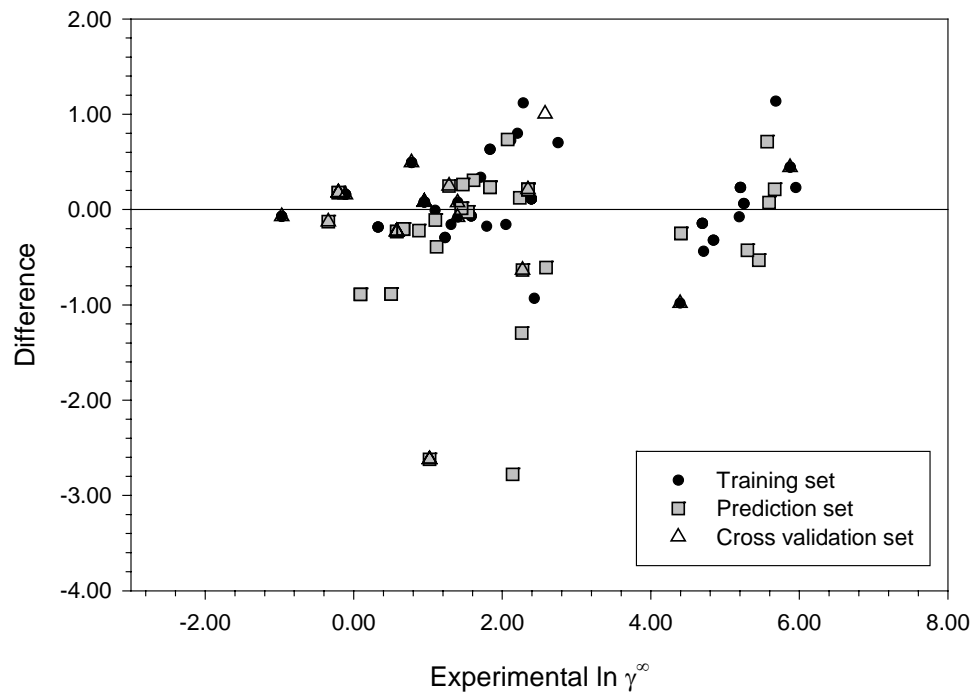


Figure J-6. Difference Plot of CS2Case Study (Type III)

