The Dissertation Committee for Hyungjoo Lee Certifies that this is the approved version of the following dissertation:

Estimation of In-situ Fluid Properties from the Combined Interpretation of Nuclear, Dielectric, Optical, and Magnetic Resonance Measurements

Committee:

Carlos Torres-Verdín, Supervisor

Hugh Daigle

Zoya Heidari

Ryosuke Okuno

Mark Raizen

Estimation of In-situ Fluid Properties from the Combined Interpretation of Nuclear, Dielectric, Optical, and Magnetic Resonance Measurements

by

Hyungjoo Lee

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December, 2018

Dedication

To my wife, Kit-Kwan Chiu, for her unconditional support and excruciating tolerance.

Acknowledgements

I would like to express my sincere gratitude to my advisor, Dr. Carlos Torres-Verdín for his consistent support and motivation. I am grateful to him for shaping my thought processes during my graduate studies. Special thanks to Dr. Hugh Daigle and Dr. Zoya Heidari for their guidance, encouragement and sharing of their laboratory equipment. The technical advances in this dissertation would not have been possible without them. To Dr. Mark Raizen and Dr. Ryosuke Okuno, I say thank you for the technical discussions, wonderful insights, and reviews of my dissertation.

The work reported in this dissertation was funded by The University of Texas at Austin's Research Consortium on Formation Evaluation, jointly sponsored by Afren, Anadarko, Apache, Aramco, Baker-Hughes, BG, BHP Billiton, BP, Chevron, ConocoPhillips, COSL, ENI, ExxonMobil, Halliburton, Hess, Maersk, Marathon Oil Corporation, Mexican Institute for Petroleum, Nexen, ONGC, OXY, Petrobras, PTT Exploration and Production, Repsol, RWE, Schlumberger, Shell, Statoil, Total, Weatherford, Wintershall, and Woodside Petroleum Limited.

Special thanks to Rey Casanova for his administrative support at the Formation Evaluation Research Consortium. I'm grateful to Frankie Hart and Amy Douglas Stewart for their help and support in the petroleum engineering department. To Roger Terzian, thank you for all your support and efforts toward the management of my desktop and cluster computers. I am appreciative of the Petroleum Engineering Department and College of Engineering for bestowing me a Graduate Student Professional Development Award in 2014.

I was fortunate to spend two summers as an intern both Schlumberger in Sugar Land, and BHP Billiton in Houston. I am immensely grateful to these companies for the internship opportunities and technical contributions toward the developments documented in this dissertation. Sincere thanks to John Rasmus and Alexander Kostin, for their support during my internships. Their expert contributions toward some of the advances in this research have added tremendous value to the dissertation. Additionally, special thanks to Hojoong Youn at Daewoo Shipbuilding and Marine Engineering (DSME). My prior work experiences at DSME were invaluable toward the advancement of my career. I am especially grateful to Borak Kim for leading me to science. I am also indebted to Sheng-Che Yang for his technical guidance for molecular dynamics simulations. My sincere gratitude is reserved for Dr. David Medellin Salas for his insights and clarifications for this research.

I would like to thank my friends, colleagues, and office mates: Oyinkansola Ajayi, Haryanto Adiguna, Mathilde Luycx, Siddharth Mishra, Shan Huang, Colin Schroeder, Joshua Bautista, Edwin Ortega, Tatyana Torskaya, Eva Vinegar, Chicheng Xu, Ben Voss, Paul Linden, Hamid Beik, Essi Kwabi, Vahid Shabro, Amir Frooqnia, Dipo Olukitibi, Rebecca Gao, Andy Popielski, Elton Ferreria, Philippe Marouby, Kanay Jerath, Bruce Klappauf, Joaquin Ambia Garrido, Ayaz Mehmani, Shaina Kelly, Wilberth Herrera, Adam McMullen, Mohamed Bennis, Daria Olszowska, Elsa Maalouf, Juan Diego Escobar Gomez, Lei Wang, Tianqi Deng, You Wang, Mohamed Albusairi, Mauro Palavecino, Abdollah Pour, Antoine Montaut, Olabode Ijasan, Jun Ma, Qinshan Yang, Pawel Matuszyk, Javier Miranda, Artur Posenato Garcia, Abhishek Bansal, Wei Li, David Gonzalez, Paul Sayar, Ruijia Wang, Marisa Jasso, Saurabh Tandon, Marwane Smaiti, Matt Montgomery, Vivek Ramachandrann, David Pardo, Angel Rodriguez, Camilo Gelvez, Cameron McLenaghan, Gabriel Gallardo Giozza, Ali Eghbali, Jingxuan

Liu, Naif Rubaie, James Carter, Xiaowen Zhang, Wen Pan, Pierre Aerens, and Valeriia Sobolevskaia.

Finally, I'm immensely grateful to my family for their support and patience throughout this journey. To Dongseok Lee, Samsoon Lee, Hyungbae Lee, LingKang Chiu, Adela Chiu, and Enoch Chiu, thank you so much for all your perseverance, prayers, and yearn for my success.

Estimation of In-Situ Fluid Properties from the Combined Interpretation of Nuclear, Dielectric, Optical, and Magnetic Resonance Measurements

Hyungjoo Lee, Ph.D.

The University of Texas at Austin, 2018

Supervisor: Carlos Torres-Verdín

During the last few decades, the quantification of hydrocarbon pore volume from borehole measurements has been widely studied for reservoir descriptions. Relatively less effort has been devoted to estimating in-situ fluid properties because (1) acquiring fluid samples is expensive, (2) reservoir fluids are a complex mixture of various miscible and non-miscible phases, and (3) they depend on environmental factors such as temperature and pressure. This dissertation investigates the properties of fluid mixtures based on various manifestations of their electromagnetic properties from the MHz to the THz frequency ranges. A variety of fluids, including water, alcohol, alkane, aromatics, cyclics, ether, and their mixtures, are analyzed with both laboratory experiments and numerical simulations.

A new method is introduced to quantify in-situ hydrocarbon properties from borehole nuclear measurements. The inversion-based estimation method allows depth-continuous assessment of compositional gradients at in-situ conditions and provides thermodynamically consistent interpretations of reservoir fluids that depend greatly on phase behavior. Applications of this interpretation method to measurements acquired in

two field examples, including one in a gas-oil transition zone, yielded reliable and verifiable hydrocarbon compositions.

Dielectric properties of polar liquid mixtures were analyzed in the frequency range from 20 MHz to 20 GHz at ambient conditions. The Havriliak-Negami (HN) model was adapted for the estimation of dielectric permittivity and relaxation time. These experimental dielectric properties were compared to Molecular Dynamics (MD) simulations. Additionally, thermodynamic properties, including excess enthalpy, density, number of hydrogen bonds, and effective self-diffusion coefficient, were computed to cross-validate experimental results. Properties predicted from MD simulations are in excellent agreement with experimental measurements.

The three most common optical spectroscopy techniques, i.e. Near Infrared (NIR), Infrared, and Raman, were applied for the estimation of compositions and physical properties of liquid mixtures. Several analytical techniques, including Principal Component Analysis (PCA), Radial Basis Functions (RBF), Partial Least-Squares Regression (PLSR), and Artificial Neural Networks (ANN), were separately implemented for each spectrum to build correlations between spectral data and properties of liquid mixtures. Results show that the proposed methods yield prediction errors from 1.5% to 22.2% smaller than those obtained with standard multivariate methods. Furthermore, the errors can be decreased by combining NIR, Infrared, and Raman spectroscopy measurements.

Lastly, the ^{1}H NMR longitudinal relaxation properties of various liquid mixtures were examined with the objective of detecting individual components. Relaxation times and diffusion coefficients obtained via MD simulations for these mixtures are in agreement with experimental data. Also, the ^{1}H-^{1}H dipole-dipole relaxations for fluid mixtures were decomposed into the relaxations emanate from the intramolecular and

intermolecular interactions. The quantification of intermolecular interactions between the same molecules and different molecules reveals how much each component contributes to the total NMR longitudinal relaxation of the mixture as well as the level of interactions between different fluids.

Both experimental and numerical simulation results documented in this dissertation indicate that selecting measurement techniques that can capture the physical property of interest and maximize the physical contrasts between different components is important for reliable and accurate in-situ fluid identification.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

This dissertation introduces new methods to analyze the in-situ properties of fluid mixtures using nuclear, dielectric, optical, and magnetic resonance measurements. These measurements are governed by different physics phenomena and use different ranges of electromagnetic radiations from MHz to THz. Fluids with various degrees of complexity are analyzed with each of the techniques and their responses compared to assess which of them provides the best differentiation contrast. First, I develop a thermodynamically consistent linear inversion technique to quantify in-situ hydrocarbon compositions using borehole nuclear measurements. Second, the dielectric properties of various fluid mixtures are measured and cross-validated with Molecular Dynamics (MD) simulations. Third, spectra of Near-Infrared (NIR), Infrared, and Raman optical measurements spectroscopies are acquired and compared for the estimation of compositions and properties of various fluid mixtures. Lastly, Nuclear Magnetic Resonance (NMR) properties of fluids are investigated with MD simulations to quantify the relaxation time originating from each hydrogen-based molecule.

1.1 BACKGROUND

Quantification of reservoir fluid properties is crucial for optimizing facilities and maximizing production. Hydrocarbons are the most complex fluids present in subsurface reservoirs and they are usually described in terms of density and viscosity. More detailed information including molecular compositions, Gas-Oil Ratio (GOR), bubble and dew points, and compressibility are related to Pressure, Volume, and Temperature (PVT) behavior. Therefore, it is necessary to consider phase behavior for reservoir fluid identification.

Conventional well logs are not adequate for fluid typing. Gamma-Ray (GR) logs are sensitive to mineral composition, while resistivity and SP logs are related to total pore volume and salinity of water. Sonic logs are sensitive to fluids present in rocks but the slowness response is dominated by the rock matrix. Although nuclear logs are responsive to both reservoir matrix and fluids, some of them are relatively insensitive to the rock matrix. Among the nuclear properties, migration length and Hydrogen Index (HI) can be good candidates for fluid characterizations because they are mainly affected by hydrogen atoms present in the fluids.

The Downhole Fluid Analyzer (DFA) tool was introduced to determine hydrocarbon composition (Mullins et al., 2004) in open-hole wireline logging. A major application of the DFA tool is the use of optical spectroscopy measurements at a limited number of discrete depths. Integration of DFA measurements with NMR measurements could improve the accuracy of fluid typing (Steene et al., 2012) in that NMR provides longitudinal relaxation time, T_1, transverse relaxation time, T_2, and mass diffusion coefficient (D), which can lead to the indirect estimation of hydrocarbon viscosity. Although NMR measurements are affected by mud filtrate invasion due to a shallow radial depth of investigation (DOI), NMR provides unique information about in-situ fluids because it only measures the magnetization of hydrogen nuclei present in fluids.

Dielectric measurements have been extensively studied for applications in the assessment of downhole (Hizem et al., 2008), water/hydrocarbon saturation (Chen et al., 2016), heavy oil characterization (Yaarubi et al., 2012; Heaton et al., 2012), and wettability (Garcia et al., 2018). Most of the dielectric studies focus on the time relaxation of polar liquids. Because the physics of dielectric measurements is based on polarization phenomena, dielectric measurements are favorable for the characterization of polar liquids. In particular, mixtures of water and alcohol are part of the simplest model

containing hydrophobic groups. They are completely miscible but exhibit non-ideal thermodynamic behaviors when present in fluid mixtures. Also, these mixtures exhibit nonlinear relationships of dielectric permittivity and relaxation time as a function of concentrations. Comprehensive dielectric measurements have been reported (Sato et al., 1999, 2003, and 2005) for water, methanol, and 2-propanol mixtures; the authors suggested that positive excess enthalpy of the mixtures at water-rich concentrations are related to a reduction of the local density of the hydrogen bond acceptor and donor sites.

Optical spectroscopy measurements have been used in the oil industry for the last two decades. The DFA tool was introduced in open-hole wireline logging operations to assess compositional grading in real time (Fujisawa et al., 2008). Recently, Raman spectroscopy was applied to evaluate the resource potential of reservoirs (Bryndzia et al., 2016). Downhole Reservoir Raman System (DRRS) contains a Raman spectrometer and sensor platform attached to a wireline tool that measures hydrocarbons dissolved in water. Attenuated Total Reflectance (ATR) Fourier Transform Infrared (FTIR) spectroscopy is also an effective technique for the assessment of crude oil in terms of °API values (Filgueiras et al., 2014). Contrary to NIR absorptions, Infrared and Raman bands are rich in information concerning optical vibrational stretching and bending modes. Also, compared to NIR broad peaks, Infrared and Raman spectra exhibit sharp peaks with less blending. Despite these advantages, Infrared and Raman spectroscopy measurements are not widely used in field applications because of thermal noise, fiber optical materials, and fluorescence. However, some laboratory attempts were made to compare these spectroscopic techniques to analyze heavy petroleum products (Chung et al., 2000), Benzene, Toluene, Ethylbenzene and Xylene (BTEX) in petroleum fuels (Cooper et al., 1997), diesel properties (Santos et al., 2005), and citrus oils (Schulz et al., 2002).

NMR time relaxations have been widely used in the oil industry to characterize the properties of reservoir fluids. These NMR properties have been extensively studied both experimentally and theoretically. Recently, MD simulations became a reliable method with improved computing power to obtain numerical solutions of the classical equations of motion. There have been many studies of NMR relaxations using MD simulations with RNA (Villa et al., 2006), proteins (Caballero-Manrique et al., 2007; Ollila et al., 2018), polymers (Markelov et al., 2015), lipids (Lindahl et al., 2001), and hydrocarbons (Singer et al., 2018). Recent studies for short alkanes (Singer et al., 2017a) and polymer and heptane mixtures (Singer et al., 2017b) indicated that the relaxation process is mainly governed by both intramolecular and intermolecular dipole-dipole interactions between hydrogen atoms.

Along with these conventional well logs and spectroscopic measurements, various interpretation techniques are typically applied to quantify the properties and concentrations of fluid mixtures. The complementary physics of different measurements can improve the evaluation of fluid properties.

1.2 PROBLEM STATEMENT

Characterizing in-situ fluid properties in reservoirs is important to prevent potential hazards, optimize production, and estimate hydrocarbon quality. A direct approach to analyzing hydrocarbon properties is to acquire fluid samples in boreholes using a wireline formation tester. This technique enable accurate and reliable estimations of in-situ hydrocarbon properties, but it is impossible to acquire samples from all reservoirs due to stability problems and financial constraints. SARA (Saturated, Aromatics, Resins, and Asphaltenes) analysis (Aske et al., 2001) is

another common method to separate oils into components. However, it normally takes two days to analyze the sample while leftover solvents are harmful to the environment.

Evaluation of fluid properties from well logs is highly uncertain because fluids are typically a mixture of multiple components. Also, different components such as alkanes and aromatics can provide very similar log responses, making the assessment non-unique; PVT conditions also affect the phase behavior of reservoir fluids, and their response can vary accordingly. Therefore, identifying hydrocarbon components based on conventional well logs may lead to incorrect assessments because of the thermodynamic behavior of reservoir fluids.

Dielectric measurements are widely used in downhole applications; their polarization mechanisms have been broadly studied. There have been several attempts to describe the dielectric properties at an atomic level with MD simulations. Some of the previous works could not reproduce experimental results and may need better simulation algorithms. Also, how the dipole moment and polarizability of individual liquid molecules are accumulated into the ensemble average of macroscopic dielectric properties needs to be quantitatively described.

The Beer-Lambert law is the most common way to assess optical spectral absorption measurements. It states that the transmittance of the material sample is related to its thickness, L, and to its absorbance, A, as

$$\log_{10}\left(\frac{I}{I_0}\right) = \sum_i \varepsilon_i c_i L \equiv A, \tag{1.1}$$

where I_0 is the incident intensity, I is the intensity after traversing the sample, ε_i is the

molar absorptivity of component i, and c_i is the concentration of the i-th component in the sample. However, the measured optical spectra of various hydrocarbons often exhibit nonlinear absorption for various reasons, whereby, traditional linear multivariate methods can give rise to unreliable estimations of fluid properties.

Magnetic resonance logging is broadly used in subsurface formations to determine its porosity and permeability. Previous works on NMR relaxations using MD simulations have successfully reproduced the available experimental data and calculated intramolecular and intermolecular interactions separately. However, no quantitative comparison has been published of dipole-dipole interactions between different types of hydrocarbons. Also, most fluid mixtures exhibit nonlinear relationships between NMR relaxation times and diffusion coefficients as a function of concentrations, and it is unclear how much each fluid contributes to the total time relaxation.

1.3 RESEARCH OBJECTIVES

The main purpose of the dissertation is to develop quantitative interpretation methods to estimate the compositions and properties of various in-situ fluid mixtures using nuclear, dielectric, optical, and magnetic resonance measurements. Secondary objectives of the dissertation are as follows:

- To develop an algorithm to estimate the concentrations of hydrocarbon components using nuclear measurements. To achieve this objective, inversion techniques are implemented separately on different combinations of nuclear measurements. In order to account for phase behavior of reservoir hydrocarbons,

a thermodynamically consistent interpretation of nuclear measurements needs to be considered to enable continuous assessments of fluid compositional gradients.

- To compare the viscosity calculated from the inversion of borehole nuclear measurements to the viscosity assessed from the NMR measurements. This comparison can cross-validate the hydrocarbon compositions estimated from borehole nuclear measurements.
- To develop nonlinear inversion algorithms to appraise dielectric permittivity and relaxation time for liquid mixtures. Forward modeling can use the Havriliak-Negami's (HN) equation, the Cole-Davidson's (CD) equation, or the Cole-Cole (CC) equation to reconstruct the real and imaginary parts of dielectric relaxations.
- To reproduce dielectric properties of fluids using MD simulations and compare the simulation results to inversion-based calculations. The MD simulations can compute various dielectric and thermodynamic microscopic properties to explain the macroscopic ensemble average of dielectric properties obtained from experiments.
- To build training and calibration models to appraise the concentrations and in-situ properties of liquid mixtures using optical spectroscopy measurements. These models can be constructed by implementing Radial Basis Functions (RBF) and Artificial Neural Networks (ANN). The Root-Mean-Square Errors (RMSE) associated with these methods are compared to the traditional multivariate method.
- To combine multiple optical spectroscopy measurements in order to improve the accuracy of in-situ fluid property estimations. Careful selection of spectral bands is required to avoid integrating redundant features of the spectra.

- To quantify intramolecular and intermolecular ^{1}H-^{1}H dipole-dipole interactions between binary mixtures of various liquids and to appraise their contributions to the total NMR relaxation process. Similarly, mass diffusion coefficients for each component are computed separately and compared to the effective mass diffusion coefficient of the liquid mixtures.
- To diagnose the advantages and limitations of each measurement and to identify which measurement provides the best contrast of physical properties of individual fluids present in a mixture.

1.4 OUTLINE OF THE DISSERTATION

This dissertation is composed of the current introductory chapter, four technical chapters, and a final chapter that summarizes the findings and provides conclusions and recommendations for future work.

The first technical part of the dissertation is Chapter 2. This chapter presents a thermodynamically consistent interpretation of nuclear measurements for the assessment of hydrocarbon compositions. Viscosities computed from one synthetic example and two field data yield good agreements to the viscosity derived from NMR measurements.

Chapter 3 discusses the dielectric and thermodynamic properties of polar liquids. These properties are investigated with laboratory measurements, while nonlinear inversions provide estimates of the dielectric permittivity and relaxation time of various fluid mixtures. The calculated properties are validated and described with MD simulation results.

In Chapter 4, three types of optical spectroscopy measurements of fluid mixtures are analyzed and compared. The main contribution is that the proposed methods provide

much smaller errors in the assessment of the compositional and in-situ properties of the samples compared to traditional multivariate methods. Estimations are further improved with the combination of multiple spectral measurements.

Chapter 5 describes the successful quantification of NMR time relaxations of liquid mixtures via MD simulations. The distance and angle between hydrogen atom pairs provide the quantification of relaxational contributions of each atom and molecule. A new finding obtained from this chapter is the identification of NMR relaxations of individual molecules which are hard to measure with experimental approaches.

Following the technical chapters, the last chapter summarizes the dissertation's contributions, conclusions, and recommendations for future research.

1.6 LIST OF PUBLICATIONS

Several journal and conference papers based on the research documented in this dissertation have been published or will be submitted for review. They are listed below.

1.6.1 Refereed Journal Publications

Lee, H. and Torres-Verdín, C., 2019a, Estimation of liquid mixture compositions from near-Infrared spectrum data using radial basis functions and artificial neural network, (to be submitted for publication).

Lee, H. and Torres-Verdín, C., 2019b, Comparison of NIR, Infrared, and Raman spectroscopies for the estimation of density and viscosity using PLSR and artificial neural network analysis, (to be submitted for publication).

Lee, H. and Torres-Verdín, C., 2019c, Investigation of dielectric and thermodynamic properties of polar liquids by dielectric spectroscopy and molecular dynamics simulation, (to be submitted for publication).

Lee, H., Medellin, D., and Torres-Verdín, C., 2019d, NMR relaxation times and diffusion coefficients of the liquid mixtures from molecular dynamics simulations, (to be submitted for publication).

1.6.2 Refereed Conference Proceedings

Lee, H. and Torres-Verdín, C., 2015, Thermodynamically consistent estimation of hydrocarbon composition from nuclear and magnetic resonance measurements, SPWLA Annual Conference, Long Beach, California, July 18-22.

Chapter 2: Thermodynamically Consistent Estimation of Hydrocarbon Composition from Nuclear and Magnetic Resonance Measurements[1]

Accurate assessment of hydrocarbon compositions is critical for optimizing oil and gas production. However, in-situ assessment of hydrocarbon properties is difficult because of environmental conditions and fluid sample quality. Fluid samples can be acquired using Wireline Formation Testers (WFT) for laboratory analysis, but the number of samples is limited to discrete depths. Reservoir fluid samples tested in the laboratory may not exhibit the same properties as in-situ samples because the phase behavior of hydrocarbons varies significantly with temperature and pressure. Thermodynamically consistent fluid interpretations are crucial for obtaining accurate estimates of in-situ hydrocarbon properties and composition.

I introduce a new method to quantify in-situ hydrocarbon properties from borehole nuclear measurements. These measurements are influenced by fluid and rock matrix. After separating matrix, water, and hydrocarbon effects on borehole nuclear measurements, I use the hydrocarbon-dependent portion of the nuclear response to assess composition. Applying the new hydrocarbon composition method to two field examples, including one in a gas-oil transition zone, yielded reliable and verifiable results. Oil-viscosity values derived from NMR T_2 distributions and WFT pressure data confirmed the estimated hydrocarbon composition. This estimation method allows continuous assessment of compositional gradients at in-situ conditions and provides thermodynamically consistent interpretations of reservoir fluids that depend greatly on

[1] This chapter is based on the following paper and I am a primary author.

Lee, H. and Torres-Verdín, C., 2015, Thermodynamically consistent estimation of hydrocarbon composition from nuclear and magnetic resonance measurements, SPWLA Annual Conference, Long Beach, California, July 18-22.

phase behavior. Combining nuclear and NMR measurements identifies variations within hydrocarbon columns, diagnoses reservoir connectivity, and facilitates optimized hydrocarbon production.

2.1 INTRODUCTION

Identifying fluid properties in reservoirs is important to prevent potential hazard, optimize production, and estimate hydrocarbon quality. The conventional approach to analyzing hydrocarbon properties is to acquire fluid samples directly from boreholes using a wireline formation tester. This technique provides accurate and reliable estimation of in-situ hydrocarbon properties, but it is not possible to acquire samples from all reservoirs due to stability problems and financial constraints. Properties of fluid samples may also differ from those of in-situ fluid because of mud filtrate contamination or temperature and pressure changes.

Petrophysical assessment of hydrocarbon properties is highly non-unique because the phase behavior of hydrocarbons can vary greatly with Pressure, Volume, and Temperature (PVT). Hydrocarbon typing based on triple- or quad-combo logs can be uncertain. Gamma-Ray (GR) response is mainly affected by the matrix, and electrical resistivity is determined by the pore volume of conductive formation water. Cross-over of bulk density (ϕ_b) and neutron porosity (ϕ_N) indicates the existence of hydrocarbon components. However, identifying hydrocarbon components solely on the basis of density and neutron cross-over may lead to incorrect assessments because of shale-suppressing effects.

To mitigate this uncertainty, I introduce thermodynamically consistent interpretations of hydrocarbon properties. Fluid type is assessed using PVT flash

calculations and the Schlumberger Nuclear PARameter calculation program (SNUPAR, McKeon and Scott, 1989). SNUPAR allows us to take into account the nuclear properties of hydrocarbons and water as a function of pressure, temperature, and salinity (Ortega et al., 2013). Various nuclear properties, including density (ρ), neutron capture cross-section (Σ), Hydrogen Index (HI), inverse of migration length (ξ), PEF, and neutron porosity (ϕ_N), were used for inversions.

This chapter describes a new method for estimating hydrocarbon properties from both nuclear and NMR measurements. **Figure 2.1** indicates that nuclear properties change between ethane and pentane. To maximize their differences, hydrocarbon components are classified into three pseudo-components, C_1, $C_{2\text{-}5}$, and C_{6+}. Linear inversion using these nuclear properties estimates the volumetric concentration of each hydrocarbon component. I tested both synthetic and field examples using various combinations of nuclear properties derived from SNUPAR to estimate hydrocarbon compositions. These synthetic and field examples illustrate how hydrocarbon properties derived from nuclear measurements can be reconciled with estimated complementary NMR measurements.

2.2 METHOD

Analyzing hydrocarbon properties begins with constructing a hydrocarbon library. Computer Modeling Group (CMG) software can calculate phase behavior and corresponding properties of hydrocarbon mixtures at various temperatures and pressures. Peng-Robinson's Equation-Of-State (EOS) was implemented (Robinson and Peng, 1978) to estimate the PVT properties of hydrocarbon mixtures.

Because hydrocarbons present in a reservoir are a mixture of different components, I simplified them into three types: C_1 for methane, C_2 to C_5 for light alkanes, and C_{6+} for medium alkanes. These pseudo-hydrocarbon components are mixed in thermodynamic equilibrium. If a hydrocarbon mixture is unstable as a single-phase fluid at the pressure and temperature of interest, it will separate into two phases. In this case, flash calculations estimate the volumes of both liquid and vapor phases. By changing the composition of each hydrocarbon component with various temperatures and pressures, I constructed a library for the various properties, including density, viscosity, fugacity, molar volume, z-factor, molar weight, entropy (S), enthalpy (H), and heat capacity (C_p) for hydrocarbon mixtures. Among these properties, computed density can be used as an input for SNUPAR to assess nuclear properties.

SNUPAR yields various nuclear properties including density (ρ), migration length (L_m), Sigma (Σ), hydrogen index (HI), and PEF. Because the response of PEF from mixtures does not combine volumetrically, for interpretation purposes, a new parameter, U, named volumetric cross section (Ellis et al., 2007) was developed. This index exhibits a volumetric linear relationship for mixtures. Also, instead of migration length, inverse of migration length, ξ, is recommended for inversion as it exhibits a closer linear response with detector count rates than migration length (Mendoza et al., 2007). Migration length can be converted into neutron porosity (Edwin et al., 2014) using nonlinear relationships between neutron porosity and migration length.

Limitations of the hydrocarbon library are its ranges for temperature and pressure; 80 to 240 degrees F and 500 to 9,000 psia, respectively. It is also assumed that hydrocarbons are combinations of only alkanes without aromatic, cyclic, or inorganic components. Another limitation of this method is that mud filtrate invasion and borehole effects (wash-out and mudcake) are negligible.

The inversion-based interpretation method developed in this chapter is reliable when applied to vertical wells with high porosity and low water saturation. High salinity contrasts between borehole and formation are recommended for inversion with Sigma.

2.3 FORMULATION OF THE INVERSION

The nuclear properties derived from SNUPAR are influenced by fluid and matrix. **Figure 2.2** graphically describes the formation. For simplicity, I assumed that the matrix is composed of sand and shale for siliciclastic formations and limestone and dolomite for carbonate formations. Porosity is saturated with water and hydrocarbons.

To calculate only fluid properties, total porosity is pre-defined to eliminate the contribution of the matrix for each log measurement. Total porosity is estimated by the SNUPAR-based interactive analysis workflow (Heidari, 2011; Ijasan et al., 2013) or computed from NMR measurements. Total porosity estimation needs to be very careful because small misreading in total porosity can propagate to big errors on the estimation of fluid properties. In a similar manner, water saturation needs to be assessed precisely prior to excluding the contribution of water for each log measurement. After decomposing nuclear measurements into the matrix, water, and hydrocarbon, only the hydrocarbon portion can be input into the inversion. The forward models used to reconstruct nuclear measurements can be expressed as

$$\rho_{eff} = (x_1\rho_1 + x_2\rho_2 + x_3\rho_3)\phi(1-S_w) \; + \; \rho_w\phi S_w \; + \; \rho_m(1-\phi), \tag{2.1}$$

$$\Sigma_{eff} = (x_1\Sigma_1 + x_2\Sigma_2 + x_3\Sigma_3)\phi(1-S_w) \; + \; \Sigma_w\phi S_w \; + \; \Sigma_m(1-\phi), \tag{2.2}$$

$$HI_{eff} = (x_1 HI_1 + x_2 HI_2 + x_3 HI_3)\phi(1 - S_w) \; + \; HI_w \phi S_w \; + \; HI_m(1-\phi), \tag{2.3}$$

where ρ_{eff}, Σ_{eff}, and HI_{eff}, are the effective density, effective Sigma, and effective hydrogen index, respectively. Volumetric concentrations of each hydrocarbon component are x_1, x_2, and x_3 respectively, and each measurement with subscript numbers is the contribution of the corresponding hydrocarbon component. Measurements with subscripts w and m are the properties of water and matrix, respectively. Input measurements can be any combination of the six nuclear properties. Because each hydrocarbon component has different nuclear properties that depend on temperature and pressure, it is important to select suitable values for each reservoir.

These well logs are used to estimate volumetric concentrations of three pseudo-hydrocarbon components, x_1, x_2, and x_3. Equations can be reorganized for hydrocarbon contributions in matrix form as

$$\begin{bmatrix} \rho_1 & \rho_2 & \rho_3 \\ \Sigma_1 & \Sigma_2 & \Sigma_3 \\ HI_1 & HI_2 & HI_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \dfrac{\rho_{eff} - \rho_w \phi S_w - \rho_m(1-\phi)}{\phi(1-S_w)} \\ \dfrac{\Sigma_{eff} - \Sigma_w \phi S_w - \Sigma_m(1-\phi)}{\phi(1-S_w)} \\ \dfrac{HI_{eff} - HI_w \phi S_w - HI_m(1-\phi)}{\phi(1-S_w)} \end{bmatrix}. \tag{2.4}$$

In equation (2.4), density, Sigma, and hydrogen index can be replaced by other nuclear properties such as neutron density, inverse of migration length, or PEF. After selecting suitable variables, unknown parameters, x_1, x_2, and x_3, can be estimated using the least-squares minimization method (Aster et al., 2013), which can be written as

$$C(x) = W_d^2 \left[f(x) - d_m \right]^T \left[f(x) - d_m \right], \tag{2.4}$$

where $C(x)$ is the cost function to be minimized, W_d is the data weighting matrix, $f(x)$ is the reconstructed nuclear measurements, and d_m is the original nuclear property used in the inversion. The mass balance equation and non-negative constraints are applied to avoid unrealistic results.

2.4 ESTIMATION OF HYDROCARBON VISCOSITY

Inversion results can be used to compute hydrocarbon viscosity using the Jossi-Stiel-Thodos (JST) Correlation (Jossi et al., 1962) with the 4th order of polynomial coefficients. The viscosity of hydrocarbon mixtures can be assessed based on the assumption that total hydrocarbon viscosity is proportional to the viscosity of each hydrocarbon component with molecular volumetric concentration. The viscosity derived from inversion results can be cross-validated by NMR measurements.

Several studies (Coates et al., 1999; Bryan et al., 2005) suggest empirical correlations between hydrocarbon viscosity and NMR T_2. Other studies show that mass diffusion coefficient is proportional to the ratio of temperature to viscosity (Lo, 1999; Freed, 2005). In general, hydrocarbon viscosity exhibits inverse relationship to the temperature, T_2, and mass diffusion coefficient. However, Straley (2006) shows that universal linear correlations cannot exist for NMR parameters versus viscosity for suites of liquids. Thus, I computed NMR viscosity based on the traditional empirical correlation with fitting constants.

2.5 SYNTHETIC EXAMPLE

This section shows the application of the method to a synthetic case. **Figure 2.3** describes the synthetic vertical well example that contains 113 layers with 0.5 ft of thickness each. Well logs used in the synthetic model are simulated with UTAPWeLS (The University of Texas at Austin Petrophysical and Well-Log Simulator, Voss et al., 2009). This is a typical siliciclastic formation composed of two sands containing water, oil, and gas. **Table 1** summarizes compositions and properties of both fluids and minerals for grouped layers. Shale is a mixture of illite, chlorite, and montmorillonite. Upper and lower sands contain hydrocarbons and irreducible water. Hydrocarbon density gradually decreases from top to bottom in both sands due to gravity segregation.

Figure 2.4 discusses the earth model and inversion results using six different combinations of nuclear measurements. Show the improbable hydrocarbon values, Tracks 2 and 3, which were obtained from combinations of nuclear measurement without density. Also, combinations of nuclear measurements with Sigma, neutron porosity, and hydrogen index provide reliable results compared to those with volumetric cross section and inverse of migration length. A possible reason for the inadequacy of using U as an input for inversion is the dominant contribution of matrix on the measurement compared to fluid. Also, inversion results tend to be in disagreement with synthetic earth model values in depth of intervals with low porosity and high water saturation. Viscosities derived from NMR and inversion results are shown in Track (h) cross-validate the inversion-based interpretation method. **Figure 2.5** describes ternary diagrams of three hydrocarbon components from the earth model and inversion results using various combinations of nuclear properties. Case B and D show undesirable results due to an ill-posed condition caused by single nuclear property. The ternary diagram in case E, shows an improved estimation of in-situ hydrocarbon compositions using both density and

hydrogen index. Ternary diagrams in cases C and F indicate that inversion provides more accurate estimation of hydrocarbon compositions with combinations of multiple nuclear measurements. Inversion results can be validated by Modular Formation Dynamics Tester (MDT) samples or by optical fluorescence measurements (Fujisawa et al., 2008).

In order to investigate the reliability of this method for estimating hydrocarbon composition with noisy measurements, I added 5% zero-mean random Gaussian noise to each nuclear property. I observe how random noise propagates into inversion results. **Figure 2.6** describes the estimated hydrocarbon compositions and their error bars. As porosity or hydrocarbon saturation decreases, uncertainty increases in that relative contribution of hydrocarbon to the measurements decreases. Therefore, perturbations on nuclear measurements lead to a larger error on the inversion results where hydrocarbon volume is relatively small. Inversion results with U on Track (d) show incorrect estimations because U is mainly affected by rock matrix. Lengths of error bars decrease when inversion is performed with more than two nuclear properties. This behavior confirms the stability of the inversion-based interpretation method. From the sensitivity analysis, I found that density is essential, Sigma and neutron porosity are recommended, while volumetric cross section is optional for inversion.

2.6 FIELD CASES

The application of this method to field examples is designed to assess the reliability of inversion with nuclear properties and to reconcile inversion results with the interpretation from NMR measurements. Well data used in this example were acquired in a siliciclastic reservoir in the deep water Gulf of Mexico. This highly laminated turbidite sequence was drilled with Oil-Based Mud (OBM). Sigma measurements were acquired at

a later stage. The inversion technique is implemented in two field examples to estimate in-situ hydrocarbon properties: (I) in a gas-to-oil transition zone, where accurate thermodynamic analysis is critical for phase identification, and (II) in an oil-bearing sand, where Sigma was measured in open-hole logs with depth shifts.

Field Case I, gas and oil-bearing sand; The first case is Well α sand CC. Available measurements include density, PEF, neutron porosity, Sigma, NMR T_2 distributions, and MDT pressure samples. **Figure 2.7** shows the basic well log set together with core data and NMR T_2 distributions. Good agreement among total porosity and NMR porosity except at the upper sand is observed along the interval XXX00 – XXX50 ft, where NMR porosity shows low values compared to core data. This effect is due to the low value of hydrogen index of the gas in this region, as denoted by the low values of density porosity (ϕ_D) compared to neutron porosity (ϕ_N). Neutron porosity and density porosity were corrected for shale concentration. P- and S-wave slownesses indicate the reservoir is unconsolidated. NMR T_2 distributions shown in **Figure 2.8** also confirm the presence of gas in the upper part of the hydrocarbon column as amplitudes of T_2 distributions are suppressed due to low hydrogen index. MDT pressure samples exhibit two distinct pressure gradients, which correspond to gas and oil, respectively. Inversion results from different combinations of nuclear properties are shown in tracks (b) to (f). Although all results indicate the presence of gas mainly in the upper part of sand CC, inversion V results in Track (f) indicate gas at the bottom. This is due to the usage of U in the inversion that is mainly affected by matrix, hence may cause misleading results. Comparing inversion results for Tracks (d) and (g) also indicate that the usage of U or depth-shifted Sigma in the inversion degrades the estimation accuracy of hydrocarbon compositions. Hydrocarbon density calculated from bulk density can be cross-validated from MDT pressure samples. Hydrocarbon density and viscosity can be estimated from

the hydrocarbon library for any mixture of three pseudo-hydrocarbon components with corresponding temperature and pressure. Track (i) shows agreement between hydrocarbon density estimated from bulk density and hydrocarbon density obtained from MDT pressure samples. Track (j) shows the hydrocarbon viscosity estimated from inversion results, which is consistent with hydrocarbon viscosity obtained from both NMR T_2 distributions and pressure samples.

Field Case II, oil-bearing sand: In the lower sand of Well α, the oil column is filled with light oil. **Figure 2.9** shows field measurements with interpreted porosity, water saturation, and shale concentration. This sand contains many thin shale streaks, such that most logs are affected by shoulder beds due to their limited vertical resolution. **Figure 2.10** describes inversion results obtained using various combination of nuclear measurements. Most inversion results indicate that dominant fluids are a mixture of light and medium alkane, which is in agreement with pressure data and NMR T_2 distributions. Pressure data exhibit a constant gradient from the top to the bottom of sand EE. However, inversion results using Sigma in Track (c) show erroneous gas streaks due to local depth shifting mismatches. Also, inversion results obtained from U in Track (c) exhibit unrealistic estimations of continuous gas from the top to the bottom of the hydrocarbon column. Track (i) and (j) show the resemblance of each hydrocarbon density and viscosity derived from inversion, NMR T_2 distributions, and MDT pressure samples. Because the radial length of investigation of NMR tools is very shallow, about 1.5 to 4 inches from the borehole, NMR hydrocarbon viscosity can be affected by mud filtrate invasion.

2.7 CONCLUSIONS

Properties of fluids, especially hydrocarbons in reservoirs, are highly affected by temperature and pressure due to phase behavior. I developed a linear inversion algorithm using various nuclear properties to estimate hydrocarbon compositions. Compositional fluid analysis for multi-component mixtures can be achieved using SNUPAR and PVT flash calculations. Linear inversions render molecular volumetric compositions of three pseudo- hydrocarbon components.

The advantages of this method are (1) thermodynamic consistent interpretation of hydrocarbon compositions in complex phases, (2) reconciliation of SNUPAR-based nuclear properties with NMR measurements, and (3) continuous and reliable assessments of hydrocarbon properties without fluid sampling.

Results from a synthetic case example indicate that density is a key measurement, Sigma and neutron porosity are recommended and volumetric cross section is optional because it is controlled by matrix, not by fluid. Noise or incorrect porosity or water saturation propagates into inversion results and increase the uncertainty of results. Lastly, viscosity calculated from NMR T_2 distributions honors viscosity computed from inversion results. Discrepancies between the two viscosities can be explained by mud filtrate invasion because their DOIs are not equal.

The field examples examined in this chapter confirm that molecular volumetric compositions of in-situ hydrocarbon can be quantitatively appraised from conventional nuclear logs. Thermodynamically consistent assessments of hydrocarbon components based on the integration of inversion results, NMR T_2 distributions, and MDT pressure samples yield a better understanding of the vertical variation of fluids in the formation.

Table 2.1: Summary of solid and fluid properties for the synthetic layered earth model. Two hydrocarbon bearing sands are located between shale layers.

Layer Number		**Ranges of volumetric concentrations [%]**					
		Solids		**Fluids**			
		Quartz	**Shale**	C_1	C_2-C_5	C_{6+}	H_2O
1-17	Shale 1	43-49	40-43	0	0	0-1	9-12
18-80	Sand 1	55-63	3-12	2-24	5-15	4-14	4-11
81-88	Shale 2	47-48	32-34	0	0	0-1	17-18
89-109	Sand 2	64-69	5-10	4-9	2-4	1-4	9-19
110-113	Shale 3	55-56	33-34	0	0	0-1	10-11

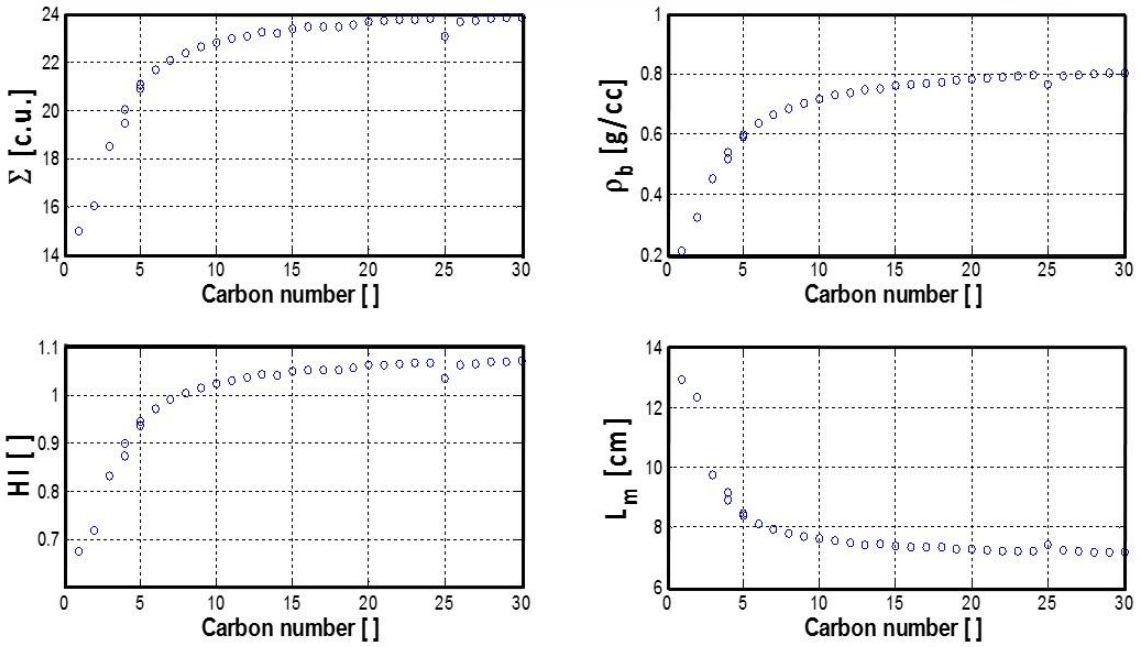


Figure 2.1: Nuclear properties with regard to hydrocarbon carbon number. Sigma, density, hydrogen index, and migration length show distinguishable changes between ethane (C_2) and pentane (C_5) with fixed temperature and pressure.

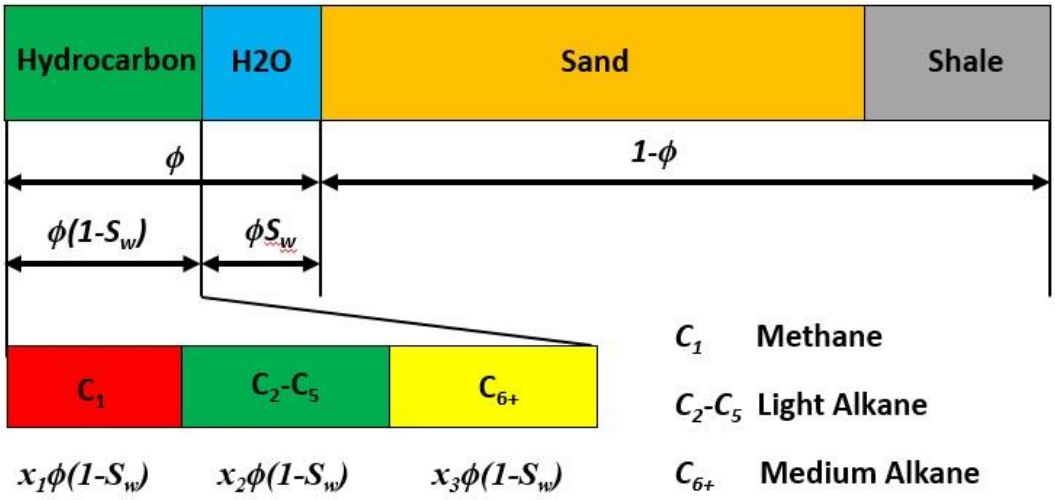


Figure 2.2: Graphical description of a formation that consists of matrix and fluid. Matrix is composed of sand and shale, fluid is filled with hydrocarbon and water, respectively. Hydrocarbon components are classified into three pseudo-components, C_1, C_{2-5}, and C_{6+}, to augment their contrasts.

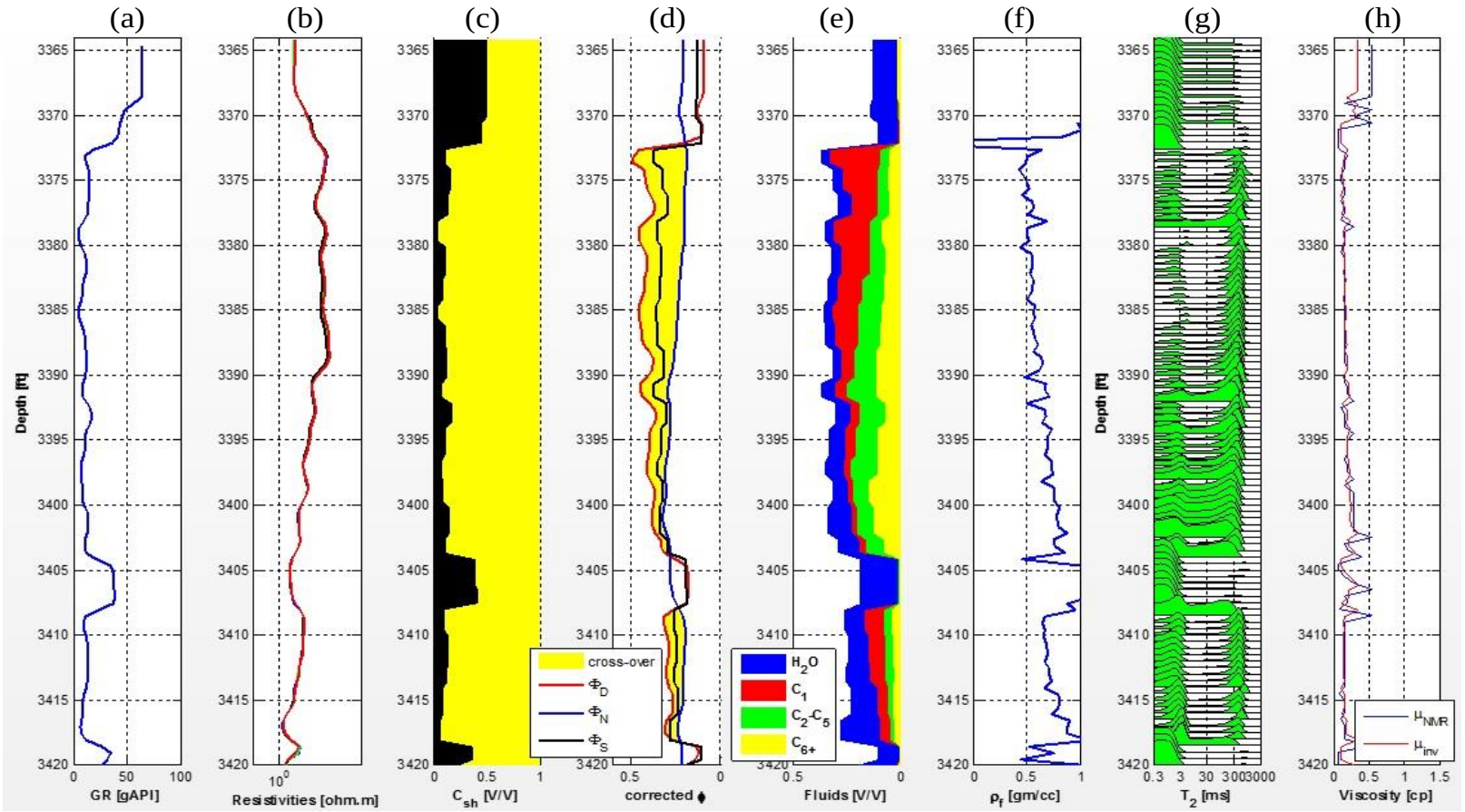


Figure 2.3: Example of a synthetic earth model. Upper sand represents a typical capillary transition zone with fluid segregation by gravity. (a): Gamma-ray log. (b): Shallow (blue curve) and deep (red curve) resistivity. (c): Volumetric shale concentration. (d): Neutron and density porosities and their cross-over. (e): Volumetric fluid distributions. (f): Fluid density. (g): Synthetic NMR T_2 distributions. (h): Hydrocarbon viscosity derived from NMR (blue curve) and inversion (red curve).

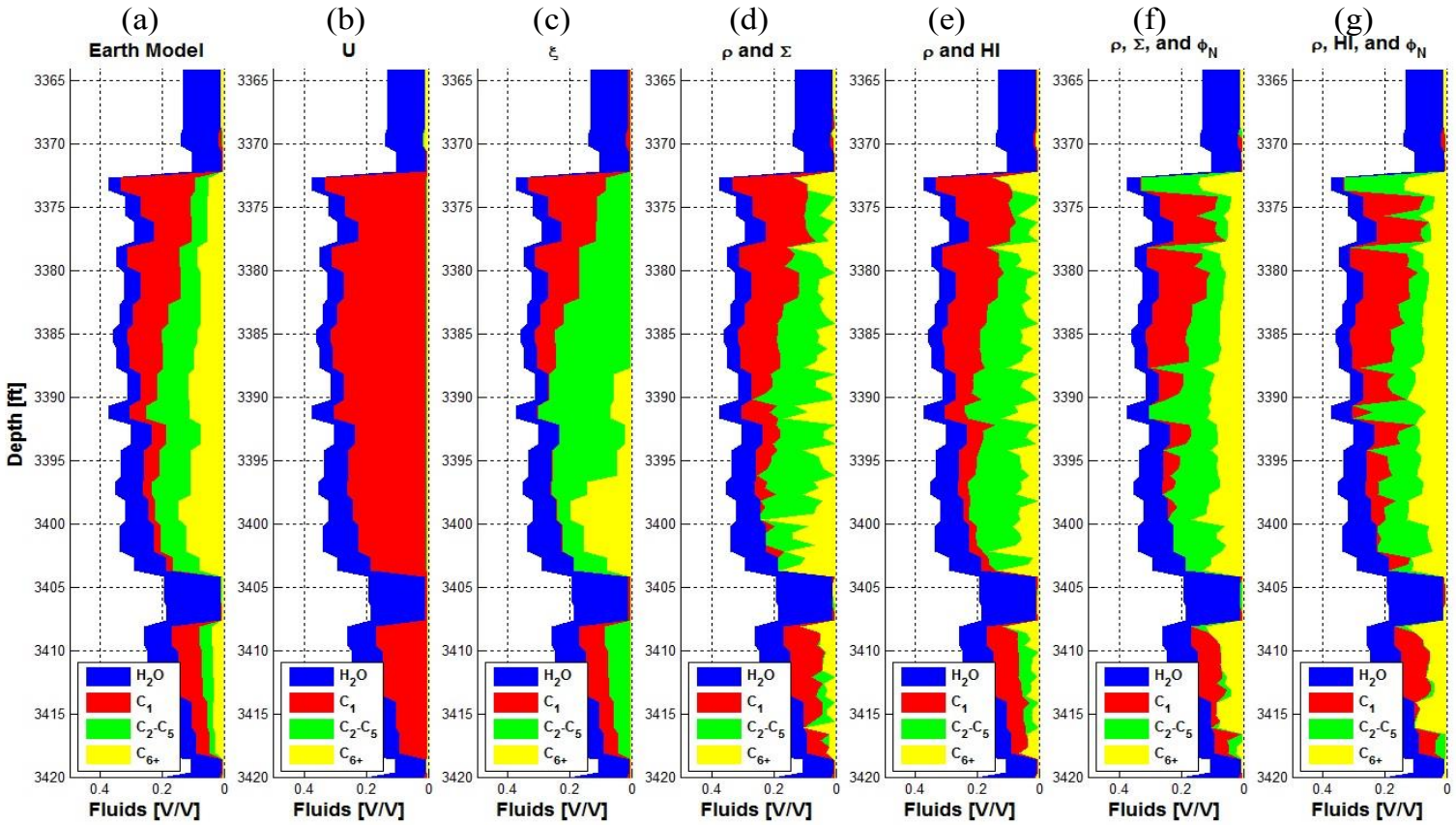


Figure 2.4: Synthetic earth model and inversion results using selected combinations of nuclear properties. Nuclear measurement combinations with single property (Tracks b and c) exhibit one or two hydrocarbon components only, thereby yielding undesirable results. (a): Earth Model. (b): Inversion results with volumetric cross section. (c): Inversion results with inverse of migration length. (d): Inversion results with density and Sigma. (e): Inversion results with density and hydrogen index. (f): Inversion results with density, Sigma, and neutron porosity. (g): Inversion results with density, hydrogen index, and neutron porosity.

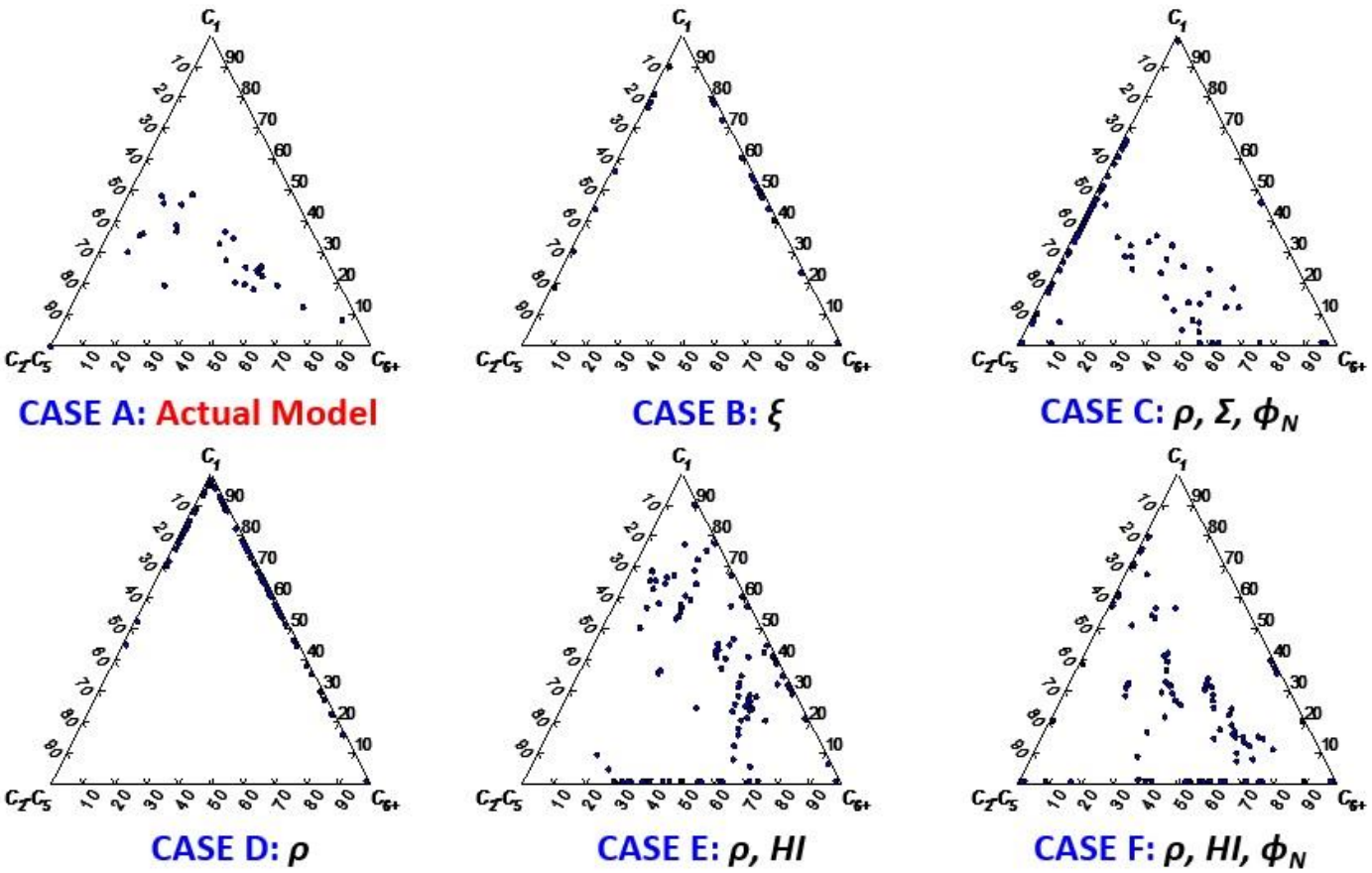


Figure 2.5: Ternary diagrams of three hydrocarbon components from the earth model and inversion results using selected combinations of nuclear properties. Whereas inversion with ill-posed conditions typically identify two types of hydrocarbons, inversion with three different nuclear measurements provide reliable estimations of in-situ hydrocarbon properties.

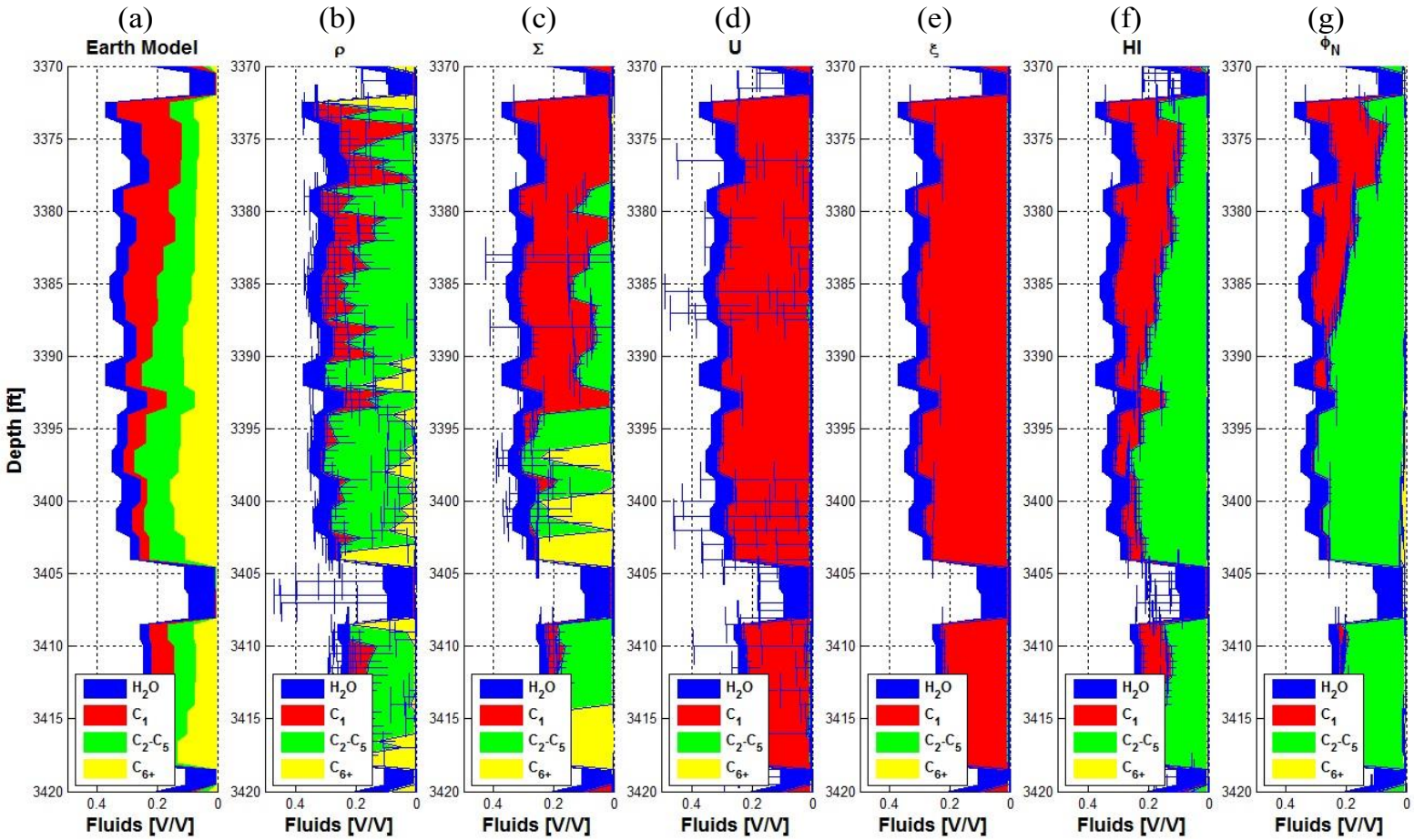


Figure 2.6: Synthetic earth model and the inversion results with 5% random Gaussian noise. (a): Earth Model. (b): Inversion results with density. (c): Inversion results with Sigma. (d): Inversion results with volumetric cross section. (e): Inversion results with inverse of migration length. (f): Inversion results with hydrogen index. (g): Inversion results with neutron porosity.

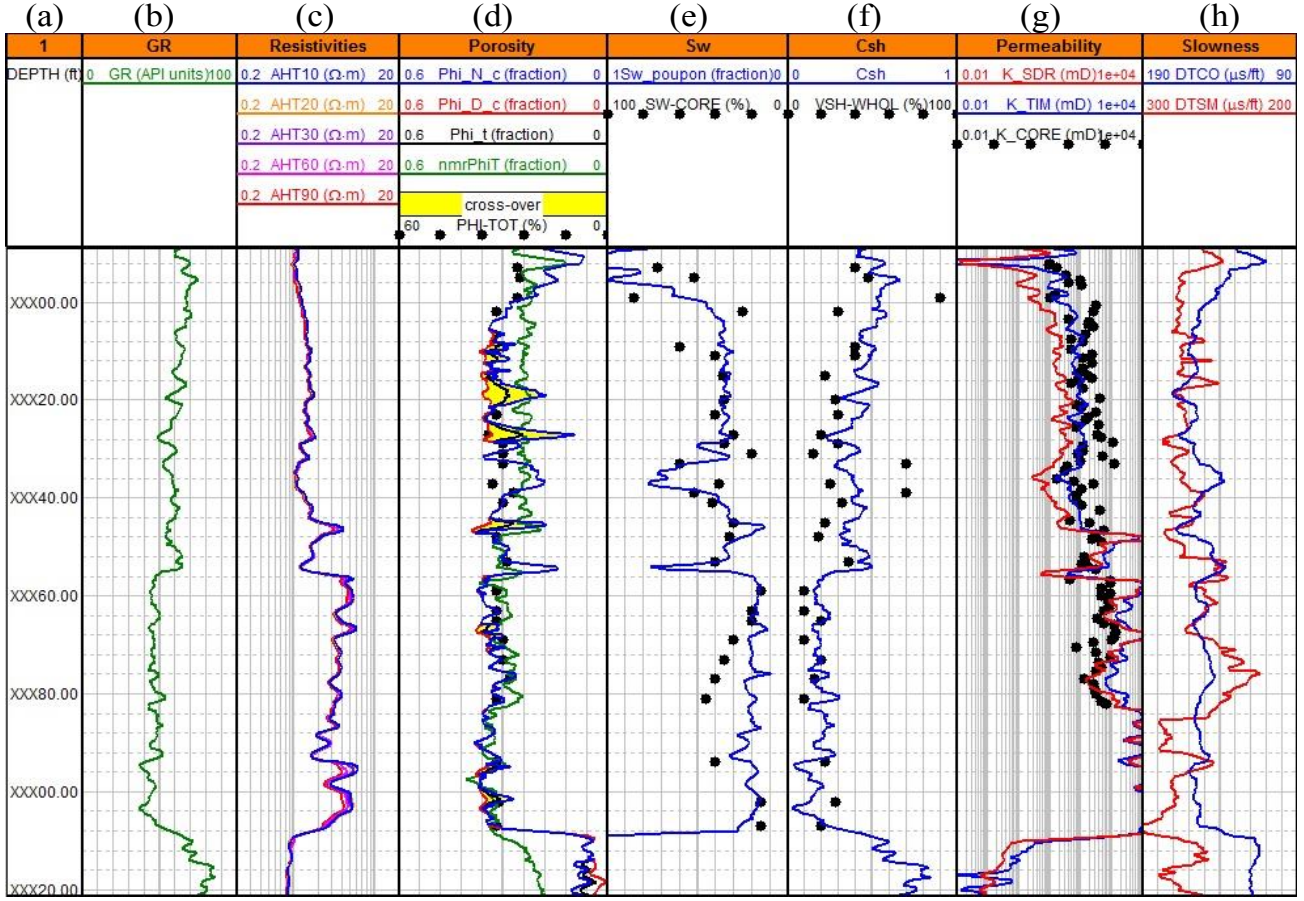


Figure 2.7: Field case No. 1: Well α Sand CC. (a): Masked depth. (b): Gamma-ray log. (c): Shallow (blue curve) and deep (red curve) resistivities. (d): Shale corrected neutron and density porosity (blue and red curves) and their cross-over, total porosity and NMR porosity (black and green curves), and core data. (e): Estimated water saturation and core data. (f): Estimated shale concentration and core data. (g): SDR and Timur-Coates permeabilities, and core data. (h): P-wave slowness (blue curve) and S-wave slowness (red curve).

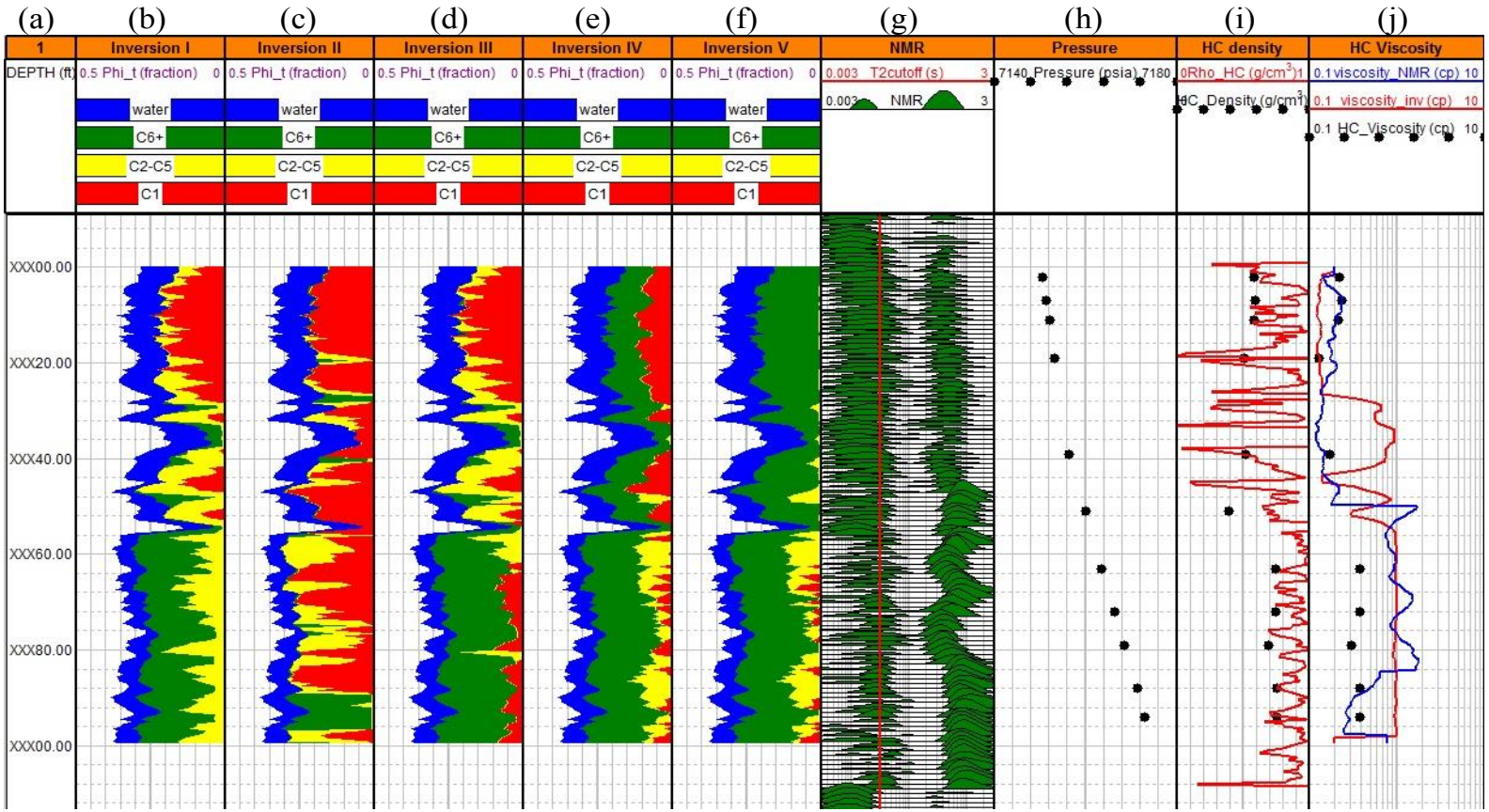


Figure 2.8: Field case No. 1: Well α Sand CC. (a): Depth. (b): Inversion result with density. (c): Inversion results with Sigma. (d): Inversion results with density and volumetric cross section. (e): Inversion results with density and neutron porosity. (f): Inversion results with neutron porosity and volumetric cross section. (g): NMR T_2 distributions and T_2 cutoff. (h): MDT pressure samples. (i): Estimated hydrocarbon densities from bulk density (red curve) and MDT pressure samples (black dots). (j): Hydrocarbon viscosities derived from NMR (blue curve), inversion (red curve), and MDT pressure samples (black dots).

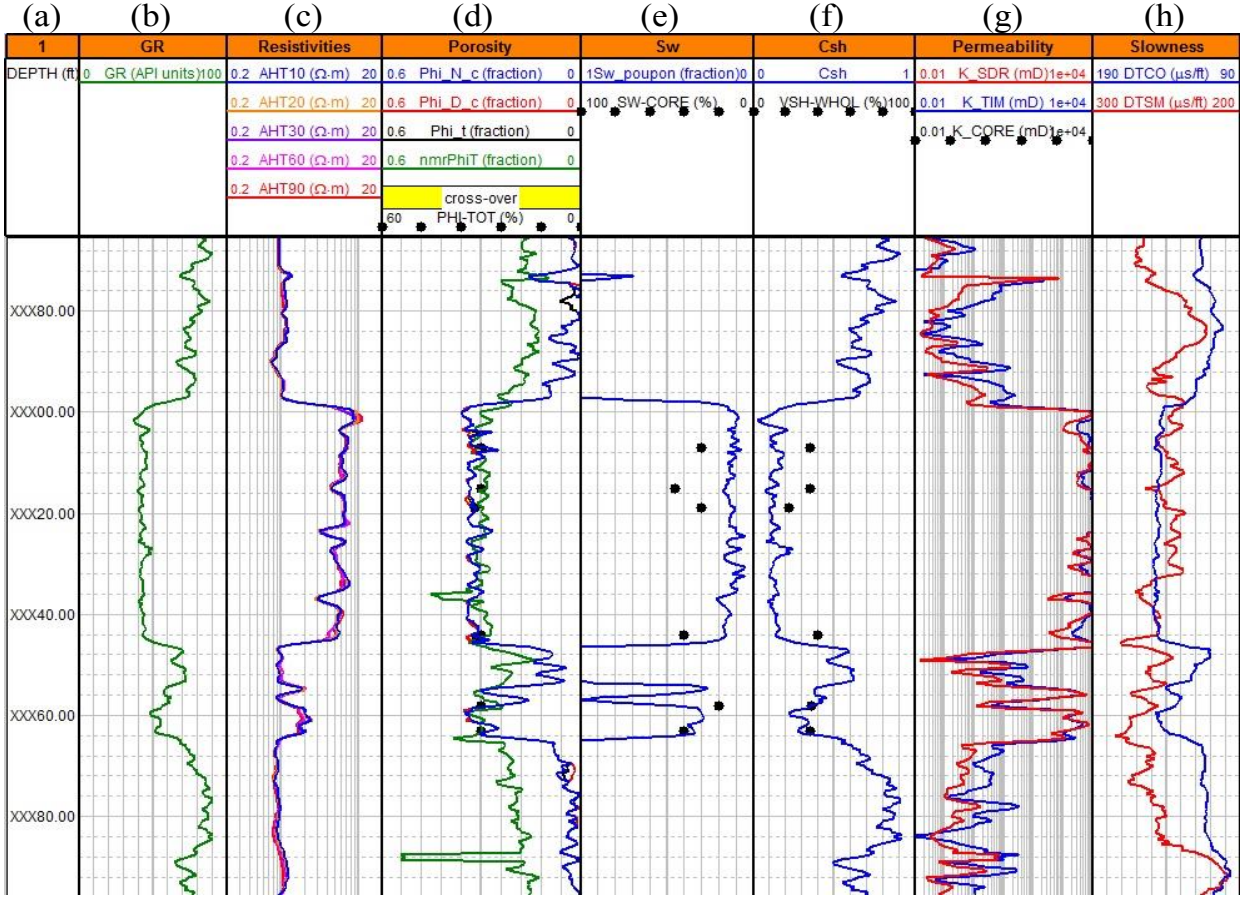


Figure 2.9: Field case No. 2: Well α Sand EE. (a): Masked depth. (b): Gamma-ray log. (c): Shallow (blue curve) and deep (red curve) resistivities. (d): Shale corrected neutron and density porosity (blue and red curves) and their cross-over, total porosity and NMR porosity (black and green curves), and core data. (e): Estimated water saturation and core data. (f): Estimated shale concentration and core data. (g): SDR and Timur-Coates permeabilities, and core data. (h): P-wave slowness (blue curve) and S-wave slowness (red curve).

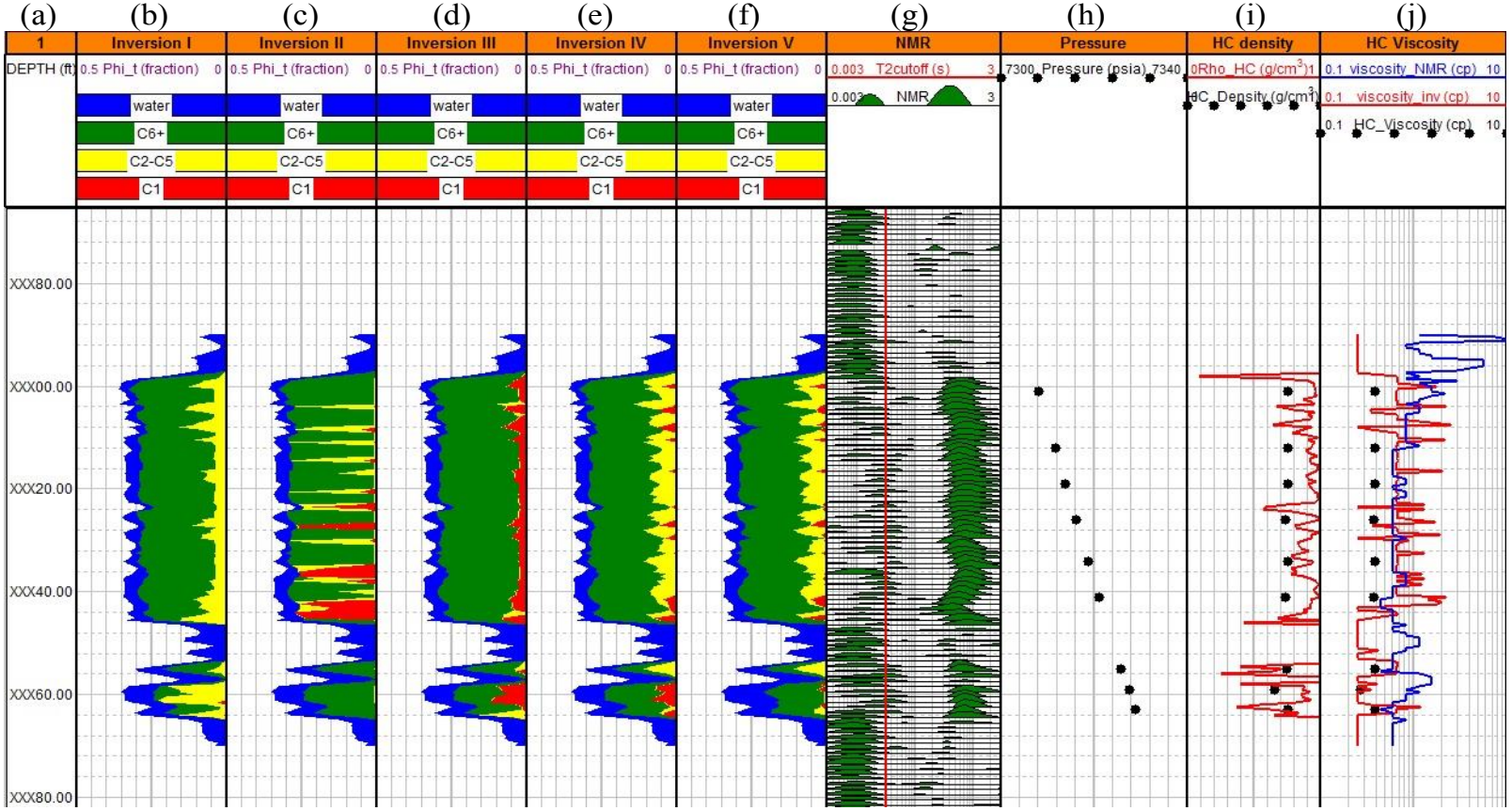


Figure 2.10: Field case No. 2: Well α Sand EE. (a): Depth. (b): Inversion result with density. (c): Inversion results with Sigma. (d): Inversion results with density and volumetric cross section. (e): Inversion results with density and neutron porosity. (f): Inversion results with neutron porosity and volumetric cross section. (g): NMR T_2 distributions and T_2 cutoff. (h): MDT pressure samples. (i): Estimated hydrocarbon densities from bulk density (red curve) and MDT pressure samples (black dots). (j): Hydrocarbon viscosities derived from NMR (blue curve), inversion (red curve), and MDT pressure samples (black dots).

Chapter 3: Investigation of Dielectric and Thermodynamic Properties of Polar Liquids by Dielectric Spectroscopy and Molecular Dynamics Simulation

In this chapter, dielectric permittivity of polar liquid mixtures were measured over an extensive composition range of 20 MHz to 20 GHz at ambient conditions. A series of fluid samples including water, 2-propanol, methanol, and their mixtures were measured for dielectric relaxation spectra. One Havriliak-Negami (HN) model is adapted for the mathematical inversion to estimate dielectric permittivity and relaxation time. These dielectric properties were calculated using molecular dynamics (MD) simulations with OPLS/AA (Optimized Potential for Liquid Simulations-All Atoms) force field and are compared with experimental results. Also, thermodynamic properties including excess enthalpy, density, number of hydrogen bonds, and effective self-diffusion coefficient are computed to cross-validate the experimental results. These properties predicted from MD simulations exhibit excellent agreements with experimental results. Furthermore, it is shown that the average number of hydrogen bonds per unit volume is strongly correlated with the dielectric relaxation time and mass diffusion coefficient. The MD simulation results show that predictions of microscopic and macroscopic properties can be improved with a better-optimized parameterization of molecules and the force field.

3.1 INTRODUCTION

There have been many studies to estimate thermodynamic and dielectric properties using molecular dynamics (MD) simulations. Investigations of water (Rønne et al., 1997) and mixtures of water and various liquids (Wensink et al., 2003; Jia et al., 2009; Pascal et al., 2012; Dzida et al., 2015; Galicia-Andrés et al., 2015; Caro et al., 2016) have been made to elucidate the microscopic and macroscopic properties of

liquids. Although several studies have focused on methanol and 2-propanol, the influence of microscopic properties on the mixture's ensemble average of macroscopic dielectric properties has not been well quantitatively described. Also, the MD simulation results of previous studies may require better modeling of simulation that honor the data. The question that I would like to answer is whether experimental dielectric static permittivity and relaxation time of liquid mixtures can be described by microscopic properties calculated from numerical simulations. Measurements of dielectric relaxation spectra are the results of ensemble averages. Therefore, if MD simulations can cross-validate the experimental observations, they can be explained in terms of microscopic properties. The microscopic properties are generated from the interaction forces between bonded and non-bonded atoms including Lennard-Jones and Coulomb potentials. Thus, from the trajectories and velocities of molecules, I can calculate various microscopic and macroscopic properties of liquid mixtures. The MD simulation method is admissible for analyzing these microscopic properties that can be computed from the dynamic evolution of the system which is the result of interactions of atoms and molecules and their trajectories. In this study, I present the mixing effects on macroscopic dielectric permittivity and relaxation time in terms of various thermodynamic properties computed from MD simulations.

This chapter is organized as follows: First, the laboratory experiment section describes the procedures of how frequency dependent dielectric spectra were measured and processed for the quantitative analysis. Following the MD simulation section, theory and simulation methods will be presented. Also, the effect of mixing on dielectric static permittivity is investigated, along with reorientational relaxation time, to develop a better understanding of polar molecule alignments in the direction of the external electric field.

The results and discussion section presents a comparison of dielectric properties based on spectroscopic measurements and MD simulations for liquid mixtures.

3.2 LABORATORY EXPERIMENTS

The laboratory experiments were designed to estimate dielectric properties of fluid mixtures. A series of fluid samples was prepared by mixing deionized (DI) water, 2-propanol, and methanol. Immediately after mixing these fluids, the samples were sealed in glass containers for spectral measurements. All measurements were taken at room temperature and ambient pressure. The dielectric spectra of fluid samples were recorded on an Agilent E5071C Network Analyzer over a frequency range of 20 MHz to 20 GHz at ambient conditions. The resolution of collected spectra was set to 1,601 points and the calibration spectra of air and water were recorded immediately before each measurement for the reference. Measured spectra of water and methanol mixtures, water and 2-propanol mixtures, and methanol and 2-propanol mixtures are shown in **Figure 3.1**.

These measurements were analyzed using least squares mathematical inversion and applied for the real and imaginary part of the dielectric spectra. To fit the asymmetric shape of the complex dielectric spectra, Havriliak-Negami (HN) equation is used as

$$\varepsilon(\omega) = \varepsilon_{\infty} + \frac{\varepsilon_0 - \varepsilon_{\infty}}{(1+(j\omega\tau)^{\alpha})^{\beta}}, \tag{3.1}$$

where ε_0 and ε_{∞} are dielectric permittivity at zero and infinity frequency, ω is angular frequency, τ is dielectric relaxation time, and the exponents α and β are asymmetric and broadness parameters of the corresponding spectra. Note that $\alpha = 1$ leads the HN equation to the Cole-Davidson equation and $\beta = 1$ causes it to the Cole-Cole equation,

respectively. The measured dielectric relaxation spectra are in GHz range with main dispersion only; the second and third relaxation terms are not necessary because they cover the motion of the end-standing alcohol monomers and the fast switching of OH-groups (Sato et al., 2005). Also, these second and third relaxations increase the number of unknown parameters, making the inversion more unstable.

The quadratic overall cost function is expressed as

$$C(x) = W_d \left\| d(\mathrm{x}) - \mathrm{d_m} \right\|_2^2 + \lambda^2 W_x \left\| x - x_0 \right\|_2^2, \tag{3.2}$$

where W_d is a data weighting matrix, d_m is the dielectric measurement which is a function of frequency, λ is the regularization parameter, W_x is a model weighting matrix, and x_0 is the initial guess for the unknown parameters, respectively. Since nonlinear minimization of the cost function is an iterative process, I use the regularized Occam's method (Aster et al., 2005; Hansen, 1994) as

$$\left[J^T W_d J + \lambda^2 W_x^T W_x \right] x = J^T W_d e, \tag{3.3}$$

$$e = d_m - d(x) + Jx + \lambda^2 x_R, \tag{3.4}$$

where *d(x)* is the vector of forward modeling constructed by HN model, *e* is the misfit between the measurements and forward modeling, and *J* is the Jacobian matrix. The corresponding Jacobian matrix has the form,

$$J = \begin{bmatrix} \frac{\partial \varepsilon_1}{\partial \varepsilon_0} & \frac{\partial \varepsilon_1}{\partial \varepsilon_\infty} & \frac{\partial \varepsilon_1}{\partial \tau} & \frac{\partial \varepsilon_1}{\partial \alpha} & \frac{\partial \varepsilon_1}{\partial \beta} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \varepsilon_N}{\partial \varepsilon_0} & \frac{\partial \varepsilon_N}{\partial \varepsilon_\infty} & \frac{\partial \varepsilon_N}{\partial \tau} & \frac{\partial \varepsilon_N}{\partial \alpha} & \frac{\partial \varepsilon_N}{\partial \beta} \end{bmatrix}, \tag{3.5}$$

where N is the number of the measured point of the dielectric spectra (1,601). The partial derivatives of the five unknown parameters are

$$\frac{\partial \varepsilon}{\partial \varepsilon_0} = \frac{1}{(1+(j\omega\tau)^\alpha)^\beta}, \tag{3.6}$$

$$\frac{\partial \varepsilon}{\partial \varepsilon_\infty} = 1 - \frac{1}{(1+(j\omega\tau)^\alpha)^\beta}, \tag{3.7}$$

$$\frac{\partial \varepsilon}{\partial \tau} = \frac{-\alpha\beta(\varepsilon_0 - \varepsilon_\infty)(j\omega\tau)^\alpha}{(1+(j\omega\tau)^\alpha)^{\beta+1}\tau}, \tag{3.8}$$

$$\frac{\partial \varepsilon}{\partial \alpha} = \frac{-\beta(\varepsilon_0 - \varepsilon_\infty)(j\omega\tau)^\alpha \log(j\omega\tau)}{(1+(j\omega\tau)^\alpha)^{\beta+1}}, \tag{3.9}$$

$$\frac{\partial \varepsilon}{\partial \beta} = \frac{(\varepsilon_0 - \varepsilon_\infty)\log(1+(j\omega\tau)^\alpha)}{(1+(j\omega\tau)^\alpha)^\beta}. \tag{3.10}$$

The linear least squares inversion is ill-conditioned, requiring reasonable initial guesses for the unknown parameters. **Table 3.1** summarizes estimated dielectric permittivities and relaxation times for liquid mixtures used in this study.

3.3 MOLECULAR DYNAMICS SIMULATIONS

I chose the software package GROMACS 5.1.4 (Abraham et al., 2017) to perform MD simulations because of its high performance and user-friendly interface. I tested similar mixtures of water and monohydroxy alcohol molecules in a cubic box with periodic boundary condition (PBC). The bulk water was prepared using an SPC/E model (Berendsen et al., 1987). For the MD simulations, selection of the proper force field is critical. A comparison (Coleman et al., 2012) between the GAFF (Wang et al., 2004), OPLS/AA (Jorgensen et al., 1996), and CHARMM (Vanommeslaeghe et al., 2010) force fields showed that OPLS/AA force field provides more accurate predictions of physical properties of various liquids than GAFF force field. This is because OPLS/AA force field was parametrized for liquids. The CHARMM force field is hardly better for the estimation of density and enthalpy of vaporization. I have compared the CHARMM36 (Best el al., 2012) and OPLS/AA force fields for the various liquid mixtures and found that the OPLS/AA force field converged slightly faster and provided more accurate microscopic properties, but the differences are very small. Thus, in this chapter, I applied the OPLS/AA force field for the MD simulations because it has been optimized to reproduce the liquid thermodynamic and structural properties at room temperature.

The numbers of water and alcohol molecules are tabulated in **Table 3.2**. The velocity-rescaling scheme (Bussi et al., 2007) was selected for the thermostat to control constant temperature. The Berendsen algorithm (Berendsen et al., 1984) was implemented for pressure control. The potential energy of the system was minimized and the box size was adjusted by applying canonical ensemble (NVT) and isothermal-isobaric ensemble (NPT). The grid spacing was defined at 0.12 nm and the radius for Lennard-Jones interactions and Particle Mesh Ewald (PME) electrostatic interactions (Essmann et al., 1995) are limited to 1.5 nm for large molecules (Jorgensen et al., 1996). After the

system was equilibrated, a production run was applied for 6 ns and the Verlet algorithm (Swope et al., 1982) was used for the equation of motion with a time step of 2.0 fs. Simulation times of more than 5 ns are essential to obtain well-converged properties of liquid mixtures. I calculated density, mass diffusion coefficient, excess enthalpy, dipole moment, dielectric permittivity, and a number of hydrogen bonds of the system and compared them to the experimental results. Simulation results show that the dielectric relaxation process can be explained as the collegial motion of the water and alcohol molecule ensemble.

3.4 RESULTS AND DISCUSSION

In this section measured dielectric spectra of liquid mixtures and their corresponding properties derived from numerical inversion results and those from MD simulations are presented, discussed, analyzed, and compared.

3.4.1 Density

Density is a key property of liquids. The density of most liquid mixtures is not linearly proportional to the concentration due to excess volume. Also, the density of liquid will affect other properties such as dielectric permittivity and mass diffusion coefficient. Accurate modeling of density and excess volume for liquid mixtures is important. The NPT simulations at ambient condition (298.15 K and 1 bar) can produce the density of liquid mixtures. Calculated densities and their error bars together with experimental data (Noda et al., 1982; Wei et al., 1984; Pang et al., 2007) are illustrated in **Figure 3.2**. The molar fraction given in the figure is the fraction of the second component. Although the density of methanol and 2-propanol is slightly overestimated by

1-2 %, the MD simulation provides a reasonable density of liquid mixtures. The reason for the overestimated density is the dipole moment and partial charges of the molecules, which will be explained in the dielectric permittivity section.