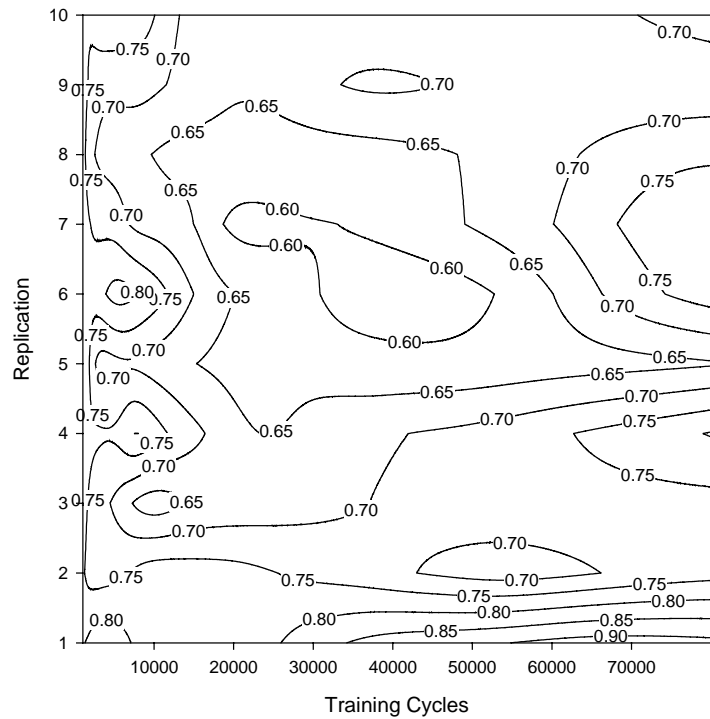


Figure J-7. Contour Plot of Cross Validation RMSE of CS3 Case Study

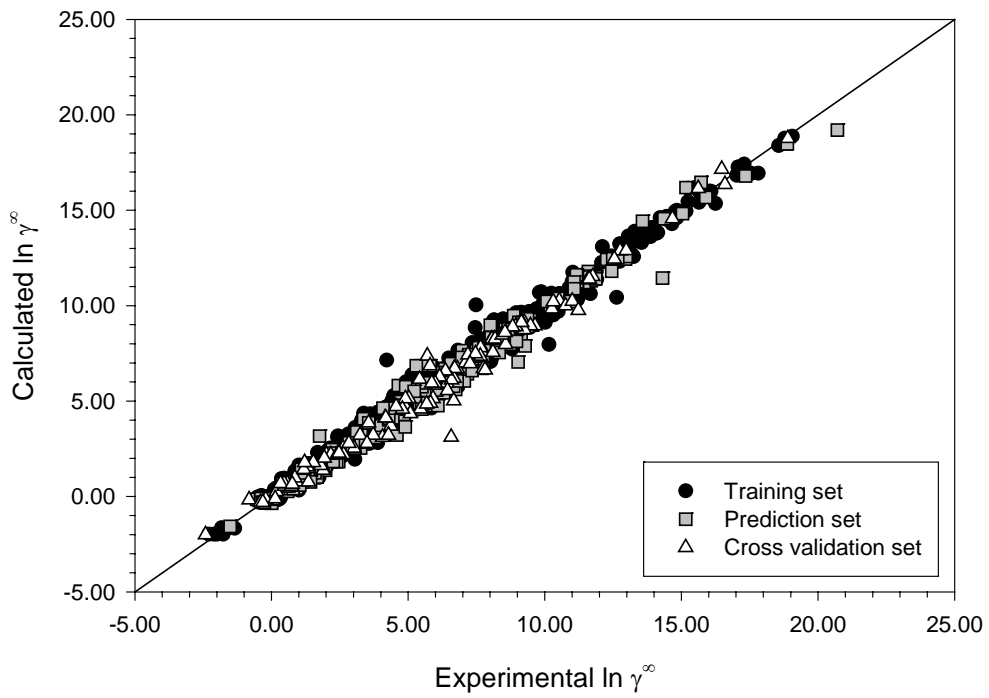


Figure J-8. Infinite-Dilution Activity Coefficients of CS3 Case Study (Type III)

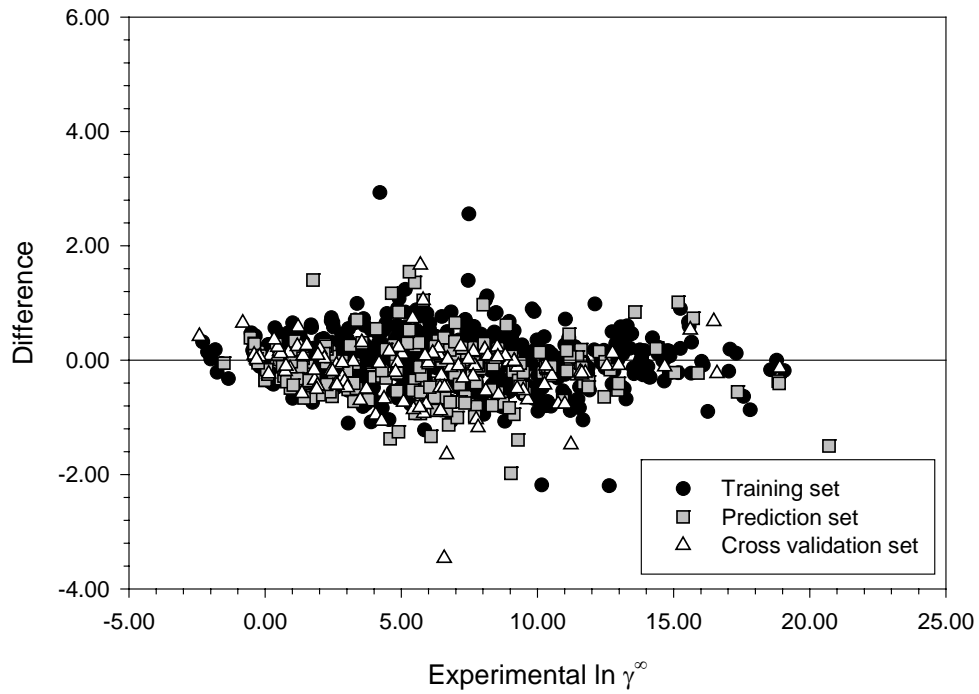


Figure J-9. Difference Plot of CS3 Case Study (Type III)