3.4.2 Enthalpy

Enthalpy changes of liquid mixtures are mainly caused by the intermolecular forces between molecules. Interactions of hydrogen bonding, dipole-dipole, and induced dipole generate strong attraction forces and exhibit a lower enthalpy of the mixture. Most liquids are non-ideal mixtures and the deviation from the ideal mixing can be expressed as an excess enthalpy,

$$\Delta H_{mix} = H_{mix} - (X_1 H_1 + X_2 H_2) \tag{3.11}$$

where ΔH_{mix} is the enthalpy of mixing, X_1 and X_2 are the molar fraction of component 1 and 2, and H_1 and H_2 are their corresponding enthalpy. Calculated excess enthalpy and the experimental data (Benjamin et al., 1963; Lama et al., 1965; Martínez-Jiménez et al., 2018) are compared in **Figure 3.3**. Excess enthalpy of water and methanol mixtures shows reasonable agreements with the experimental results. This could not be made for the previous studies (Wensink et al., 2003; Guevara-Carrion et al., 2011; Pascal et al., 2012; Galicia-Andrés et al., 2015; Caro et al., 2016; Martínez-Jiménez et al., 2018) in which either the molar fraction of the minimum ΔH_{mix} is higher or the absolute value of ΔH_{mix} is smaller than experimental data. For the water and 2-propanol mixtures, the negative trend of excess enthalpy in the water-rich side and positive on the alcohol-rich side are well predicted as experimental data. Excess enthalpy of methanol and 2-propanol

mixtures is poorly described, but the experimental values are still in the range of error bars.

3.4.3 Dielectric Permittivity

Dielectric permittivity from the MD simulation is a difficult property to emulate from experimental data. Calculation of dielectric permittivity with the MD simulation uses the dipole moment fluctuations (Neumann, 1983) of the polar liquids in Periodic Boundary Conditions (PBC) as

$$\varepsilon = 1 + \frac{\langle M^2 \rangle - \langle M \rangle^2}{3 k_B T V \varepsilon_0}, \tag{3.12}$$

where $<M^2>-<M>^2$ is dipole moment fluctuations, k_B is the Boltzmann constant, T is the absolute temperature, V is volume, and ε_0 is the vacuum permittivity, respectively. **Figure 3.4** presents the dielectric permittivity and its error bars estimated from the MD simulation compared to the experimental data as a function of the molar fraction. The dielectric permittivity estimated from the simulations showed similar slopes of the experimental data but yield generally smaller values. Previous literature (Rønne et al., 1997; Jia et al., 2009; Caleman et al., 2012; Galicia-Andrés et al., 2015; Martínez-Jiménez et al., 2018) also shows that dielectric permittivity calculated from the MD simulation is lower than the experimental values. Several possible reasons for these discrepancies are: (a) explicit polarization is not enough, (b) the size of the simulation box and number of molecules (in this study, 1,000) is too small to represent real liquid status, (c) partial charges of the OPLS force field are not optimized, and (d) the convergence of the fluctuations of total dipole moment takes a long time to reach a state

of equilibrium. However, as shown in **Figure 3.5**, the dielectric permittivity of most mixtures are well converged after 5 ns. Thus, a better-parameterized force field with a larger molecular number would be in better agreement with the experimental results. Also, the average values of multiple MD simulations can reduce statistical errors. Several attempts have been made on architectural modifications of water (Fuentes-Azcatl et al., 2014) and methanol (Salas et al., 2014) molecules to improve the estimation of dielectric permittivity by modifying the energy of non-bonded interactions and partial charges. Although their estimation of dielectric permittivity is improved, the estimation of other properties is deteriorated. A recent study (Martínez-Jiménez et al., 2018) presented a four-site potential model of methanol by incorporating a united atom of a massless methyl group into the charge distribution of the methanol monomer. Although their dielectric permittivity accurately reproduced experimental data, the agreement of mass diffusion coefficient and excess enthalpy of water and methanol mixtures are unsatisfactory. To fit the dielectric permittivity, increased dipole moment is required which causes more attractive interactions between molecules, eventually decreasing mass diffusion coefficient and increasing density. Therefore, more robust model with optimized parameters need to be developed for accurate estimations of both dielectric and thermodynamic properties of liquid mixtures.

3.4.4 Dielectric Relaxation Time

Dielectric relaxation time can be estimated by calculating interaction time between total dipole moment of molecules and applied electric field. Auto Correlation Function (ACF) of total dipole moment is considered as an exponential decay function (McQuarrie, 1976) as,

$$\Phi(t) = \frac{\langle M(0)M(\mathrm{t})\rangle}{\langle M^2\rangle} = \exp\left(-\frac{|t|}{\tau_D}\right), \tag{3.13}$$

where τ_D is a dielectric relaxation time and the angular brackets in the equation mean an average of equilibrium ensemble. The original Debye theory expressed dielectric behavior as a single exponential decay, but the correlation function could not be fully captured by one or two exponential decays due to the statistical properties. Thus, NMR-like inversion is implemented to calculate dielectric relaxation time. The autocorrelation functions of liquid mixtures calculated from MD simulations and inversion results together with experimental data are presented in **Figure 3.6**. Inverted dielectric relaxation times from both experiments and MD simulations are plotted together in **Figure 3.7**. Although the simulations yield faster relaxation time at some points, the overall trends of dielectric relaxation time for all liquid mixtures are qualitatively similar with the experimental results. Deviations are mainly due to the statistical fluctuations of ACF, which diverges from the exponential decaying behaviors.

3.4.5 Mass Diffusion Coefficient

The mass diffusion coefficients of liquid mixtures are calculated as a function of the molar fraction. As two liquids are mixed, attractive and repulsive interactions between molecules govern the effective diffusion coefficient. To quantify mass diffusion coefficient of liquid mixtures, a well-known Einstein's relation is applied as,

$$\lim_{t\to\infty}\left\langle \|r(t)-r(0)\|^2 \right\rangle = 6Dt \tag{3.14}$$

where *r(t)* is a distance a molecule travels in time *t*. The mass diffusion coefficient *D* is proportional to the slope of Mean-Square Displacement (MSD) over time, which can be calculated by tracing the trajectory of all atoms. A comparison of calculated mass diffusion coefficients from the MD simulation and experimental values (Pratt et al., 1975; Derlacki et al., 1985) is given in **Figure 3.8**. Although the overall shapes of the curves for liquid mixtures are analogous to the experimental results, there are some discrepancies between the MD simulation and experimental results. The minimum mass diffusion coefficient of the water and methanol mixtures is observed at a lower methanol molar fraction compared to the experimental data. Also, the mass diffusion coefficients of methanol and 2-propanol are underestimated especially on the alcohol-rich side. This is because of the enhanced dipole moments, which increase attracting forces between the molecules. As a result, they travel a shorter distance in a given time. Estimation of mass diffusion coefficients from previous studies (Wensink et al., 2003; Guevara-Carrion et al., 2011; Galicia-Andrés et al., 2015; Martínez-Jiménez et al., 2018) also showed qualitatively correct results with the OPLS force field, but also some discrepancies in details compared to the experimental data.

3.4.6 Hydrogen Bond

A hydrogen bond is electrostatic attraction between a hydrogen atom and other electronegative atom such as oxygen. To investigate the effect of hydrogen bond to dielectric relaxation time, numbers of hydrogen bonds between all donors and acceptors are calculated from MD simulations. Two popular geometric criteria is applied to count the number of hydrogen bonds: (1) the distance between donor and acceptor is less than 0.35 nm and (2) the angle is less than 30 degrees. **Figure 3.9** represents the average

number of hydrogen bonds per molecule as a function of molar composition. To validate simulation results, I compared the number of hydrogen bonds per water molecule. An average number of hydrogen bonds per water molecule from previous literature varies: 3.3 from X-ray absorption spectrum (Smith et al., 2004), 3.46 from NMR (Hoffmann et al., 1997), 3.58 from neutron diffraction (Soper el al., 1997), and 3.19 from Monte Carlo simulations (Kalinichev et al., 1997). These distributions are due to their different definitions of the hydrogen bond, diverse experimental or simulation techniques, and interpretational uncertainties. Thus, our simulation results of 3.48 can be considered a reasonable estimation. Panel (B) of **Figure 3.9** shows that the average number of hydrogen bonds between two components become maximum at molar fraction of 0.6 while panel (A) and (C) show the maximum at 0.5. This can be explained by the size difference between water and 2-propanol molecules, which is greater than the size differences between water and methanol and 2-propanol and methanol molecules.

The number of hydrogen bonds per unit volume decreases with an increasing dielectric relaxation time, as demonstrated by three series of liquid mixtures in **Figure 3.10**. This tendency conforms to the wait-and-switch model (Kaatze et al., 2002) that dielectric relaxation time is primarily determined by the availability of neighboring hydrogen bond sites. If larger number of alternative hydrogen bonds exist in suitable positions, a molecule needs less time to jump to an unoccupied bond to create a new hydrogen bond. This changing time is governed by the strength of hydrogen bonds and availability of a hydrogen bond donor or acceptor located nearby. Therefore, dielectric relaxation time can be considered as a period between a molecule is captured in the hydrogen bond network until it is released (Sato et al., 2005). This is supported by the previous studies of reorientation of water (Sciortino et al., 1992) and alcohol (Senthilkumar et al., 2018) molecules. Furthermore, dielectric relaxation time is related to

the size of the molecule. While monohydric alcohol contains alkyl group, water has only hydroxyl group, which occupies smaller volume. Water can generate up to four hydrogen bonds per molecule and construct a tetrahedral structure, whereas alcohols have only three possible hydrogen bonds per molecule. In terms of molar volume, methanol and 2-propanol have the same hydroxyl group, but methanol is able to build more hydrogen bonds per unit volume than 2-propanol because it has a shorter chain length. Because chain length is independent of the hydrogen bond energy (Kaatze et al., 2002), dielectric relaxation times for alcohols and aqueous solutions are substantially dependent upon the numbers of hydrogen bonds per unit volume.

The effect of hydrogen bonds on mass diffusion coefficient is similar to the dielectric relaxation time in that hydrogen bonding retards the mass diffusion of molecules (Su et al., 2010). In the case of water and alcohol mixtures, water molecules inserted into methanol or 2-propanol molecules causes a rapid rearrangement of hydrogen bonds. The average lifetime of a hydrogen bond is only several picoseconds (Luzar et al., 1996), donors and acceptors continue breaking and forming hydrogen bonds. According to the transition state theory, both relaxation and diffusion motions need to overcome the energy barrier of hydrogen bonds. Thus, a molecule with a certain amount of energy can diffuse less with a greater number of hydrogen bonds in the system. Therefore, the dielectric relaxation time and mass diffusion coefficient of liquid mixtures are mainly attributed to the collective dynamics of hydrogen bonds, dipole-dipole interactions, and hydrophobic alkyl group interactions.

3.5 LIMITATIONS OF MOLECULAR DYNAMICS SIMULATION

Despite these accurate predictions, there are limitations to the MD simulations. In order to acquire numerical stability, it requires simulation time of several ns with short time steps, which are computationally intensive. Also, most force fields including OPLS/AA are intrinsically an approximation. Additionally, one needs to repeat the same simulation to avoid statistical errors. However, the MD simulation can be considered an effective approach to investigate properties of complex liquid mixtures under various conditions. Therefore, MD simulation can serve as a bridge between theory and laboratory measurements for a better description of the dielectric behaviors.

3.6 CONCLUSIONS

I have measured the complex dielectric relaxation of water, methanol, and 2-propanol mixtures in the microwave region at ambient condition. These spectra were inverted with one HN model to estimate dielectric permittivity and relaxation time. To cross-validate experimental results, I have explored comprehensive MD simulations to estimate various dielectric and thermodynamic properties of the mixtures. These mixing properties computed from the MD simulations are able to capture the main aspects of experimental results. It is observed that dielectric relaxation time and mass diffusion coefficient of polar liquid mixtures are mainly governed by the average number of hydrogen bonds per unit volume. This is the consequence of the hydrogen bonding aggregates between alcohol hydroxyl groups and the tetrahedral water structures that hinder the collective motion and rotation of the molecules. Also, enhanced dipole moment of methanol and 2-propanol leads increased dielectric permittivity at the cost of increased density and decreased mass diffusion coefficients. This indicates that attractive interactions between molecules are slightly inflated compared to the repulsive

interactions. Thus, balanced partial charges for these molecules are required to compromise between dielectric and thermodynamic properties. Along with more refined modeling of molecules, optimized parameterization of the force fields should be scrutinized for the better estimation of liquid mixture properties.

Table 3.1: Measured dielectric permittivity and relaxation time for the three series of liquid mixtures at various concentrations at 298.15 K. The volumetric and molar fractions given in the table are the fraction of the second component.

Liquid mixtures	**X_V [V/V]**	**X_M [mol/mol]**	**Dielectric permittivity**	**Relaxation time [ps]**
Water and Methanol	0.0000	0.0000	80.2235	9.4088
	0.2500	0.1294	70.9251	20.5037
	0.4500	0.2673	62.6175	26.7417
	0.6002	0.4010	56.3463	32.3808
	0.7508	0.5732	48.7368	39.2639
	0.9009	0.8021	40.9499	55.0305
	1.0000	1.0000	34.8324	63.2194
Water and 2-propanol	0.0000	0.0000	80.2335	9.4088
	0.2857	0.0863	65.1739	27.7308
	0.4000	0.1360	56.9736	32.4633
	0.6129	0.2720	40.4826	55.0305
	0.7037	0.3592	34.3042	67.9373
	0.7997	0.4851	27.4433	106.5633
	0.8996	0.6790	23.3298	164.2381
	0.9300	0.7582	21.6109	206.8963
	1.0000	1.0000	20.7252	416.4861
Methanol and 2-propanol	0.0000	0.0000	34.8324	63.2194
	0.2500	0.1500	31.5460	83.5190
	0.4500	0.3022	28.7796	121.8446
	0.6500	0.4957	25.6558	166.3822
	0.8000	0.6792	23.4614	251.8723
	0.9000	0.8265	22.5057	285.7552
	1.0000	1.0000	20.7252	416.4861

Table 3.2: The number of molecules of water (N_W), methanol (N_M), and 2-propanol (N_P) and the corresponding box length used in the MD simulations.

Water and Methanol mixtures			**Water and 2-propanol mixtures**			**Methanol and 2-propanol mixtures**		
N_W	**N_M**	**L_{box} (nm)**	**N_W**	**N_P**	**L_{box} (nm)**	**N_M**	**N_P**	**L_{box} (nm)**
1000	0	3.11677	1000	0	3.11677	1000	0	4.04898
870	130	3.24513	940	60	3.26712	830	170	4.24661
750	250	3.36256	880	120	3.42546	660	340	4.40497
540	460	3.59335	630	370	3.98970	490	510	4.58746
360	640	3.73541	350	650	4.49634	320	680	4.75733
170	830	3.88923	220	780	4.69401	160	840	4.87248
0	1000	4.04898	0	1000	4.98465	0	1000	4.98465

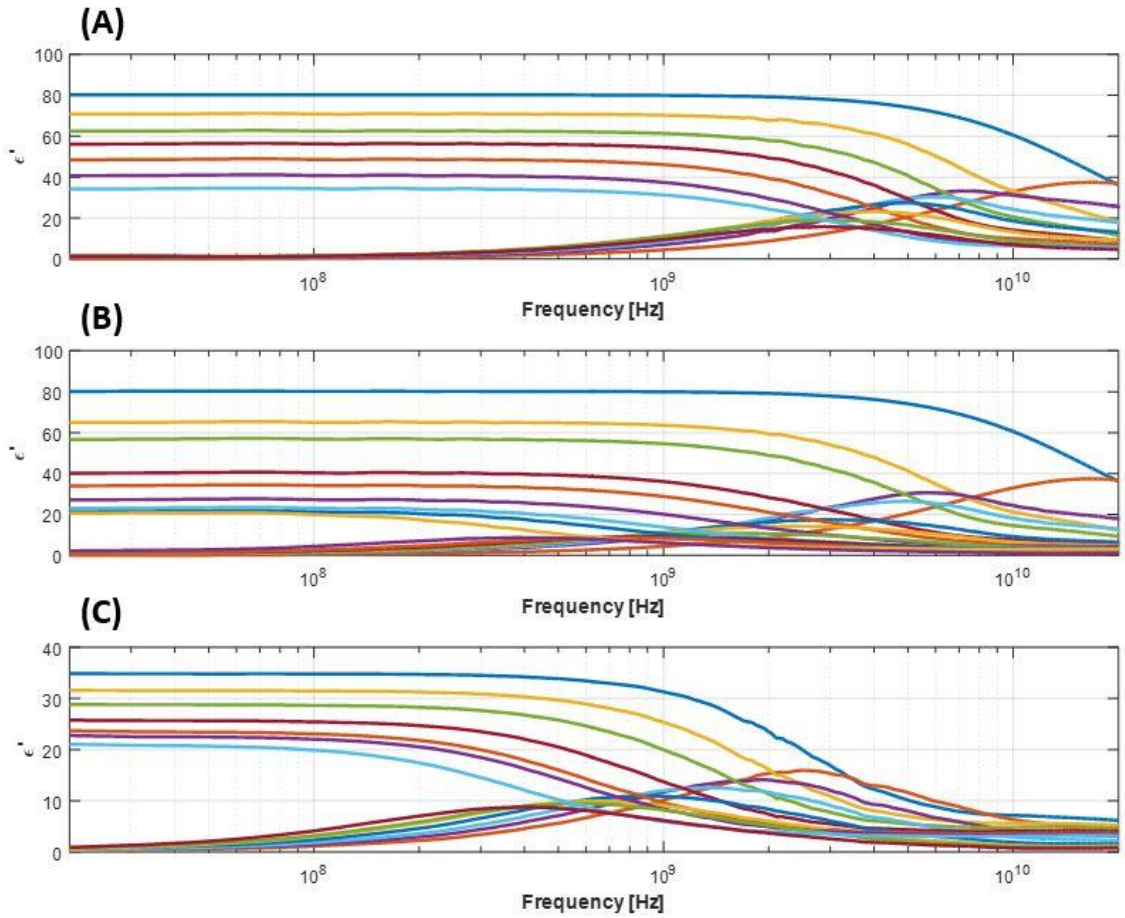


Figure 3.1: Frequency dependence of dielectric dispersion (ε′) and absorption (ε′′) spectra of the fluid mixtures of (A) water and methanol, (B) water and 2-propanol, and (C) methanol and 2-propanol at various concentrations at 298.15 K and 1 bar.

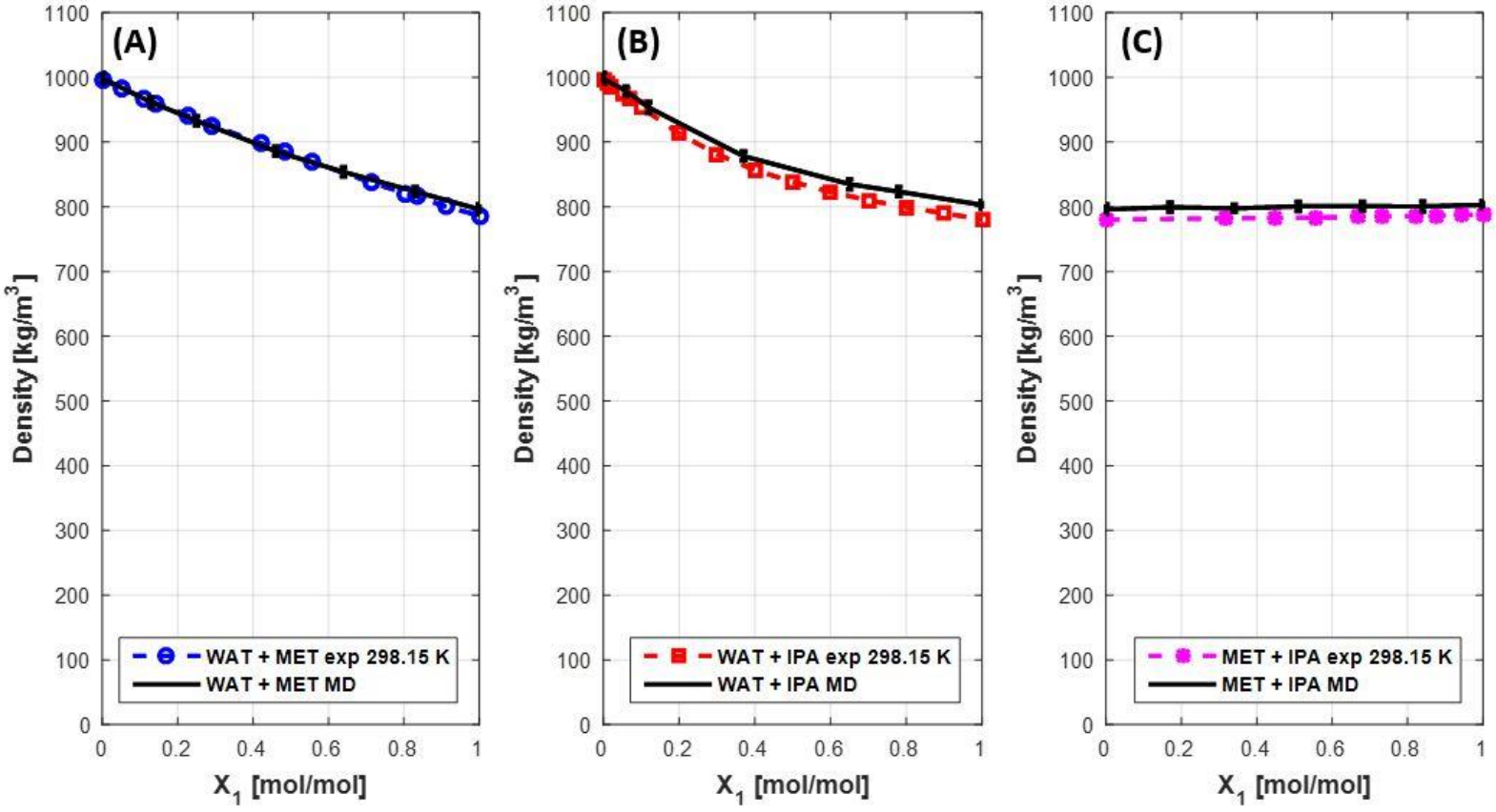


Figure 3.2: The density of (A) water and methanol, (B) water and 2-propanol, and (C) methanol and 2-propanol mixtures from NPT MD simulation results versus experimental data.

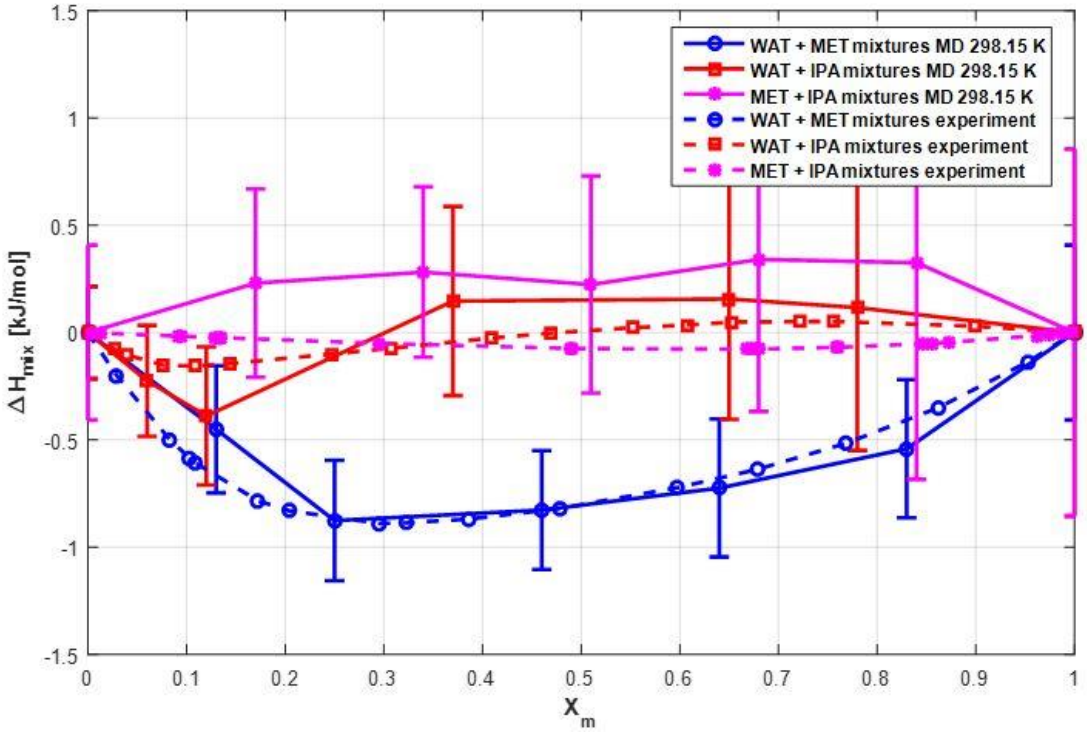


Figure 3.3: Estimated excess mixing enthalpy, ΔH_{mix} as a function of molar concentration of the second component at 298.15 K and 1 bar. The water and methanol mixtures (solid blue line with circles), water and 2-propanol mixtures (solid red line with squares), and methanol and 2-propanol mixtures (solid magenta line with stars) are compared with the experimental data (dashed lines). Simulation results show good agreement with experimental data over the entire composition range.

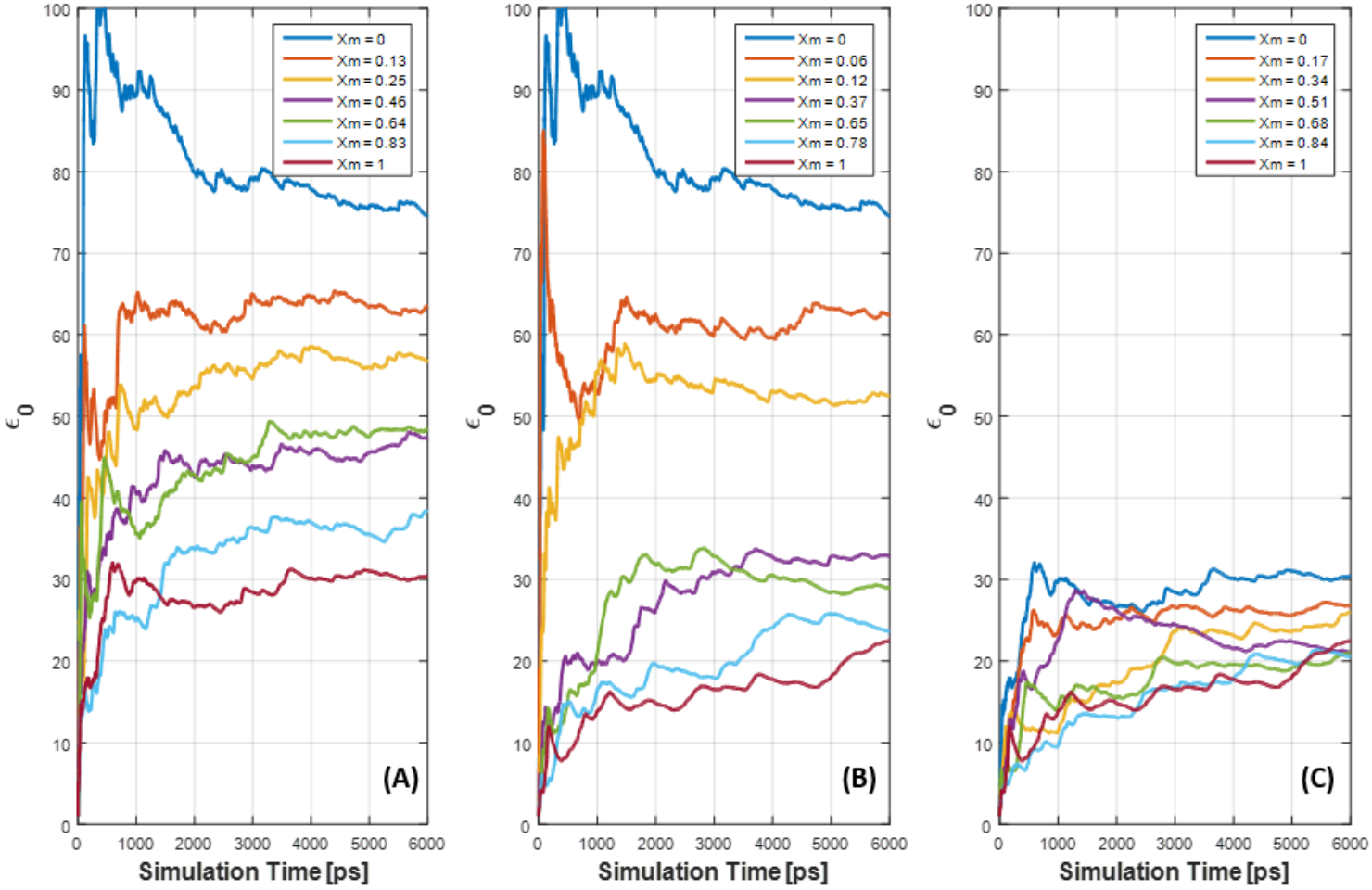


Figure 3.4: Estimated dielectric permittivity obtained by MD simulations as a function of time for (A) water and methanol, (B) water and 2-propanol, and (C) methanol and 2-propanol mixtures at different molar concentrations. The dielectric permittivity is stabilized to the asymptote values after 4 ns.

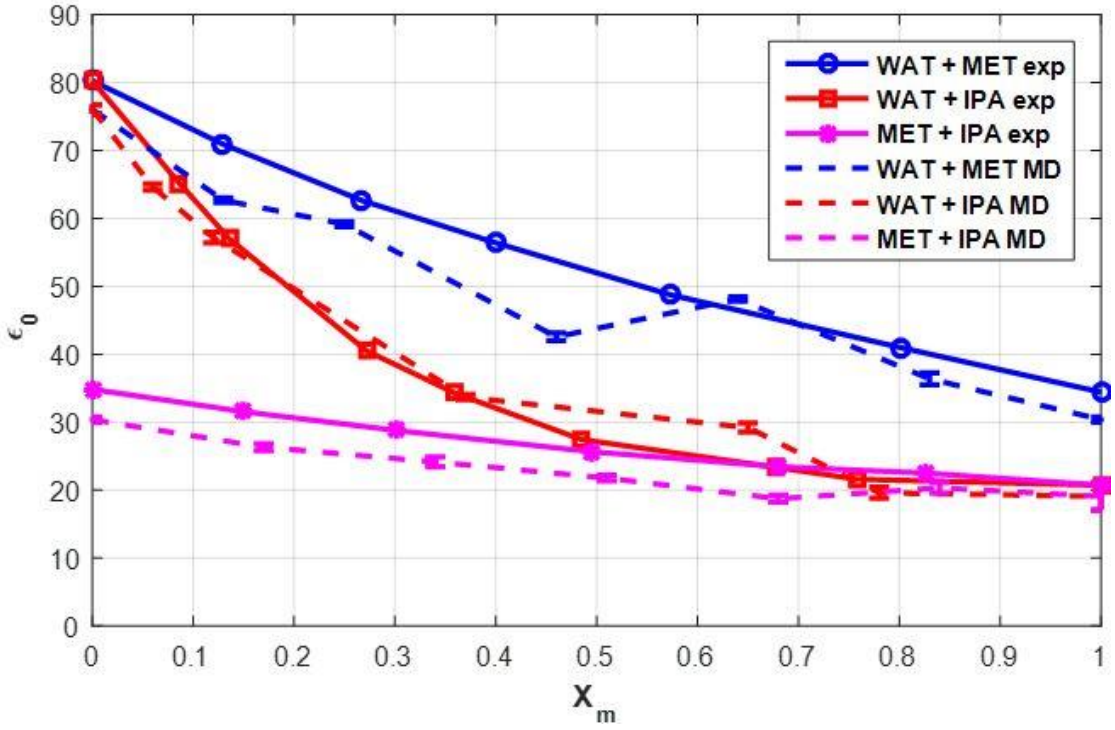


Figure 3.5: The dielectric permittivity as a function of the molar concentration of the second component at 298.15 K and 1 bar. The measured water and methanol mixtures (solid blue line with circles), water and 2-propanol mixtures (solid red line with squares), and methanol and 2-propanol mixtures (solid magenta line with stars) are compared with MD simulation results (dashed lines) with error bars.

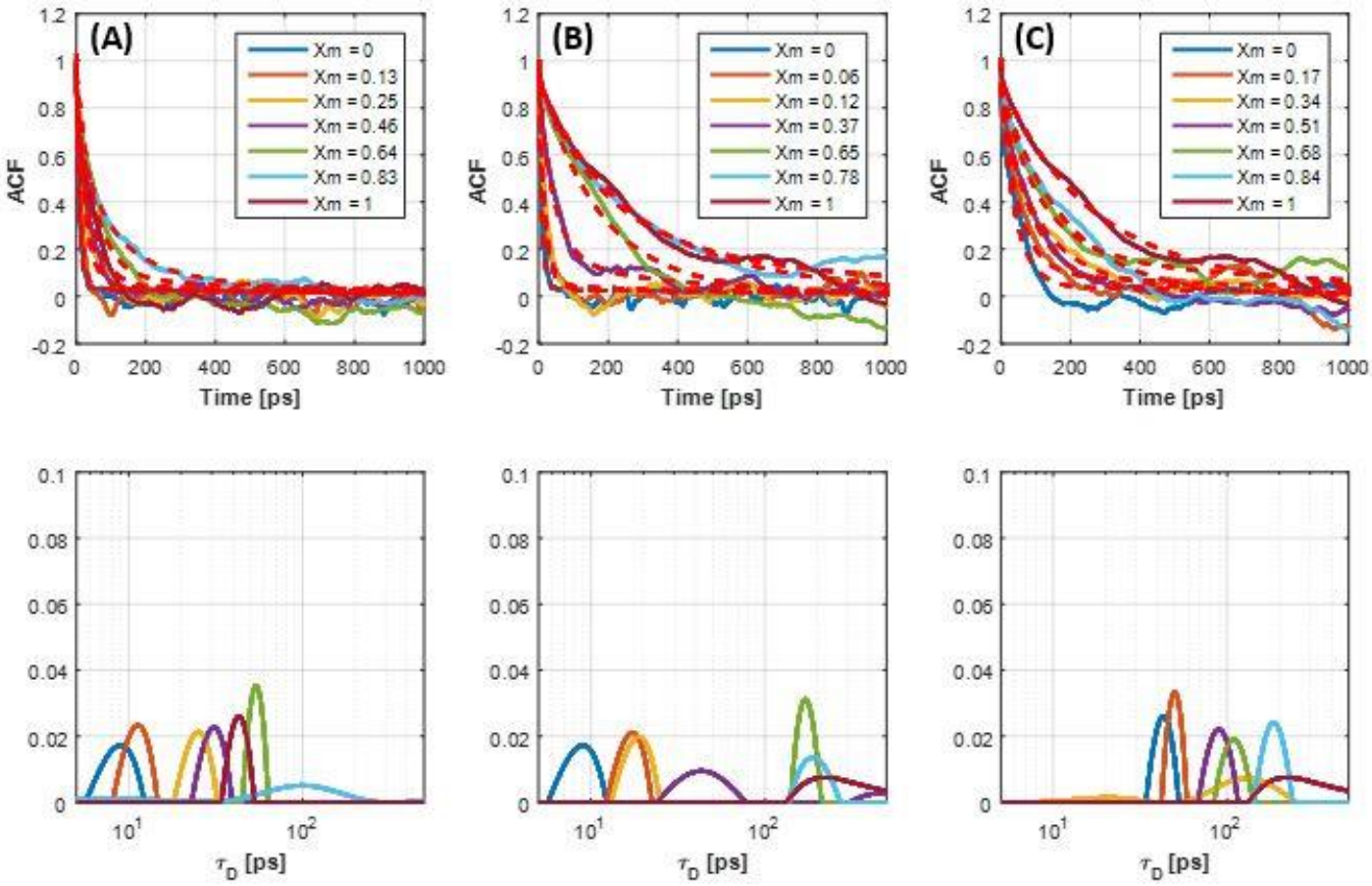


Figure 3.6: Normalized total dipole moment autocorrelation functions (ACF) for (A) water and methanol, (B) water and 2-propanol, and (C) methanol and 2-propanol mixtures at different molar concentration. The mathematical inversion is applied to the first 1,000 ps of each ACF to estimate dielectric relaxation time, and corresponding forward modeling is represented as red dashed lines. The three bottom figures represent the inversion results of the corresponding figure above.

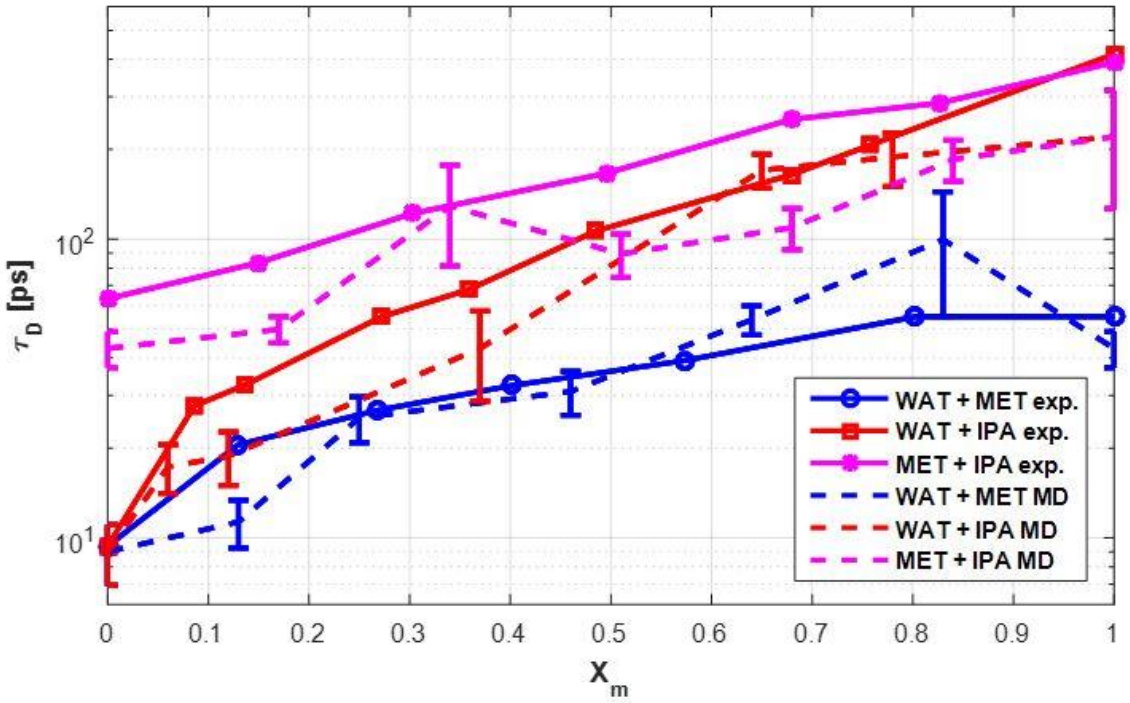


Figure 3.7: Comparison of measured dielectric relaxation times τ_D as a function of molar concentration of the second component for water and methanol mixtures (solid blue line with circles), water and 2-propanol mixtures (solid red line with squares), and methanol and 2-propanol mixtures (solid magenta line with stars) with the MD simulation results (dashed lines with error bars).

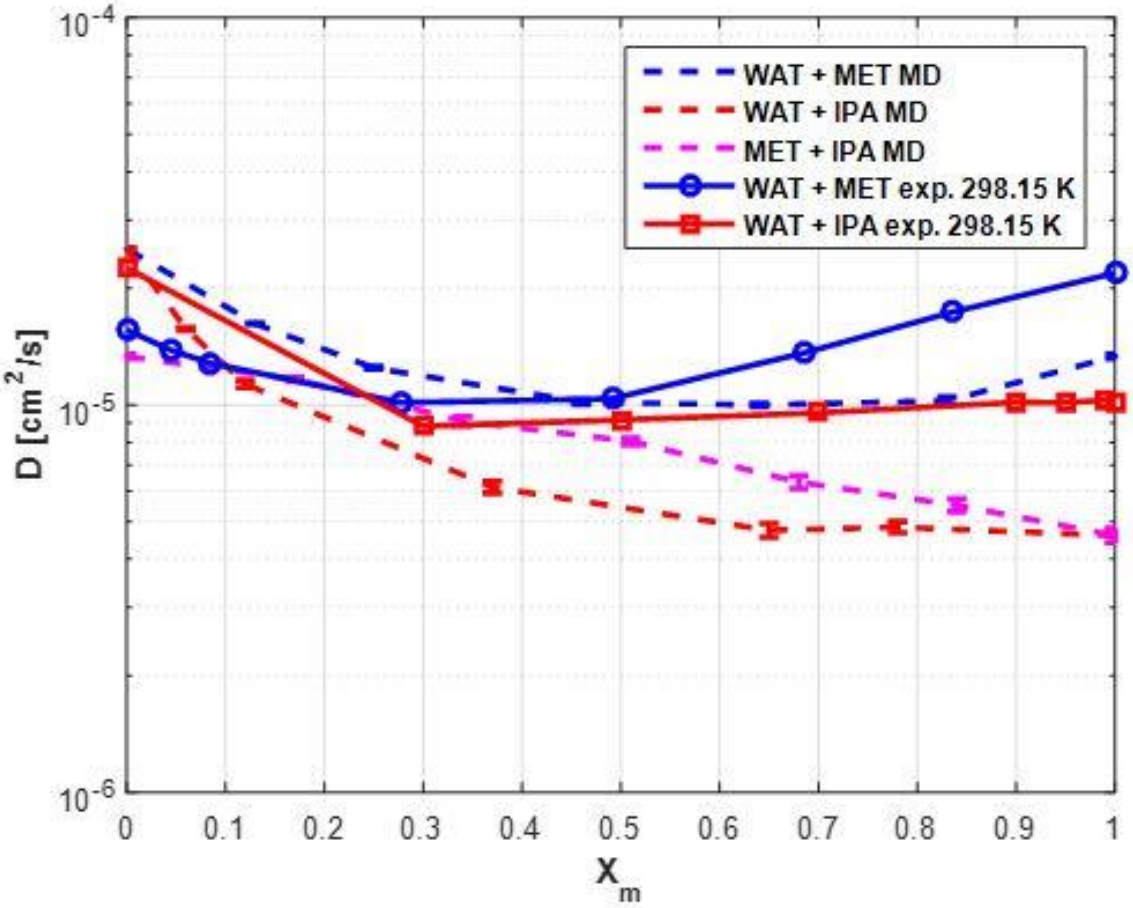


Figure 3.8: Mass diffusion coefficients calculated from the Einstein relation with MSD as a function of molar concentration of the second component at 298.15 K and 1 bar. The water and methanol mixtures (dashed blue line with error bars), water and 2-propanol mixtures (dashed red line with error bars), and methanol and 2-propanol mixtures (dashed magenta line with error bars) are compared with the experimental data (solid lines).

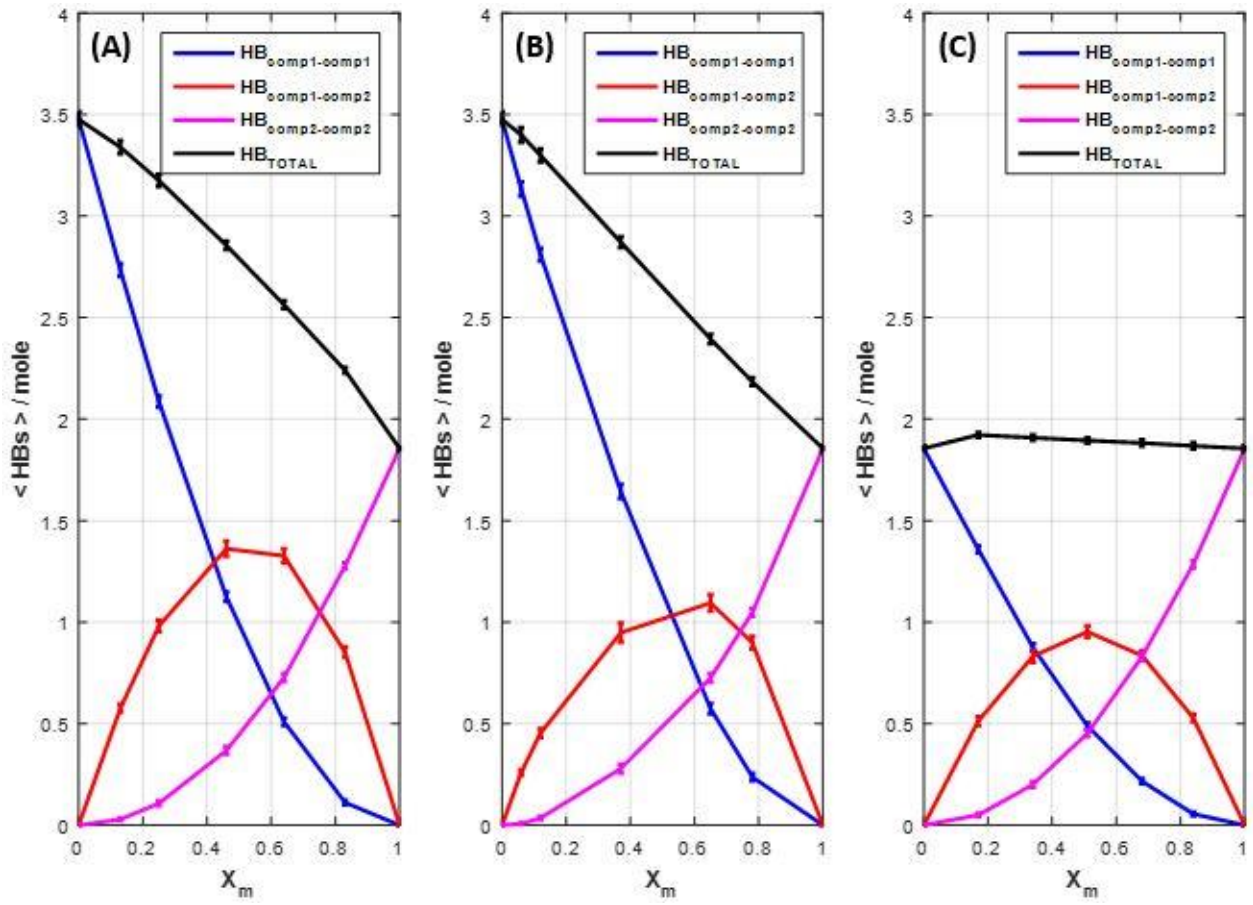


Figure 3.9: The average number of hydrogen bonds per molecule as a function of molar concentration of the second component at 298.15 K and 1 bar. With two geometric criteria (see description in the text), the average numbers of hydrogen bonds are presented in (A) water and methanol mixtures (solid blue line with error bars), (B) water and 2-propanol mixtures (solid red line with error bars), and (C) methanol and 2-propanol mixtures (solid magenta line with error bars).