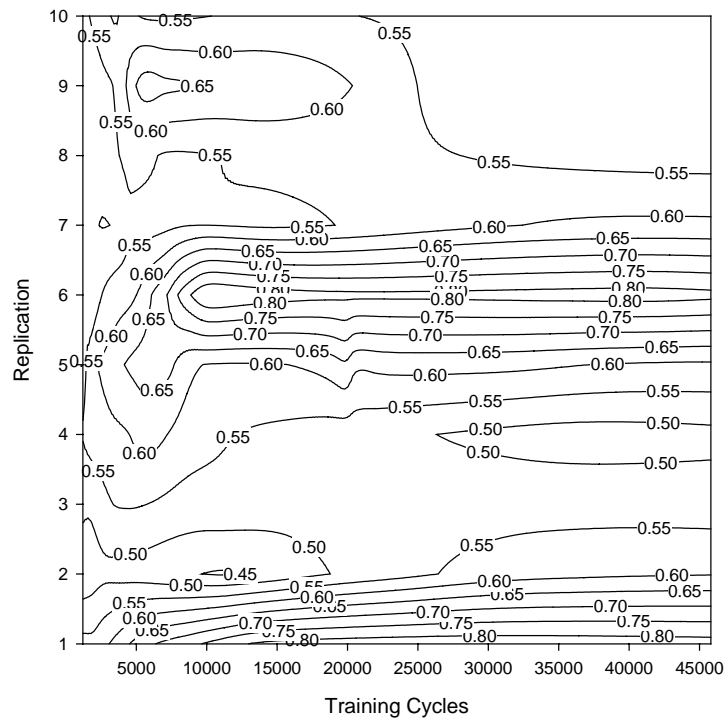


Figure J-10. Contour Plot of Cross Validation RMSE of CS4 Case Study

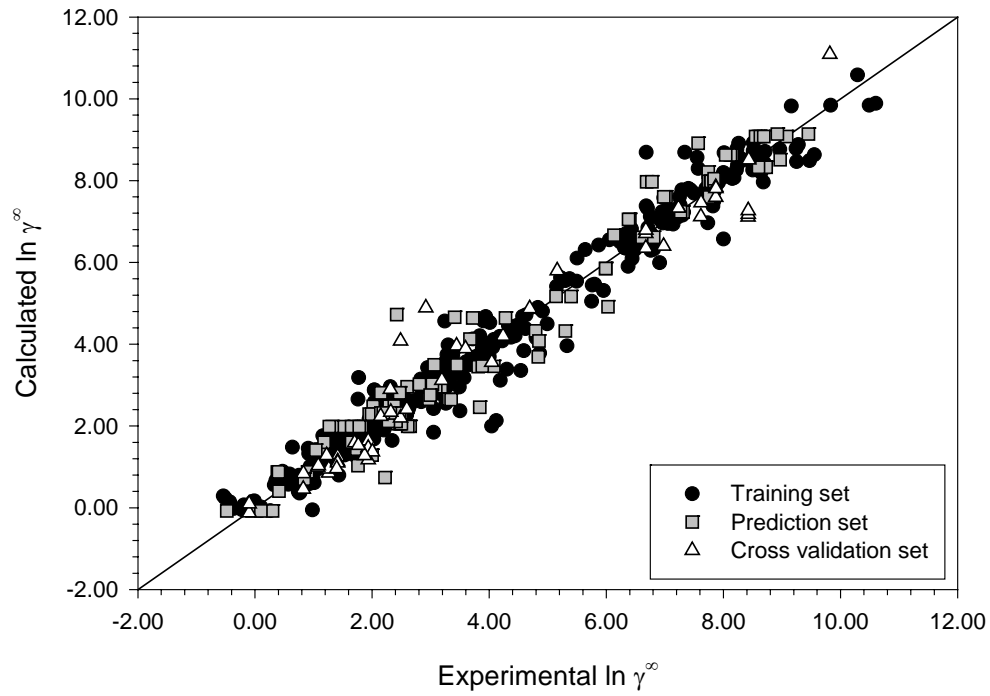


Figure J-11. Infinite-Dilution Activity Coefficients of CS4 Case Study (Type III)

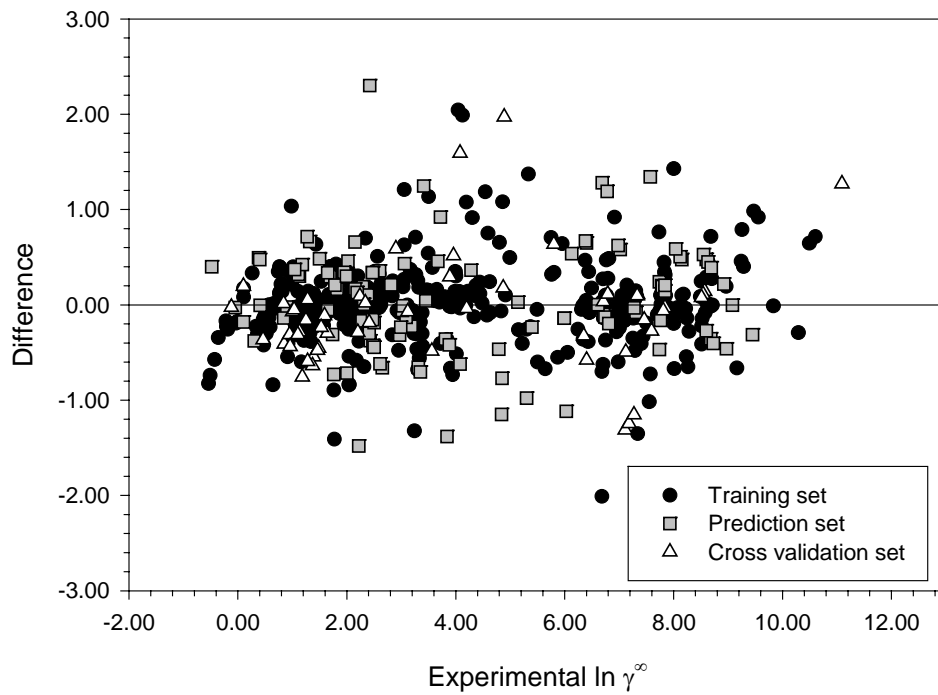


Figure J-12. Difference Plot of CS4 Case Study (Type III)

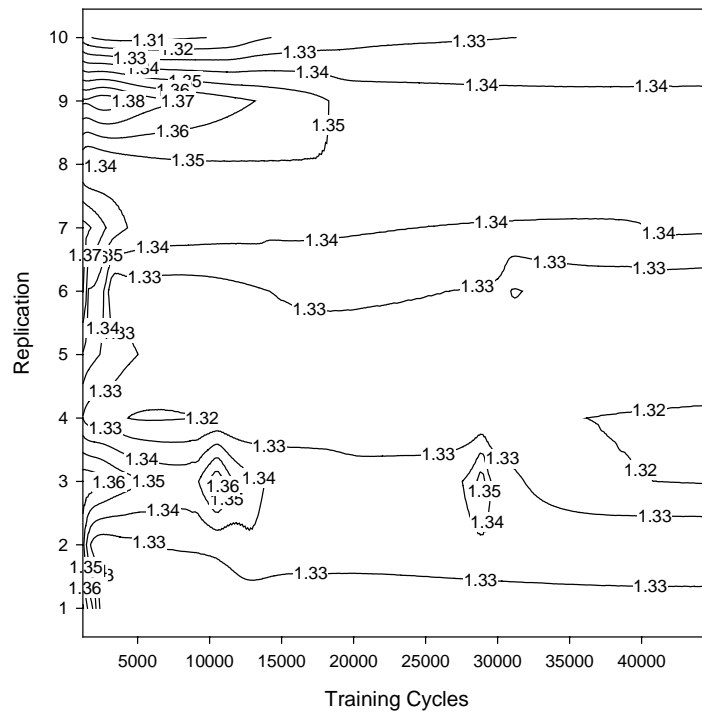


Figure J-13. Contour Plot of Cross Validation RMSE of CS5 Case Study

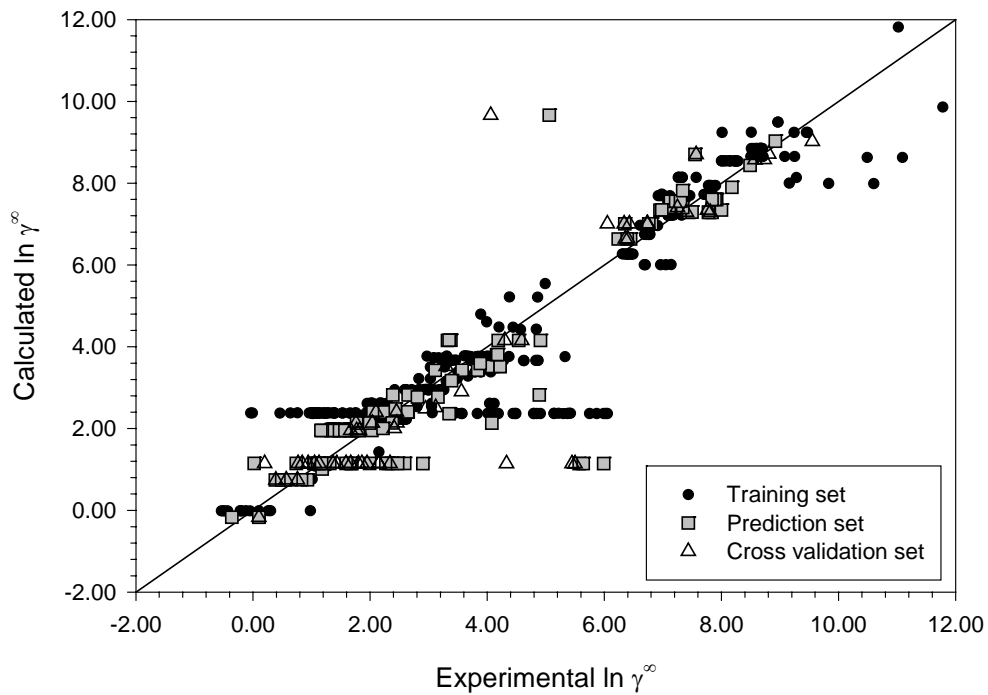


Figure J-14. Infinite-Dilution Activity Coefficients of CS5 Case Study (Type III)

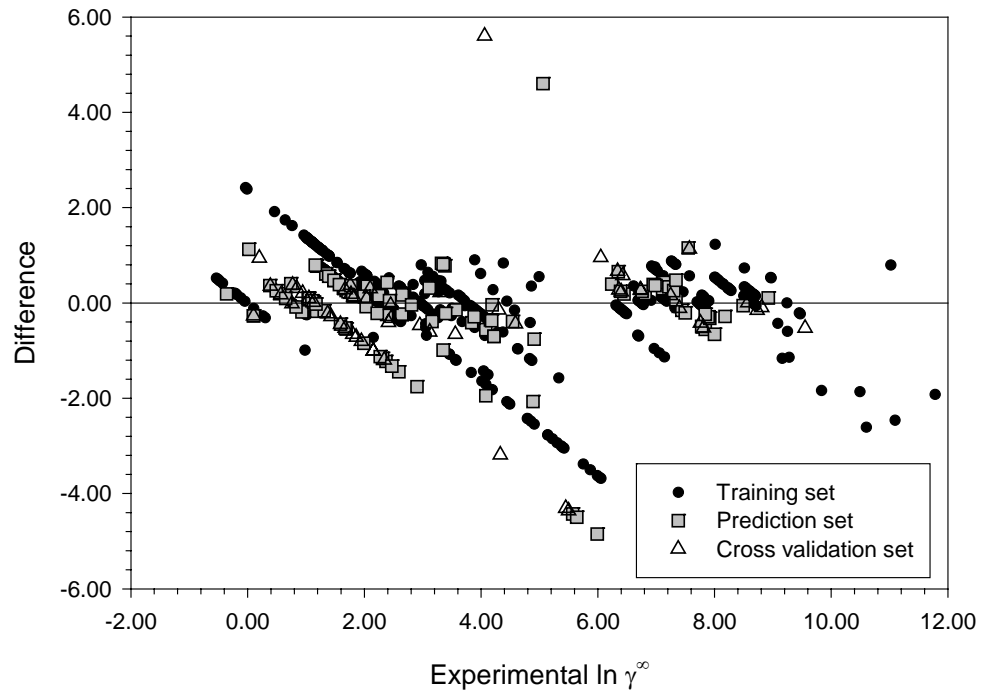


Figure J-15. Difference Plot of CS5 Case Study (Type III)