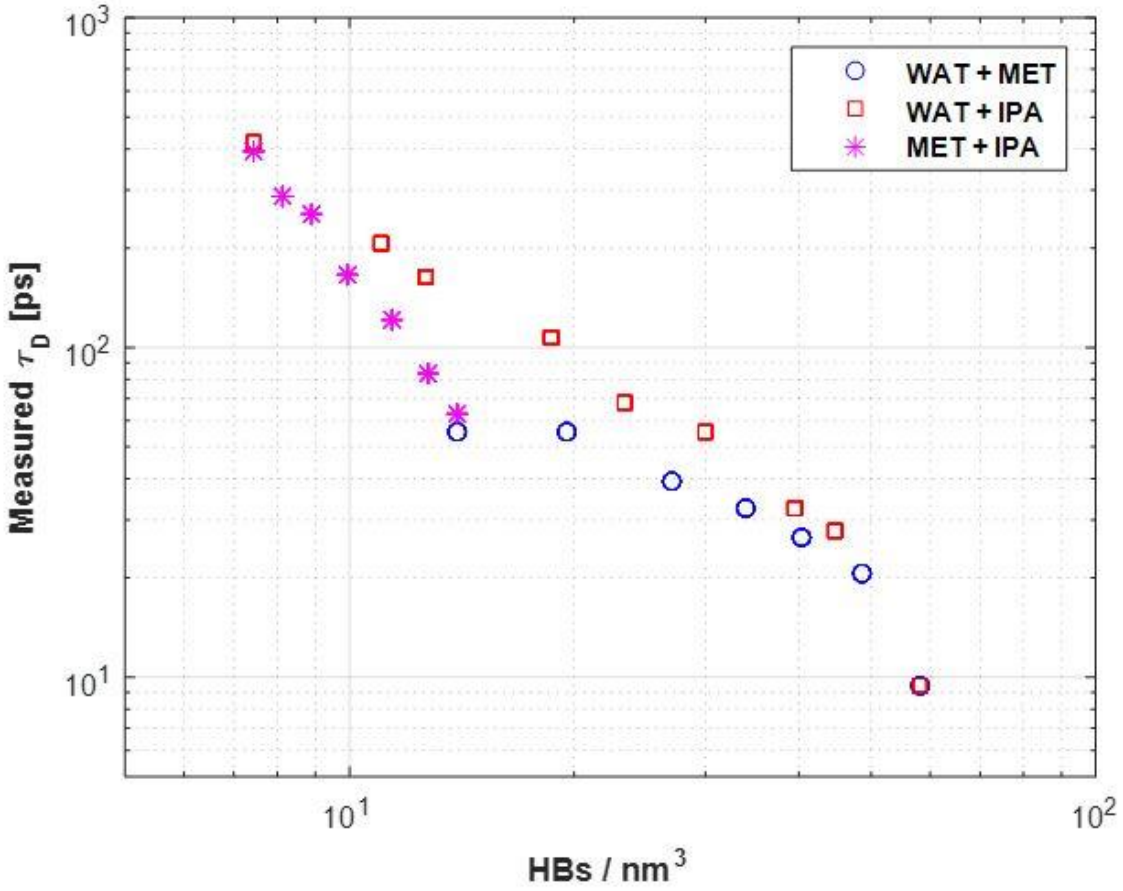


Figure 3.10: Measured dielectric relaxation time τ_D versus average number of hydrogen bonds per unit volume for water and methanol (o), water and 2-propanol alcohol (□), and methanol and 2-propanol (*) mixtures.

Chapter 4: Comparison of NIR, Infrared, and Raman Spectroscopies for the Estimation of Compositions and Physical Properties of Liquid Mixtures

This chapter proposes methodologies for the estimation of compositions and physical properties of liquid mixtures from various spectroscopic measurements. The three most common spectroscopies Near Infrared (NIR), Infrared, and Raman have been successfully applied to a series of fluid samples including Alkanes, Ethers, Alcohols, water, Cyclics, Aromatics, and their mixtures.

For NIR absorption spectra, Radial Basis Function (RBF) and Artificial Neural Network (ANN) approaches were separately applied to build correlations between spectral data and concentrations of each component. On these calibrations, the Principal Component Analysis (PCA) model was implemented. Both RBF and ANN methods were trained by the first 5 Principal Components (PCs) obtained from two hundred absorption spectra of liquid mixture samples. The trained systems were tested with 27 laboratory measurements and both results show excellent predictions of component concentrations with a Root-Mean-Square Errors (RMSE) of 2.5%. This result shows that RBF and ANN methods yield prediction error 50% less than standard multivariate methods.

I also compared the estimation results of physical properties using Partial Least Square Regression (PLSR) and ANN analysis for each spectral techniques. The PCA was implemented to each spectrum to extract PCs for the calibration and training of PLSR and ANN models. These models were tested with 51 mixture samples and the RMSE and correlation coefficient (R^2) were calculated to compare the prediction performance of these models. In the case of density estimation from ANN model, RMSE in the range of 0.7-1.1% (1.5-2.1%) for the training (test) data sets were obtained. For the viscosity

estimation with the ANN model, RMSE in the range of 4.4-5.6% (13-22.2%) were observed. ANN model decreases RMSE of 0.3-0.4% and 9.2-19.7% for the estimation of density and viscosity, respectively compared to the PLSR model. Furthermore, these errors can be decreased with a combined spectra of NIR, Infrared, and Raman. The reciprocal physics of these complementary spectroscopic techniques is key to improving analytical performance.

4.1 INTRODUCTION

Conventional techniques used to analyze Infrared spectra are multivariate analysis such as PCA (Sato, 1994; Fujisawa et al., 2008) and PLSR (Indo et al., 2015). Both PCA and PLSR provide linear calibration models for estimating components and concentrations of mixture samples. These works with NIR data are based on Beer-Lamberts law that the optical absorbance of a material is linearly proportional to the path length, absorptivity, and molecular concentration. However, the linearity of the Beer-Lambert law breaks down with

(1) Electrostatic interactions between molecules,

(2) Scattering of light,

(3) Very high or low concentrations,

(4) Chemical deviations due to pH changes.

Due to these reasons, summation of the individual component absorptions can be different from the absorption of the mixture, and conventional approaches may cause erroneous results.

To mitigate these non-linear correlations, I introduce two methodologies to estimate liquid mixture compositions using NIR spectra. The first method is RBF

interpolation that provides nonlinear approximation by using a mapping function from inputs to outputs. A linear combination of normalized Gaussian function depends on the Euclidean distance between the input points to the estimating point. The RBF method has been applied to many well-logging problems in that it can be used for any types of measurements with arbitrary dimensions of input and output data. Freedman (2006) applied RBF mapping function to nuclear measurements to predict formation Sigma. Anand et al. (2011) adapt RBF method to use porosity, irreducible water saturation, and NMR T_2 distributions to estimate effective permeability.

The second method is Artificial Neural Network (ANN). The artificial neural network is designed to mimic the neuron network of a human brain. The concept of neural networks was developed several decades ago and is widely used for pattern recognition in many areas. Hegeman et al. (2009) used ANN to estimate fluid Gas-Oil Ratio (GOR) and Falla et al. (2006) used ANN to predicted distillation properties of crude petroleum. Similar to RBF interpolations, the advantage of ANN is applicable to petrophysical problems without a priori knowledge of physical equations and forward models. The RBF and ANN calculations were performed with scripts programmed in the Matlab R2015b (Mathworks Inc.) environment.

I introduce the combined interpretation of NIR, Infrared, and Raman spectroscopies to analyze fluid properties. The three spectroscopic techniques are complementary with some materials being better suited for Infrared, while others are better suited for Raman. While Infrared spectroscopy is sensitive to the vibration of polar molecular bonds, which change the dipole moment, Raman spectroscopy relies on the inelastic scattering of photons. The energy from these photons is absorbed by non-polar bonds, which alter the polarization of the molecule, and is re-emitted at a different

frequency. Thus, Infrared spectroscopy information corresponds to the ionic bonds, and information from Raman spectroscopy is responsible to the covalent bonds of the molecule samples. For these reasons, some absorption peaks show strong amplitudes in NIR and Infrared, but they appear weak in Raman, and vice versa. These complementary techniques can improve the analytical performance of fluid property estimations.

To analyze spectroscopic measurements, I applied multivariate analysis to NIR, Infrared, and Raman spectra to estimate the density and viscosity of 51 liquid mixture samples. The PLSR is commonly used in spectral analysis and provides calibration models for estimating effective properties of mixture samples (Luinge et al., 1995; Kilner et al., 2011; Meksiarun et al., 2017). In addition to the traditional multivariate method, ANN model is applied for quantitative analysis.

This chapter is organized as follows: First, the laboratory experiment section describes the procedures of how NIR, Infrared, and Raman spectra were measured and prepared for the quantitative analysis. In the Radial basis function interpolation and Artificial neuron network sections, concepts and mathematical equations for model training are presented. The Results and discussion section compares the estimated component concentrations and RMSE for the assessment of compositions using each method. Also, PLSR calibration and ANN training models based on three spectroscopic measurements for the estimation of fluid density and viscosity will be presented. Additionally, the feasibility of the integrated spectroscopic measurements is investigated, along with multivariate and artificial neural network analysis, for the improvement of density and viscosity estimations. Finally, the prediction results obtained with NIR, Infrared, Raman, and their combined spectroscopies and two chemometric methods will be compared based on their corresponding RMSE and correlation coefficient squared R^2.

4.2 LABORATORY EXPERIMENT

Laboratory experiments were designed to appraise the performance of estimating components and concentrations, and petrophysical properties of fluid mixtures. A series of fluid samples were prepared by mixing DI water, Acetone, Cyclohexane, n-Decane, n-Heptane, 2-propanol, n-Pentane, Tetrahydrofuran, Toluene, and Methanol. These fluids are purchased from Fisher Scientific with 99% purity or higher. Density and viscosity for these fluid mixtures vary from 0.7 to 0.9 g/cm^3 and from 0.3 to 1.8 cP, respectively. These samples were chosen because they have a variety of chemical bonds representing their molecular structures and corresponding optical properties with absorption and scattering. For instance, the molecular structure of Tetrahydrofuran is similar to that of Cyclohexane, but it exhibits distinct absorption peaks due to COC bond. Immediately after mixing the fluids, samples were sealed in glass containers for spectral measurements. All measurements were taken at room temperature and ambient pressure. A set of 51 samples with independently varying concentrations of each component are presented in **Table 4.1**.

The NIR spectra of fluid samples were recorded on an Agilent Cary 5000 spectrometer at room temperature and ambient pressure. Samples were injected into Starna cells Quartz cuvettes with 1 mm optical path length and the spectral resolution was set to 1 nm. An empty cuvette filled with air was measured as the reference. The measured wavenumbers ranged from 1,000 to 2,500 nm (4,000 to 10,000 cm^{-1}). The background spectrum of air was recorded immediately before each measurement of the single-beam sample spectrum. This range covers the first and second overtones, and combination regions of chemical bonds of interests. The first overtone was used as the main target for spectrum analysis because the second overtone has a much lower intensity, due to excitation probability. Combination mode is much more complex to

interpret in that multiple excitations from stretching and bending blend with each other.

Confocal Raman spectra were obtained with a Witec Alpha 300 for Stokes scatterings. Samples were dropped onto the premium plain glass microscope slides specially coated with aluminum using the Cooke Ebeam/Sputter Deposition System. Aluminum coatings provide stable background scattering signals compared to commercial Reynolds aluminum foils. The spectral resolution was set to 1 cm^{-1}. An average number of accumulations of 5 scans and a typical integration time of 2 seconds was designed to minimize the noise level. An excitation wavelength of the laser was 488 nm. Aluminum background and fluorescence signals were removed manually.

For the Infrared spectra, Mattson Infinity Gold FTIR spectrometer from Thermo Mattson was used to measure absorption spectra of the samples. The spectrometer is equipped with a liquid nitrogen cooled Narrow Band MCT detector and Attenuated Total Reflectance (ATR) cell equipped with a GATR Germanium crystal. The spectral resolution was set to 2 cm^{-1} covering from 700 to 4,000 cm^{-1} spectral range and each spectrum has an average of 256 scans to ensure a high signal-to-noise ratio. Like NIR and Raman, background spectra of air was measured before each measurement for the baseline correction.

Measured NIR, Infrared, and Raman spectra require pre-treatment to minimize variations of spectra irrelevant to the change of concentrations. Such undesirable variations are irrelevant to the chemical property, which can mislead interpretations. Thus, it is important to apply appropriate pre-treatment to reduce such unrelated effects in order to construct more robust input data for the chemometric analysis.

The spectral truncation is the simplest and most practical procedure because the elimination of uninformative spectral regions can significantly improve the

performance of the calibration model (Jiang et al., 2002). After truncation, baseline corrections were applied to the selected wavelength region. Among commonly used pre-treatment methods such as Derivatives, Standard Normal Variate (SNV, Barnes et al., 1989), Multiplicative Scatter Correction (MSC, Dhanoa et al., 1994), and Extended Multiplicative Signal Correction (EMSC, Martens et al., 2003), I choose SNV due to the effectiveness in scattering correction. The derivatives are not used because of noise enhancement. These pre-treatments allow more robust input data for RBF and ANN models in that it minimizes variations of absorption spectra irrelevant to the change of concentrations.

For NIR spectral measurements, I applied PCA method for dimension-reduction. It is a mathematical procedure that uses an orthogonal transformation to convert correlated variables, which is measured spectra, into a set of linearly uncorrelated variables called principal components. The first five PC account for 99.02% of the variability in the spectrum. Thus, instead of the entire spectrum, these five PCs can be used as input data for the RBF and ANN analysis to reduce computational time. The input vector $\boldsymbol{x}$ has 5 PCs extracted from NIR spectra and output vector $\boldsymbol{y}$ has 6 concentrations of Acetone, Cyclohexane, Decane, 2-propanol, Tetrahydrofuran, and Toluene. The data is divided into two sets, training data and test data. Training data is prepared with mixtures of random combinations of component concentrations constructed by Beer-Lamberts law. Laboratory measured absorption spectra and extracted principal components used for the test data is presented in **Figure 4.1**. The distinct features of each component and corresponding NIR spectral peaks are summarized in **Table 4.2**.

The three sets of optical spectra, NIR absorption, Infrared absorption, and Raman Stokes scattering, have different characteristic bands. Observed NIR absorption spectra

correspond to first overtones of CH and OH stretches (5,000 - 6,000 cm^{-1}) and combination regions (4,000 - 5,000 cm^{-1}). On the other hand, Infrared and Raman scattering spectra show fundamental stretching (2,800 - 3,100 cm^{-1}) and bending (500 - 1,800 cm^{-1}) modes associated with carbon, hydrogen, and oxygen bonds. Baseline corrections were applied to each measured spectrum to reduce various instrumental and scattering effects. Experimental results show that each of these spectra can establish calibration models for fluid mixture samples.

4.3 METHODS

To analyze measured optical spectra, several techniques are applied including Principal Component Analysis (PCA), Partial Least Squares Regression (PLSR), Radial Basis Function (RBF), and Artificial Neural Networks (ANN). This section describes multiple approaches to investigate optical spectra measured from three different spectroscopies.

4.3.1 Principal Component Analysis

Principal Component Analysis (Pearson, 1901) is a widely used statistical method to transform the data of possibly correlated variables into a set of linearly uncorrelated variables called Principal Components (PC). This method is an analogue of the Singular Value Decomposition (SVD) or Eigenvalue Decomposition (EVD) technique in that it is based on the covariance and eigenvector analysis. Variance and covariance are important terms in describing PCA. Variance can be explained as a measure of the data spread, and covariance is expressed as a joint variability of two data. Given a data of n points with p variables as,

$$x = (x_1,\ x_2,\ ...,\ x_p). \tag{4.1}$$

The *k*th PC of the data is defined by the linear transformation as,

$$z_k = a_k^T x \quad (k = 1,\ 2,\ ...,\ p), \tag{4.2}$$

where a_k is an eigenvector of the covariance matrix S corresponding to eigenvalue λ_k for the variable x. The eigenvector is chosen such that the variance of the z_k is maximized. In other words, PCA finds a normalized direction in p dimensional space that makes the variance of data is maximized. The benefit of PCA is a dimension reduction or data compression. Because a large amount of variances are mainly associated with the first several principal components, one can discard the other insignificant components.

4.3.2 Partial Least Square Regression

Another statistical method applied to optical spectroscopy measurements is Partial Least Squares Regression (PLSR, Wold et al., 2001) that finds a linear regression model by projecting the predicted variables and the measured variables to a new space. A PLSR model finds the multidimensional direction in the X space that explains the maximum multidimensional variance direction in the Y space as,

$$X = T P^T + E, \tag{4.3}$$

$$Y = U\,Q^T + F, \tag{4.4}$$

where X and Y are matrices of predictors and responses, T and U are projections of X and Y, P and Q are orthogonal loading matrices for the projected X and Y scores, and E and F are the residual terms for X and Y, respectively. The decompositions of X and Y are evaluated to maximize the covariance between T and U. Therefore, PLSR is identifying the features that explain the most variance between the predictors and responses; use these features to predict the responses. Both PLSR and PCA are linear decomposition techniques to reduce dimensions but construct their components differently. While components created from PCA explain the variability in the predictor variables (X) without considering the response variables (Y), PLSR takes account the response variables (Y) to build models.

4.3.3 Radial Basis Function

Radial Basis Function (RBF) is used to build up approximations of the form

$$F(x) \equiv y = \sum_{i=1}^{N} w_i \phi\left(\left\|x - x_i\right\|\right), \tag{4.5}$$

where the interpolation function *y(x)* is a sum of N radial basis functions. w_i is a weighting coefficient, which can be calculated using linear least square method. ϕ is a type of radial basis functions with Gaussian function as,

$$\phi(x, x_i) = \exp\left(-\frac{\left\|x - x_i\right\|^2}{2\sigma^2}\right). \tag{4.6}$$

The distance between two features x and x_i is recognized as the squared Euclidean distance, and σ is a standard deviation that controls the width of the distributions. The most important feature of RBF is a strong dependence on the distance. This means the points far from x have negligible contributions to RBF. This localized Gaussian RBF can be normalized as

$$\phi(x, x_i) = \frac{\exp\left(-\frac{\|x - x_i\|^2}{2\sigma_i^2} \right)}{\sum_{j=1}^{N} \exp\left(-\frac{\|x - x_j\|^2}{2\sigma_j^2} \right)}. \tag{4.7}$$

Using Nadaraya-Watson Regression Estimator (NWRE, Nadaraya, 1964) with a simple approximation of neglecting the overlap of RBFs, the weighting coefficient w_i in equation (4.5) can be replaced to y_i and the estimator becomes a linear smoother. Coupling the equation (4.5) and (4.7) yields

$$F(x) = \frac{\sum_{i=1}^{N} y_i \exp\left(-\frac{\|x - x_i\|^2}{2\sigma_i^2} \right)}{\sum_{i=1}^{N} \exp\left(-\frac{\|x - x_i\|^2}{2\sigma_i^2} \right)}, \tag{4.4}$$

which is the NWRE mapping function. This equation is a weighted average of the output data with RBF governed by the Euclidean distance. The equation (4.4) was used for data training and the trained interpolation surface is shown in **Figure 4.2**.

For visualization, the figure plots only PC1 and PC2 but it can be expanded to an arbitrary dimension.

4.3.4 Artificial Neural Networks

A neural network with three layers used in this study is represented in **Figure 4.3**. The first layer, called input layer, has n_{i+1} neurons, where n_i is the number of input units and the additional neuron is called the bias unit, which always has a value of one. This bias value is important for training because it allows the activation functions to shift. Similarly, the numbers of neurons in the hidden layer and output layer are represented as n_h and n_o, respectively.

The regularized cost function for the neural network is given by

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{n_o}\left[-y_k^{(i)}\log((h_\theta(\mathrm{x}^{(i)}))_\mathrm{k}) \; - \; (1-y_k^{(i)})\log(1-(h_\theta(\mathrm{x}^{(i)}))_\mathrm{k})\right] \; + \; \frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{j=1}^{n_h}\sum_{k=1}^{n_i}(\Theta_{j,k}^{(l)})^2 \,, \quad (3.5)$$

where m is a number of input data, L is a total number of layers, y_k is the concentration of a kth component, λ is a regularization parameter, $\Theta^{(l)}_{j,k}$ is a weight from a k-th unit of l layer to a j-th unit of the $l+1$ layer. These weights are randomly initialized for symmetry breaking and are computed to the optimized values through the training process. Also, $h_\Theta(x)_k$ is the activation of the k-th output unit as

$$h(x) = \frac{1}{1+e^{-x}} \,. \quad (3.6)$$

The sigmoid function was applied to an activation function (Cybenko, 1989); it enables a smooth transition as input values vary. I implemented feedforward

propagation for neural networks to predict component concentrations. Feedforward neural network computes $h_\Theta(x)_k$ for every input data, and summation of all costs. Once cost functions are calculated, backpropagation can be implemented to calculate gradients (Rumelhart et al., 1986). The idea of backpropagation is that activations calculated from feedforward can be compared to the true output values and measure how much each neuron is subject to the difference. This is an iterative process of gradient descent to minimize the overall cost that simultaneously calculates weights for neurons and gradients. The difference between gradients calculated from backpropagation and numerical gradients is very small, showing that the backpropagation provides a good approximation of gradients with much less computation time.

To avoid an overfitting problem, an optimal regularization parameter was calculated by trying a range of different regularization parameters. Comparing the costs on a cross-validation data, selection of a regularization parameter that yields the minimum error is shown in **Figure 4.4**. Learning curves represented in **Figure 4.5** can diagnose whether training is suffering from an overfitting problem. It shows training and cross-validation errors as a function of training data size. As the training data size increases, average training errors increase and average cross-validation errors decrease because training with larger number of data ensures better generalizing to new data.

4.4 RESULTS AND DISCUSSION

Characterizing the performance of trained models requires a test dataset. The test data is independent of the training process but has the same probability distributions.

Thus, cost functions calculated from test data can provide an unbiased evaluation of the trained model**. Figure 4.6** and **figure 4.7** show the comparisons between estimated and actual concentrations for RBF and ANN methods, respectively. The excellent prediction capability is observed for both methods by comparing the actual and predicted concentrations of test data. To assess the accuracy, Root Mean Squared Error (RMSE) was calculated for the predictions. **Figure 4.8** shows RMSE of individual components calculated for both training data and test data with RBF and ANN methods. While RBF could predict the concentration with an average RMSE of 2.01%, ANN shows similar prediction with an average RMSE of 2.27%. Traditional multivariate methods provide an average RMSE of 5.17% with the same test data and be considered as outliers because they deviate from Beer-Lambert law. However, RBF and ANN can successfully adapt those data with errors of less than half compared to the traditional method.

It is worth mentioning that the RMSE of Cyclohexane and Decane estimated from RBF method is larger than others. This is because these components have comparable chemical structures of methylene. The proximity of methylene absorption peaks of Cyclohexane and Decane leads to similar principal component values. Corresponding Euclidean distance is closer than other components and this degrades the estimation of Cyclohexane and Decane concentrations. Decreasing grid size can improve the accuracy of RBF estimation compared to coarser interpolation. Interpolation with a smaller grid size requires more CPU time, so certain compromises are required between computation time and accuracy. Neural networks with increased number of hidden layers were also tested. However, a MultiLayer Perceptron (MLP) with multiple hidden layers tends to overfit the training data with the structure of 5 PC inputs and 6 component concentration outputs. For a number of neurons in the hidden layer, it is recommended to have a similar size of input and output layer to avoid overfitting problem.

The three spectral measurements of NIR, Infrared, and Raman spectra of liquid mixtures are presented in **Figure 4.9**. While the absorption bands of NIR are broad and overlap, Infrared absorption and Raman scattering signals are sharp with a lower signal-noise ratio. The main reason of this broad absorption peaks is collisional broadening. I attempted property estimations with the full spectral range, but the calibration and training results were poor. Therefore, spectral truncations are applied to eliminate uninformative spectral regions, which can significantly improve the estimation accuracy. For NIR spectra, I chose the first overtone as a main target. This is because the second overtone has a much lower intensity than the first due to excitation probability. Also, the combination region has multiple peaks blending with each other, which make interpretations more complex. For the Infrared and Raman spectra, bending vibrational modes (500 - 1,500 cm^{-1}) are chosen over the fundamental stretching region (2,700 - 3,100 cm^{-1}) because they have well-resolved peaks of various vibrational motions such as rocking, scissoring, wagging, and twisting. Details of selected wavenumber ranges and corresponding functional groups for NIR, Infrared, and Raman spectra are summarized in **Table 4.3**. These bands deliver the most significant information about molecular structures; thus, they can discriminate different physical properties of liquid mixtures. Note that the OH band is strong in the NIR and Infrared spectra but weak on Raman scattering.

To compare NIR, Infrared, and Raman spectra, each spectral data are L_2-norm normalized to make sure intensities are on the same scale for comparison. Standard Normal Variate (SNV) was applied (Barnes et al., 1989) to normalized spectra due to the effectiveness of the scattering correction. The derivatives are not used because of noise enhancement. Thereafter, PCA is applied for dimension-reduction (Pearson, 1901). The PCA is a mathematical procedure that uses an orthogonal transformation to convert

correlated variables (measured spectra) into a set of linearly uncorrelated variables (principal components). The first seven scores account for 99.29% of the total variation in the spectra. Thus, instead of the whole spectrum, these seven scores can be used as input data to reduce computational time. The PCA score plot with the first three principal components, shown in **Figure 4.10**, describes fluid mixtures categorized by the variance of measured optical spectra.

For quantitative analysis of spectroscopy data, PLSR and ANN were implemented to determine the density and viscosity of liquid mixtures. The performance of PLSR and ANN models with NIR, Infrared, Raman, and the combinations of these spectra are compared to estimate density and viscosity of various liquid mixtures. Density and viscosity of liquid mixtures were prepared based on the CoolProp (Bell et al., 2014) database, which is an open-source thermophysical property library. These thermodynamic properties, obtained from multiparameter Helmholtz-energy-explicit-type formulations, provide an equivalent accuracy level as the REFPROP (Lemmon et al., 2013), which is the most generally used commercial library from the United States National Institutes of Standards and Technology (NIST). All data processing, PLSR and ANN calculations were performed with scripts programmed in the Matlab R2015b (Mathworks Inc.) environment.

The input vector has seven scores extracted from spectra, and output vector includes petrophysical properties such as density and viscosity of the corresponding fluid mixture. Measured data is divided into training data and test data. The training data consists of pure component sample spectra and their mixtures with random combinations of concentrations. Characterizing the performance of calibrated PLSR and trained ANN models requires a test dataset. Test data is prepared with 51 optical spectra of liquid mixtures measured from laboratory, which are independent of the calibration or training

process but have the same probability distributions. Thus, the cost function with test data can provide an unbiased evaluation.

To find the fundamental relations between input and output data, ANN algorithms are applied. Details of ANN cost function, feedforward (Svozil et al., 1997), and backpropagation (Rumelhart et al., 1986) are presented in **Appendix A**. To avoid an overfitting problem, the optimal regularization parameter was calculated by trying a range of different regularization parameters and observing which gives the minimum costs on cross-validation data. As the training data size increases, average training error increases and average cross-validation error decreases because training with more data ensures better generalizing to new data. To compare two methods, I used the same number of samples to calibrate PLSR and to train ANN.

Seven different input data are considered to assess and compare NIR, Infrared, and Raman spectroscopy measurements. The input data and their corresponding PLSR and ANN training and test results are tabulated in **Table 4.4**. Comparisons between estimated and actual densities for PLSR and ANN methods are illustrated in **Figure 4.11** and **Figure 4.12**, respectively. The excellent prediction capability is observed for both methods by comparing the actual and estimated values of test data. Note that NIR, Infrared, and Raman inputs for the density estimation presents RMSE values less than 2.6, for both PLSR and ANN models. Viscosity estimations with PLSR and ANN are presented in **Figure 4.13** and **Figure 4.14**, respectively. In general, estimated density and viscosity obtained with the ANN model have moderately smaller RMSE values than those obtained with the PLSR model. This can be explained by the fact that ANN can account for the nonlinearity of measured spectra, which is explicitly excluded in traditional approaches. Estimation results of viscosity with PLSR and ANN are summarized in values in **Table 4.5**.

Upon comparing three spectroscopic measurements, Infrared and Raman spectra achieved slightly better prediction results than NIR in terms of RMSE values for the estimation of density and viscosity. NIR alone provides the highest RMSE values because Infrared and Raman spectra contain information of bending vibrations directly related to the molecular structures. On the other hand, the first overtone of NIR spectra have only stretching vibrations, which highly overlap each other. Also, Infrared spectra work better than Raman spectra in the estimation of density and viscosity for both PLSR and ANN models.

The results show that the prediction accuracy can be improved by combining the multiple spectra. The combined spectra of NIR, Infrared, and Raman for the estimation of density and viscosity are more accurate than the single spectral input. This is important for downhole applications because NIR and Infrared spectra are highly vulnerable to the contamination of Water-Based Mud (WBM) filtrate due to its strong dipole moment. Raman spectra are relatively less influenced because they are only sensitive to a molecule polarizability. Also, the combination of Infrared and Raman spectra is more accurate than the combination of NIR and Infrared or NIR and Raman spectra except the training with ANN model. Also, there is no notable difference in the density estimations between the combination of NIR and Infrared spectra, and the combination of Infrared and Raman spectra.

To examine the prediction capability of these models for the multiple liquid mixtures, the number of components used in the liquid mixtures was gradually increased. As the number of components increases from 2 to 10 by subsequently adding a new liquid, resulting each time in a discrete mixture, RMSE increases and corresponding correlation coefficient R^2 decreases. These results are presented in **Figure 4.15** and **Figure 4.16** for the density and viscosity, respectively. The abrupt increases of prediction

errors in component 9 and 10 are mainly related to the spectral similarity of n-Decane to those of n-Heptane and n-Pentane.

Despite these accurate predictions, there are some limitations for the downhole application. Sample contaminations with water can deteriorate the prediction performance because substantial absorption peaks caused by strong dipole moment of water will blend with other absorption peaks. Also, Infrared cannot detect homonuclear diatomic molecules because the change in the dipole moment with respect to a change in the vibration is zero. In this case, Raman scattering can be an alternative option. However, strong fluorescence needs to be properly handled for Raman scattering spectra. Additionally, in order to apply ATR FTIR into downhole wireline tools, the detector must be maintained at a low temperature to avoid thermal noise. Another drawback is that ANN training process takes a longer time than multivariate methods and is computationally expensive. Nevertheless, NIR, Infrared, and Raman spectroscopies combined with PLSR, RBF or ANN analysis is suitable for predicting petrophysical properties of liquid mixtures.

4.4 CONCLUSIONS

Developing a calibration model to estimate component compositions from nonlinear NIR absorption spectra has proved challenging. Contrary to the traditional multivariate methods, this nonlinearity is explicitly included in RBF and ANN approaches. In order to have accurate predictions, building a robust training model is a key factor; requires pre-treatments of NIR absorption spectra, optimization of the regularization parameter, and constructing a mapping function. I have demonstrated that well-trained RBF and ANN models can successfully estimate concentrations of each

component and the prediction error for test data shows remarkable reduction compared to the traditional methods.

Also, I demonstrated quantitative comparisons between NIR, Infrared, and Raman spectral data for the prediction of density and viscosity. Both calibrated PLSR and trained ANN models were successfully associated with physical properties of liquid mixtures to corresponding spectral data. The spectral pre-treatment including baseline corrections, truncations, and normalizations is important to maximize the correlations between spectral features and liquid properties. These spectra and their combinations associated with PLSR and ANN models yield accurate estimations of fluid properties with low prediction RMSE.

Comparisons of density and viscosity assessments conclude that spectroscopic measurements associated with ANN model yield considerable reduction of RMSE compared to the traditional methods. This is because nonlinearity is explicitly included in the ANN approach, which is not in PLSR methods. Also, complementary spectroscopic techniques can be incorporated together to improve analytical performance of liquid characterizations. Hence, Infrared and Raman detectors can be useful tools to characterize downhole in-situ fluids. Although the spectral analysis presented in this chapter is focused on the estimation of composition, density and viscosity, the same methodology can be used for evaluating other petrophysical and geophysical properties.

Table 4.1: The list of liquid mixture samples with corresponding components and concentrations. Water (WAT), acetone (ACE), cyclohexane (CYC), decane (DEC), heptane (HEP), 2-propanol (IPA), pentane (PEN), Tetrahydrofuran (THF), toluene (TOL), and methanol (MET) are mixed with different concentrations.

	WAT	**ACE**	**CYC**	**DEC**	**HEP**	**IPA**	**PEN**	**THF**	**TOL**	**MET**
Sample 1	0	0.86	0.14	0	0	0	0	0	0	0
Sample 2	0	0.60	0	0	0	0.40	0	0	0	0
Sample 3	0	0.53	0	0	0	0	0	0.47	0	0
Sample 4	0	0	0	0	0	0.16	0	0.84	0	0
Sample 5	0.13	0	0	0	0	0	0	0	0	0.87
Sample 6	0	0.42	0	0	0	0	0	0	0	0.58
Sample 7	0	0	0.59	0	0	0	0	0	0	0.41
Sample 8	0	0	0	0	0	0.18	0	0	0	0.82
Sample 9	0	0	0	0	0	0	0	0.47	0	0.53
Sample 10	0	0	0	0	0	0	0	0	0.22	0.78
Sample 11	0	0	0.96	0	0	0	0	0	0	0
Sample 12	0.11	0	0.89	0	0	0	0	0	0	0
Sample 13	0	0	0.86	0	0.14	0	0	0	0	0
Sample 14	0	0	0.54	0	0	0	0.46	0	0	0
Sample 15	0	0	0	0.21	0	0.79	0	0	0	0
Sample 16	0	0	0	0	0	0.94	0.06	0	0	0
Sample 17	0	0	0	0	0	0.48	0	0	0.52	0
Sample 18	0	0	0.35	0	0	0	0	0	0.65	0
Sample 19	0	0	0	0.06	0	0	0	0	0.94	0
Sample 20	0	0	0	0	0.35	0	0	0	0.65	0
Sample 21	0	0	0	0	0	0	0.17	0	0.83	0
Sample 22	0	0	0	0	0	0	0	0.56	0.44	0
Sample 23	0	0	0.36	0	0	0	0	0.64	0	0

Table 4.1 continued.

	WAT	ACE	CYC	DEC	HEP	IPA	PEN	THF	TOL	MET
Sample 24	0	0	0	0	0	0	0.59	0.41	0	0
Sample 25	0	0	0.43	0	0.57	0	0	0	0	0
Sample 26	0	0	0	0.79	0	0	0.21	0	0	0
Sample 27	0	0	0	0	0	0.82	0	0.18	0	0
Sample 28	0	0.56	0	0	0	0	0.44	0	0	0
Sample 29	0	0.99	0	0	0	0	0	0	0.01	0
Sample 30	0	0	0	0	0	0.25	0	0	0	0.75
Sample 31	0	0	0	0.20	0	0.80	0	0	0	0
Sample 32	0	0	0	0	0.98	0	0	0	0.02	0
Sample 33	0	0	0	0	0.99	0	0	0	0.01	0
Sample 34	0	0.14	0	0	0.86	0	0	0	0	0
Sample 35	0	0.09	0	0	0	0	0.91	0	0	0
Sample 36	0	0.94	0	0	0	0	0	0	0.06	0
Sample 37	0	0	0.47	0.53	0	0	0	0	0	0
Sample 38	0	0	0.35	0	0.65	0	0	0	0	0
Sample 39	0	0	0.78	0	0	0	0.22	0	0	0
Sample 40	0	0	0.34	0	0	0	0	0.66	0	0
Sample 41	0	0	0.55	0	0	0	0	0	0.45	0
Sample 42	0	0	0	0.52	0	0	0.48	0	0	0
Sample 43	0	0	0	0.34	0	0	0	0	0.66	0
Sample 44	0	0	0	0	0.44	0.56	0	0	0	0
Sample 45	0	0	0	0	0.67	0	0.33	0	0	0
Sample 46	0	0	0	0	0.19	0	0	0.81	0	0
Sample 47	0	0	0	0	0	0.36	0.64	0	0	0
Sample 48	0	0	0	0	0	0.29	0	0	0.71	0
Sample 49	0	0	0	0	0	0	0.23	0.77	0	0
Sample 50	0	0	0	0	0	0	0.12	0	0.88	0
Sample 51	0	0	0	0	0	0	0	0.74	0.26	0

Table 4.2: The diagnostic absorption ranges of NIR spectra for functional groups of components used in the study.

Functional Group	Band position in nm	Component
CH_3	1690-1700	Decane, Cyclohexane, Toluene, 2-propanol, Acetone
CH_2	1720-1760	Decane, Cyclohexane, Tetrahydrofuran
OH	1370-1400	2-propanol
Benzene Ring	1680	Toluene
C=O	1880-1950	Acetone

Table 4.3: Spectral regions of NIR, Infrared, and Raman spectroscopy measurements used in the PLS calibration and ANN training models.

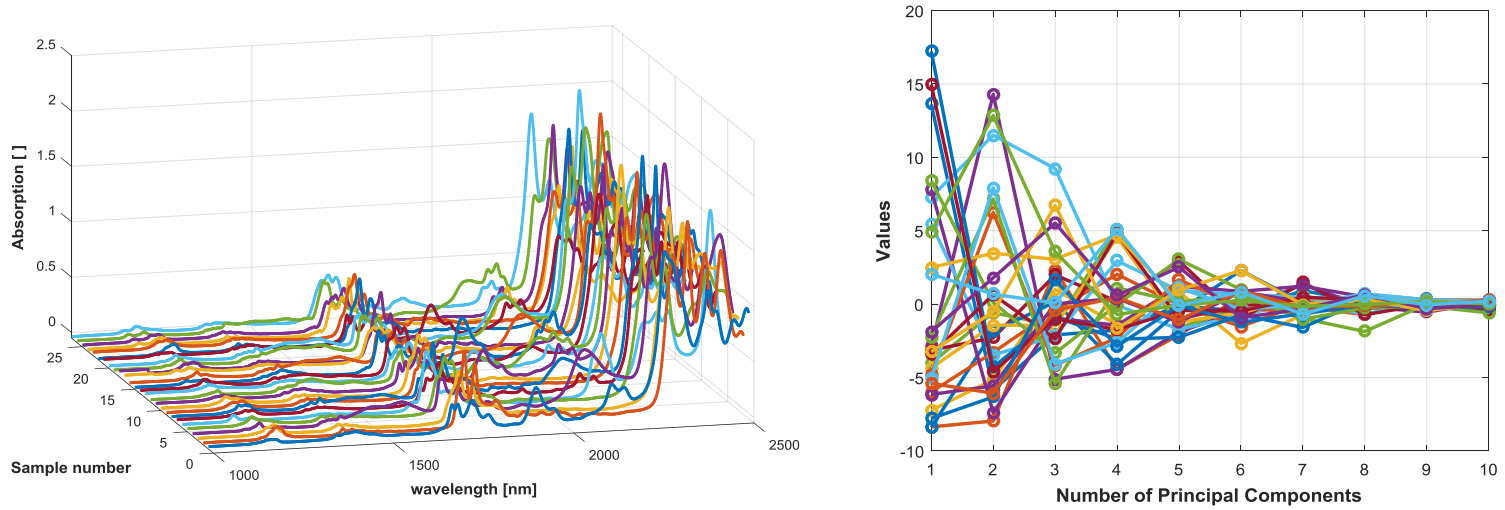
Spectroscopic Technique	Spectral regions [cm^{-1}]	Functional Group
NIR	1,440-1,500	OH stretching
	1,680-1,800	CH_2 and CH_3 stretching
Infrared	1,020-1,040	CO stretching
	1,065-1,075	COC out-phase bending
	1,100-1,165	CCO out-phase bending and CH_3 rocking
	1,215-1,240	CCC out-phase bending
	1,340-1,420	CH_3 in-phase bending
Raman	515-530	C=O ricking
	780-820	Ring in-phase and CCC in-phase stretching
	890-930	5 membered ring in-phase stretching
	1,000-1,050	CO stretching
	1,200-1,240	CCC out-phase bending
	1,280-1,320	CH_2 in-phase twist
	1,410-1,490	CH_2 and CH_3 bending

Table 4.4: Summary of density estimation results using PLSR and ANN methods for the mixture of 7 components (Acetone, Cyclohexane, n-Decane, 2-propanol, Tetrahydrofuran, Toluene, and Methanol).

Applied spectra	PLSR				ANN			
	Calibration		Test		Training		Test	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
NIR	0.647	0.986	2.421	0.795	0.861	0.984	1.676	0.873
Infrared	0.647	0.986	2.316	0.868	0.509	0.991	2.066	0.865
Raman	0.647	0.986	2.596	0.898	0.448	0.993	2.157	0.884
NIR+Infrared	0.647	0.986	2.074	0.896	0.448	0.993	1.741	0.895
NIR+Raman	0.647	0.986	2.468	0.903	0.417	0.994	2.067	0.887
Infrared+Raman	0.647	0.986	1.996	0.921	0.436	0.994	1.684	0.908
NIR+Infrared+Raman	0.647	0.986	1.815	0.928	0.399	0.994	1.576	0.919

Table 4.5: Summary of viscosity estimation results using PLSR and ANN methods for the mixture of 7 components (Acetone, Cyclohexane, n-Decane, 2-propanol, Tetrahydrofuran, Toluene, and Methanol).

Applied spectra	PLSR				ANN			
	Calibration		Test		Training		Test	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
NIR	3.534	0.991	117.2	0.459	2.916	0.992	33.73	0.820
Infrared	3.534	0.991	20.26	0.918	1.891	0.995	13.86	0.921
Raman	3.534	0.991	27.29	0.887	2.009	0.995	22.21	0.880
NIR+Infrared	3.534	0.991	23.70	0.922	1.815	0.995	13.94	0.929
NIR+Raman	3.534	0.991	25.59	0.897	1.884	0.995	20.32	0.892
Infrared+Raman	3.534	0.991	22.65	0.939	2.026	0.995	11.62	0.959
NIR+Infrared+Raman	3.534	0.991	23.62	0.939	1.750	0.995	11.40	0.964



(a) **(b)**

Figure 4.1: A series of measured NIR spectra. Panel (a) shows complicated absorption peaks of overtones and combination regions over the wavelength range from 1000 to 2500 nm for 27 liquid mixtures. Panel (b) represents that the first several principal components can explain the most variance of the measured optical spectra.