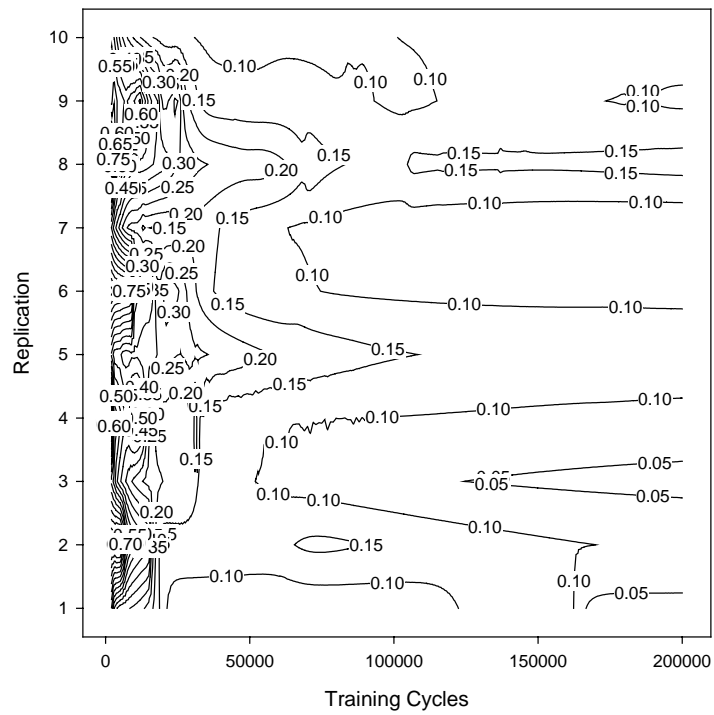


Figure J-16. Contour Plot of C_{12} from Bader-Gasem EOS Case Study

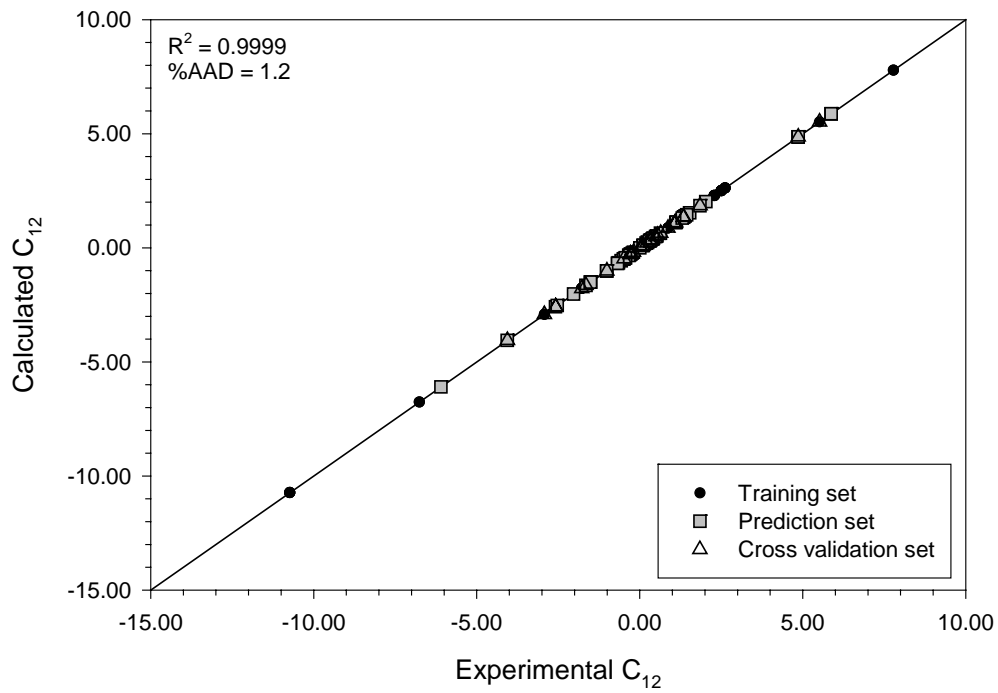


Figure J-17. C_{12} Parameter from Bader-Gasem EOS Case Study

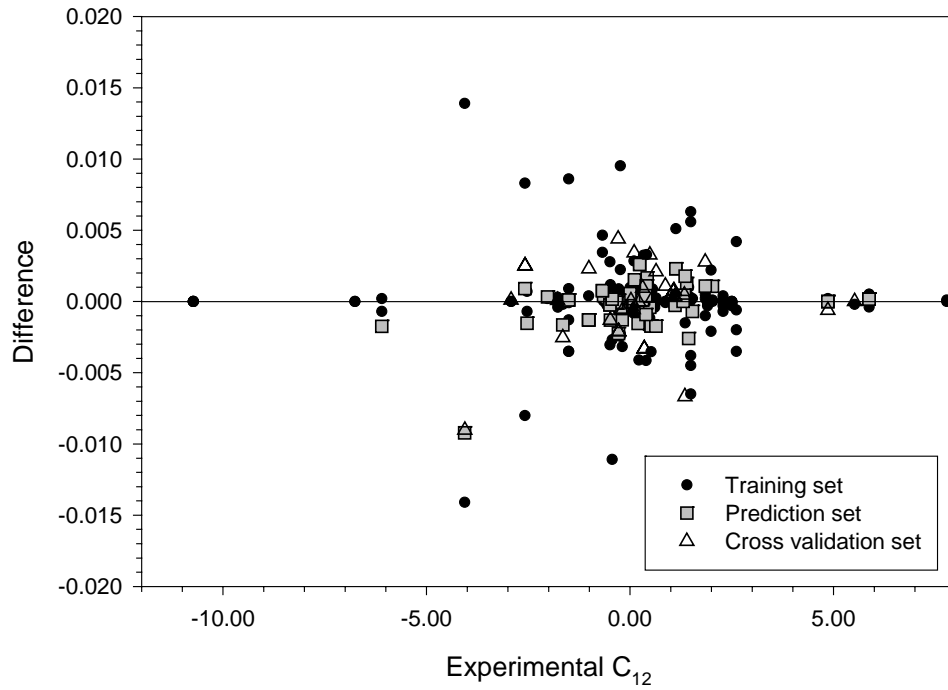


Figure J-18. Difference Plot of C_{12} from Bader-Gasem EOS Case Study

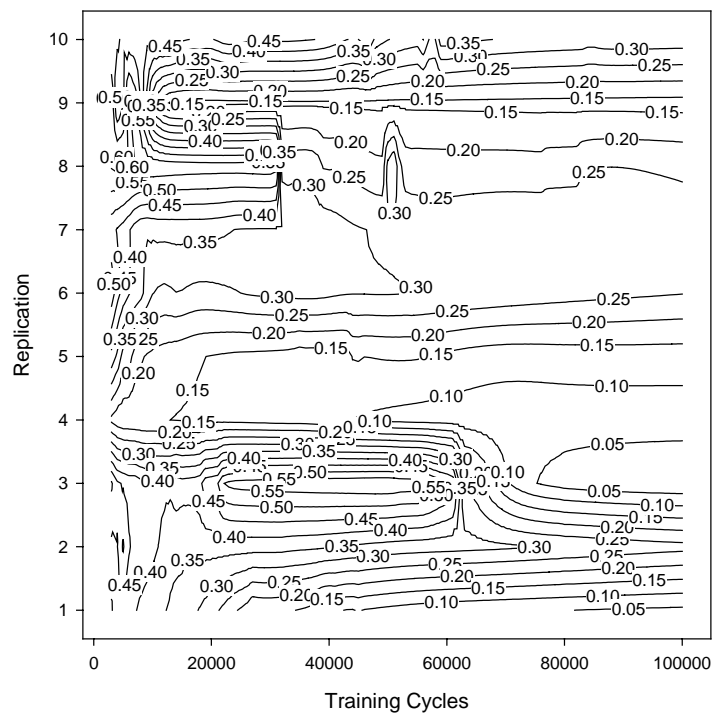


Figure J-19. Contour Plot of D_{12} from Bader-Gasem EOS Case Study

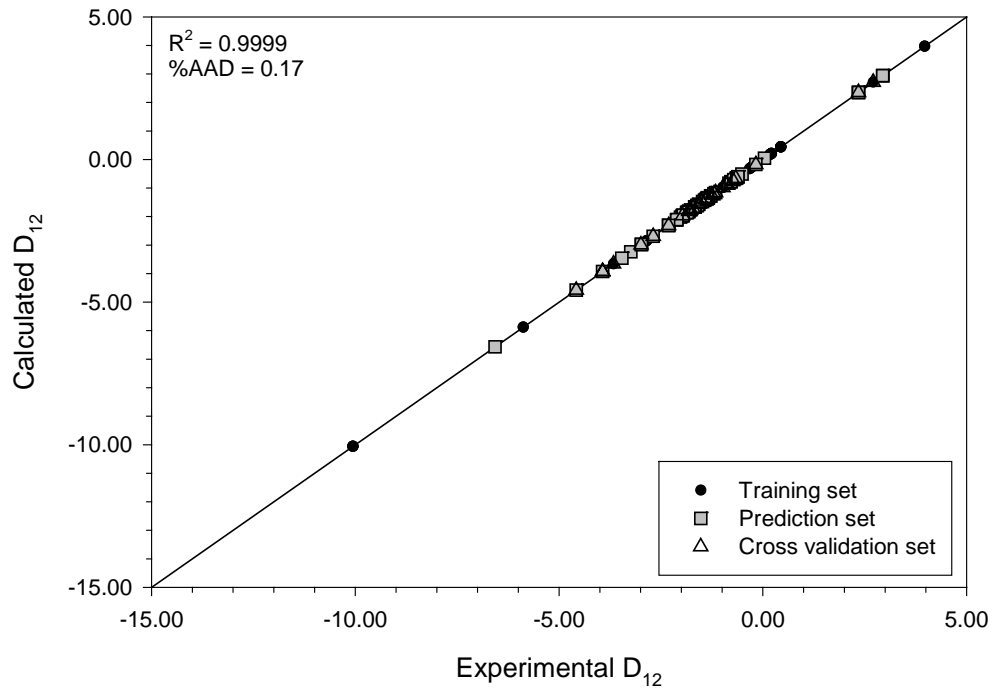


Figure J-20. D_{12} Parameter from Bader-Gasem EOS Case Study