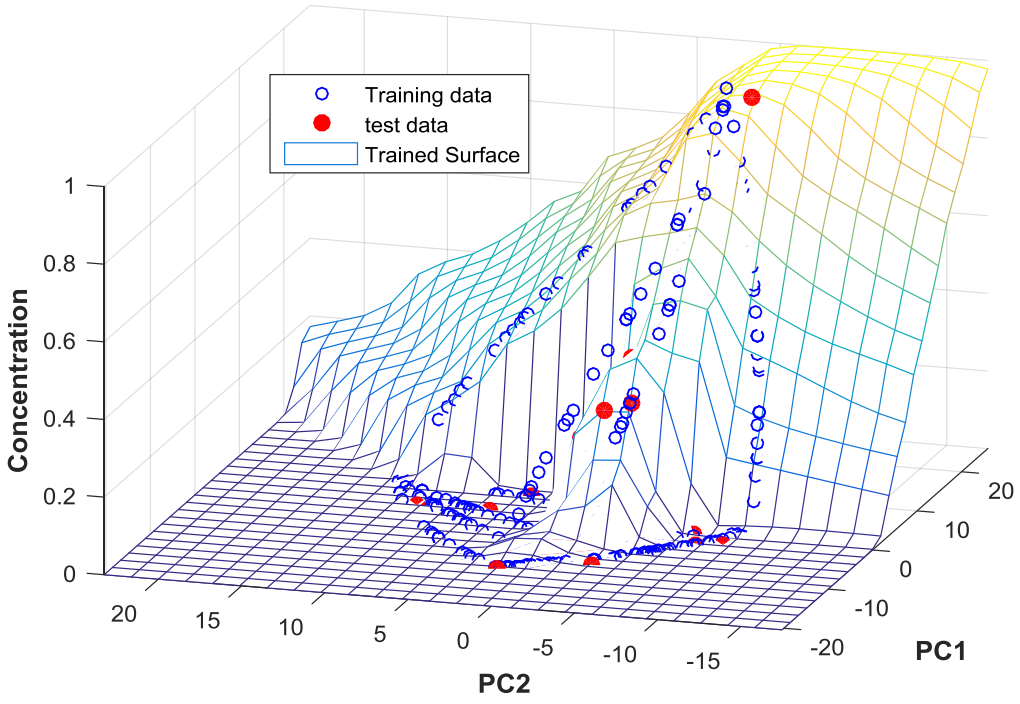


Figure 4.2: Two-dimensional plot showing the estimated concentration surface trained by training data over the first two principal components. Training data (blue circles) used for the mapping function and test data (red circles) used for verification are agreed to the estimated surface.

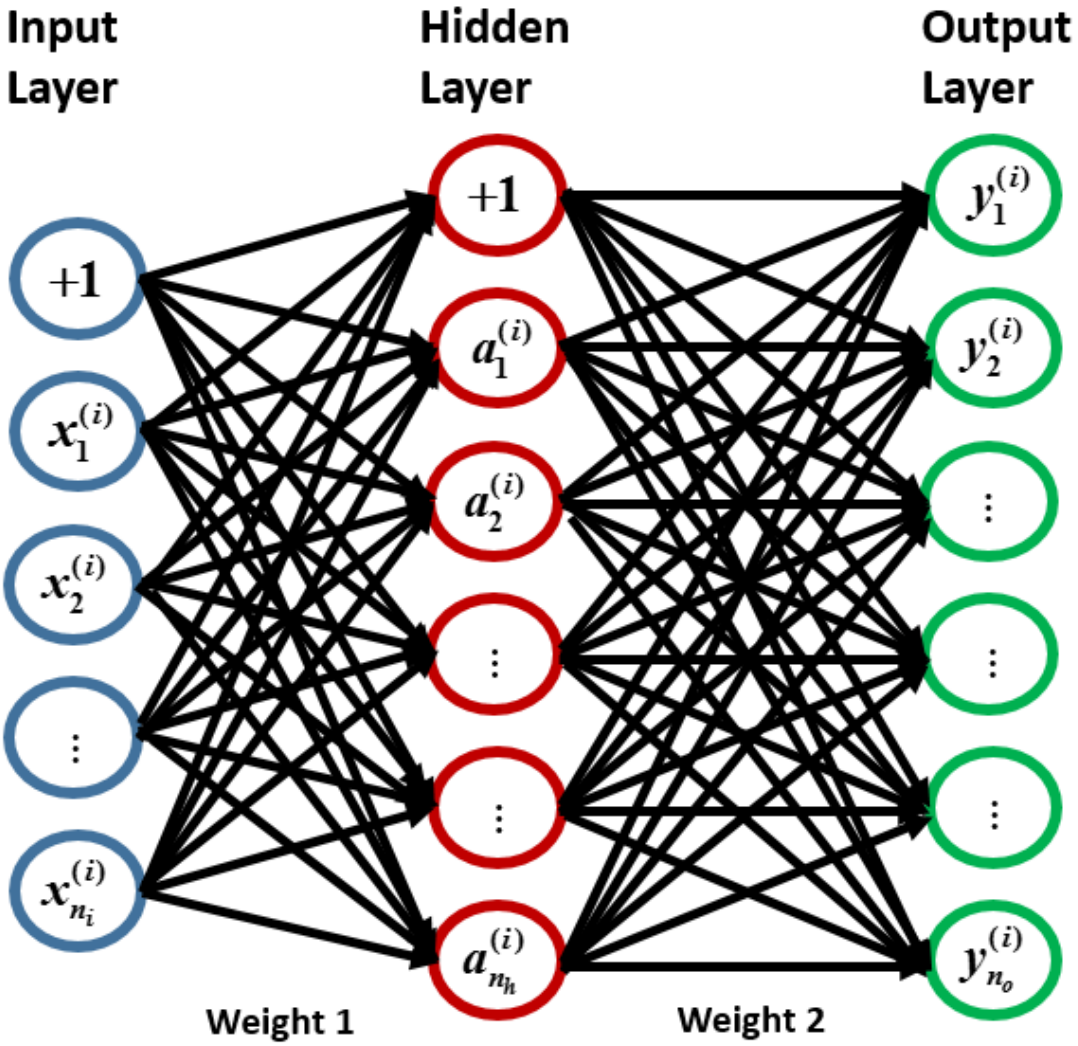


Figure 4.3: A feedforward neural network with three layers. The first column of blue circles are the input layer and the last column of green circles are the output layer. A hidden layer marked as red circles is consist of n neurons plus one offset biased unit that transforms the sum of weighted input data using activation function. Connections between Input and hidden layers and connections between hidden and output layers have corresponding weights that express the strength of connections between two neurons.

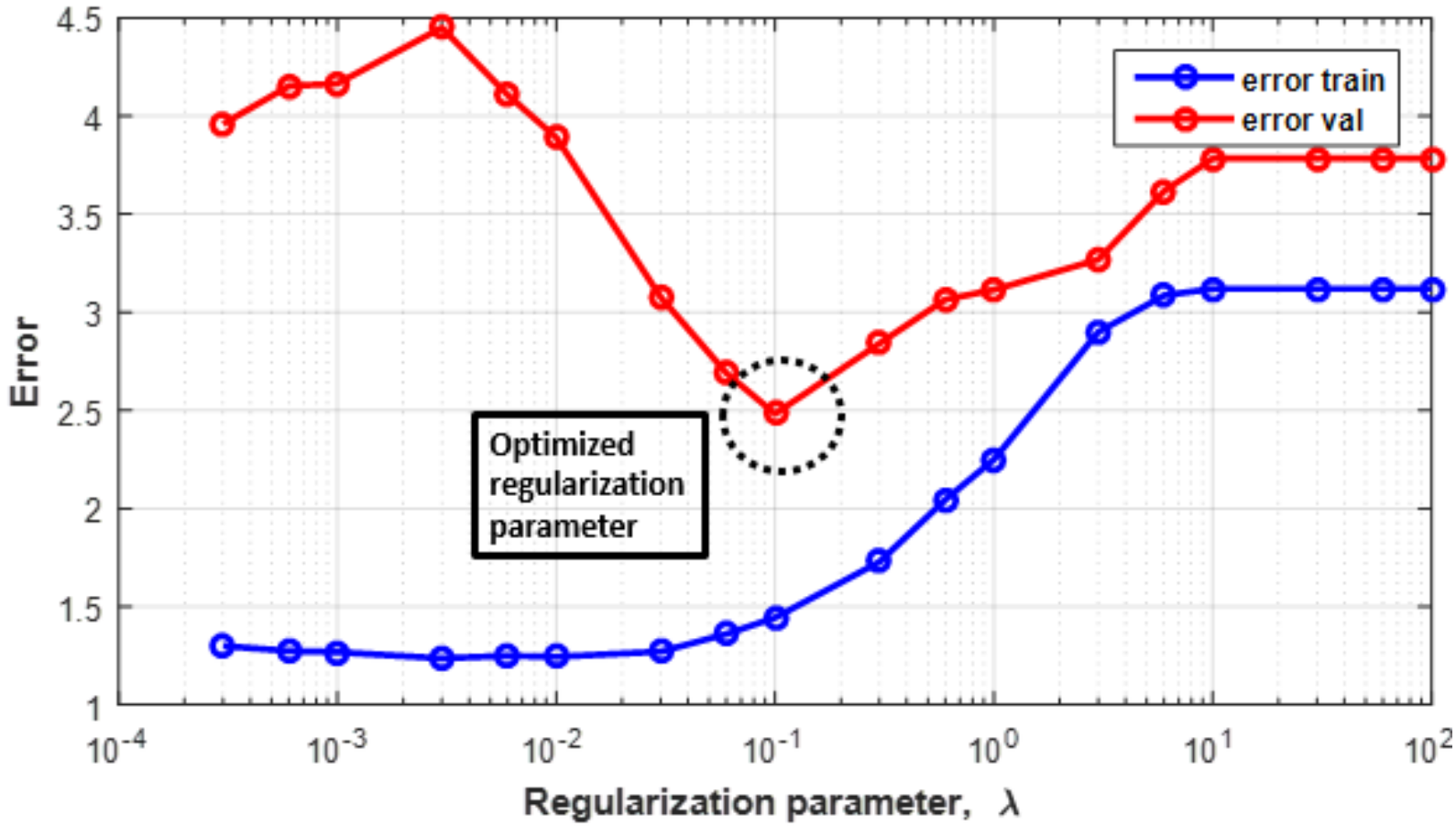


Figure 4.4: A selection of an optimized regularization parameter. While the regularization parameter is increased, training error is gradually increased but validation error becomes minimum at the optimum regularization. This indicates the regularization parameter is effectively alleviate overfitting from the training data.

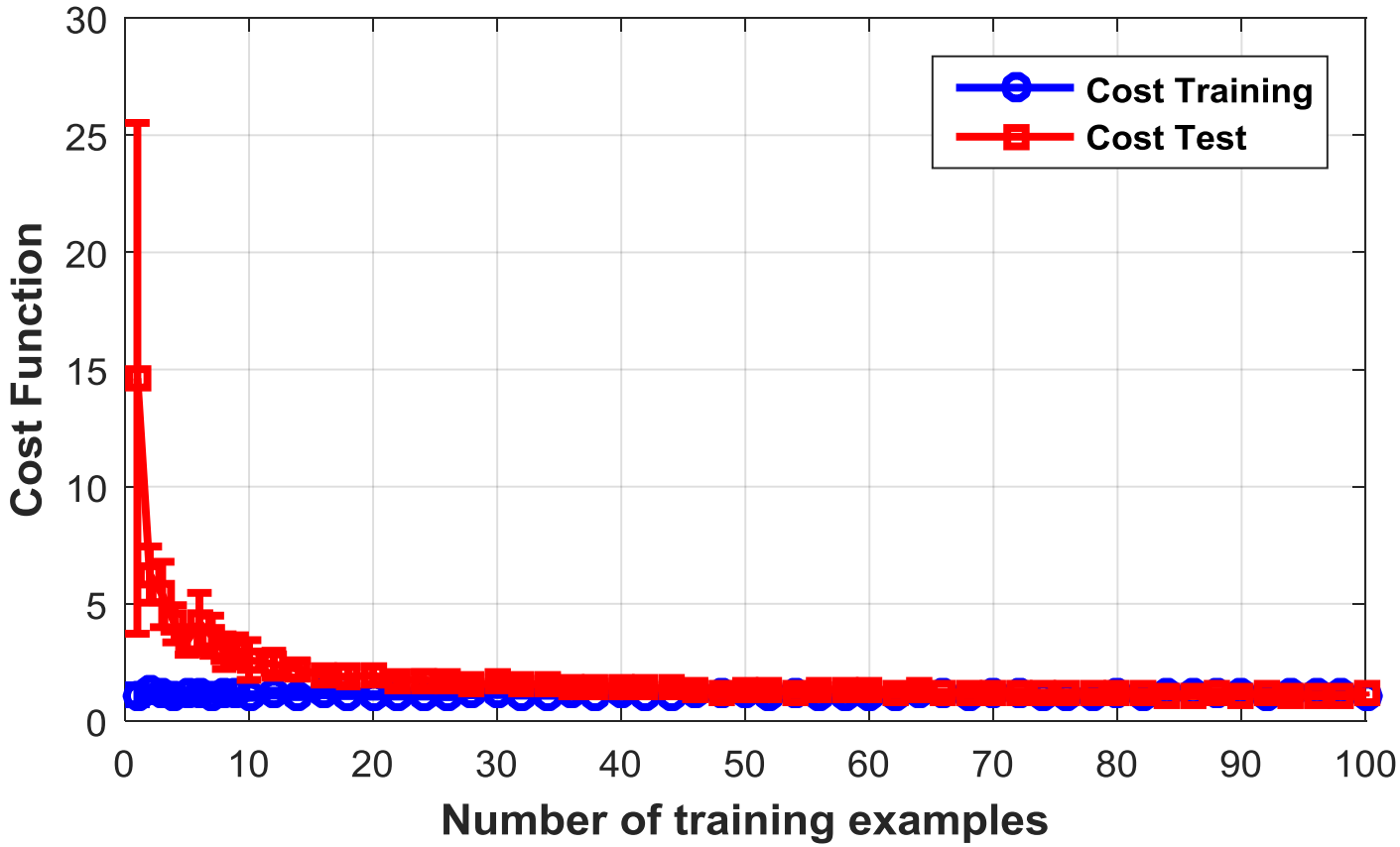


Figure 4.5: A learning curve plot with training and test error as a function of training data size. While the number of training data is increased, training error is increased and test error is decreased respectively. This indicates the regularization parameter is effectively alleviate overfitting from the training data.

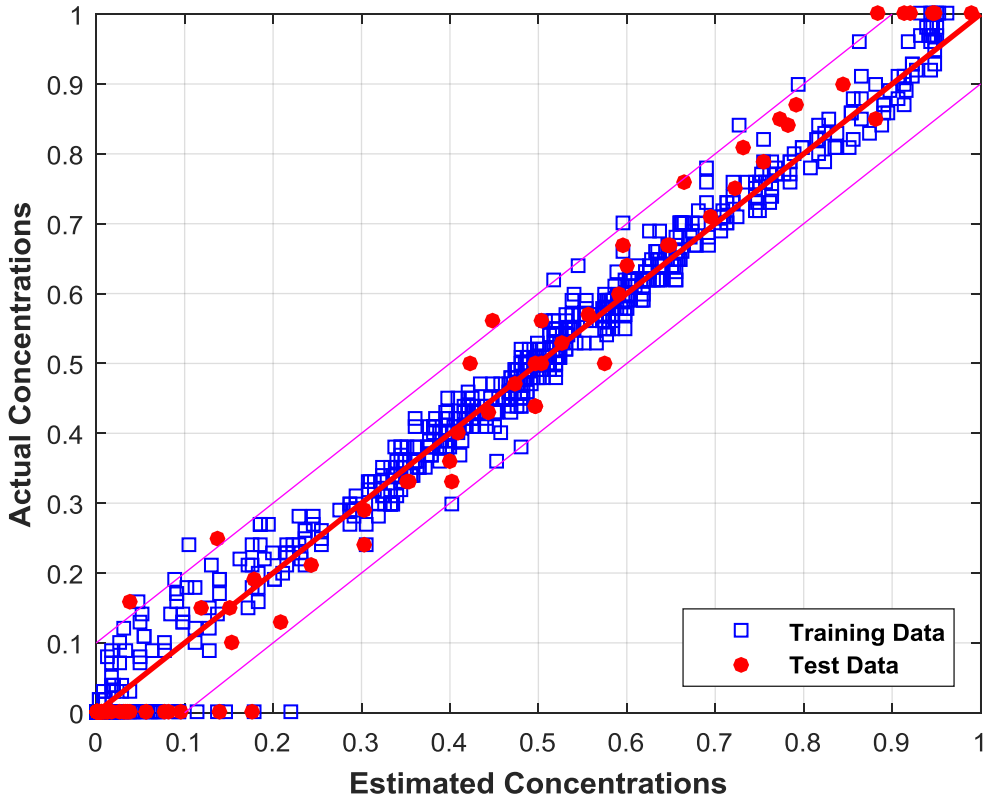


Figure 4.6: Cross-plot for the comparison of concentrations predicted using RBF method with the actual values. The training data are marked with blue squares and the test data are marked as red circles. The solid line is the ideal fit line, and the dashed lines indicate an absolute deviation of ± 10% concentrations.

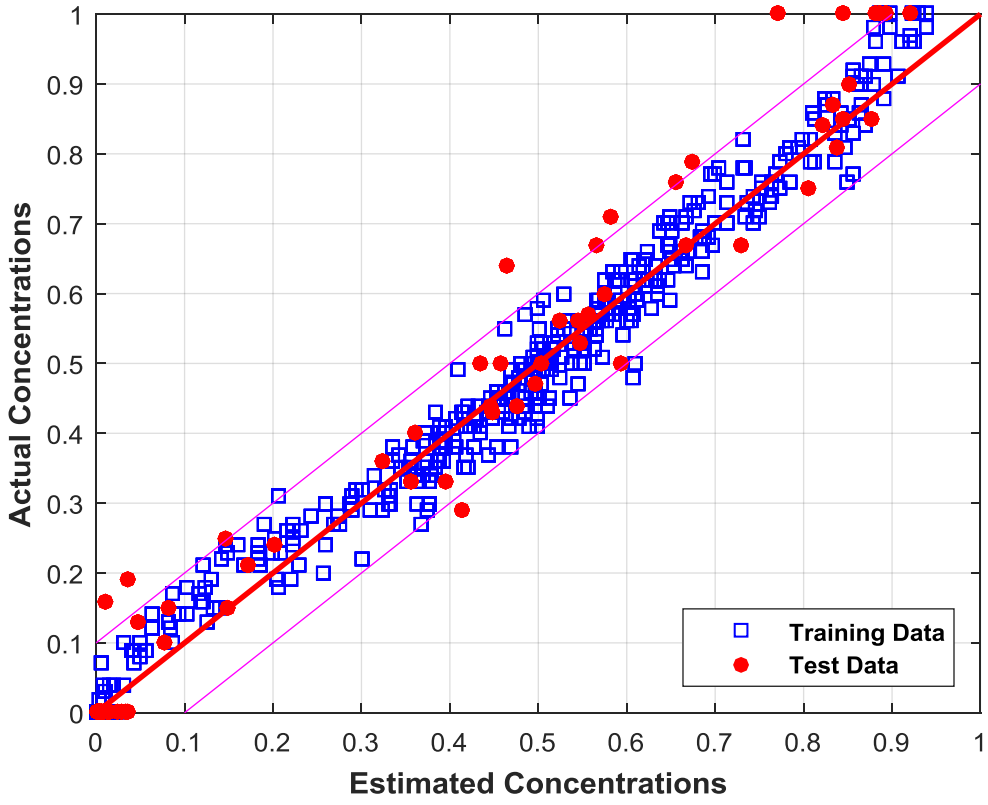


Figure 4.7: Cross-plot for the comparison of concentrations predicted using ANN method with the actual values. The training data are marked with blue squares and the test data are marked as red circles. The solid line is the ideal fit line, and the dashed lines indicate an absolute deviation of ± 10% concentrations.

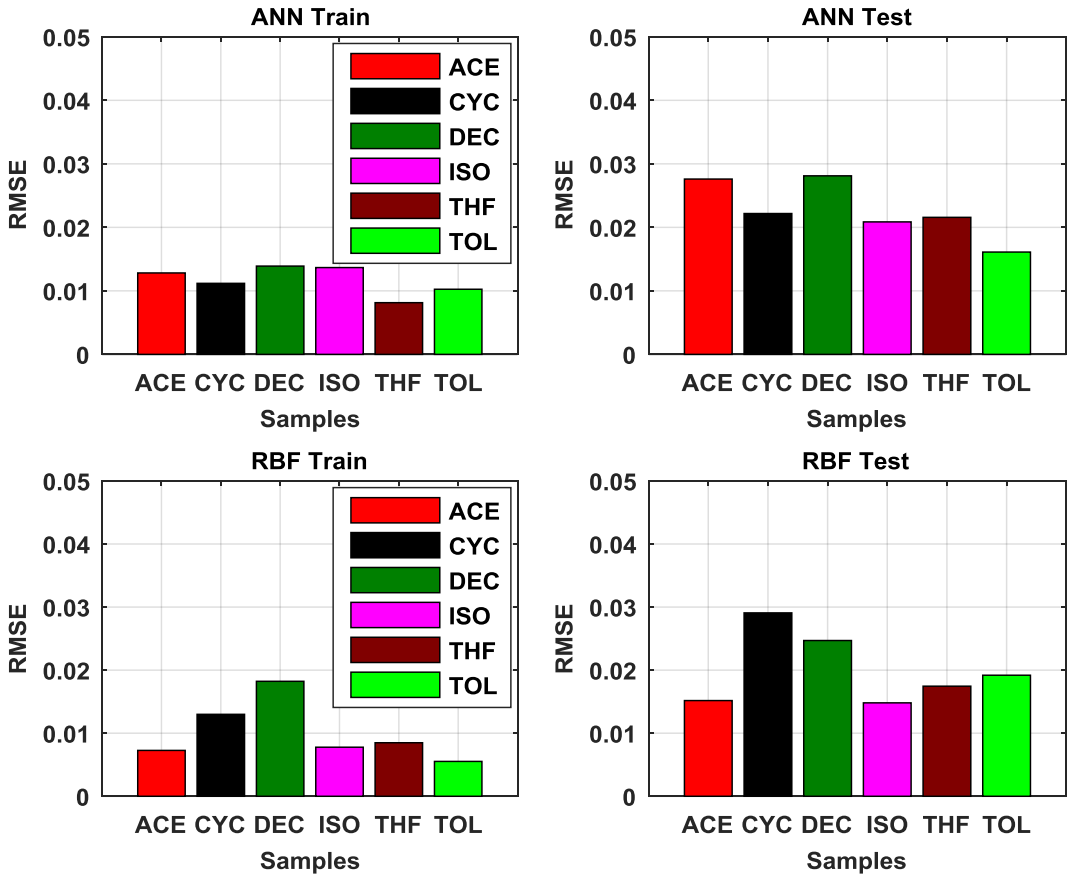


Figure 4.8: Root mean square error of estimated concentrations to corresponding actual concentrations for each sample. The top figures are RMSE of training data and test data for ANN method. The bottom figures are RMSE of training data and test data for RBF methods. The performance of the methods is compared by assessing the error function using an independent test data.

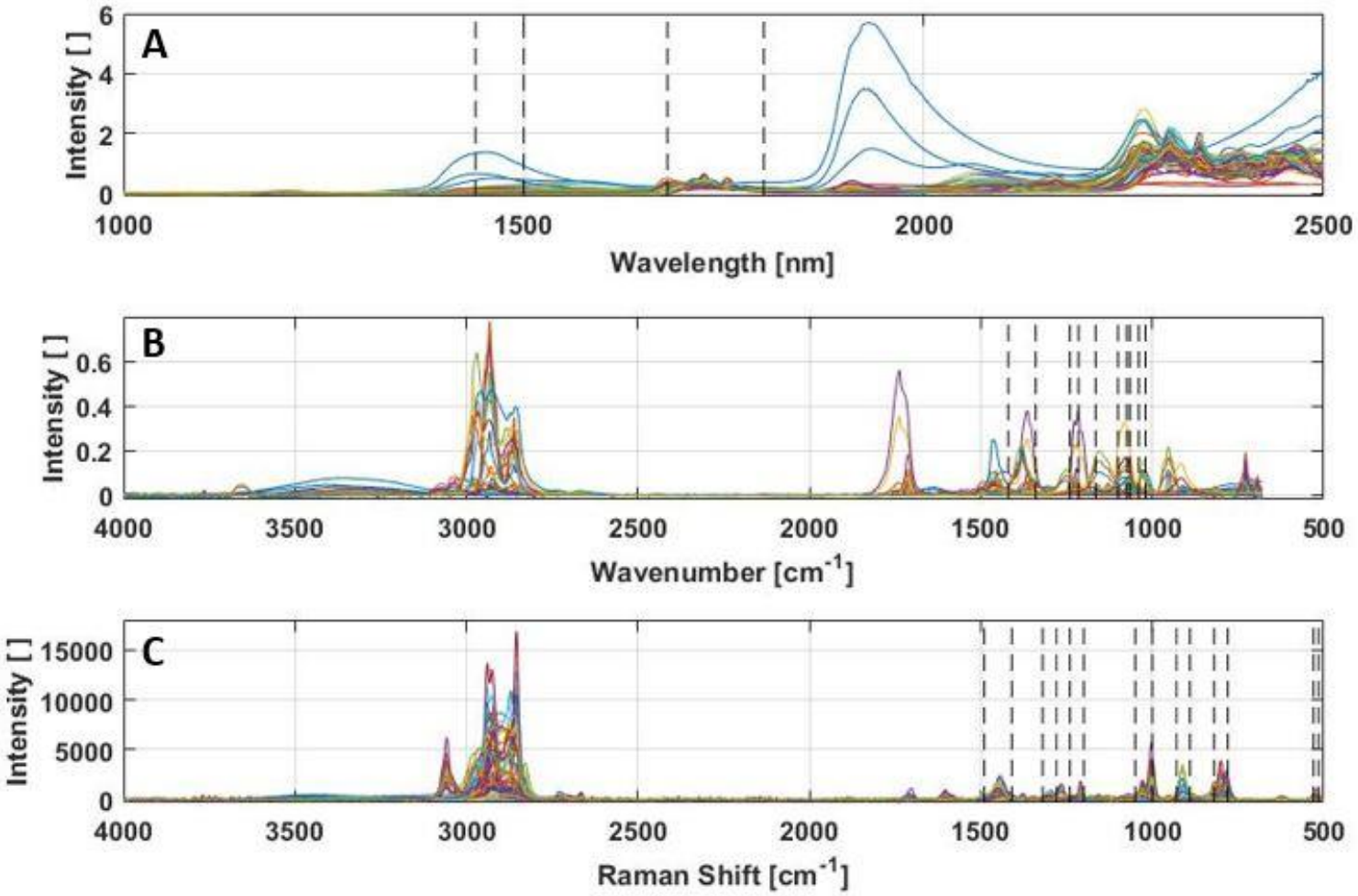


Figure 4.9: Measured spectra of the fluid mixtures of (A) NIR, (B) Infrared, and (C) Raman; Selected bands used for the quantitative analysis are marked with vertical black dash lines.

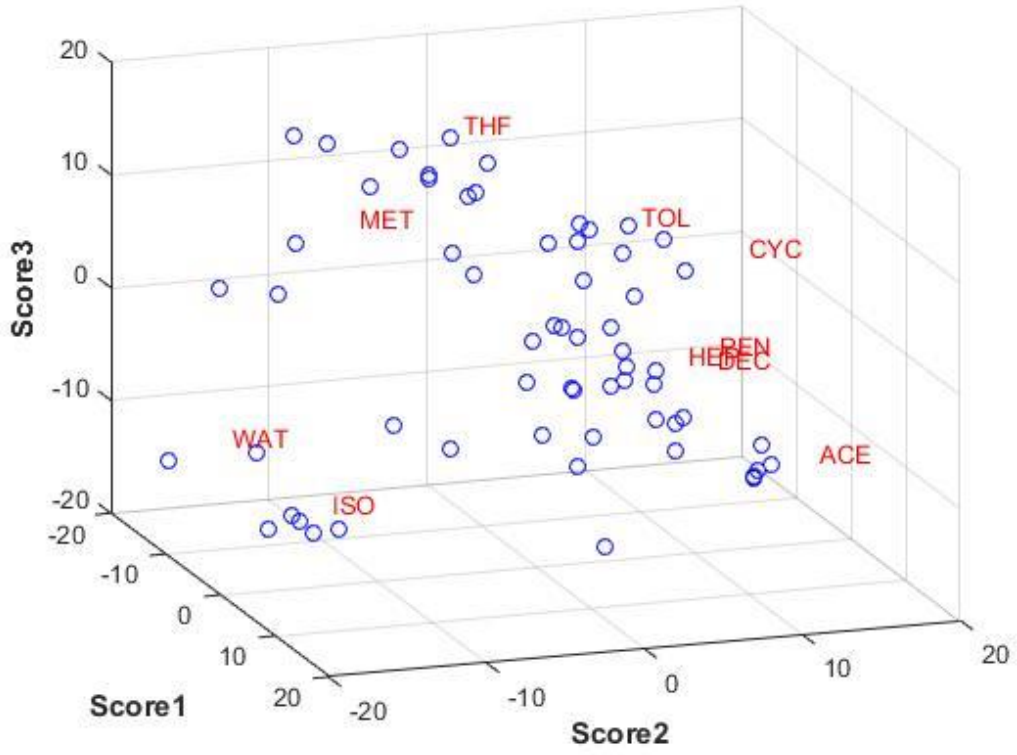


Figure 4.10: Three dimensional PCA scores based on the combined spectra of NIR, Infrared, and Raman. Open blue circles correspond to scores of fluid mixtures. Different fluid components can be discriminated except three normal alkanes. This is due to the similarity of molecular structures of n-Decane, n-Heptane, and n-Pentane.

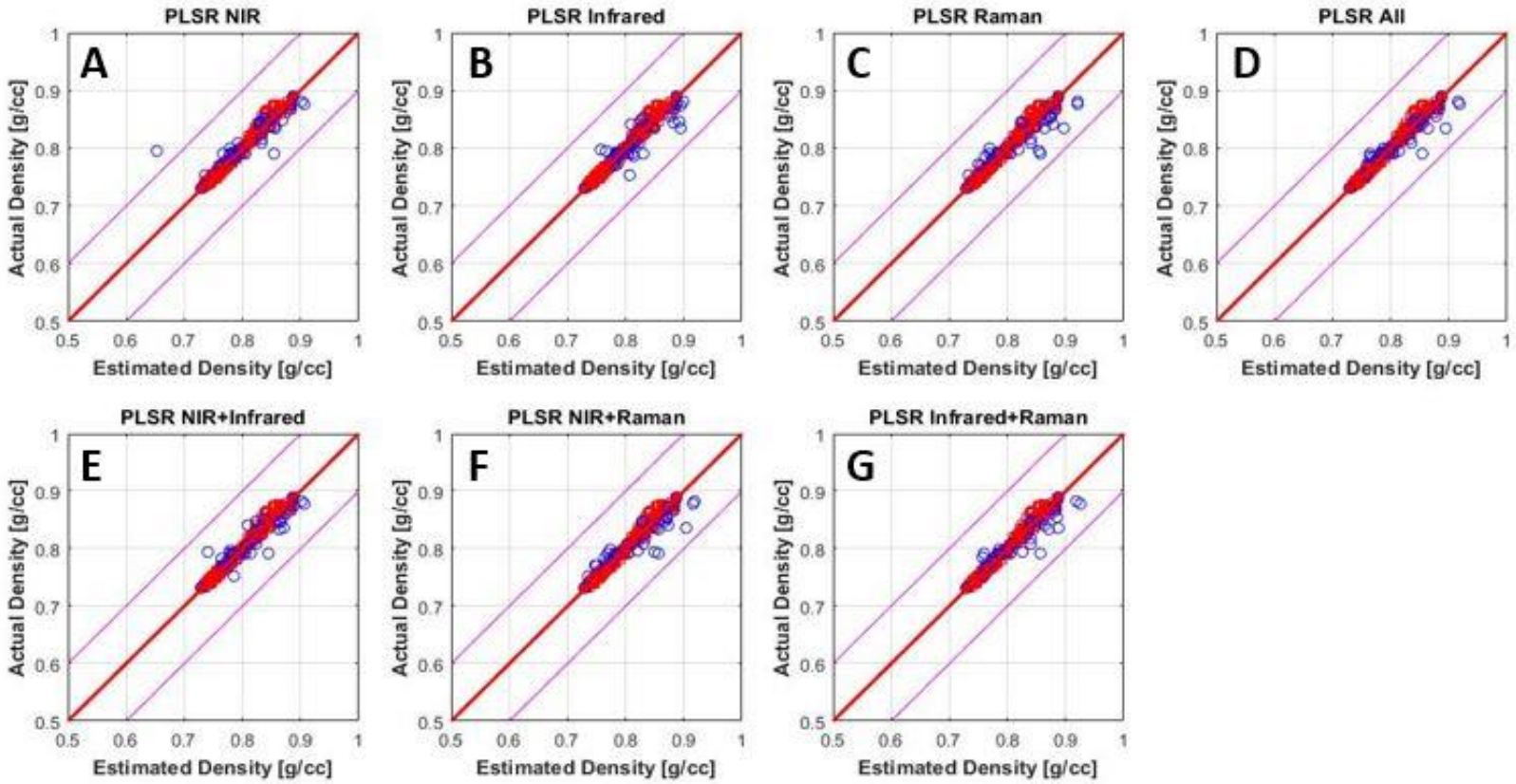


Figure 4.11: Estimated density values obtained using PLSR versus actual density values. Calibration uses seven different input data: (A) NIR, (B) Infrared, (C) Raman, (D) combined spectra of NIR, Infrared, and Raman, (E) combined spectra of NIR and Infrared, (F) combined spectra of NIR and Raman, and (G) combined spectra of Infrared and Raman. The training data are marked with red squares and the test data are marked as blue circles. The solid red line is the ideal fit line, and the magenta lines indicate an absolute deviation of ± 10% density.

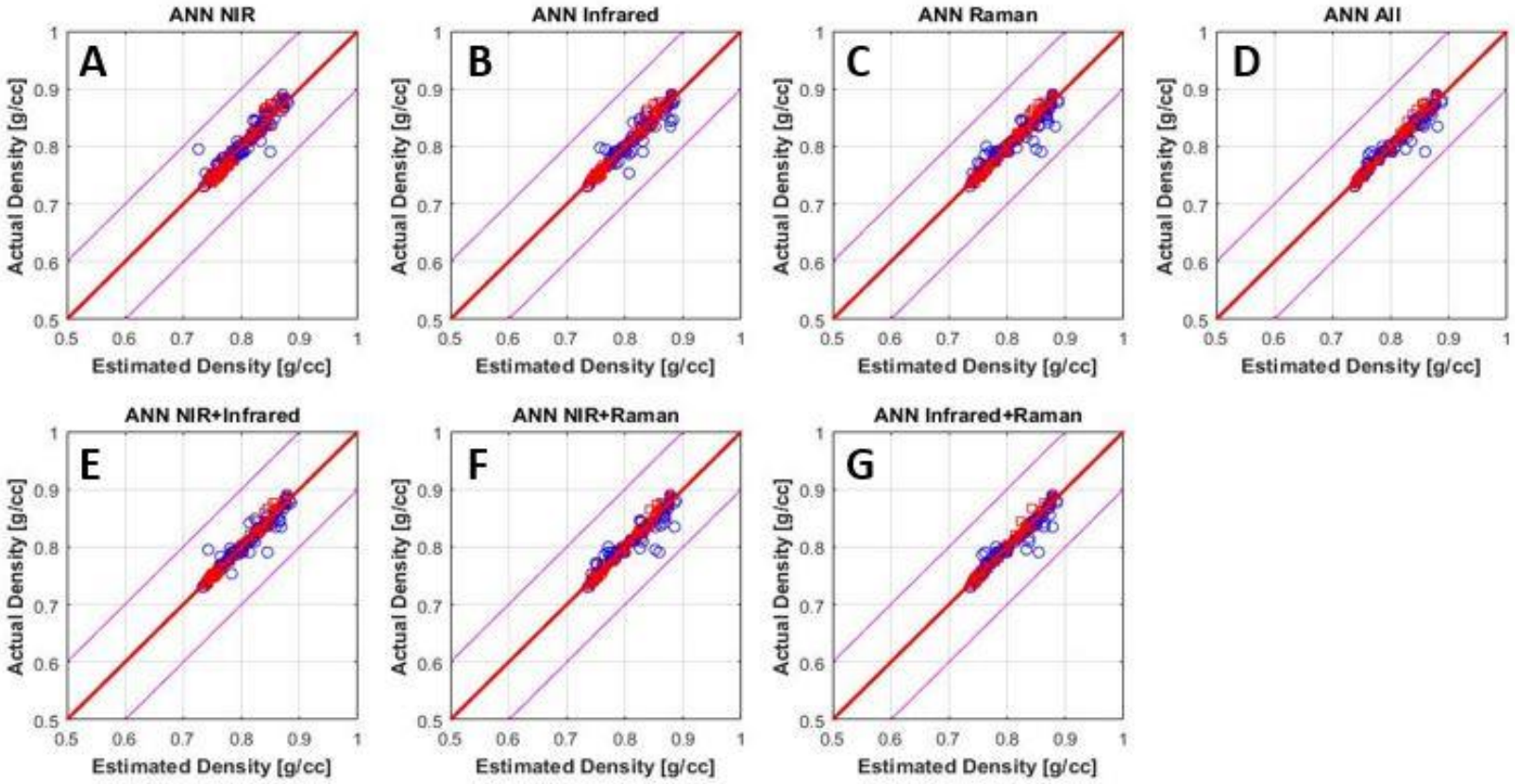


Figure 4.12: Estimated density values obtained using ANN versus actual density values. Training uses seven different input data: (A) NIR, (B) Infrared, (C) Raman, (D) combined spectra of NIR, Infrared, and Raman, (E) combined spectra of NIR and Infrared, (F) combined spectra of NIR and Raman, and (G) combined spectra of Infrared and Raman. The training data are marked with red squares and the test data are marked as blue circles. The solid red line is the ideal fit line, and the magenta lines indicate an absolute deviation of ± 10% density.

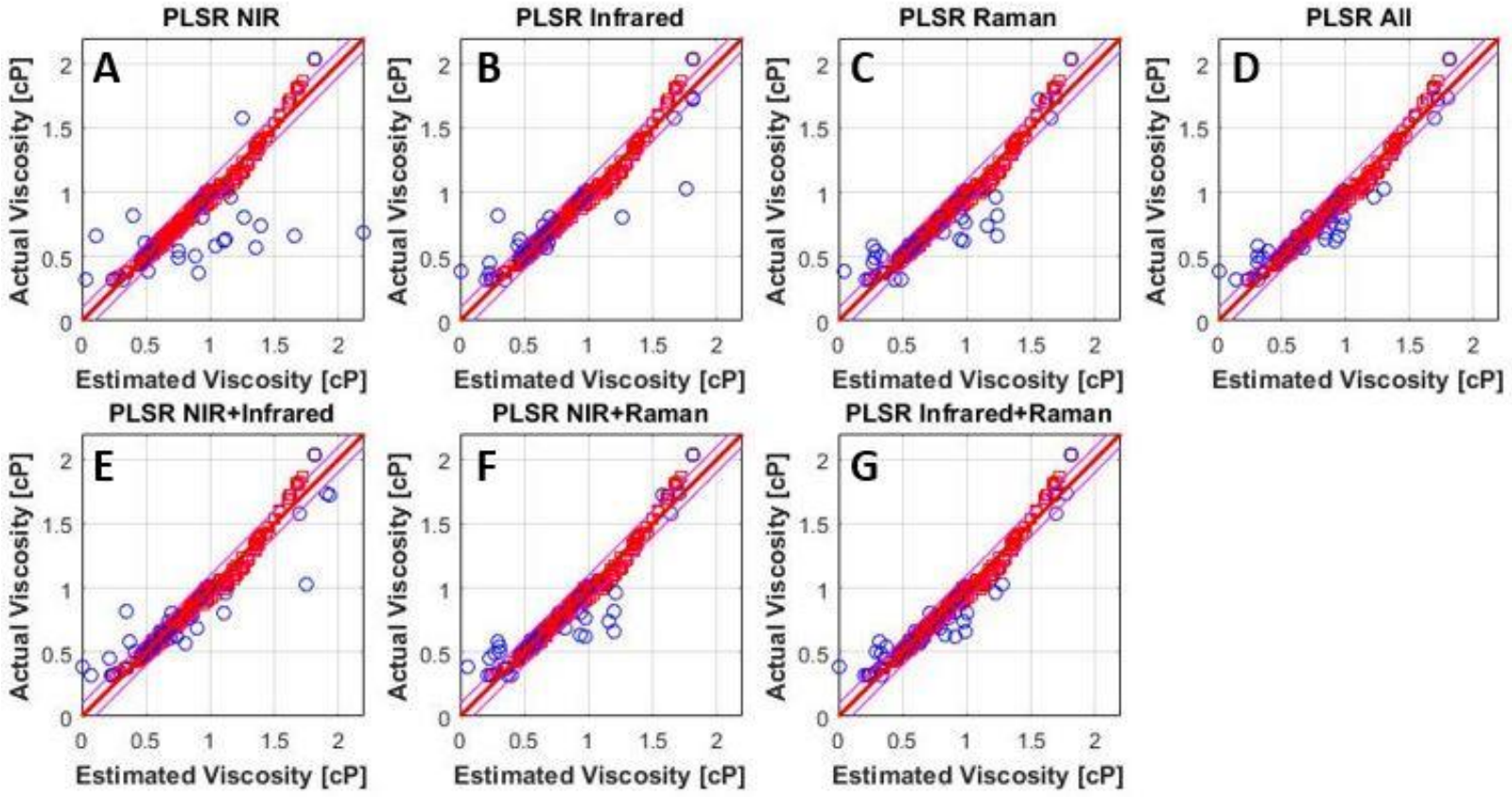


Figure 4.13: Estimated viscosity values obtained using ANN versus actual viscosity values. Calibration uses seven different input data: (A) NIR, (B) Infrared, (C) Raman, (D) combined spectra of NIR, Infrared, and Raman, (E) combined spectra of NIR and Infrared, (F) combined spectra of NIR and Raman, and (G) combined spectra of Infrared and Raman. The training data are marked with red squares and the test data are marked as blue circles. The solid red line is the ideal fit line, and the magenta lines indicate an absolute deviation of ± 10% viscosity.

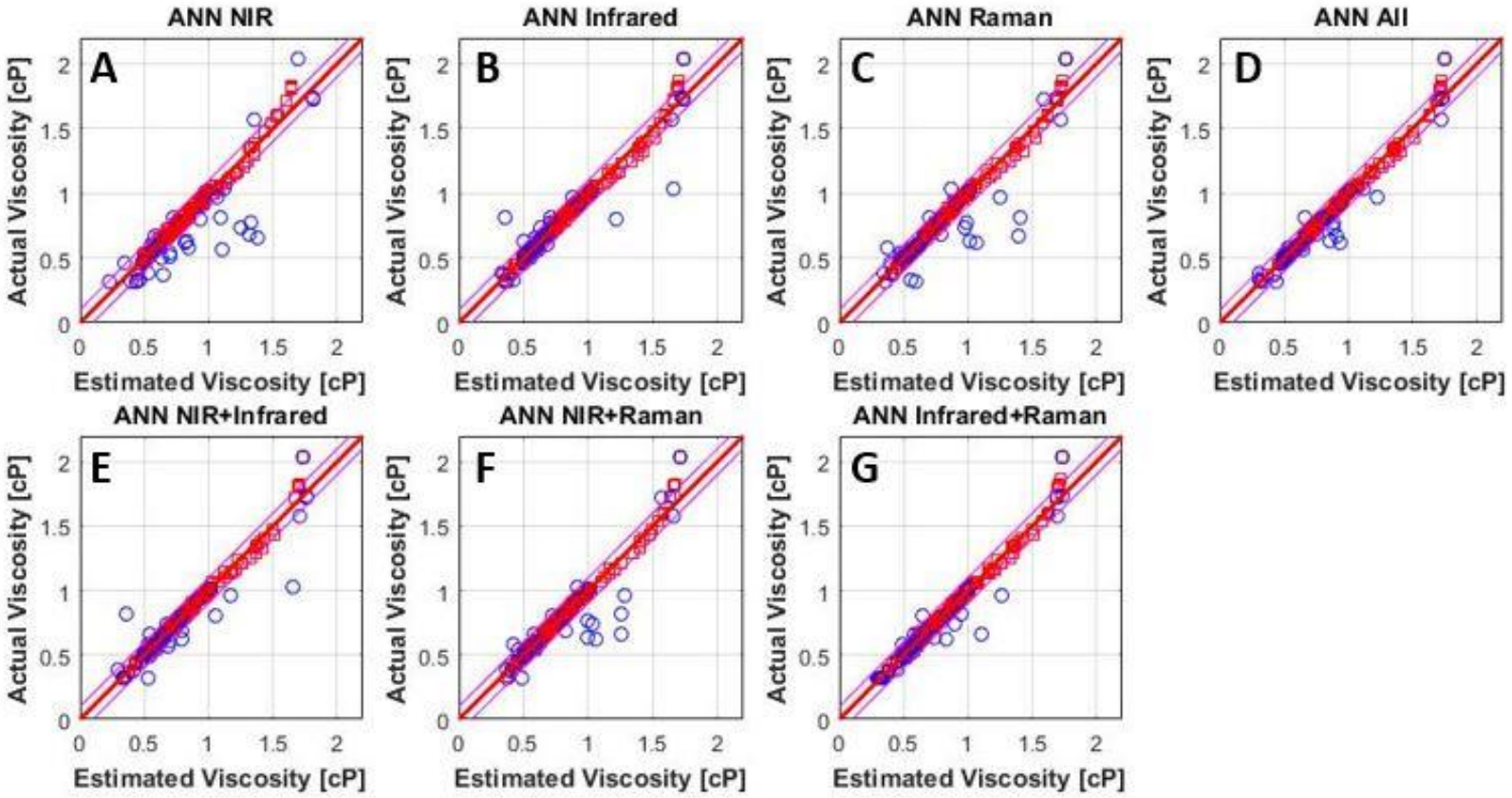


Figure 4.14: Estimated viscosity values obtained using ANN versus actual viscosity values. Training uses seven different input data: (A) NIR, (B) Infrared, (C) Raman, (D) combined spectra of NIR, Infrared, and Raman, (E) combined spectra of NIR and Infrared, (F) combined spectra of NIR and Raman, and (G) combined spectra of Infrared and Raman. The training data are marked with red squares and the test data are marked as blue circles. The solid red line is the ideal fit line, and the magenta lines indicate an absolute deviation of ± 10% viscosity.