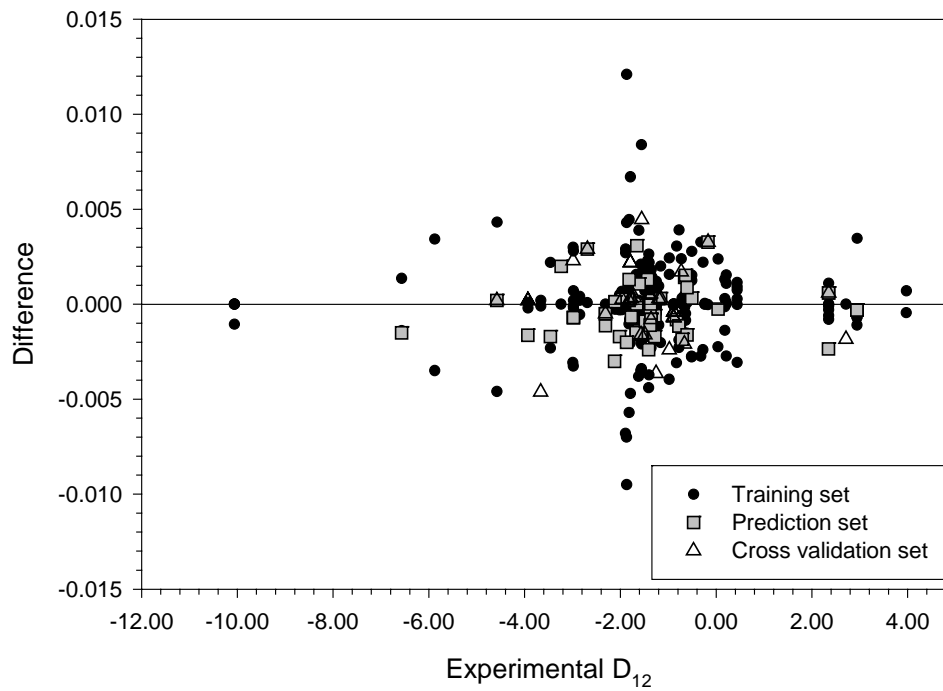


Figure J-21. Difference Plot of D_{12} from Bader-Gasem EOS Case Study

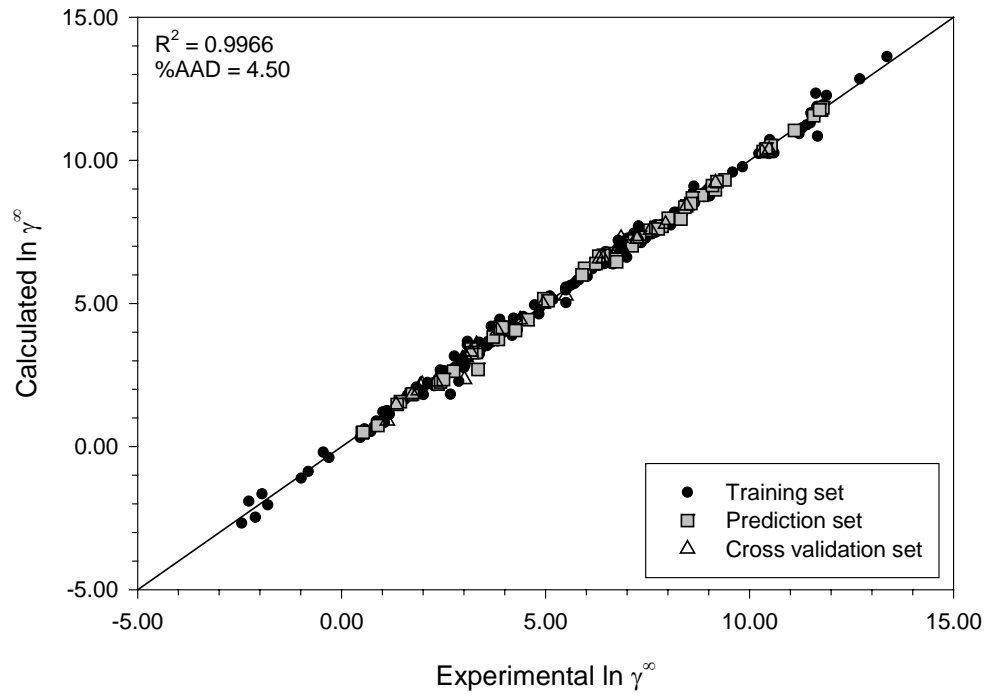


Figure J-22. Infinite-Dilution Activity Coefficients of Bader-Gasem EOS Case Study

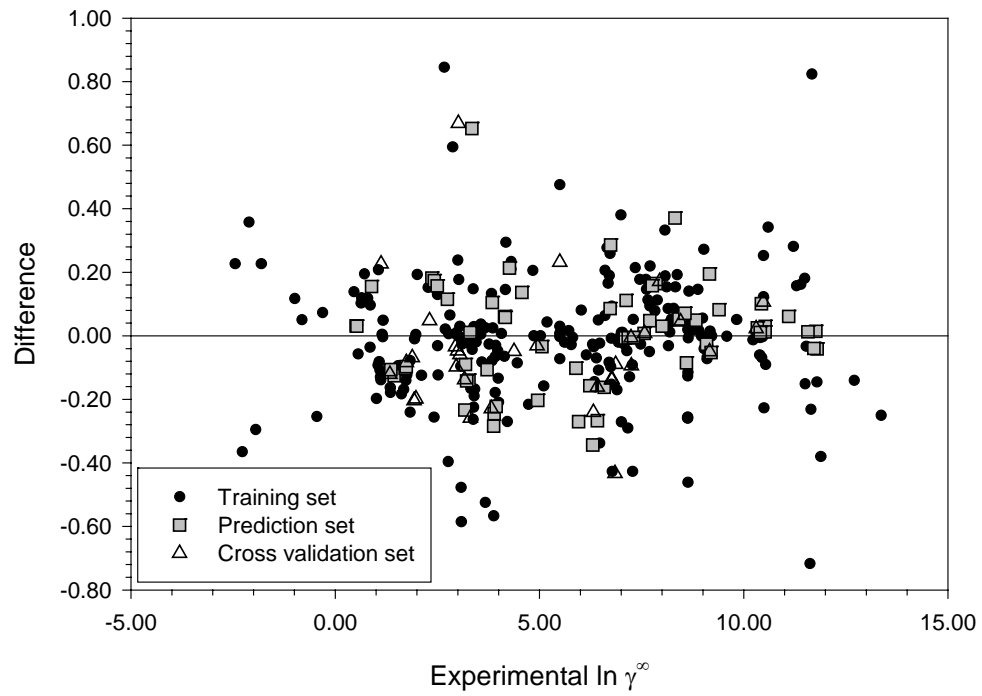


Figure J-23. Difference Plot of Bader-Gasem EOS Case Study

VITA

Brian J. Neely

Candidate for the Degree of

Doctor of Philosophy

Thesis: AQUEOUS HYDROCARBON SYSTEMS: EXPERIMENTAL
MEASUREMENTS AND QUANTITATIVE STRUCTURE-PROPERTY
RELATIONSHIP MODELING

Major Field: Chemical Engineering

Biographical:

Personal Data: Born in Minot, North Dakota, on March 16, 1965, the son of Jerry J. and Bonnie J. Neely.

Education: Graduated from Rye High School, Rye, Colorado in May 1983; received Associate of Science degree in Chemistry from Ricks College, Rexburg, Idaho in May 1985; received Bachelor of Science degree in Biochemistry with a Psychology minor from Brigham Young University, Provo, Utah in May 1990; received Master of Science degree in Chemical Engineering from Oklahoma State University. Completed the requirements for the Doctor of Philosophy degree in Chemical Engineering at Oklahoma State University, Stillwater, Oklahoma in May 2007.

Experience: Employed by Brigham Young University as a research assistant, 1987-1990; employed by American Veterinary Products, Ft. Collins, Colorado as a formulator, 1990-1993; employed by Oklahoma State University, School of Chemical Engineering as a graduate teaching assistant, 1993, 2000-2006; employed by Oklahoma State University, School of Chemical Engineering as a graduate research assistant, 1993-present.

Professional Memberships: American Institute of Chemical Engineers

Name: Brian J. Neely

Date of Degree: May, 2007

Institution: Oklahoma State University

Location: Stillwater, Oklahoma

Title of Study: AQUEOUS HYDROCARBON SYSTEMS: EXPERIMENTAL MEASUREMENTS AND QUANTITATIVE STRUCTURE-PROPERTY RELATIONSHIP MODELING

Pages in Study: 298

Candidate for the Degree of Doctor of Philosophy

Major Field: Chemical Engineering

Scope and Method of Study: The experimental objectives of this work were to (a) evaluate existing mutual hydrocarbon-water liquid-liquid equilibrium (LLE) data, and (b) develop an experimental apparatus capable of measuring accurately the hydrocarbon-water (LLE) mutual solubilities. The hydrocarbon-water systems studied included benzene-water, toluene-water, and 3-methylpentane water. The modeling efforts in this study focused on developing quantitative structure-property relationship (QSPR) models for the prediction of infinite-dilution activity coefficient values (γ_i^∞) of hydrocarbon-water systems. Specifically, case studies were constructed to investigate the efficacy of (a) QSPR models using multiple linear regression analyses and non-linear neural networks; and (b) theory-based QSPR model, where the Bader-Gasem activity coefficient model derived from a modified Peng-Robinson equation of state (EOS) is used to model the phase behavior, and QSPR neural networks are used to generalize the EOS binary interaction parameters. The database used in the modeling efforts consisted of 1400 infinite-dilution activity coefficients at temperatures ranging from 283 K to 373 K.

Findings and Conclusions: A continuous flow apparatus was utilized to measure the LLE mutual solubilities at temperatures ranging from ambient to 500 K, which is near the three-phase critical end point of the benzene-water and toluene-water systems. The well-documented benzene-water system was used to validate the reliability of the sampling and analytical techniques employed. Generally, adequate agreement was observed for the benzene-water, toluene-water, and 3-methylpentane-water systems with literature data. An error propagation analysis for the three systems indicated maximum expected uncertainties of 4% and 8% in the water phase and organic phase solubility measurements, respectively. In general, the use of non-linear QSPR models developed in this work were satisfactory and compared favorably to the majority of predictive models found in literature; however, these model did not account for temperature dependence. The Bader-Gasem activity coefficient model fitted with QSPR generalized binary interactions was capable of providing accurate predictions for the infinite-dilution activity coefficients of hydrocarbons in water. Careful validation of the model predictions over the full temperature range of the data considered yielded absolute average deviations of 3.4% in $\ln \gamma_i^\infty$ and 15% in γ_i^∞ , which is about twice the estimated experimental uncertainty. This study provides valuable LLE mutual solubility data and further demonstrates the effectiveness of theory-framed QSPR modeling of thermophysical properties.

ADVISER'S APPROVAL: K. A. M. Gasem