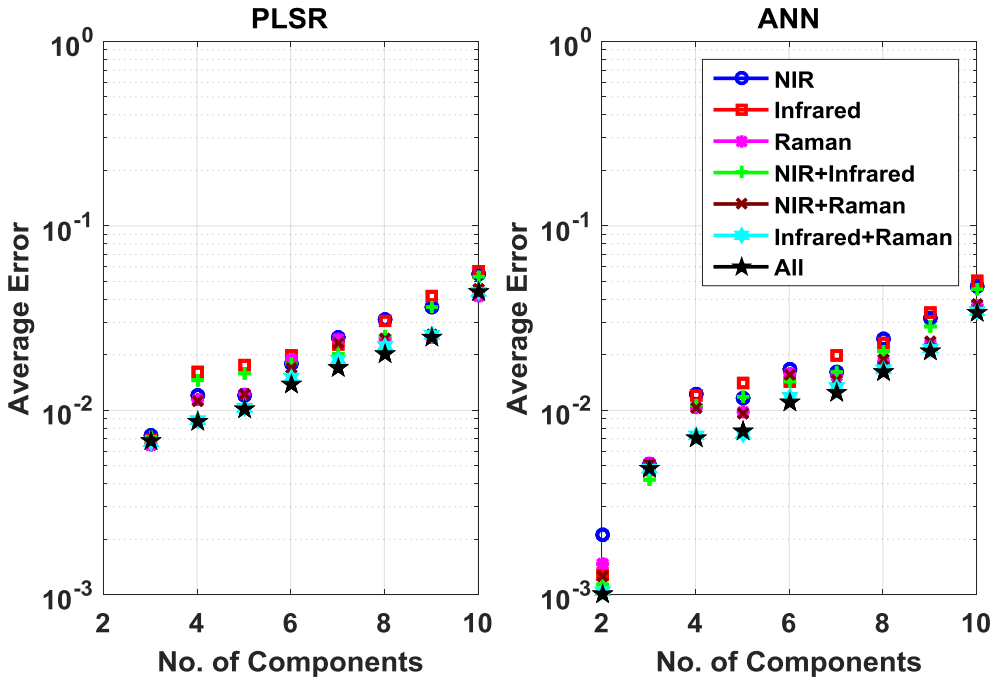


Figure 4.15: A comparison of density estimations with seven different input data. As the number of components used in the analysis increases, RMSE continuously increases for both PLSR and ANN methods. Ten components are subsequently added to the calibration and training process in the order of 2-propanol, Cyclohexane, Acetone, Tetrahydrofuran, Methanol, Toluene, n-Decane, DI water, n-Heptane, and n-Pentane.

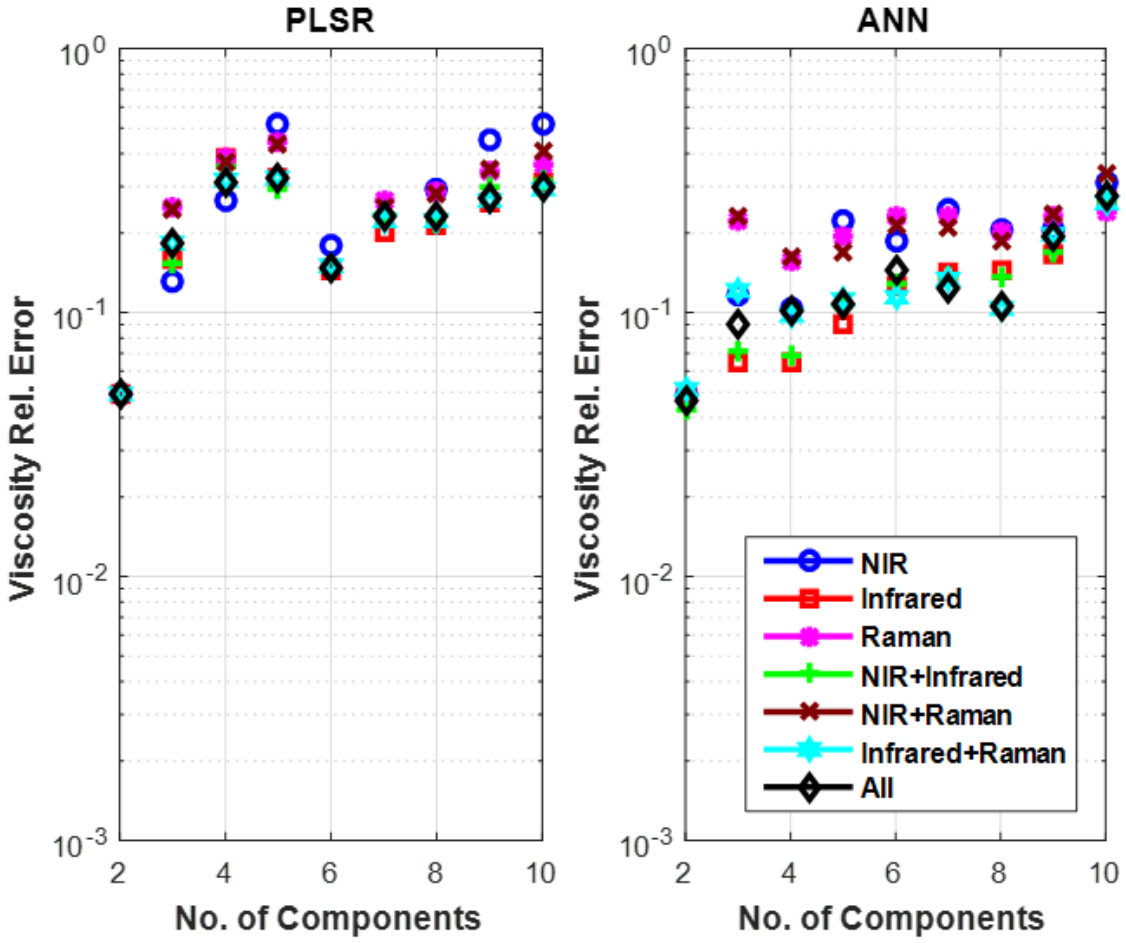


Figure 4.16: A comparison of viscosity estimations with seven different input data. As the number of components used in the analysis increases, RMSE continuously increases for both PLSR and ANN methods. Ten components are subsequently added to the calibration and training process in the order of 2-propanol, Cyclohexane, Acetone, Tetrahydrofuran, Methanol, Toluene, n-Decane, DI water, n-Heptane, and n-Pentane.

Chapter 5: NMR Relaxation Time and Diffusion Coefficient of Liquid Mixtures Calculated with Molecular Dynamics Simulations

In this chapter, the ^{1}H NMR relaxation time and the mass diffusion coefficient of various liquid mixtures were studied over an extensive composition range at ambient conditions. A series of fluid samples including water, alcohol, alkane, cycloalkane, aromatic and their mixtures were tested using molecular dynamics (MD) simulations with an OPLS/AA (Optimized Potential for Liquid Simulations-All Atoms) force field. The simulated relaxation time and the mass diffusion coefficient for these mixtures agree with the experimental data. Also, the quantification of ^{1}H-^{1}H dipole-dipole relaxations for these fluid mixtures reveals the extents to which each component contributes to the total relaxation process and to which interactions occur between different fluids. These results suggest that MD simulation can serve as a bridge between theory and laboratory measurements for a better description of NMR relaxation process.

5.1 INTRODUCTION

The calculation of longitudinal relaxation time, T_1 and transverse relaxation time, T_2 can be defined by

$$\frac{1}{T_1} = \frac{1}{T_{1bulk}} + 3\frac{\rho_1}{r}, \tag{5.1}$$

$$\frac{1}{T_2} = \frac{1}{T_{2bulk}} + 3\frac{\rho_2}{r} + \frac{D(\gamma GTE)^2}{12}, \tag{5.2}$$

where ρ_1 and ρ_2 are surface relaxivity for T_1 and T_2 and, T_{1bulk} and T_{2bulk} are bulk longitudinal relaxation time and bulk transverse relaxation time, respectively.

Additionally, r represents the radius of spherical pores. D is a fluid diffusivity coefficient and γ is the gyromagnetic ratio of a hydrogen proton. G is a field-strength gradient and TE is inter-echo spacing used in the Carr-Purcell-Meiboom-Gill (CPMG) sequence (Carr and Purcell 1954; Meiboom and Gill 1958). Recently, researchers have published several studies on NMR relaxations specifically concerning pore-scale evaluation of organic-rich mudrocks (Tandon et al., 2017), surface relaxation of kerogen (Zhang et al., 2017), and fluid substitutions of hydrocarbon and water (Medellin et al., 2018). However, these studies do not offer quantitative description of the hydrogen spin pair interactions between different fluids or their contributions to the total relaxations. These bulk relaxations are important because most reservoir fluids are mixtures of multiple components and distinctive relaxations occur between various combinations of different hydrocarbon types.

NMR relaxations can be measured in the laboratory, but they are the results of ensemble averages. As such, if the ensemble average of hydrogen relaxations calculated from MD simulation can cross-validate the experimental observations, then they can be explained in terms of the relaxation of each hydrogen atom. The relaxations between a pair of hydrogen atoms are generated from the interaction forces between bonded and non-bonded atoms including Lennard-Jones and Coulomb potentials. Because atomic and molecular interactions cause a dynamic evolution of the system, the MD simulation method is admissible for analyzing NMR properties. Thus, from XYZ trajectories of molecules, distance between hydrogen atoms and angle to the magnetic field (z-direction) of liquid mixtures can be calculated. This study presents the mixing effects on

macroscopic NMR relaxation time and mass diffusion coefficient in terms of various properties computed from MD simulations.

This chapter is organized as follows: First, it introduces theories and equations of NMR relaxations and mass diffusion coefficients. Following the NMR theories, several systems of liquid mixtures and MD simulation methods are presented. The results and discussion section describes the estimations of NMR properties based on MD simulations and compares them to the experimental results for various liquid mixtures. From these liquid mixtures, relaxation times and mass diffusion coefficients contributed from each component are analyzed.

5.2 METHODS

This section explains the formulations and procedures to calculate NMR relaxation time, density, and mass diffusion coefficient of liquid mixtures from MD simulations.

5.2.1 NMR relaxation time

The spin-lattice and spin-spin ^{1}H NMR relaxations have been extensively studied both experimentally and theoretically. Mathematical derivation for quantum mechanical analysis from Hamiltonian is beyond the scope of this study. Instead, some important equations to calculate NMR relaxation time are here presented. According to the NMR relaxation theory (Abragam, 1961; McConnel, 1987; Cowan, 1997; and Ernst et al., 2004), the NMR relaxation times T_1 and T_2 are given by

$$\frac{1}{T_1} = \sum_{m=-2}^{2} m^2 \int_0^{\infty} G_m(t) \exp(\mathrm{im}\,\omega_0 \mathrm{t})\ \mathrm{dt} \quad , \tag{5.3}$$

and

$$\frac{1}{T_2} = \sum_{m=-2}^{2} [3 - m^2 / 2] \int_0^{\infty} G_m(t) \exp(\mathrm{im}\,\omega_0 \mathrm{t})\ \mathrm{dt} \quad , \tag{5.4}$$

where *m* is the spin-flip index, *G(t)* is the autocorrelation function, and ω_0 is the Larmor frequency, respectively. The dipolar spectral density function *J(ω)* is the Fourier transformation of an autocorrelation function as

$$J(\omega) = \int_{-\infty}^{\infty} G(t) \exp(i\omega t)\ dt \quad , \tag{5.5}$$

and the NMR relaxation times become

$$\frac{1}{T_1} = J_1(\omega_0) + 4J_2(2\omega_0) \quad , \tag{5.6}$$

and

$$\frac{1}{T_2} = \frac{3}{2} J_0(0) + \frac{5}{2} J_1(\omega_0) + J_2(2\omega_0) \quad . \tag{5.7}$$

Considering the fast motion regime where the correlation time is much smaller than the inverse of the Larmor frequency, the spectral density function *J(ω)* is approximated as *J(0)*. Then, the spectral density functions J_0, J_1, and J_2 are identical to *J(0)* which is twice the area under *G(t)*. Thus, the Eq. (5.6) and (5.7) give

$$\frac{1}{T_1} = \frac{1}{T_2} = 5J_0(0) \quad . \tag{5.8}$$

The autocorrelation function *G(t)* of dipole-dipole interactions is expressed as

$$G(t) = \frac{3}{16}\left(\frac{\mu_0}{4\pi}\right)^2 \hbar^2\gamma^4 \frac{1}{N}\sum_{i\neq j}^{N}\left\langle \frac{3\cos^2\theta_{ij}(t+\tau)-1}{r_{ij}^3(t+\tau)} \frac{3\cos^2\theta_{ij}(\tau)-1}{r_{ij}^3(\tau)} \right\rangle_\tau , \tag{5.9}$$

where μ_0 is the vacuum permeability, $\hbar$ is the reduced Planck constant, γ is the gyromagnetic ratio for ^{1}H atom, N is the number of ^{1}H-^{1}H dipole-dipole interactions, r and θ are the separation distance and corresponding angle to the magnetic field between two Hydrogen atoms, and τ and t are the time and lag time, respectively. The r and θ can be calculated from the trajectory of a MD simulation and the autocorrelation function can be computed using the FFT (Fast Fourier Transform) function. The second-moment M of dipole-dipole interaction is related to the autocorrelation function with 0 lag time as

$$M = 3G(0) = \frac{9}{16}\left(\frac{\mu_0}{4\pi}\right)^2 \hbar^2\gamma^4 \frac{1}{N}\sum_{i\neq j}^{N}\left\langle \frac{\left(3\cos^2\theta_{ij}(\tau)-1\right)^2}{r_{ij}^6(\tau)} \right\rangle_\tau . \tag{5.10}$$

I assume that the variations of r and θ are uncorrelated and independent of each other. Thus, the angular term of Eq. (5.10) can be integrated separately as

$$\frac{1}{4\pi}\int_0^{\pi}\left(3\cos^2\theta_{ij}(\tau)-1\right)^2\sin\theta\,d\theta=\frac{4}{5}, \tag{5.11}$$

and it can be expressed as

$$M=\frac{9}{20}\left(\frac{\mu_0}{4\pi}\right)^2\hbar^2\gamma^4\frac{1}{N}\sum_{i\neq j}^{N}\left\langle\frac{1}{r_{ij}^6(\tau)}\right\rangle_{\tau}. \tag{5.12}$$

The correlation time τ_c is the time scale of the autocorrelation decay function and is expressed as

$$\tau_c=\frac{1}{G(0)}\int_0^{\infty}G(t)\;dt, \tag{5.13}$$

which can be explained as the width of the autocorrelation function or its area normalized by its height. Combining Eq. (5.6), (5.7), (5.8), (5.10), (5.12), and (5.13), it becomes

$$J(0)=\frac{2}{3}M\tau_c, \tag{5.14}$$

and Eq. (5.8) becomes

$$\frac{1}{T_1}=\frac{1}{T_2}=\frac{10}{3}M\tau_c. \tag{5.15}$$

These relaxation times have two types of interactions: (1) intramolecular ^{1}H-^{1}H dipole-dipole interactions; and (2) intermolecular ^{1}H-^{1}H dipole-dipole interactions. Then, Eq. (5.115) is expressed as

$$\frac{1}{T_1} = \frac{1}{T_2} = \frac{10}{3} M_{\text{intra}} \tau_{c,\text{intra}} + \frac{10}{3} M_{\text{inter}} \tau_{c,\text{inter}} \quad . \tag{5.16}$$

The computations of NMR relaxation times with MD simulations are based on the several assumptions: (1) Fluids used in the analysis have low viscosity that longitudinal and transverse relaxation times are equal; (2) The spectral density functions at low Larmor frequency and zero angular frequency are approximately same; and (3) The fluids are in the fast motion regime with low correlation time.

If the system has a binary mixture with two different molecules, the five different types of interactions can be categorized as

(1) The partial ensemble of intramolecular dipole-dipole interactions between hydrogens in a molecule of component 1.
(2) The partial ensemble of intermolecular dipole-dipole interactions between hydrogens in different molecules of component 1.
(3) The partial ensemble of intramolecular dipole-dipole interactions between hydrogens in a molecule of component 2.
(4) The partial ensemble of intermolecular dipole-dipole interactions between hydrogens in different molecules of component 2.
(5) The partial ensemble of intermolecular dipole-dipole interactions between hydrogen in a molecule of component 1 and a molecule of component 2.

Finally, considering these interactions, following equation is obtained.

$$\frac{1}{T_1} = \frac{1}{T_2} = \frac{10}{3}\left(M_{\text{intraC1}}\tau_{c,\text{intraC1}} + M_{\text{interC1}}\tau_{c,\text{interC1}} + M_{\text{intraC2}}\tau_{c,\text{intraC2}} + M_{\text{interC2}}\tau_{c,\text{interC2}} + M_{\text{interC1C2}}\tau_{c,\text{interC1C2}} \right) \quad (5.18)$$

Thus, contributions of intramolecular and intermolecular interactions can be quantified for each component, and intermolecular interactions between molecules of two different components can be assessed.

5.2.2 Density

The effective densities of the binary mixtures are evaluated in the MD simulations. The density is computed by the mass of all atoms divided by the total volume of the system. The results are in agreement with the experimental data and are plotted in **Figure 5.1**.

5.2.3 Mass Diffusion Coefficient

The mass diffusion coefficients can be considered as the distance a molecule traveled in a given time. The dominant forces for these movements are the attractive and repulsive interactions between molecules. To quantify the mass diffusion coefficient of liquid mixtures, a well-known Einstein's relation is applied as

$$\lim_{t\to\infty}\left\langle \left\| r(t) - r(0) \right\|^2 \right\rangle = 6Dt \quad (5.19)$$

where *r(t)* is the distance a molecule travels in time *t*. The mass diffusion coefficient *D* is proportional to the slope of Mean-Square Displacement (MSD) over time, which can be calculated by tracing the trajectory of all atoms. The MSD and calculated mass diffusion coefficient from MD simulations for five binary mixtures are given in **Figure 5.2**. I also calculated the molecular movements of Component 1 and Component 2 separately and the corresponding mass diffusion coefficients are plotted together in **Figure 5.3** with the overall mass diffusion coefficients. It is clear that compared to the pure state, the movements of a molecule in the mixtures can be enlarged or restricted due to the intermolecular attractive and repulsive interactions between different molecules.

5.3 MOLECULAR DYNAMICS SIMULATION

The physical theory of MD simulations is a combination of quantum mechanics and statistical mechanics. The Schrödinger equation in quantum mechanics plays a role of Newton's law and conservation of energy in classical mechanics. It predicts the future behavior of a dynamic system and the corresponding kinetic and potential energies are converted into the Hamiltonian to generate the evolution of the wave function in time and space domains. Also, the statistical mechanics have an effect on NMR properties because it can be calculated from a wide ensemble of all different states with their own probabilities.

Force field is a function that describes the interactions between atoms and molecules as,

$$V = \sum_{bonds} \frac{1}{2} k_{bonds} (r - r_0)^2 + \sum_{angles} \frac{1}{2} k_{angle} (\theta - \theta_0)^2$$
$$+ \sum_{torsions} \frac{1}{2} k_{bonds} (\mathrm{V}_1(1+\cos\varphi) + \mathrm{V}_2(1-\cos 2\varphi) + \mathrm{V}_3(1+\cos 3\varphi) + \mathrm{V}_1(1-\cos 4\varphi))$$
$$+ \sum_{Coulomb}^{i<j} \frac{q_i q_j}{r_{ij}} + \sum_{vdW}^{i<j} \left[4\varepsilon_{ij} \left\{ \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right\} \right] . \quad (5.20)$$

The potential energy function can be expressed as a sum of five contributions. The first two terms are the oscillations of the equilibrium bond length and bond angle. The next is called dihedral potential of torsions which are basically the rotations between two groups of bonds. The last two terms are related to non-bonded interactions; coulomb potential and Lennard-Jones potential, the latter of which calculates the Van der Waals interaction between the atoms. Thus, the force is exerted to the direction of the negative gradient of the potential energy, and corresponding atoms and molecules interact accordingly. The limitation of MD simulation is the reduction from a fully quantum description to a classical potential. This potential is an approximation in that quantum mechanical atoms are accelerated to the direction of the force by Newton's second law which is classical mechanics. More details of MD simulations are described in Allen et al. (1987).

I chose the software package GROMACS 5.1.4 (Abraham et al., 2017) to perform MD simulations because it provides high performance with a user-friendly interface. I tested the water, methanol, octane, pentane, cyclohexane, toluene, and their mixtures in a cubic box with the Periodic Boundary Condition (PBC). For the water molecule, I tested one of the popular models SPC/E (Berendsen et al., 1987) but it makes intramolecular interactions weaker. This weakness exist because it has longer distance and the wider angle between hydrogen atoms than the structure of the ideal water model. Another popular water model of TIP3P provides very high mass diffusion coefficient which is

unrealistic. Therefore, I prepared the water molecule manually, and the details of structure and topology for all molecules used in this study are presented in **Appendix B**. For MD simulations, selection of the proper force field is critical. Caleman et al. (2012) compared the force fields of the GAFF (Wang et al., 2004), OPLS/AA (Jorgensen et al., 1996), and CHARMM (Vanommeslaeghe et al., 2010) and found that the OPLS/AA force field is suitably parametrized for most liquids. Thus, in this study, I applied the OPLS/AA force field for all MD simulations because it has been optimized to reproduce the thermodynamic and structural liquid properties at various temperatures.

I tested five binary mixtures with four concentrations each in the system of a cubic box with a 3.0 nm length. The number of molecules and their corresponding molar and volumetric fractions are tabulated in **Table 5.1**. They are miscible but show nonlinear behaviors when mixed. The velocity-rescaling scheme (Bussi et al., 2007) was selected for the thermostat to control the constant temperature of 298.15 K. The Berendsen algorithm (Berendsen et al., 1984) was implemented for pressure control. The potential energy of the system was minimized and the box size was adjusted by applying canonical ensemble (NVT) and isothermal-isobaric ensemble (NPT) ensembles. The grid spacing was defined at 0.12 nm and the radius for Lennard-Jones interactions and Particle Mesh Ewald (PME) electrostatic interactions (Essmann et al., 1995) are limited to 1.4 nm for all molecules (Jorgensen et al., 1996). The LINCS algorithm was adapted to constrain bond distances between atoms (Hess et al., 1997). After the system was equilibrated, the production run was applied for 2 ns and the Verlet algorithm (Swope et al., 1982) was used for the equation of motion with a time step of 2.0 femtoseconds. A graphical description of MD simulation is presented in **Figure 5.4**. The output trajectory of all atoms are recorded every 0.1 ps for NMR analysis. A longer simulation time (up to 500 ns) and a larger cubic box (4 nm length) were tested but the estimation results are similar.

Thus, the current simulation setup is considered a good compromise between computational load and desired accuracy. Density, NMR relaxation time, and mass diffusion coefficient of each system were calculated, and found that intramolecular ^{1}H-^{1}H dipole-dipole interaction plays a dominant role in the NMR relaxation process.

5.4 RESULTS AND DISCUSSIONS

The NMR longitudinal relaxation times and mass diffusion coefficient of liquid mixtures are computed from their molecular trajectories. I have considered water and alcohol mixtures (water and methanol), alcohol and alkane mixtures (octane and methanol), alkane and alkane mixtures (octane and pentane), alkane and cycloalkane mixtures (pentane and cyclohexane), and cycloalkane and aromatic mixtures (cyclohexane and toluene). All of these mixtures are stabilized and equilibrated at 298.15 K and 1 bar before the trajectory acquisition. Since the molecular correlation times of liquid mixtures are smaller than the inverse Larmor angular frequency, the estimated longitudinal relaxation time is almost equal to the transverse relaxation time.

Estimated autocorrelation functions, $G(t)$ of intramolecular dipole-dipole interactions of the first component are presented in **Figure 5.5**. As time proceeds, the original arrangement of hydrogen spin at $t = 0$ in the sample becomes more distracted and the spins change their positions. Thus, autocorrelation function decays to zero at $t = \infty$. A molecule with a short correlation time decays faster because its hydrogen spin arrangement will be disturbed rapidly. A decay of autocorrelation function can be expressed as an exponential function as

$$G(t) = \exp\left(-\frac{t}{\tau_c}\right) \tag{5.21}$$

and its Fourier transform is Lorentzian spectral density function *J(ω)* as

$$J(\omega) = G(0)\frac{2\tau_c}{1+\omega^2\tau_c^2}. \tag{5.22}$$

According to time-dependent perturbation theory, a probability of spin transitions between two states is evaluated by spectral density function at the frequency of inverse correlation time. Therefore, a shorter correlation time leads to a smaller rate of spin transitions; NMR relaxation time becomes longer respectively.

Effective NMR relaxation times of liquid mixtures are presented in **Figure 5.6**. The estimated NMR relaxation time of water from MD simulations showed a slightly larger value than data from the experiment (Krynicki, 1966). A comparative study on NMR relaxation time of liquid water from MD simulations (Calero et al., 2015) showed that most water models are overestimating NMR relaxation time; 12.4 sec (TIP3P), 7.0 sec (SPC/E), 5.0 sec (TIP4P), and 3.8 sec (TIP4P/2005). Relaxation times of methanol and cyclohexane showed excellent agreements to the experimental data. The estimated relaxation time of octane and pentane are slightly underestimated compared to the experimental results (Shikhov et al., 2016) but they are in the same order of magnitude. The difference between the relaxation times estimated from MD simulations and experimental data can be explained by (1) the length of autocorrelation functions, (2) number of hydrogen spin pairs considered for autocorrelation functions, (3) simulation time, (4) size of a cubic box, (5) molecular structures and distributions of partial charges, and (6) force field parameters.

Contributions of intramolecular and intermolecular dipole-dipole interactions are compared in **Figure 5.7**. Except for water and methanol mixtures, most fluid mixtures exhibit 60 to 70% of total relaxations arising from intramolecular interactions. Hydrogen bonds provide a possible explanation for lower intramolecular interactions between water and methanol mixtures. Because water molecules located between methanol molecules generate hydrogen bonds; average distance between intermolecular hydrogen spins decreases and corresponding NMR relaxation time decreased. Intermolecular dipole-dipole interactions between different molecules are presented in **Figure 5.8**. The second moment can be explained by square of the local magnetic field induced by dipole-dipole interactions. As a molar fraction of the first component increases, intermolecular interactions between the first component increases, and those between the second component decreases, respectively. Molar fractions where the interactions between the first and second components become maximum depend on viscosity, chemical bonds, size, and partial charge of the molecules.

5.5 CONCLUSIONS

I have estimated NMR time relaxations and mass diffusion coefficients of water, methanol, octane, pentane, cyclohexane, toluene, and their mixtures of various concentrations at an ambient condition. These NMR properties computed from MD simulations successfully reproduced the main aspects of experimental results. The approach presented in this chapter is used to quantify contributions of each component of the ensemble average of overall NMR properties. NMR time relaxations are mainly governed by the intramolecular ^{1}H-^{1}H dipole-dipole interactions. Also, MD simulations indicate that intermolecular interactions are governed by not only the distance and angle

between hydrogen spins but also molar fractions of each component. This understanding shows that an increased number of spin pairs of a component leads to a higher probability that hydrogen spins interact each other, and those increased spin pairs also contribute to the total relaxation process. Mass diffusion coefficients computed from Einstein's relation showed nonlinear behaviors as a function of molar fractions. Mass diffusion coefficients of each component do not proportionally contribute to the effective diffusion coefficients. The prediction accuracy of NMR properties can be improved with more molecule samples, refined modeling of molecular structure and partial charge, optimized parameterization of the force fields, and longer simulation time.

Table 5.1: Number of molecules for each component and their volumetric and molar concentrations of the five series of liquid mixtures at 298.15 K. The volumetric and molar concentrations given in the table are the fractions of the first component.

Liquid mixtures	X_V **[V/V]**	X_M **[mol/mol]**	**Number of component 1 molecule**	**Number of component 2 molecule**	**Length of box [nm]**
Methanol and Water	0.0000	0.0000	0	900	2.99668
	0.4470	0.2647	180	500	2.98043
	0.6160	0.4167	250	350	2.98719
	0.7683	0.5962	310	210	2.98690
	0.9122	0.8222	370	80	3.00510
	1.0000	1.0000	402	0	3.00784
Octane and Methanol	0.0000	0.0000	0	402	3.00784
	0.5010	0.2000	50	200	3.01981
	0.7009	0.3684	70	120	3.02177
	0.8505	0.5862	85	60	3.01979
	0.9502	0.8261	95	20	3.01428
	1.0000	1.0000	100	0	3.01675
Octane and Pentane	0.0000	0.0000	0	140	2.98139
	0.2513	0.1923	25	105	2.98569
	0.5018	0.4167	50	70	2.99527
	0.6868	0.6087	70	45	3.00965
	0.8570	0.8095	85	20	2.99691
	1.0000	1.0000	100	0	3.01675
Cyclohexane and Pentane	0.0000	0.0000	0	140	2.98139
	0.1965	0.2069	30	115	3.01220
	0.3982	0.4138	60	85	2.99567
	0.5704	0.5862	85	60	2.98720
	0.7895	0.8000	120	30	3.00468
	1.0000	1.0000	150	0	3.00307
Toluene and Cyclohexane	0.0000	0.0000	0	150	3.00307
	0.1974	0.2000	30	120	2.99839
	0.3961	0.4000	60	90	2.99090
	0.5881	0.5921	90	62	2.99964
	0.7815	0.7843	120	33	3.01627
	1.0000	1.0000	153	0	3.00811

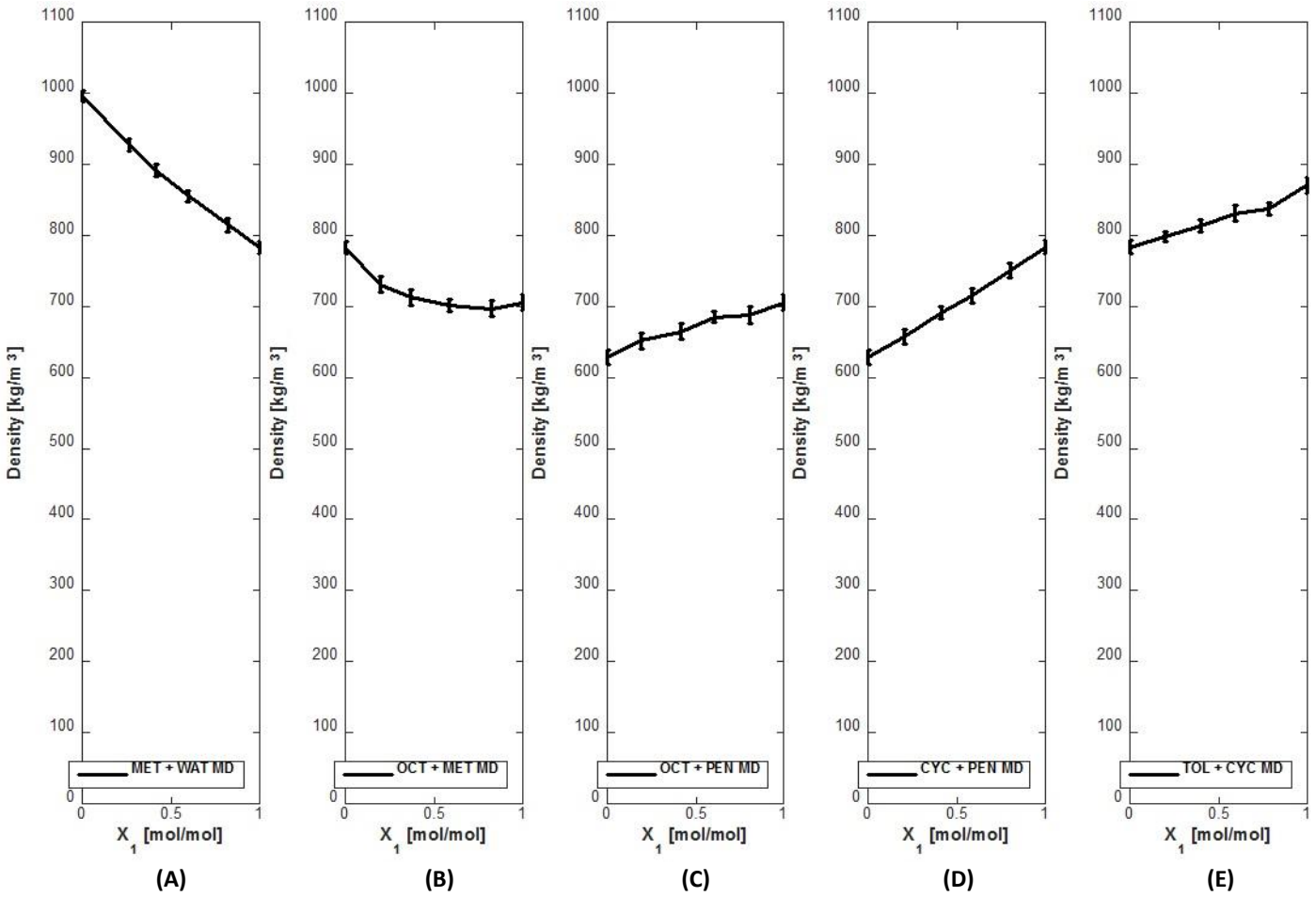


Figure 5.1: The density of (A) methanol and water, (B) octane and methanol, (C) octane and pentane, (D) cyclohexane and pentane, and (E) toluene and cyclohexane mixtures from MD simulation results with corresponding error bars.

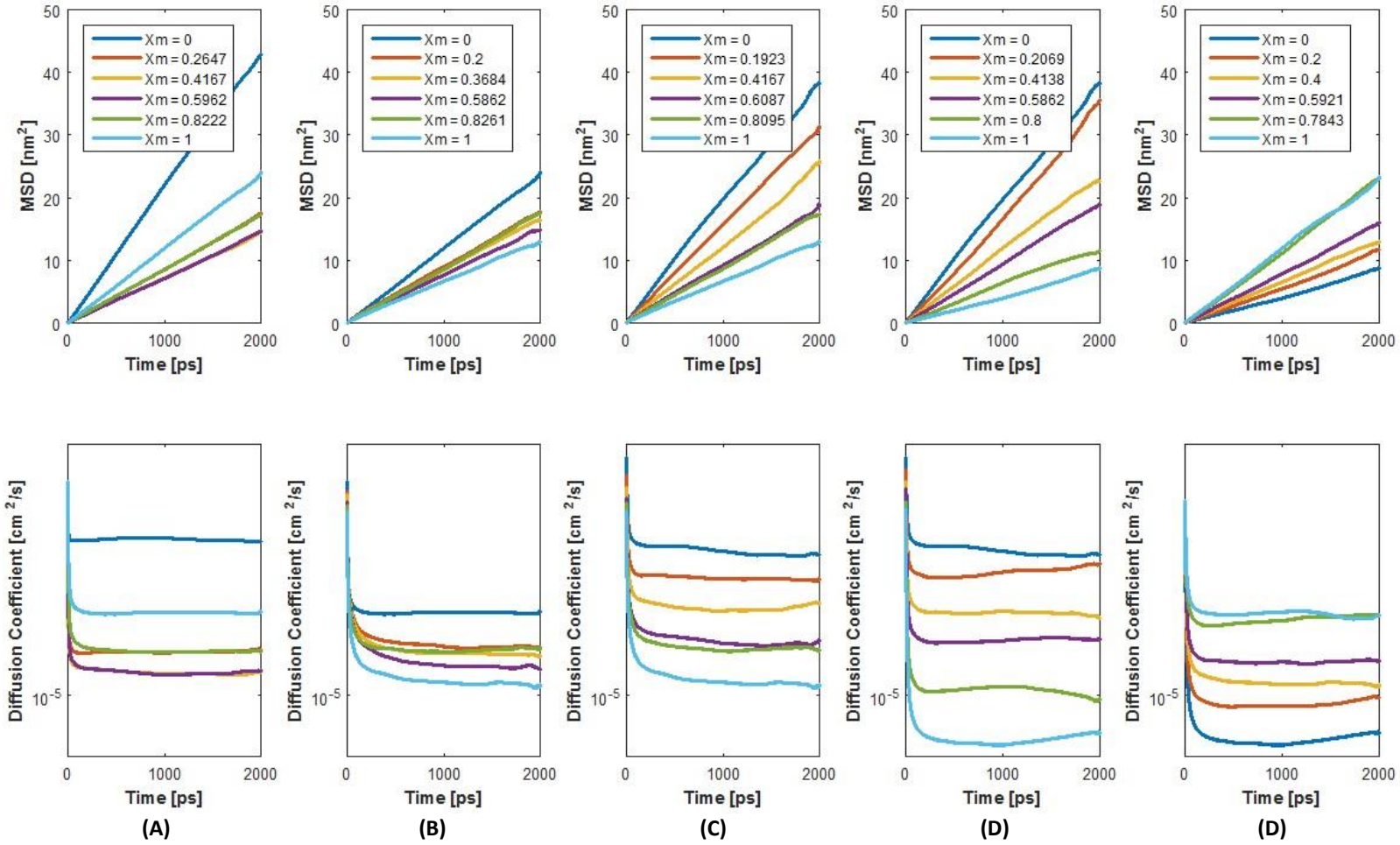


Figure 5.2: The mean square displacements (upper figures) and mass diffusion coefficients (lower figures) of (A) methanol and water, (B) octane and methanol, (C) octane and pentane, (D) cyclohexane and pentane, and (E) toluene and cyclohexane mixtures from MD simulation results.

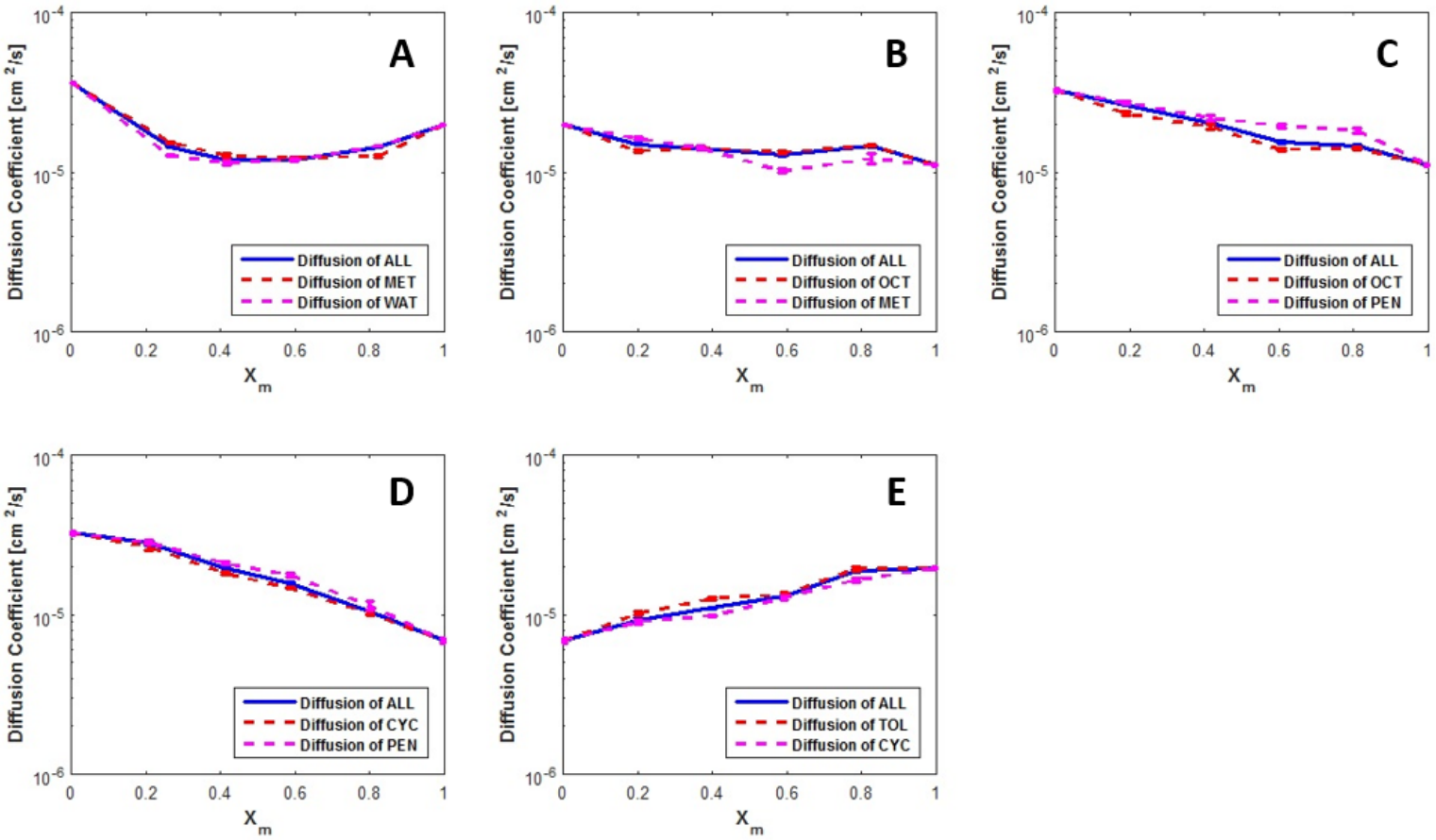


Figure 5.3: Mass diffusion coefficients calculated from the Einstein relation as a function of the molar concentration of the first component at 298.15 K and 1 bar. The mass diffusion coefficients of all molecules (continuous blue line), the first component (dashed red line), and the second component (dashed magenta line) are marked with error bars on (A) methanol and water, (B) octane and methanol, (C) octane and pentane, (D) cyclohexane and pentane, and (E) toluene and cyclohexane mixtures.

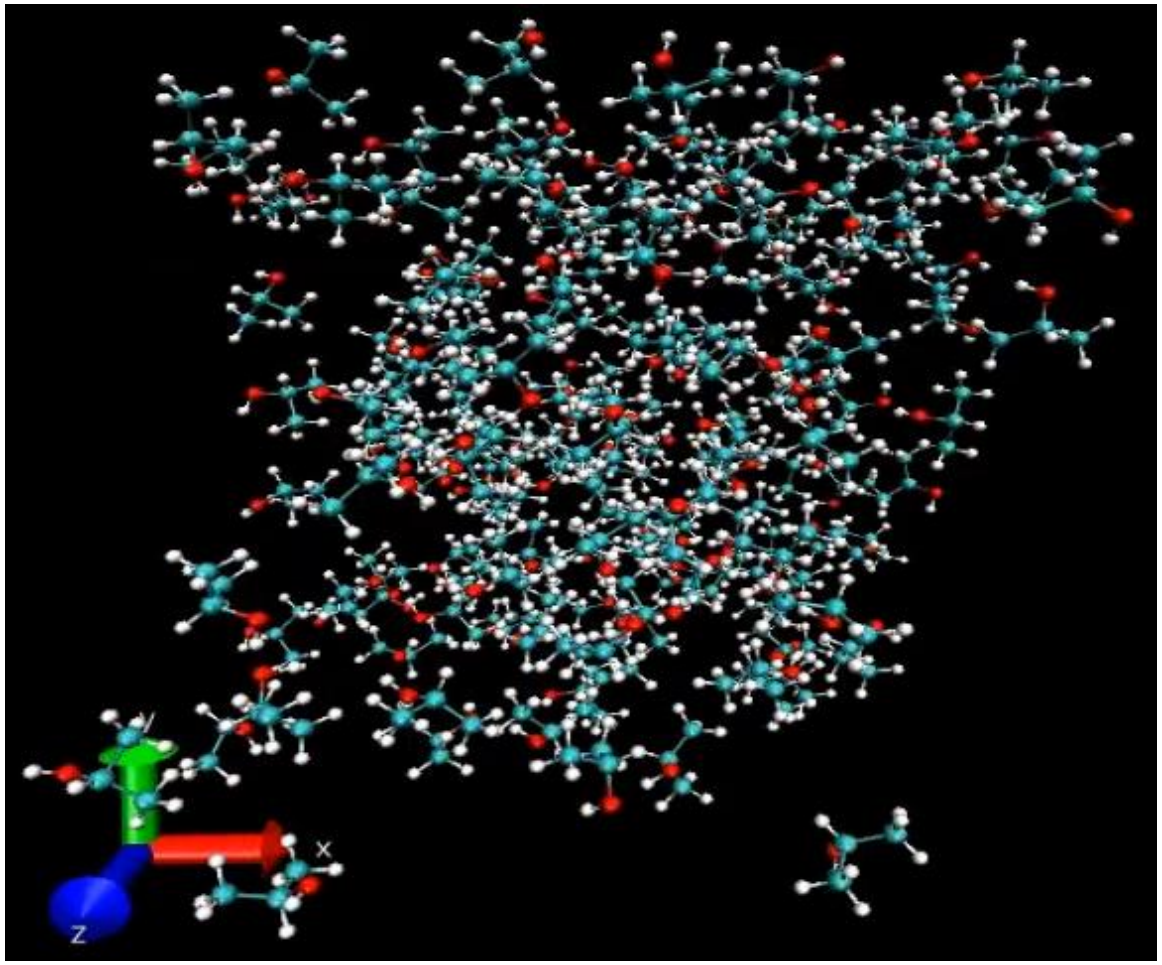


Figure 5.4: A typical visualization of MD simulation. Red, white, and cyan colors represent oxygen, hydrogen, and carbon atoms, respectively. These molecules are moving inside the cubic box with PBC at given temperature and pressure.

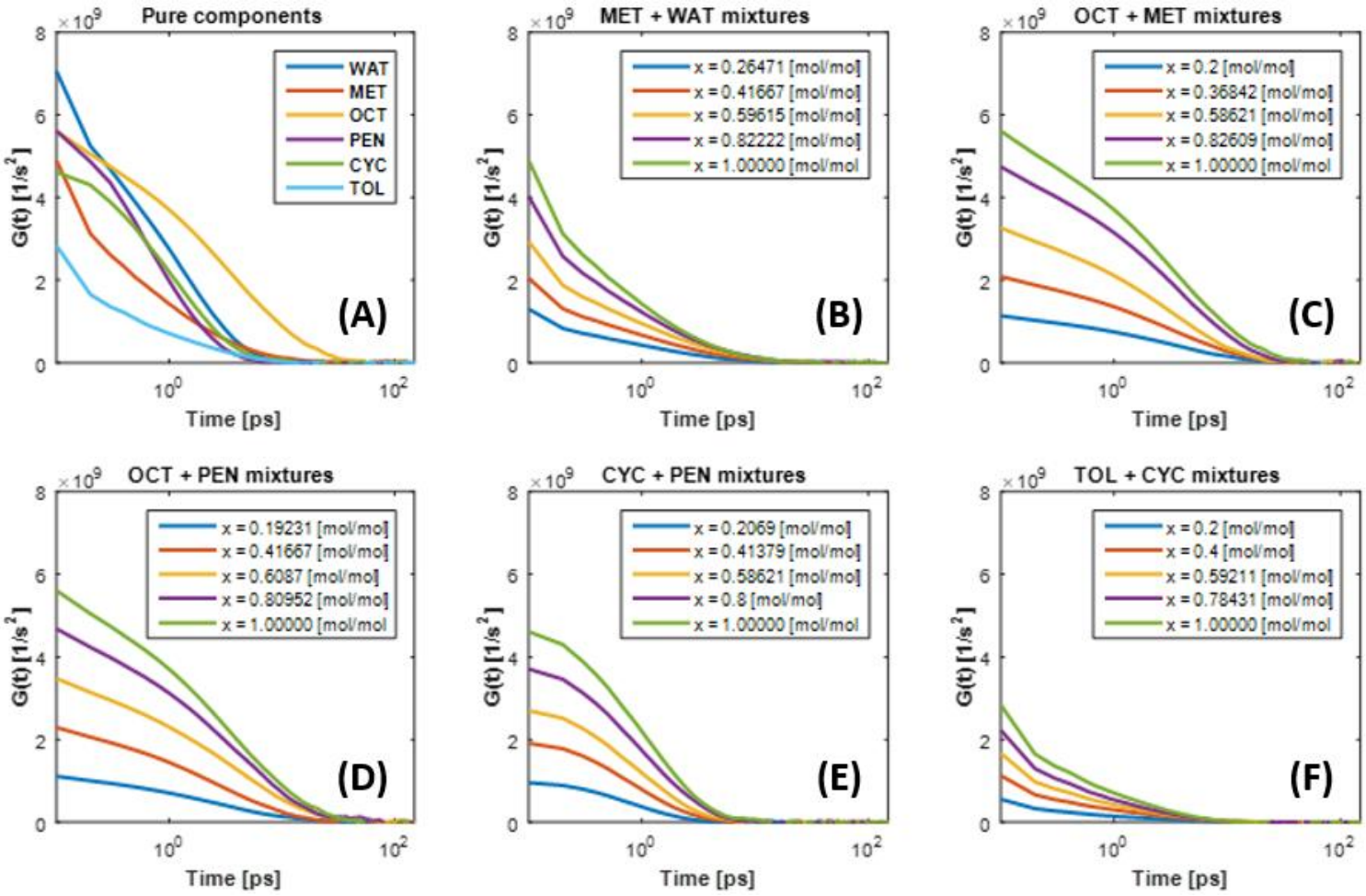


Figure 5.5: The estimated autocorrelation function of intramolecular dipole moment interactions of the first component of (A) pure components, (B) methanol and water, (C) octane and methanol, (D) octane and pentane, (E) cyclohexane and pentane, and (F) toluene and cyclohexane mixtures from MD simulation results.

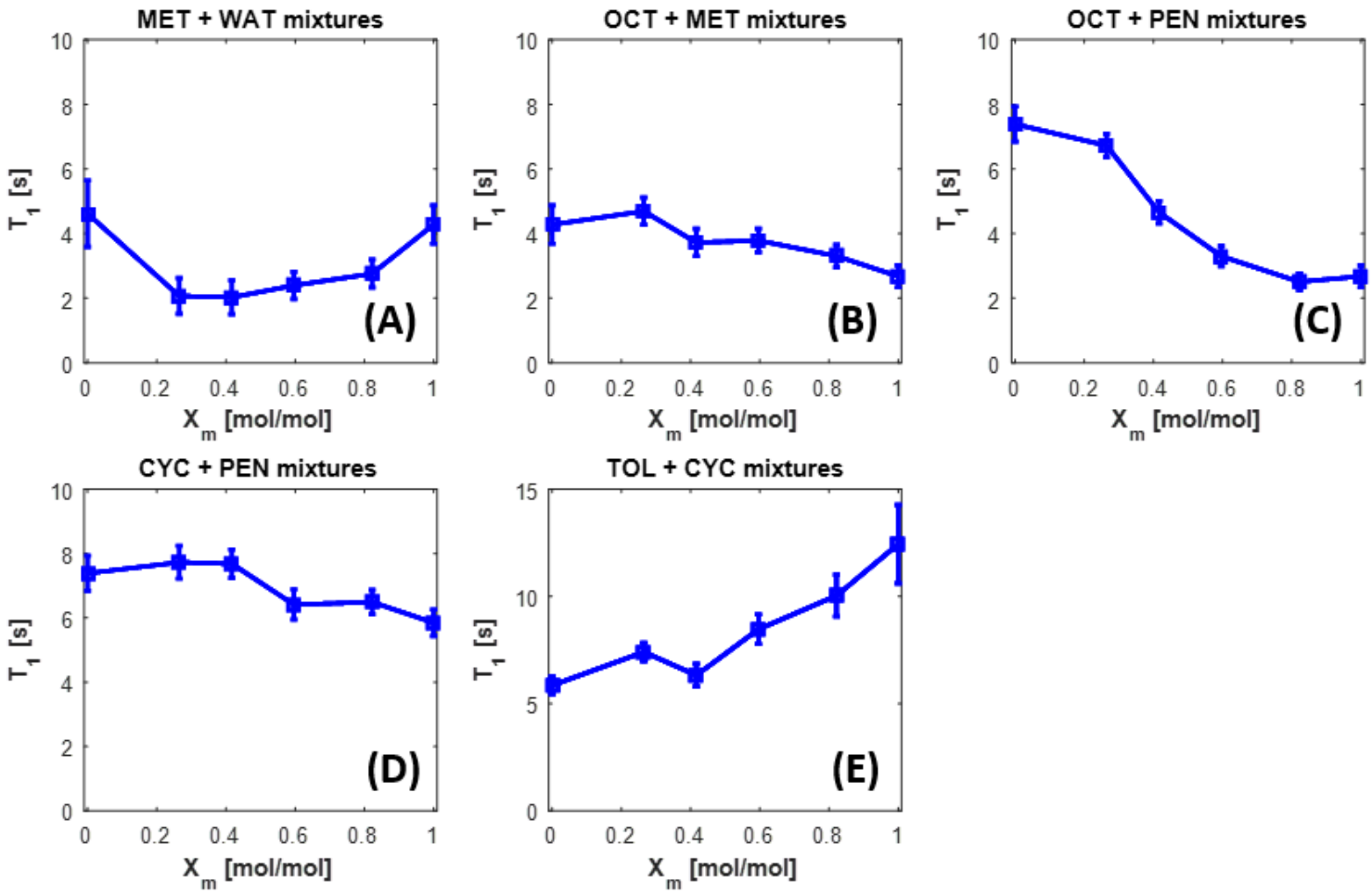


Figure 5.6: The estimated NMR relaxation times and corresponding error bars of (A) methanol and water, (B) octane and methanol, (C) octane and pentane, (D) cyclohexane and pentane, and (E) toluene and cyclohexane mixtures from MD simulation results.

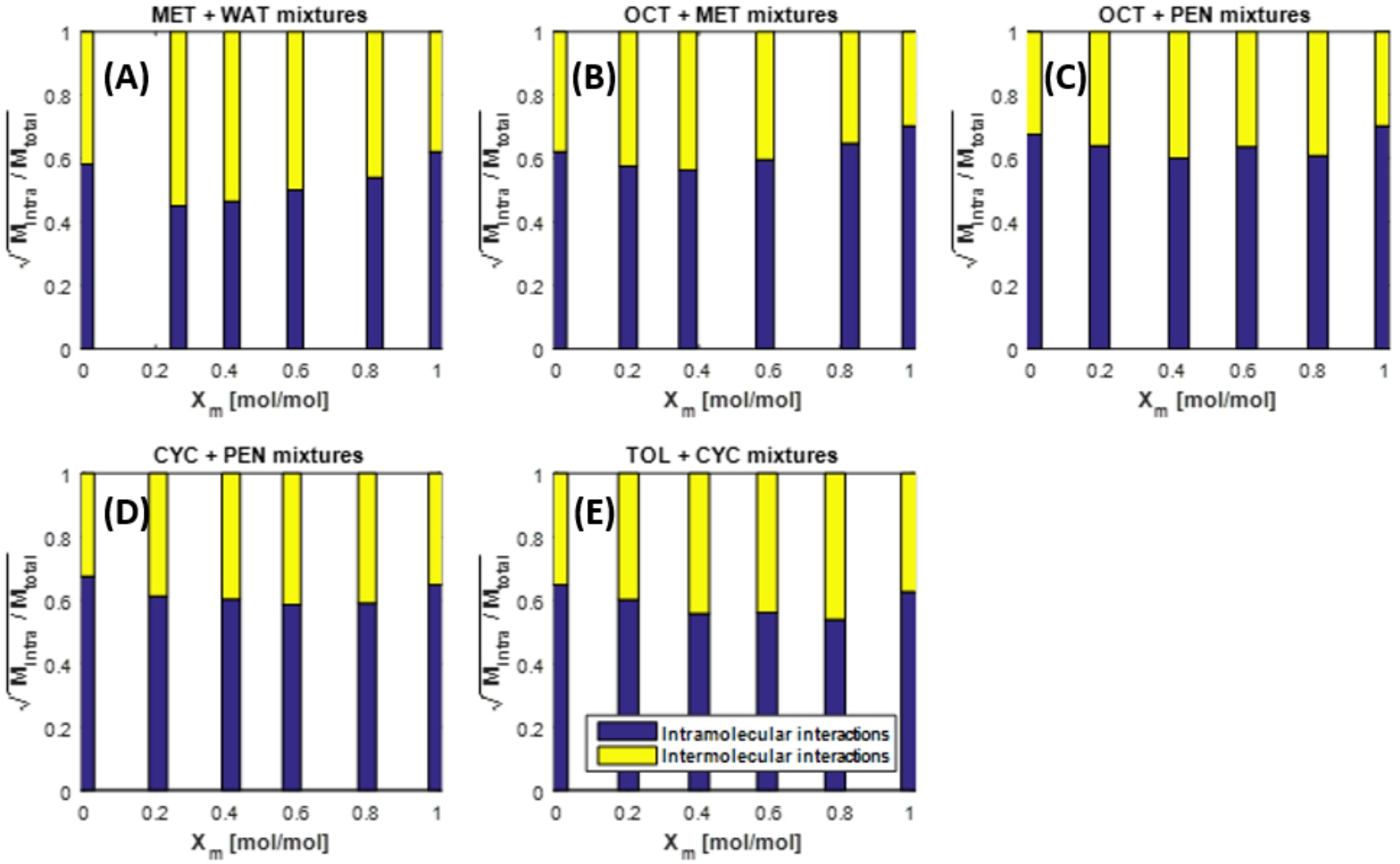


Figure 5.7: The estimated NMR relaxation times of intra (blue) and inter (yellow) molecular interactions of (A) methanol and water, (B) octane and methanol, (C) octane and pentane, (D) cyclohexane and pentane, and (E) toluene and cyclohexane mixtures from MD simulation results.

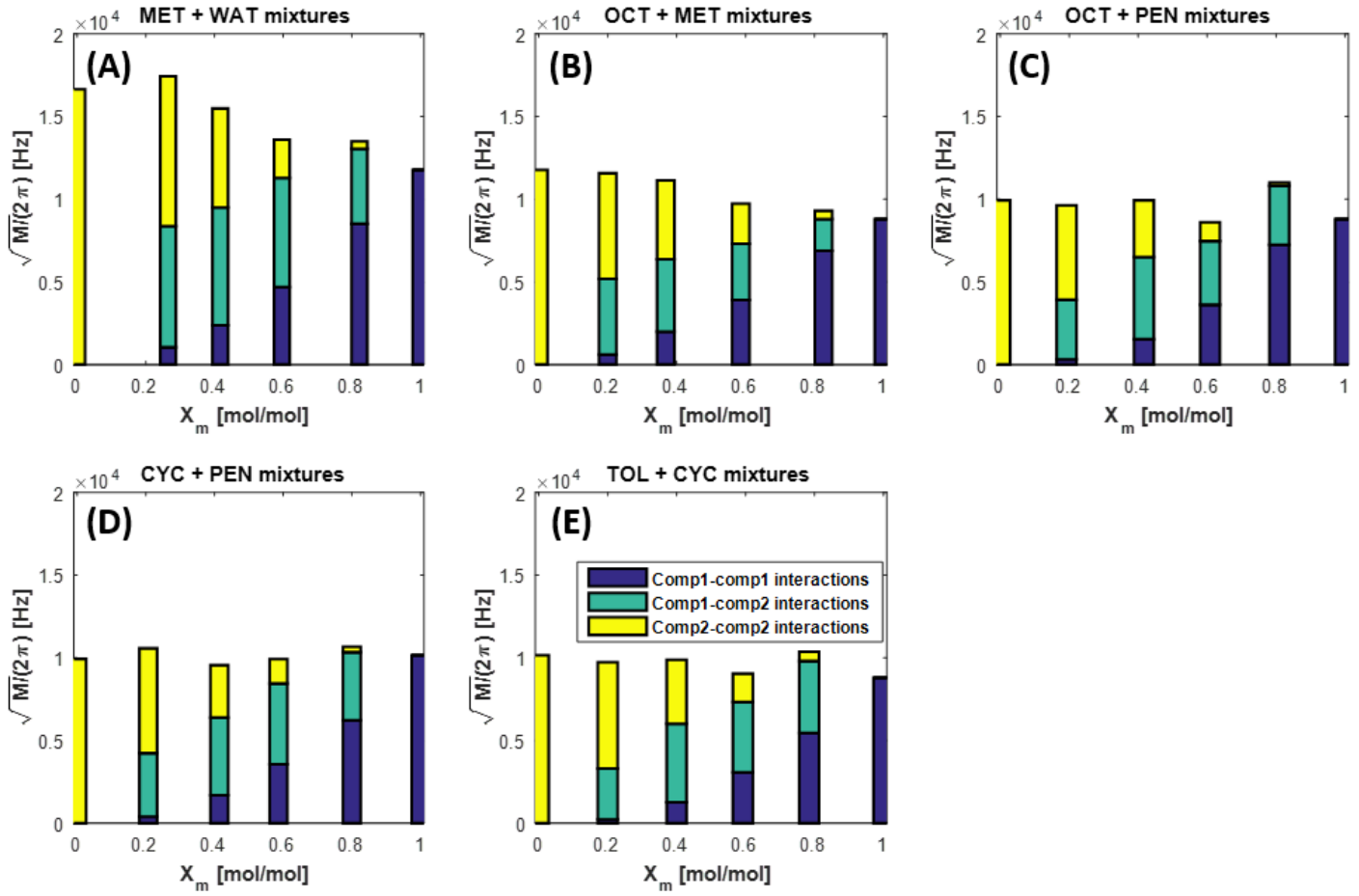


Figure 5.8: The estimated square root of the second moment of intermolecular interactions between component 1 molecules (blue), component 1 and 2 molecules (green), and component 2 molecules (yellow) of (A) methanol and water, (B) octane and methanol, (C) octane and pentane, (D) cyclohexane and pentane, and (E) toluene and cyclohexane mixtures from MD simulation results.

Chapter 6: Summary, Conclusions, and Recommendations

This final chapter summarizes the technical developments and contributions stemming from the dissertation. It also gives general conclusions from the results, and provides recommendations for future study.

6.1 SUMMARY

The primary objective of this dissertation is to develop new methods for the assessment of in-situ fluid properties using various borehole measurements. To achieve this goal, I considered nuclear, dielectric, optical, and magnetic resonance measurements in the analysis. Along with well logs and laboratory measurements, I took varied approaches for the detection and concentration quantification of individual components included in a complex fluid mixture. Approaches included linear and nonlinear inversions, molecular dynamics simulations, multivariate analysis, radial basis functions, and artificial neural networks. To assess fluid properties, the methods developed in this dissertation have several advantages compared to conventional techniques, namely:

- The inversion algorithm using various nuclear properties enables a thermodynamically consistent assessment of three pseudo-hydrocarbon components in complex phases. A cooperative implementation of SNUPAR and PREOS flash calculation enables the robust estimations of in-situ fluid properties.
- The forward and inverse modeling of dielectric relaxation spectra estimates accurate dielectric permittivity and relaxation time of polar

liquid mixtures. These properties were cross-validated with MD simulations and confirmed that macroscopic ensemble average of dielectric properties can be explained with microscopic features such as a number of hydrogen bonds per unit volume.

- Interpretations of optical spectroscopy measurements associated with RBF and ANN models exhibited a substantial reduction of RMSE compared to traditional multivariate methods. Additionally, combining optical spectroscopy measurements with complementary physics reduce the uncertainty of liquid identification.
- The MD simulations successfully described the main features of experimental NMR time relaxations. They were able to quantify individual intramolecular and intermolecular ^{1}H-^{1}H dipole-dipole interactions, which are arduous and challenging tasks in traditional experimental analysis. Also, mass diffusion coefficients of each component were assessed and successfully compared to the effective diffusion coefficient of binary liquid mixtures.

In Chapter 2, I introduced a linear inversion algorithm to estimate in-situ fluid properties from borehole nuclear measurements. In conjunction with SNUPAR and PREOS, inversions enabled depth-continuous and stable estimations of hydrocarbon compositions. The application of the proposed method to one synthetic and two field examples yielded the thermodynamically consistent interpretation of nuclear measurements. Hydrocarbon viscosities calculated from nuclear measurements were in agreement with the viscosity independently computed from magnetic resonance data. This validation verified the robustness of the proposed interpretation method.

In Chapter 3, I introduced laboratory measurements of dielectric relaxation spectra of various polar liquid mixtures. A nonlinear inversion algorithm with the HN model was developed to estimate dielectric permittivity and relaxation time of fluid mixtures. Also, MD simulations were introduced to assess thermodynamic and dielectric properties of fluids. Dielectric properties estimated from MD simulations exhibited fair agreements with inversion-based calculations performed with laboratory measurements.

Chapter 4 described three optical spectroscopy measurements. I measured NIR, Infrared, and Raman spectra of fluid mixture samples and identified their characteristic bands. Each spectrum was preprocessed for chemometric analysis. Several techniques were applied to maximize the correlations between measured optical spectra and physical properties of fluid mixtures. Compared to the conventional PLSR approach, RBF and ANN methods yield better predictions of fluid mixture compositions. Also, the combination of multiple spectroscopic techniques was introduced to enhance the accuracy of the estimated physical properties.

Finally, I described MD simulations to assess NMR time relaxations of liquid mixtures in Chapter 5. Water, methanol, octane, pentane, cyclohexane, toluene, and their binary mixtures were tested and compared to experimental data. The MD approach can calculate dipole-dipole interactions of each hydrogen spin pair. It also enables quantification of intramolecular and intermolecular interactions of each component. Therefore, NMR time relaxations arising from both intramolecular and intermolecular interactions contribute to total NMR time relaxation, and their changes in intermolecular interaction were assessed with different molar fractions of binary liquid mixtures.

6.2 CONCLUSIONS

This section lists the pivotal conclusions deriving from the four technical chapters of the dissertation.

6.2.1 Part One: Thermodynamically consistent estimation of hydrocarbon composition from nuclear and magnetic resonance measurements.

i. The main contribution from this part is the development of inversion-based interpretation techniques in combination with thermodynamic phase behaviors. Complex variations in fluid phases (from oil to gas phase) could cause erroneous identification of reservoir fluids. The proposed inversion algorithms use borehole nuclear measurements to calculate hydrocarbon composition of reservoir fluids for oil, gas, and transitional saturated sandstones.

ii. Reconciliation of SNUPAR and PVT flash calculations yielded density, Sigma, neutron porosity, inverse of migration length, and volumetric cross section from borehole nuclear measurements. For this approach, the appraisal of porosity and water saturation is critical because small errors on these petrophysical properties can propagate to large uncertainty in the estimation of hydrocarbon compositions.

iii. When comparing nuclear properties, density is a key property for the estimation of hydrocarbon compositions. Also, Sigma, inverse of migration length, and neutron porosity are helpful to quantify hydrocarbon properties. Unlike the above properties, volumetric cross section or PEF will degrade the inversion results because it primarily depends on matrix properties instead of fluids. The inversion technique based on nuclear measurement introduced in this dissertation is reliable when oil and gas coexist in complex phases.

iv. The joint interpretation of hydrocarbon viscosity from nuclear measurements and NMR relaxation times was successfully applied to one synthetic case perturbed with 5% Gaussian noise and two field examples. Estimated hydrocarbon compositions and viscosities honored both well logs and core data.

6.2.2 Part Two: Investigation of dielectric and thermodynamic properties of polar liquids by dielectric spectroscopy and molecular dynamics simulation.

i. Mixtures of water, methanol, and 2-propanol exhibit nonlinear variations of dielectric permittivity and relaxation time as a function of molar fractions. The newly developed inversion algorithms successfully estimate dielectric permittivity and relaxation time from dielectric relaxation measurements. Forward modeling with CC, DC, and HN models reproduced the real and imaginary parts of the measured dielectric spectra.

ii. MD simulations offer a reliable approach to emulate various thermodynamic and physical properties of liquids. The density and enthalpy of polar liquid mixtures agree with the experimental data. Also, dielectric properties estimated via inversion algorithms were successfully benchmarked with MD simulations. In comparison with experimental data, I found that average errors of dielectric properties calculated from MD simulations were lower than 4%.

iii. Mass diffusion coefficients of liquid mixtures were evaluated using the MSD of each molecule and Einstein's relation. Furthermore, the number of hydrogen bonds between molecules of equal and different components were calculated. Both mass diffusion coefficient and number of hydrogen bonds are in agreement with experimental data.

iv. A strong correlation exists between the number of hydrogen bonds per unit volume and the dielectric relaxation time of liquid mixtures. This behavior is due to the fact that dielectric relaxation times are associated with the rapid process of hydrogen bond generations and decompositions. The probability of interactions between water and other alcohol molecules is largely governed by the dynamics of hydrogen bonds and their structures. Therefore, larger numbers of hydrogen bonds per unit volume allow a molecule to find more available hydrogen bond acceptors nearby; thus, it takes less time for them to align in the direction of the external electric field.

v. The selection of the force field and parameters for MD simulations has great effects on the estimation of physical properties. Specifically, partial charges in the structure of a molecule govern its dipole moment which is directly related to the interactions with the Coulomb potential and dielectric permittivity. Optimized parameterizations of molecular structures and force fields are necessary to accurately estimate dielectric properties with MD simulations.

6.2.3 Part Three: Comparison of NIR, Infrared, and Raman spectroscopies for the estimation of liquid mixture compositions, density, and viscosity.

i. I successfully developed calibration models to estimate component concentrations from nonlinear NIR absorption spectra. Unlike traditional multivariate approaches, the introduced RBF and ANN methods account for these nonlinear features. A series of spectral pre-treatments including baseline corrections, spectral truncations, and normalization generated a better relationship between measured optical spectra and component concentrations. The first seven PCA

scores, which contain 99.29% of the total spectral variance, were used for spectral analysis.

ii. Comparisons between NIR, Infrared, and Raman spectral measurements were introduced for the analysis of fluid components. These optical spectroscopy measurements were processed with PLSR and ANN methods to calculate physical properties such as density and viscosity of liquid mixtures. Infrared and Raman spectra provided better prediction results in comparison with NIR spectra.

iii. Finding the spectral bands that are most sensitive to variations of fluid component concentrations is important to construct robust training models. For NIR spectra, the first overtone was chosen over the combination region to avoid the superposition of multiple broad absorption peaks. For Infrared and Raman spectra, bending vibration modes were selected because they are rich in information concerning molecular structures.

iv. I achieved improved predictions of physical properties by integrating three optical spectra together. Combining spectral measurements with the first overtones and bending vibrations provided detailed information about molecular structures and their chemical bonds. It was found that the complementary physics of optical spectra can reduce uncertainty in hydrocarbon identifications.

v. I applied an ANN method to analyze optical spectra and compared estimation results to results obtained with traditional methods. The ANN model yielded a substantially reduced RMSE compared to the PLSR method for the estimation of density (1.5-2.1%) and viscosity (13-22.2%).

6.2.4 Part Four: NMR relaxation time and diffusion coefficient of liquid mixtures calculated with molecular dynamics simulations.

i. I successfully developed fluid mixture models with MD simulations for atomic scale interpretations of NMR relaxation properties. Water, methanol, octane, pentane, cyclohexane, toluene, and their mixtures of various concentrations were simulated at 298.15 K and 1 bar. From the trajectory of these molecules, the distance and angle to the z-direction were calculated between hydrogen spin pairs.

ii. The developed method can quantify NMR time relaxations of not only each hydrogen spin pair but also the ensemble average of all hydrogen atoms. This feature enables the appraisal of each component's capacity to influence the overall NMR time relaxations. Similarly, the MSD of each component was separately computed to estimate the mass diffusion coefficient of the corresponding components.

iii. NMR properties computed from MD simulations successfully reproduced the main features of experimental results. However, simulated NMR relaxation times were larger than the relaxation times measured at downhole conditions. Because no surface and gradient relaxations apply to MD simulations, the theoretical bulk relaxation times of hydrocarbons tend to have longer relaxation times in comparison with NMR logging measurements. Also, the absence of dissolved paramagnetic materials in the numerical simulations increases the NMR relaxation times.

Given the above experiments and numerical simulations, this study determines that the dielectric technique best distinguishes between polar and nonpolar liquids. Maximum property contrasts between alkanes and aromatics were observed with optical

spectra. Similarly, NMR measurements can be used to differentiate between oil and gas or heavy and light hydrocarbon mixtures. In the presence of mud filtrate contamination in hydrocarbons, none of these techniques reliably discriminates individual fluid components. Careful selection of the measurement technique is necessary, which depends on the property contrasts between hydrocarbons and mud filtrate.

6.3 RECOMMENDATIONS

Following is a list of recommendations that could advance the research topics documented in this dissertation:

i. Nuclear properties are calculated with SNUPAR. However, SNUPAR is outdated and provides inaccurate results compared to simulations obtained with the Monte Carlo N-Particle code (MCNP, X-5 Monte Carlo Team, 2005). Instead of SNUPAR, UT-NuPro can be adapted with flash calculations to yield more reliable and accurate calculations of nuclear properties.

ii. In this dissertation, PREOS was used to calculate the phase behavior of hydrocarbons. Other EOS models such as those advanced by Redlich-Kwong (Murdock, 1993), Soave-Redlich-Kwong (Soave, 1972), Elliott-Suresh-Donohue (Elliott et al., 1990), and Benedict-Webb-Rubin-Starling (Starling, 1973) can be applied for better predictions of vapor-liquid equilibria.

iii. The effects of temperature, frequency of electric field, and sodium chlorite on dielectric properties need to be investigated further. As temperature increases, interactions between dipoles and the applied electric field will decrease due to thermal agitations. Also, electric fields in the THz range can provide valuable information about the ultrafast relaxation mode in femtosecond range. The

addition of salt ions causes hydration, and the corresponding polarization and permittivity will decrease. These effects can be quantified with both experiments and MD simulations.

iv. A new dielectric mixing law can be formulated based on MD simulations. Those MD simulations facilitate the appraisal of both polarization orientation and electronic polarization of molecules. These total polarizations, estimated from MD simulations, can enable a better description of dielectric mixing laws.

v. The ANN model used in this dissertation was designed with one or two hidden layers, and the number of neurons was determined by trial and error. However, the architecture of the neural network can be optimized by considering the complexity of the learning task. The Multi-Particle Collision Algorithm (MPCA) or Dynamically Expandable Networks (DEN) may be applied to achieve this goal.

vi. Other supervised learning algorithms including decision trees, Support Vector Machine (SVM), and naive Bayesian algorithms can be used to analyze the measured optical spectra for the estimation of fluid properties.

vii. The effects of WBM or OBM on optical spectroscopy measurements need to be further investigated. I acquired measurements of hydrocarbons mixed with WBM or OBM (not documented in this dissertation) and found that approximately 10% or more mud contamination makes the NIR interpretations unreliable. The robustness of the estimations in the presence of mud contamination can be analyzed and compared for the three optical spectroscopic techniques considered in this dissertation.

viii. The nonlinear inversion problem with dielectric measurements introduced in this dissertation is unstable when applying a nonrealistic initial guess or very low regularization parameters. The remedy for this problem is the inclusion of a

reference vector in the cost function. Better initial guesses or reference vectors can be obtained from the total polarizations of dipole moment and polarizability of fluids measured in the laboratory.

ix. Furthermore, the effects of temperature, pressure, and paramagnetic materials such as oxygens on NMR relaxation time and mass diffusion coefficient can be analyzed.

Appendices: Artificial Neural Network and Molecular Dynamics Simulations

APPENDIX A: ARTIFICIAL NEURAL NETWORK

This section will describe mathematical equations and calculation routines. A typical neural network with three layers used in this study is represented in **Figure A1**. The first layer is called the input layer that has n_{i+1} neurons, where n_i is the number of input units. Likewise, the second layer is a hidden layer with n_{h+1} neurons and the output layer has no neurons, respectively. A conventional feedforward propagation is implemented for neural networks. Feedforward neural network computes these parameters for every input data *x* as follows:

A.1 Calculation of Activation Functions

Calculate activations of unit *j* in the hidden layer as

$$a_j^{(2)}(\mathrm{x}) = \sum_{k=0}^{n_i} g(\Theta_{j,k}^{(1)} x_k^{(1)}) = \sum_{k=0}^{n_i} \frac{1}{1+e^{-\Theta_{j,k}^{(1)} x_k^{(1)}}}, \qquad \text{(A.1)}$$

where sigmoid function *g* was adapted to the activation function (Cybenko, 1989) since it provides a smooth transition as input values vary. $\Theta_{j,k}^{(1)}$ is a weight from a *k*-th unit of input layer to a *j*-th unit of the hidden layer. These weights are randomly initialized for symmetry breaking and are computed to the optimized values through the training process. Note that additional neuron $x_0^{(1)}$ is called the bias unit, which always has a value

of one. This bias value is important for training because it allows the activation function to shift.

A.2 Calculation of the Outputs

Similarly, the estimated output can be calculated by the activation function applied to the sum of the hidden layer activation values, which have been multiplied by the corresponding weight as,

$$\mathrm{h}_\theta(\mathrm{x}) = \sum_{j=0}^{n_h} \frac{1}{1+e^{-\Theta_{i,j}^{(2)} a_j^{(2)}}} . \tag{A.2}$$

A.3 Regularized Cost Function

Also, a bias neuron $a_0^{(2)}$ was added to the hidden layer. There is only one hidden layer but it can be expanded to an arbitrary number of hidden layers by repeating the calculation of activation function. Adding these hidden layers in the networks captures more complex and nonlinear features of the relationship between input data and output data.

The regularized cost function for the neural network is given by

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \sum_{l=1}^{n_o} \left[-y_k^{(i)} \log((\mathrm{h}_\theta(\mathrm{x}^{(i)}))_l) - (1-y_l^{(i)}) \log(1-(\mathrm{h}_\theta(\mathrm{x}^{(i)}))_l) \right] + \frac{\lambda}{2m} \sum_{l=1}^{L-1} \sum_{k=1}^{n_l} \sum_{j=1}^{n_h} (\Theta_{j,k}^{(l)})^2 , \tag{A.3}$$

where m is a number of input data, L is a total number of layers, y_l is the actual property of a l-th component, λ is a regularization parameter. The regularization parameter is essential to avoid overfitting or memorizing problem, which is common in the training

process. This regularization parameter can be optimized by comparing the errors on the training set and the errors on the validation set.

A.4 Calculation of the Backpropagation

Once cost functions are calculated, backpropagation can be implemented to calculate gradients. The idea of backpropagation is that activations calculated from feedforward can be compared to the true output values and measure how much each neuron is subject to the difference. This is an iterative process of gradient descent to minimize the overall cost that simultaneously calculates weights for neurons and gradients. The difference between each estimated output and the actual output unit can be computed as

$$\delta^{(3)} = (\mathrm{h}_\theta(\mathrm{x}^{(\mathrm{i})})) - y \,. \tag{A.4}$$

This error will be back propagated from the output layer to the hidden layer as

$$\delta^{(2)} = (\Theta^{(l)})\delta^{(3)} g'(\Theta^{(1)} a^{(1)}) \,, \tag{A.5}$$

where the sigmoid function gradient can be represented as

$$g'(\mathrm{z}) = \frac{d}{dz}\left(\frac{1}{1+e^{-z}}\right) = \left(\frac{1}{1+e^{-z}}\right)\left(1-\frac{1}{1+e^{-z}}\right) \,. \tag{A.6}$$

This gradient is accumulated for all output data as

$$\Delta^{(2)} = \sum_{i=1}^{m} \delta^{(3)} a^{(2)} \tag{A.7}$$

$$\Delta^{(1)} = \sum_{i=1}^{m} \delta^{(2)} a^{(1)} \tag{A.8}$$

Finally, regularized gradients of the cost function between the output layer and the hidden layer and those between the hidden layer and the input layer can be obtained by

$$\frac{\partial J(\Theta)}{\partial \Theta_{lj}^{(2)}} = \frac{1}{m} \Delta_{lj}^{(2)} + \frac{\lambda}{m} \Theta_{lj}^{(2)}, \tag{A.9}$$

$$\frac{\partial J(\Theta)}{\partial \Theta_{jk}^{(1)}} = \frac{1}{m} \Delta_{jk}^{(1)} + \frac{\lambda}{m} \Theta_{jk}^{(1)}. \tag{A.10}$$

Note that the regularization should be applied to all weights except the bias unit. The difference between gradients calculated from backpropagation and numerical gradients is very small, showing that the backpropagation provides a good approximation of gradients with much less computation time.

By repeating these steps from 1 to 4 for each input data, minimum cost function and optimized weights can be computed with gradient descent in conjunction with feedforward and backpropagation. To avoid the local minima of the cost function, one can repeat this process with different initial weights several times and choose the parameters with the minimum overall costs.

APPENDIX B: MOLECULAR DYNAMICS SIMULATION

B.1 Introduction

The first MD simulation was performed (Alder et al., 1957) using the hard-sphere model. The simulation imitates physical motions of molecules and atoms in a real environment. Each atom is interacting with nearby atoms for a given time and moves toward the direction of exerting forces. The trajectory of all atoms in the system is the output of MD simulation, and various interpretations can be made based on their movements. There are dozens of MD simulation packages, and some popular ones are GROMACS (Abraham et al., 2017), NAMD (Phillips et al., 2005), DL POLY (Smith et al., 2010), LAMMPS (Plimpton et al., 1995), and AMBER (Case et al., 2005).

Force fields contain information about the potential energy of a simulation system. As previously explained in Chapter 3 and Chapter 5, choosing a suitable force field is a critical factor for successful MD simulation. Each force field has its own parameters which are determined experimental and quantum mechanical studies of selected molecules. Brief formulation of the force field potential is described in Chapter 5. Some popular force fields are OPLS-AA (Jorgensen et al., 1996), CHARMM27 (Foloppe et al., 2000), AMBER94 (Cornell et al., 1995), and GROMOS 96 (Schuler et al., 2001).

B.2 Simulation methods

A general flowchart describing MD simulation steps are shown in **Figure B1**. MD simulation starts with preparations of molecular information. The detail structures and topologies of molecules used in this dissertation are attached at the end of **Appendix B**.

After preparing input files, molecules are inserted into the system at random positions. To stabilize the system, these molecules need to be equilibrated to release strain. Using the gradient of potential energy, an iteration of the steepest descent method is applied to minimize the energy. This is an essential procedure because unequilibrated molecules can be crashed or even broken by the forces. The next step is coupling with temperature and pressure. As temperature changes gradually from 0 K to the designed temperature, velocity of atoms are moderately changed to the corresponding temperature at each time step. Depending on experimental environments or physical property of interest, canonical ensemble (NVT), isothermal-isobaric ensemble (NPT), isoenthalpic-isobaric ensemble (NPH), or microcanonical ensemble (NVE) can be considered. To conserve mass and number of atoms in the system, periodic boundary conditions (PBC) can be implemented. A graphical description of the PBC concept is presented in **Figure B2**. When an atom goes out of the system, an identical atom enters from the opposite side of the boundary.

The last step of MD simulation is a production run. Every atom generates XYZ trajectory in every time step until the end of the simulation time. From this trajectory information, various physical and thermodynamic properties can be calculated. A system with several thousand atoms on several nanoseconds can be manageable in the workstation computing capability. However, if the size of simulation system is larger or the time scale is longer, it is recommended to use a supercomputer. Most computational tasks presented in Chapter 5 was performed by Texas Advanced Computing Center (TACC).

MD simulation can be a bridge between theoretical and experimental petrophysicists. While experiments find some interesting observations, theorists can make a hypothesis and explanation of experimental data and provide an approximated prediction of next experiments. Then experimentalist can verify the proposed theory, and

these interactions continue until the results from both sides can be cross-validated. On top of dielectric and NMR relaxations, a variety of physical phenomena on X-ray scattering (Moore, 1980), neutron scattering (Hansen et al., 1990), and atomic force microscopy (AFM, Kobayashi et al., 2016) have been studied with MD simulations.

B.3 Structure and topology inputs for MD simulations

Below information is structures and topologies of the molecules used in the MD simulations. Detailed parameter settings and input commands are available upon request.

<Structure of a water molecule>

```
One water
  3
    1SOL     OW    1   0.000   0.000   0.000 -0.5236  0.3981
0.1442
    1SOL    HW1    2  -0.096   0.000   0.000  1.8112 -1.0518 -
0.9491
    1SOL    HW2    3   0.024   0.093   0.000  1.3601  2.0733
3.6058
   0.00000   0.00000   0.00000
```

<Structure of a methanol molecule>

```
One methanol
    6
    0MET     H1    4  -0.111  -0.052   0.089
    0MET     O1    2   0.071   0.006  -0.001
    0MET     C1    3  -0.072  -0.001  -0.000
    0MET     H2    5  -0.109   0.102   0.000
    0MET     H3    6  -0.111  -0.052  -0.089
    0MET     H4    1   0.104  -0.085   0.001
   0.00000   0.00000   0.00000
```

<Structure of an 2-propanol molecule>

```
One IPA
   12
    0ISO     H1    3  -0.131  -0.176   0.028
    0ISO     C1    2  -0.127  -0.073  -0.010
    0ISO     H2    4  -0.216  -0.019   0.024
    0ISO     H3    1  -0.131  -0.079  -0.120
```

```
    0ISO     C2    5   0.000  -0.002   0.036
    0ISO     H4    6   0.000   0.003   0.146
    0ISO     O1    7  -0.001   0.136  -0.005
    0ISO     H5   10   0.131  -0.078  -0.120
    0ISO     C3    9   0.128  -0.072  -0.010
    0ISO     H6   11   0.216  -0.018   0.024
    0ISO     H7   12   0.133  -0.175   0.029
    0ISO     H8    8  -0.001   0.136  -0.103
   0.00000   0.00000   0.00000
```

<Structure of an octane molecule>

```
One Octane
   26
    0OCT     H1    1   0.160   0.076  -0.826
    0OCT     C1    2   0.214   0.140  -0.893
    0OCT     H2    3   0.175   0.239  -0.888
    0OCT     H3    4   0.204   0.102  -0.993
    0OCT     C2    5   0.363   0.142  -0.854
    0OCT     H4    6   0.417   0.205  -0.921
    0OCT     H5    7   0.373   0.179  -0.754
    0OCT     C3    8   0.420  -0.001  -0.861
    0OCT     H6    9   0.366  -0.065  -0.794
    0OCT     H7   10   0.410  -0.038  -0.961
    0OCT     C4   11   0.569   0.001  -0.822
    0OCT     H8   12   0.579   0.039  -0.720
    0OCT     H9   13   0.624   0.066  -0.890
    0OCT     C5   14   0.626  -0.142  -0.829
    0OCT    H10   15   0.616  -0.180  -0.931
    0OCT    H11   16   0.572  -0.206  -0.761
    0OCT     C6   17   0.775  -0.140  -0.790
    0OCT    H12   18   0.785  -0.102  -0.688
    0OCT    H13   19   0.830  -0.075  -0.858
    0OCT     C7   20   0.832  -0.282  -0.797
    0OCT    H14   21   0.822  -0.320  -0.899
    0OCT    H15   22   0.777  -0.347  -0.729
    0OCT     C8   23   0.981  -0.280  -0.758
    0OCT    H16   24   1.021  -0.381  -0.763
    0OCT    H17   25   1.036  -0.216  -0.826
    0OCT    H18   26   0.991  -0.242  -0.656
   0.00000   0.00000   0.00000
```

<Structure of a pentane molecule>

```
One pentane
   17
    0PEN     H1    1   0.160   0.076  -0.826
    0PEN     C1    2   0.214   0.140  -0.893
    0PEN     H2    3   0.175   0.239  -0.888
    0PEN     H3    4   0.204   0.102  -0.993
```

```
    0PEN     C2    5   0.363   0.142  -0.854
    0PEN     H4    6   0.417   0.205  -0.921
    0PEN     H5    7   0.373   0.179  -0.754
    0PEN     C3    8   0.420  -0.001  -0.861
    0PEN     H6    9   0.366  -0.065  -0.794
    0PEN     H7   10   0.410  -0.038  -0.961
    0PEN     C4   11   0.569   0.001  -0.822
    0PEN     H8   12   0.579   0.039  -0.720
    0PEN     H9   13   0.624   0.066  -0.890
    0PEN     C5   14   0.626  -0.142  -0.829
    0PEN    H10   15   0.616  -0.180  -0.931
    0PEN    H11   16   0.572  -0.206  -0.761
    0PEN    H12   17   0.732  -0.140  -0.801
   0.00000   0.00000   0.00000
```

<Structure of a cyclohexane molecule>

```
One cyclohexane
   18
    0CYC     H1    1   0.208  -0.139  -0.014
    0CYC     C1    2   0.122  -0.082   0.023
    0CYC     H2    3   0.127  -0.085   0.133
    0CYC     C2    4   0.132   0.065  -0.023
    0CYC     H3    5   0.224   0.110   0.014
    0CYC     H4    6   0.138   0.068  -0.133
    0CYC     C3    7  -0.010  -0.146  -0.023
    0CYC     H5    8  -0.010  -0.153  -0.133
    0CYC     H6    9  -0.017  -0.249   0.014
    0CYC     C4   10  -0.132  -0.065   0.023
    0CYC     H7   11  -0.224  -0.110  -0.014
    0CYC     H8   12  -0.138  -0.068   0.133
    0CYC     C5   13  -0.122   0.082  -0.023
    0CYC     H9   14  -0.127   0.085  -0.133
    0CYC    H10   15  -0.208   0.139   0.014
    0CYC     C6   16   0.010   0.146   0.023
    0CYC    H11   17   0.017   0.249  -0.014
    0CYC    H12   18   0.010   0.153   0.133
   0.00000   0.00000   0.00000
```

<Structure of a toluene molecule>

```
One toluene
   15
    0TOL    H1    3  -0.287   0.000   0.102
    0TOL    C1    2  -0.247  -0.000   0.000
    0TOL    H2    4  -0.287   0.089  -0.050
    0TOL    H3    1  -0.287  -0.089  -0.050
    0TOL    C2    5  -0.095   0.000  -0.000
    0TOL    C3    8  -0.023  -0.120  -0.000
    0TOL    H4    9  -0.077  -0.215  -0.001
```

```
    0TOL    C4   10   0.116  -0.121   0.000
    0TOL    H5   11   0.170  -0.215   0.000
    0TOL    C5   12   0.187  -0.000  -0.000
    0TOL    H6   13   0.296  -0.000  -0.000
    0TOL    C6   14   0.116   0.121   0.000
    0TOL    H7   15   0.170   0.215   0.000
    0TOL    C7    6  -0.023   0.120  -0.000
    0TOL    H8    7  -0.077   0.215  -0.001
   0.00000   0.00000   0.00000
```

<Topology of a water molecule>

```
[ moleculetype ]
; molname  nrexcl
SOL        2

[ atoms ]
;   nr   type  resnr residue  atom   cgnr     charge       mass
     1  opls_116   1    SOL     OW      1      -0.8740
     2  opls_117   1    SOL    HW1      1       0.4370
     3  opls_117   1    SOL    HW2      1       0.4370

#ifndef FLEXIBLE
[ settles ]
; OW  funct doh   dhh
1     1     0.09584    0.15150

[ exclusions ]
1     2     3
2     1     3
3     1     2
#else
[ bonds ]
; i   j     funct length     force.c.
1     2     1     0.09584    345000     0.09584 345000
1     3     1     0.09584    345000     0.09584 345000

[ angles ]
; i   j     k     funct angle force.c.
2     1     3     1     104.45     383   104.45     383
#endif
```

<Topology of an 2-propanol molecule>

```
[ moleculetype ]
; Name            nrexcl
ISO             3

[ atoms ]
```

```
;   nr       type  resnr residue  atom   cgnr     charge
mass  typeB    chargeB      massB
         1  opls_140   1   ISO      H1    1   0.09630     1.008
         2  opls_135   1   ISO      C1    1   -0.4212     12.011
         3  opls_140   1   ISO      H2    1   0.09400     1.008
         4  opls_140   1   ISO      H3    1   0.08645     1.008
         5  opls_158   1   ISO      C2    2   0.62020     12.011
         6  opls_140   1   ISO      H4    2   -0.0069     1.008
         7  opls_154   1   ISO      O1    2   -0.7156     15.9994
         8  opls_155   1   ISO      H5    2   0.39120     1.008
         9  opls_135   1   ISO      C3    3   -0.4212     12.011
        10  opls_140   1   ISO      H6    3   0.09630     1.008
        11  opls_140   1   ISO      H7    3   0.09400     1.008
        12  opls_140   1   ISO      H8    3   0.08645     1.008

[ bonds ]
1 2 1   0.110  284512.0
2 3 1   0.110  284512.0
2 4 1   0.109  284512.0
2 5 1   0.153  224262.4
5 6 1   0.110  284512.0
5 7 1   0.144  267776.0
5 9 1   0.153  224262.4
7 8 1   0.097  462750.4
9 10 1   0.110  284512.0
9 11 1   0.109  284512.0
9 12 1   0.110  284512.0

[ angles ]
1 2 3 1  107.712  276.144
1 2 4 1  108.238  276.144
1 2 5 1  110.680  313.800
2 5 6 1  108.934  313.800
2 5 7 1  110.897  418.400
2 5 9 1  112.779  488.273
3 2 4 1  108.460  276.144
3 2 5 1  111.160  313.800
4 2 5 1  110.484  313.800
5 7 8 1  106.931  460.240
5 9 10 1  110.666  313.800
5 9 11 1  110.510  313.800
5 9 12 1  111.094  313.800
6 5 7 1  103.979  292.880
6 5 9 1  108.911  313.800
7 5 9 1  110.934  418.400
10 9 11 1  108.254  276.144
10 9 12 1  107.726  276.144
11 9 12 1  108.488  276.144
```

```
[ dihedrals ]
1 2 5 6 3
3 2 5 6 3
4 2 5 6 3
1 2 5 7 3
3 2 5 7 3
4 2 5 7 3
1 2 5 9 3
3 2 5 9 3
4 2 5 9 3
2 5 7 8 3
6 5 7 8 3
9 5 7 8 3
2 5 9 10 3
6 5 9 10 3
7 5 9 10 3
2 5 9 11 3
6 5 9 11 3
7 5 9 11 3
2 5 9 12 3
6 5 9 12 3
7 5 9 12 3

[ dihedrals ]

[ pairs ]
6 1 1
6 3 1
6 4 1
7 1 1
7 3 1
7 4 1
9 1 1
9 3 1
9 4 1
8 2 1
8 6 1
8 9 1
10 2 1
10 6 1
10 7 1
11 2 1
11 6 1
11 7 1
12 2 1
12 6 1
12 7 1
```

<Topology of a methanol molecule>

```
[ moleculetype ]
; Name            nrexcl
MET             3

[ atoms ]
;   nr       type  resnr residue  atom   cgnr     charge
mass  typeB    chargeB      massB
        1  opls_155   1   MET      H1     1   0.41800     1.008
        2  opls_154   1   MET      O1     1   -0.6830     15.9994
        3  opls_157   1   MET      C1     1   0.14500     12.011
        4  opls_156   1   MET      H2     1   0.04000     1.008
        5  opls_156   1   MET      H3     1   0.04000     1.008
        6  opls_156   1   MET      H4     1   0.04000     1.008

[ bonds ]
1 2 1   0.097  462750.4
2 3 1   0.143  267776.0
3 4 1   0.110  284512.0
3 5 1   0.109  284512.0
3 6 1   0.110  284512.0

[ angles ]
1 2 3 1  107.283  460.240
2 3 4 1  112.419  292.880
2 3 5 1  106.852  292.880
2 3 6 1  112.256  292.880
4 3 5 1  108.353  276.144
4 3 6 1  108.539  276.144
5 3 6 1  108.267  276.144

[ dihedrals ]
1 2 3 4 3
1 2 3 5 3
1 2 3 6 3

[ dihedrals ]

[ pairs ]
4 1 1
5 1 1
6 1 1
```

<Topology of an octane molecule>

```
[ moleculetype ]
; Name            nrexcl
OCT             3
```

```
[ atoms ]
;   nr       type  resnr residue  atom   cgnr     charge
mass  typeB    chargeB      massB
         1  opls_140   1   OCT      H1     1   0.18220     1.008
         2  opls_135   1   OCT      C1     1   -0.5800     12.011
         3  opls_140   1   OCT      H2     1   0.1884     1.008
         4  opls_140   1   OCT      H3     1   0.1910     1.008
         5  opls_136   1   OCT      C2     2   -0.3058     12.011
         6  opls_140   1   OCT      H4     2   0.1594     1.008
         7  opls_140   1   OCT      H5     2   0.1689     1.008
         8  opls_136   1   OCT      C3     3   -0.3522     12.011
         9  opls_140   1   OCT      H6     3   0.17540     1.008
        10  opls_140   1   OCT      H7     3   0.16060     1.008
        11  opls_136   1   OCT      C4     4   -0.3317     12.011
        12  opls_140   1   OCT      H8     4   0.17220     1.008
        13  opls_140   1   OCT      H9     4   0.17240     1.008
        14  opls_136   1   OCT      C5     5   -0.3311     12.011
        15  opls_140   1   OCT      H10    5   0.17190     1.008
        16  opls_140   1   OCT      H11    5   0.17230     1.008
        17  opls_136   1   OCT      C6     6   -0.3523     12.011
        18  opls_140   1   OCT      H12    6   0.17530     1.008
        19  opls_140   1   OCT      H13    6   0.16070     1.008
        20  opls_136   1   OCT      C7     7   -0.3062     12.011
        21  opls_140   1   OCT      H14    7   0.15970     1.008
        22  opls_140   1   OCT      H15    7   0.16900     1.008
        23  opls_135   1   OCT      C8     8   -0.58     12.011
        24  opls_140   1   OCT      H16    8   0.18820     1.008
        25  opls_140   1   OCT      H17    8   0.18240     1.008
        26  opls_140   1   OCT      H18    8   0.19120     1.008

[ bonds ]
1 2 1   0.110  284512.0
2 3 1   0.110  284512.0
2 4 1   0.110  284512.0
2 5 1   0.154  224262.4
5 6 1   0.110  284512.0
5 7 1   0.110  284512.0
5 8 1   0.155  224262.4
8 9 1   0.110  284512.0
8 10 1   0.110  284512.0
8 11 1   0.154  224262.4
11 12 1   0.110  284512.0
11 13 1   0.110  284512.0
11 14 1   0.154  224262.4
14 15 1   0.110  284512.0
14 16 1   0.110  284512.0
14 17 1   0.154  224262.4
```

```
17 18 1   0.110  284512.0
17 19 1   0.110  284512.0
17 20 1   0.155  224262.4
20 21 1   0.110  284512.0
20 22 1   0.110  284512.0
20 23 1   0.154  224262.4
23 24 1   0.110  284512.0
23 25 1   0.110  284512.0
23 26 1   0.110  284512.0

[ angles ]
1 2 3 1  107.597  276.144
1 2 4 1  107.648  276.144
1 2 5 1  111.003  313.800
2 5 6 1  107.801  313.800
2 5 7 1  109.736  313.800
2 5 8 1  115.202  488.273
3 2 4 1  107.421  276.144
3 2 5 1  110.709  313.800
4 2 5 1  112.259  313.800
5 8 9 1  109.263  313.800
5 8 10 1  108.565  313.800
5 8 11 1  116.361  488.273
6 5 7 1  105.671  276.144
6 5 8 1  108.615  313.800
7 5 8 1  109.371  313.800
8 11 12 1  107.545  313.800
8 11 13 1  108.912  313.800
8 11 14 1  115.764  488.273
9 8 10 1  105.694  276.144
9 8 11 1  109.378  313.800
10 8 11 1  107.036  313.800
11 14 15 1  109.168  313.800
11 14 16 1  108.977  313.800
11 14 17 1  115.774  488.273
12 11 13 1  106.121  276.144
12 11 14 1  109.140  313.800
13 11 14 1  108.929  313.800
14 17 18 1  109.425  313.800
14 17 19 1  107.127  313.800
14 17 20 1  116.379  488.273
15 14 16 1  106.107  276.144
15 14 17 1  107.523  313.800
16 14 17 1  108.861  313.800
17 20 21 1  108.650  313.800
17 20 22 1  109.357  313.800
17 20 23 1  115.190  488.273
18 17 19 1  105.616  276.144
```

```
18 17 20 1  109.181  313.800
19 17 20 1  108.563  313.800
20 23 24 1  110.701  313.800
20 23 25 1  111.045  313.800
20 23 26 1  112.256  313.800
21 20 22 1  105.677  276.144
21 20 23 1  107.834  313.800
22 20 23 1  109.692  313.800
24 23 25 1  107.618  276.144
24 23 26 1  107.364  276.144
25 23 26 1  107.650  276.144

[ dihedrals ]
1 2 5 6 3
3 2 5 6 3
4 2 5 6 3
1 2 5 7 3
3 2 5 7 3
4 2 5 7 3
1 2 5 8 3
3 2 5 8 3
4 2 5 8 3
2 5 8 9 3
6 5 8 9 3
7 5 8 9 3
2 5 8 10 3
6 5 8 10 3
7 5 8 10 3
2 5 8 11 3
6 5 8 11 3
7 5 8 11 3
5 8 11 12 3
9 8 11 12 3
10 8 11 12 3
5 8 11 13 3
9 8 11 13 3
10 8 11 13 3
5 8 11 14 3
9 8 11 14 3
10 8 11 14 3
8 11 14 15 3
12 11 14 15 3
13 11 14 15 3
8 11 14 16 3
12 11 14 16 3
13 11 14 16 3
8 11 14 17 3
12 11 14 17 3
```

```
13 11 14 17 3
11 14 17 18 3
15 14 17 18 3
16 14 17 18 3
11 14 17 19 3
15 14 17 19 3
16 14 17 19 3
11 14 17 20 3
15 14 17 20 3
16 14 17 20 3
14 17 20 21 3
18 17 20 21 3
19 17 20 21 3
14 17 20 22 3
18 17 20 22 3
19 17 20 22 3
14 17 20 23 3
18 17 20 23 3
19 17 20 23 3
17 20 23 24 3
21 20 23 24 3
22 20 23 24 3
17 20 23 25 3
21 20 23 25 3
22 20 23 25 3
17 20 23 26 3
21 20 23 26 3
22 20 23 26 3

[ dihedrals ]

[ pairs ]
6 1 1
6 3 1
6 4 1
7 1 1
7 3 1
7 4 1
8 1 1
8 3 1
8 4 1
9 2 1
9 6 1
9 7 1
10 2 1
10 6 1
10 7 1
11 2 1
```

```
11 6 1
11 7 1
12 5 1
12 9 1
12 10 1
13 5 1
13 9 1
13 10 1
14 5 1
14 9 1
14 10 1
15 8 1
15 12 1
15 13 1
16 8 1
16 12 1
16 13 1
17 8 1
17 12 1
17 13 1
18 11 1
18 15 1
18 16 1
19 11 1
19 15 1
19 16 1
20 11 1
20 15 1
20 16 1
21 14 1
21 18 1
21 19 1
22 14 1
22 18 1
22 19 1
23 14 1
23 18 1
23 19 1
24 17 1
24 21 1
24 22 1
25 17 1
25 21 1
25 22 1
26 17 1
26 21 1
26 22 1
```

<Topology of a pentane molecule>

```
[ moleculetype ]
; Name            nrexcl
PEN             3

[ atoms ]
;   nr       type  resnr residue  atom   cgnr     charge
mass  typeB    chargeB      massB
         1  opls_140   1   PEN      H1     1   0.19030     1.008
         2  opls_135   1   PEN      C1     1   -0.5779     12.011
         3  opls_140   1   PEN      H2     1   0.18940     1.008
         4  opls_140   1   PEN      H3     1   0.18250     1.008
         5  opls_136   1   PEN      C2     2   -0.3137     12.011
         6  opls_140   1   PEN      H4     2   0.16050     1.008
         7  opls_140   1   PEN      H5     2   0.16580     1.008
         8  opls_136   1   PEN      C3     3   -0.3468     12.011
         9  opls_140   1   PEN      H6     3   0.16730     1.008
        10  opls_140   1   PEN      H7     3   0.17510     1.008
        11  opls_136   1   PEN      C4     4   -0.3026     12.011
        12  opls_140   1   PEN      H8     4   0.16960     1.008
        13  opls_140   1   PEN      H9     4   0.16610     1.008
        14  opls_135   1   PEN      C5     5   -0.5871     12.011
        15  opls_140   1   PEN      H10    5   0.18910     1.008
        16  opls_140   1   PEN      H11    5   0.18390     1.008
        17  opls_140   1   PEN      H12    5   0.18850     1.008

[ bonds ]
1 2 1   0.110  284512.0
2 3 1   0.110  284512.0
2 4 1   0.110  284512.0
2 5 1   0.153  224262.4
5 6 1   0.110  284512.0
5 7 1   0.110  284512.0
5 8 1   0.154  224262.4
8 9 1   0.110  284512.0
8 10 1   0.110  284512.0
8 11 1   0.154  224262.4
11 12 1   0.110  284512.0
11 13 1   0.110  284512.0
11 14 1   0.153  224262.4
14 15 1   0.110  284512.0
14 16 1   0.110  284512.0
14 17 1   0.110  284512.0

[ angles ]
1 2 3 1  107.620  276.144
1 2 4 1  107.430  276.144
```

```
1  2  5  1   111.959   313.800
2  5  6  1   108.786   313.800
2  5  7  1   109.519   313.800
2  5  8  1   114.302   488.273
3  2  4  1   107.660   276.144
3  2  5  1   110.947   313.800
4  2  5  1   111.028   313.800
5  8  9  1   108.915   313.800
5  8  10 1   109.029   313.800
5  8  11 1   114.725   488.273
6  5  7  1   106.042   276.144
6  5  8  1   108.733   313.800
7  5  8  1   109.131   313.800
8  11 12 1   109.907   313.800
8  11 13 1   108.965   313.800
8  11 14 1   112.987   488.273
9  8  10 1   106.004   276.144
9  8  11 1   108.543   313.800
10 8  11 1   109.267   313.800
11 14 15 1   111.045   313.800
11 14 16 1   111.503   313.800
11 14 17 1   111.269   313.800
12 11 13 1   106.073   276.144
12 11 14 1   109.158   313.800
13 11 14 1   109.516   313.800
15 14 16 1   107.661   276.144
15 14 17 1   107.479   276.144
16 14 17 1   107.695   276.144

[ dihedrals ]
1  2  5  6  3
3  2  5  6  3
4  2  5  6  3
1  2  5  7  3
3  2  5  7  3
4  2  5  7  3
1  2  5  8  3
3  2  5  8  3
4  2  5  8  3
2  5  8  9  3
6  5  8  9  3
7  5  8  9  3
2  5  8  10 3
6  5  8  10 3
7  5  8  10 3
2  5  8  11 3
6  5  8  11 3
7  5  8  11 3
```

```
5 8 11 12 3
9 8 11 12 3
10 8 11 12 3
5 8 11 13 3
9 8 11 13 3
10 8 11 13 3
5 8 11 14 3
9 8 11 14 3
10 8 11 14 3
8 11 14 15 3
12 11 14 15 3
13 11 14 15 3
8 11 14 16 3
12 11 14 16 3
13 11 14 16 3
8 11 14 17 3
12 11 14 17 3
13 11 14 17 3

[ dihedrals ]

[ pairs ]
6 1 1
6 3 1
6 4 1
7 1 1
7 3 1
7 4 1
8 1 1
8 3 1
8 4 1
9 2 1
9 6 1
9 7 1
10 2 1
10 6 1
10 7 1
11 2 1
11 6 1
11 7 1
12 5 1
12 9 1
12 10 1
13 5 1
13 9 1
13 10 1
14 5 1
14 9 1
```

```
14 10 1
15 8 1
15 12 1
15 13 1
16 8 1
16 12 1
16 13 1
17 8 1
17 12 1
17 13 1
```

<Topology of a cyclohexane molecule>

```
[ moleculetype ]
; Name            nrexcl
CYC             3

[ atoms ]
;   nr       type  resnr residue  atom   cgnr     charge
mass  typeB    chargeB      massB
        1  opls_140   1   CYC      H1    1   0.17500     1.008
        2  opls_136   1   CYC      C1    1   -0.3600     12.011
        3  opls_140   1   CYC      H2    1   0.18500     1.008
        4  opls_136   1   CYC      C2    2   -0.3600     12.011
        5  opls_140   1   CYC      H3    2   0.17500     1.008
        6  opls_140   1   CYC      H4    2   0.18500     1.008
        7  opls_136   1   CYC      C3    3   -0.3600     12.011
        8  opls_140   1   CYC      H5    3   0.18500     1.008
        9  opls_140   1   CYC      H6    3   0.17500     1.008
       10  opls_136   1   CYC      C4    4   -0.3600     12.011
       11  opls_140   1   CYC      H7    4   0.17500     1.008
       12  opls_140   1   CYC      H8    4   0.18500     1.008
       13  opls_136   1   CYC      C5    5   -0.3600     12.011
       14  opls_140   1   CYC      H9    5   0.18500     1.008
       15  opls_140   1   CYC      H10   5   0.17500     1.008
       16  opls_136   1   CYC      C6    6   -0.3600     12.011
       17  opls_140   1   CYC      H11   6   0.17500     1.008
       18  opls_140   1   CYC      H12   6   0.18500     1.008

[ bonds ]
1 2 1   0.110  284512.0
2 3 1   0.110  284512.0
2 4 1   0.154  224262.4
2 7 1   0.154  224262.4
4 5 1   0.110  284512.0
4 6 1   0.110  284512.0
4 16 1   0.154  224262.4
7 8 1   0.110  284512.0
7 9 1   0.110  284512.0
```

```
7 10 1   0.154  224262.4
10 11 1   0.110  284512.0
10 12 1   0.110  284512.0
10 13 1   0.154  224262.4
13 14 1   0.110  284512.0
13 15 1   0.110  284512.0
13 16 1   0.154  224262.4
16 17 1   0.110  284512.0
16 18 1   0.110  284512.0

[ angles ]
1 2 3 1  106.518  276.144
1 2 4 1  110.283  313.800
1 2 7 1  110.240  313.800
2 4 5 1  110.241  313.800
2 4 6 1  109.093  313.800
2 4 16 1  111.457  488.273
2 7 8 1  109.083  313.800
2 7 9 1  110.304  313.800
2 7 10 1  111.461  488.273
3 2 4 1  109.112  313.800
3 2 7 1  109.101  313.800
4 2 7 1  111.449  488.273
4 16 13 1  111.461  488.273
4 16 17 1  110.215  313.800
4 16 18 1  109.120  313.800
5 4 6 1  106.535  276.144
5 4 16 1  110.286  313.800
6 4 16 1  109.091  313.800
7 10 11 1  110.286  313.800
7 10 12 1  109.091  313.800
7 10 13 1  111.457  488.273
8 7 9 1  106.520  276.144
8 7 10 1  109.120  313.800
9 7 10 1  110.215  313.800
10 13 14 1  109.112  313.800
10 13 15 1  110.283  313.800
10 13 16 1  111.449  488.273
11 10 12 1  106.535  276.144
11 10 13 1  110.241  313.800
12 10 13 1  109.093  313.800
13 16 17 1  110.304  313.800
13 16 18 1  109.083  313.800
14 13 15 1  106.518  276.144
14 13 16 1  109.101  313.800
15 13 16 1  110.240  313.800
17 16 18 1  106.520  276.144
```

```
[ dihedrals ]
1  2  4  5  3
3  2  4  5  3
7  2  4  5  3
1  2  4  6  3
3  2  4  6  3
7  2  4  6  3
1  2  4  16  3
3  2  4  16  3
7  2  4  16  3
1  2  7  8  3
3  2  7  8  3
4  2  7  8  3
1  2  7  9  3
3  2  7  9  3
4  2  7  9  3
1  2  7  10  3
3  2  7  10  3
4  2  7  10  3
2  4  16  13  3
5  4  16  13  3
6  4  16  13  3
2  4  16  17  3
5  4  16  17  3
6  4  16  17  3
2  4  16  18  3
5  4  16  18  3
6  4  16  18  3
2  7  10  11  3
8  7  10  11  3
9  7  10  11  3
2  7  10  12  3
8  7  10  12  3
9  7  10  12  3
2  7  10  13  3
8  7  10  13  3
9  7  10  13  3
7  10  13  14  3
11  10  13  14  3
12  10  13  14  3
7  10  13  15  3
11  10  13  15  3
12  10  13  15  3
7  10  13  16  3
11  10  13  16  3
12  10  13  16  3
10  13  16  4  3
14  13  16  4  3
```

```
15 13 16 4 3
10 13 16 17 3
14 13 16 17 3
15 13 16 17 3
10 13 16 18 3
14 13 16 18 3
15 13 16 18 3

[ dihedrals ]

[ pairs ]
5 1 1
5 3 1
5 7 1
6 1 1
6 3 1
6 7 1
16 1 1
16 3 1
16 7 1
8 1 1
8 3 1
8 4 1
9 1 1
9 3 1
9 4 1
10 1 1
10 3 1
10 4 1
13 2 1
13 5 1
13 6 1
17 2 1
17 5 1
17 6 1
18 2 1
18 5 1
18 6 1
11 2 1
11 8 1
11 9 1
12 2 1
12 8 1
12 9 1
13 2 1
13 8 1
13 9 1
14 7 1
```

```
14 11 1
14 12 1
15 7 1
15 11 1
15 12 1
16 7 1
16 11 1
16 12 1
4 10 1
4 14 1
4 15 1
17 10 1
17 14 1
17 15 1
18 10 1
18 14 1
18 15 1
```

<Topology of a toluene molecule>

```
[ moleculetype ]
; Name            nrexcl
TOL             3

[ atoms ]
;   nr       type  resnr residue  atom   cgnr     charge
mass  typeB    chargeB      massB
        1  opls_140   1   TOL      H1     1   0.02100     1.008
        2  opls_148   1   TOL      C1     1   -0.0770     12.011
        3  opls_140   1   TOL      H2     1   0.02600     1.008
        4  opls_140   1   TOL      H3     1   0.02600     1.008
        5  opls_145   1   TOL      C2     2   0.20000     12.011
        6  opls_145   1   TOL      C3     3   -0.2310     12.011
        7  opls_146   1   TOL      H4     3   0.11800     1.008
        8  opls_145   1   TOL      C4     4   -0.2310     12.011
        9  opls_146   1   TOL      H5     4   0.11800     1.008
       10  opls_145   1   TOL      C5     5   -0.0610     12.011
       11  opls_146   1   TOL      H6     5   0.08900     1.008
       12  opls_145   1   TOL      C6     6   -0.1090     12.011
       13  opls_146   1   TOL      H7     6   0.08300     1.008
       14  opls_145   1   TOL      C7     7   -0.0610     12.011
       15  opls_146   1   TOL      H8     7   0.08900     1.008

[ bonds ]
1 2 1   0.110  284512.0
2 3 1   0.110  284512.0
2 4 1   0.109  284512.0
2 5 1   0.151  265265.6
5 6 1   0.140  392459.2
```

```
5 8 1   0.140  392459.2
6 7 1   0.109  307105.6
6 14 1   0.140  392459.2
8 9 1   0.109  307105.6
8 10 1   0.140  392459.2
10 11 1   0.109  307105.6
10 12 1   0.140  392459.2
12 13 1   0.109  307105.6
12 14 1   0.140  392459.2
14 15 1   0.109  307105.6

[ angles ]
1 2 3 1  107.194  276.144
1 2 4 1  108.087  276.144
1 2 5 1  111.381  292.880
2 5 6 1  120.936  585.760
2 5 8 1  120.936  585.760
3 2 4 1  107.162  276.144
3 2 5 1  111.413  292.880
4 2 5 1  111.389  292.880
5 6 7 1  119.318  292.880
5 6 14 1  121.059  527.184
5 8 9 1  119.318  292.880
5 8 10 1  121.059  527.184
6 5 8 1  118.128  527.184
6 14 12 1  120.186  527.184
6 14 15 1  119.704  292.880
7 6 14 1  119.622  292.880
8 10 11 1  119.704  292.880
8 10 12 1  120.186  527.184
9 8 10 1  119.622  292.880
10 12 13 1  120.310  292.880
10 12 14 1  119.380  527.184
11 10 12 1  120.108  292.880
12 14 15 1  120.108  292.880
13 12 14 1  120.310  292.880

[ dihedrals ]
1 2 5 6 3
3 2 5 6 3
4 2 5 6 3
1 2 5 8 3
3 2 5 8 3
4 2 5 8 3
2 5 6 7 3
8 5 6 7 3
2 5 6 14 3
8 5 6 14 3
```

```
2 5 8 9 3
6 5 8 9 3
2 5 8 10 3
6 5 8 10 3
5 6 14 12 3
7 6 14 12 3
5 6 14 15 3
7 6 14 15 3
5 8 10 11 3
9 8 10 11 3
5 8 10 12 3
9 8 10 12 3
8 10 12 13 3
11 10 12 13 3
8 10 12 14 3
11 10 12 14 3
10 12 14 6 3
13 12 14 6 3
10 12 14 15 3
13 12 14 15 3

[ dihedrals ]
2 6 5 8 1
5 7 6 14 1
5 9 8 10 1
8 11 10 12 1
10 13 12 14 1
6 12 14 15 1

[ pairs ]
6 1 1
6 3 1
6 4 1
8 1 1
8 3 1
8 4 1
7 2 1
7 8 1
14 2 1
14 8 1
9 2 1
9 6 1
10 2 1
10 6 1
12 5 1
12 7 1
15 5 1
15 7 1
```

```
11 5 1
11 9 1
12 5 1
12 9 1
13 8 1
13 11 1
14 8 1
14 11 1
6 10 1
6 13 1
15 10 1
15 13 1
```

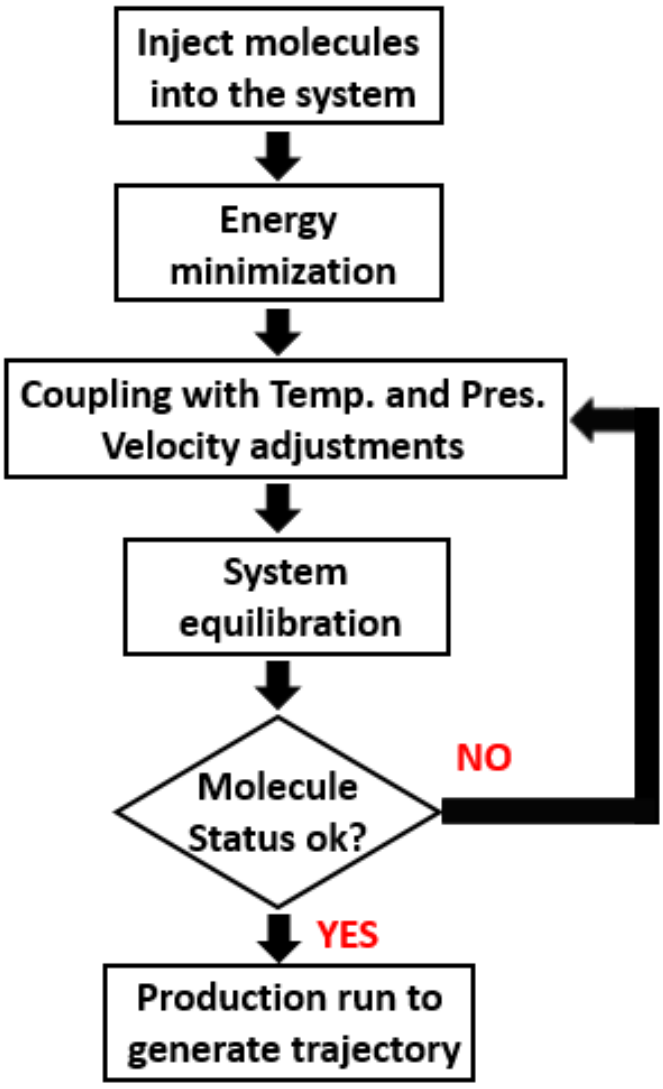


Figure B.1: A flowchart describing steps of MD simulations. Molecules are injected into the system and are equilibrated for the minimization of kinetic and potential energies. Using the desired ensembles, the system is coupled with the temperature and pressure and the atoms start to alter their velocities according to the temperature of each time step. Production run can produce trajectory of each atom after checking all molecules and atoms are properly synced with the environmental status and yield the correct physical properties.

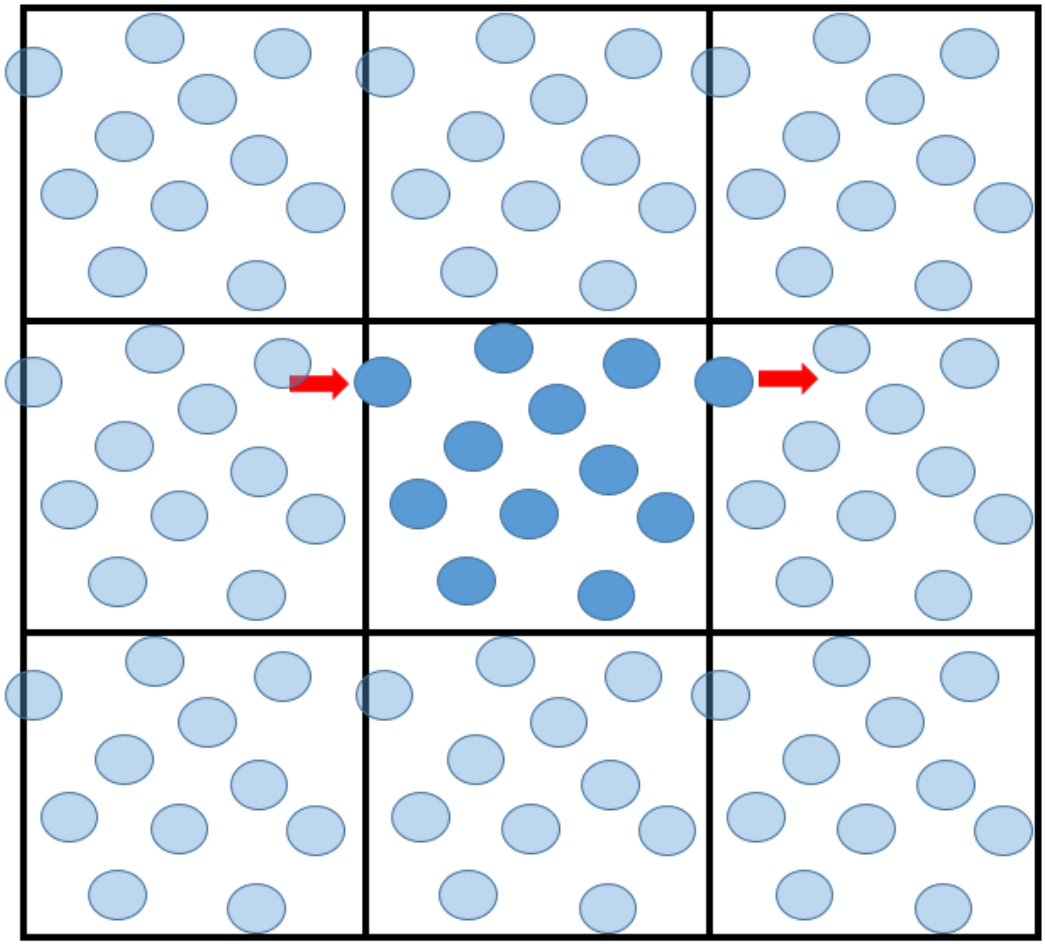


Figure B.2: A graphical description of periodic boundary conditions in two dimensions. The MD simulation box located in the center is replicated in all directions surrounding it. All the features of atoms are identical in each of the boxes. Once an atom is moving outside of the box, it will replicate in all surrounding boxes and thus it enters to the opposite side. This maintains the same number of atoms in the system.

APPENDIX C: LIST OF ACRONYMS AND LIST OF SYMBOLS

C.1 List of Acronyms

ACF	*Autocorrelation function*
ANN	*Artificial neural network*
API	*American Petroleum Institute*
ATR	*Attenuated Total Reflectance*
BTEX	*Benzene, Toluene, Ethylbenzene and Xylene*
CD	*Cole-Davidson*
CC	*Cole-Cole*
CHARMM	*Chemistry at HARvard Macromolecular Mechanics*
CMG	*Computer Modeling Group*
DFA	*Downhole Fluid Analyzer*
DI	*Deionized*
DOI	*Depth of Investigation*
DRRS	*Downhole Reservoir Raman System*
EMSC	*Extended Multiplicative Signal Correction*
EVD	*Eigenvalue Decomposition*
FFT	*Fast Fourier Transform*
FTIR	*Fourier Transform Infrared*
GAFF	*General Amber Force Field*
GATR	*Grazing angle Attenuated Total Reflectance*
GOR	*Gas-Oil Ratio*
GR	*Gamma-Ray*
GROMACS	*GROningen MAchine for Chemical Simulations*
HI	*Hydrogen Index*
HN	*Havriliak-Negami*
JST	*Jossi-Stiel-Thodos*
MCT	*Mercury Cadmium Telluride*

MD	*Molecular Dynamics*
MDT	*Modular Formation Dynamics Tester*
MLP	M*ultiLayer Perceptron*
MSC	*Multiplicative Scatter Correction*
MSD	*Mean-square displacement*
NIR	*Near-Infrared*
NIST	*National Institutes of Standards and Technology*
NMR	*Nuclear Magnetic Resonance*
NPT	*Number-Pressure-Temperature*
NVE	*Number-Volume-Energy*
NVT	*Number-Volume-Temperature*
NWRE	*Nadaraya-Watson regression estimator*
OBM	*Oil-Based Mud*
OPLS/AA	*Optimized Potential for Liquid Simulations-All Atoms*
PBC	*Periodic Boundary Condition*
PC	*Principal Component*
PCA	*Principal component analysis*
PEF	*Photoelectric Factor*
PLSR	*Partial least square regression*
PME	*Particle Mesh Ewald*
PREOS	*Peng-Robinson's equation-of-state*
PVT	P*ressure, volume, and temperature*
RBF	*Radial basis function*
RMSE	*Root-Mean-Square Error*
RNA	*Ribonucleic Acid*
R^2	*Correlation coefficient*
SARA	*Saturated, Aromatics, Resins, and Asphaltenes*
SP	*Spontaneous Potential*
SPC/E	*Extended Simple Point Charge*
SNUPAR	*Schlumberger Nuclear Parameters*

SNV	*Standard Normal Variate*
SVD	*Singular Value Decomposition*
TIP3P	*Transferable Intermolecular Potential with 3 Points*
UTAPWeLS	*The University of Texas at Austin Petrophysical and Well-Log Simulator*
WBM	*Water-Based Mud*
WFT	*Wireline Formation Testers*

C.2 List of Symbols

a_j	Activation of a *j*th neuron
α	Asymmetric exponent of the dielectric relaxation spectra [ ]
β	Broadness exponent of the dielectric relaxation spectra [ ]
C_p	Heat capacity [J/K]
C_{sh}	Volumetric shale concentration [ ]
C(x)	Cost function
D	Mass diffusion coefficient [μm^2/sec]
d_m	Vector of measured data
G	Autocorrelation function [1/sec^2]
G	Magnetic gradient [Gauss/cm]
g	Sigmoid activation function
J	Dipolar spectral density function [1/sec]
H	Enthalpy [J]
HI	Hydrogen index [ ]
J	Jacobian matrix
k_B	Boltzmann constant [J/K]
L	Total number of layers [ ]
L_m	Migration length [cm]
M	Dipole moment [Debye]
m	Total number of input data [ ]
N	Number of ^{1}H-^{1}H dipole-dipole interactions [ ]

n_i Number of neurons in input layer [ ]

n_h Number of neurons in hidden layer [ ]

n_o Number of neurons in output layer [ ]

r Separation distance between two hydrogen atoms [nm]

r Radius of spherical pores [μm]

r Distance a molecule travels in time t [nm]

S Entropy [J/K]

S_w Water saturation [v/v]

T Absolute temperature [K]

T_{2bulk} NMR T_2 bulk transverse relaxation time [ms]

T_{1bulk} NMR T_1 bulk longitudinal relaxation time [ms]

TE Inter-echo time in CPMG sequence [ms]

T_1 Longitudinal relaxation time [ms]

T_2 Transverse relaxation time [ms]

T_{2LM} Logarithmic mean transverse relaxation time [ms]

$T_{2cutoff}$ Transverse relaxation time cutoff [ms]

t Lag time of autocorrelation [ps]

U Volumetric cross section [cross section/cm^3]

V Volume [m^3]

W_d Data weighting matrix

ω Angular frequency [Hz]

ω_0 Larmor frequency [Hz]

x_0 Bias unit [ ]

x_1 Volumetric concentration of methane [ ]

x_2 Volumetric concentration of light alkane [ ]

x_3 Volumetric concentration of medium alkane [ ]

y_l Actual data of a lth component [ ]

θ Angle between two hydrogen atoms to the magnetic field [degree]

Σ Macroscopic neutron capture cross-section [c.u.]

ρ_1 Surface relaxivity for T_1 [μm/sec]

ρ_2 Surface relaxivity for T_2 [μm/sec]

ρ Density [g/cm^3] or [kg/m^3]

ρ_b Bulk density [g/cm^3]

ρ_f Fluid density [g/cm^3]

ϕ Porosity [v/v]

ϕ_D Density porosity [v/v]

ϕ_N Neutron porosity [v/v]

ξ Inverse of migration length [1/cm]

ε_0 Dielectric permittivity at zero frequency [ ]

ε_∞ Dielectric permittivity at infinity frequency [ ]

$\hbar$ Reduced Planck constant [J·sec]

h_Θ Estimated output at the output layer [ ]

Δ Accumulated sigmoid function gradient [ ]

δ Difference between estimated output and the actual data [ ]

γ Gyromagnetic ratio of hydrogen atom [MHz/T]

τ_D Dielectric relaxation time [ps]

$\Theta_{j,k}$ Weight from a *k*th unit of previous layer to a *j*th unit of current layer [ ]

τ Time [ps]

μ Viscosity [cp]

μ_0 Vacuum magnetic permeability [T·m/A]

λ Regularization parameter [ ]

References

Abraham, M. J., Van der Spoel D., Lindahl, E., Hess, B., and the GROMACS Development Team, 2017, GROMACS User Manual version 2016.3, www.gromacs.org.

Abragam, A., 1961, The Principles of Nuclear Magnetism, Oxford University Press, London.

Alder, B.J., and Wainwright, T.E., 1957, Phase Transition for a Hard Sphere System, Journal of Chemical Physics, vol. 27, pp. 1208-1209

Allen, M.P., and Tildesley, D.J., 1987, Computer Simulation of Liquids; Oxford Science Publications, New York.

Al-Yaarubi, A., Al-Bulushi, N. I., and Al Mjeni, R., 2012, Downhole Heavy Oil Characterization Using 3D NMR and Modern Dielectric Logs. SPE Heavy Oil Conference, Calgary, Alberta, Canada, June 12-14.

Anand, V., Freedman, R., Crary, S., Minh, C.C., and Terry, R.L., 2011, Predicting Effective Permeability to Oil in Sandstone and Carbonate Reservoirs from Well-Logging Data, SPE Reservoir Evaluation & Engineering, vol. 14, issue 6, pp. 750-762.

Aske, N., Kallevik, H., Sjoblom, J., 2001, Determination of Saturate, Aromatic, Resin, and Asphaltenic (SARA) Components in Crude Oils by Means of Infrared and Near-Infrared Spectroscopy, Energy & Fuels vol. 15, No. 5, pp. 1304-1312.

Aster, R.C., Borchers, B., and Thurber, C.H., 2013, Parameter Estimation and Inverse Problems, Second Edition: Academic Press, Burlington, Massachusetts.

Barnes R.J., Dhanoa M.S., Lister S.J., 1989, Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. Applied Spectroscopy, vol. 43, issue 5, pp. 772-777.

Bell, I.H., Wronski, J., Quoilin, S., and Lemort, V., 2014, Pure and Pseudo-pure Fluid Thermophysical Property Evaluation and the Open-Source Thermophysical Property Library CoolProp, Industrial & Engineering Chemistry Research, vol. 53, no. 6, pp. 2498-2508.

Benjamin, L., Benson, G.C., 1963, A Deuterium Isotope Effect on the Excess Enthalpy of Methanol-Water Solutions, The Journal of physical chemistry, vol. 67, No. 4, pp. 858-861.

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., and Haak, J.R., 1984, Molecular Dynamics with Coupling to an External Bath, The Journal of Chemical Physics, vol. 81, issue 8, pp. 3684-3690.

Berendsen, H.J.C., Grigera, J.R., and Straatsma, T.P., 1987, The Mixing Term in Effective Pair Potentials, The Journal of physical chemistry, vol. 91, no. 24, pp. 6269-6271.

Best, R.B., Zhu. X., Shim, J., Lopes, P.E.M., Mittal, J, Feig, M., MacKerell, A.D. Jr., 2012, Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-chain $\chi 1$ and $\chi 2$ Dihedral Angles, Journal of Chemical Theory and Computation, vol. 8, No. 9, pp. 3257-3273.

Bryan, J., Kantzas, A., and Bellehumeur, C., 2005, Oil-Viscosity Predictions from Low-Field NMR Measurements: SPE Journal vol. 8, no. 1, pp. 44-52.

Bryndzia, L.T., Braunsdorf. N.R., Hofmann, R., Morgan, Q., Christian, P., Toews, A., Parkins, J., Myers, G., and Pope, J., 2016, Down-Hole Reservoir Raman System: A Novel New Technology for the Rapid Appraisal of Shale Gas Resource Potential, Unconventional Resource Technology Conference, San Antonio, Texas, USA, August 1-3.

Bussi, G., Donadio, D., Parrinello, M., 2007, Canonical Sampling through Velocity Rescaling, The Journal of Chemical Physics, vol. 126, issue 1, pp. 014101:1-7

Caballero-Manrique, E., Bray, J.K., Deutschman, W.A., Dahlquist, F.W., and Guenza, M.G., 2007, A Theory of Protein Dynamics to Predict NMR Relaxation, Biophysical Journal, vol. 93, issue 12, pp. 4128-4140.

Caleman, C., van Maaren, P.J., Hong, M., Hub, J.S., Costa, L.T., and van der Spoel, D., 2012, Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant, Journal of Chemical Theory and Computation, vol. 8, no. 1, pp. 61-74.

Calero, C., Martí, J., and Guàrdia, E., 2015, ^{1}H Nuclear Spin Relaxation of Liquid Water from Molecular Dynamics Simulations, The Journal of Physical Chemistry, vol. 119, issue 5, pp. 1966-1973.

Caro, M.A., Laurila, T., and Lopez-Acevedo, O., 2016, Accurate Schemes for Calculation of Thermodynamic Properties of Liquid Mixtures from Molecular Dynamics Simulations, The Journal of chemical physics, vol 145, issue 24, pp. 24504: 1-11.

Carr, H. Y. and Purcell, E. M., 1954, Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments, Physical Review, vol. 94, no. 3, pp. 630-638.

Case, D.A., Cheatham, T.E. 3rd, Darden T., Gohlke, H., Luo, R., Merz, K.M. Jr., Onufriev, A., Simmerling, C., Wnag, B., and Woods, R.J., 2005, The Amber Biomolecular Simulation Programs. Journal of Computational Chemistry, vol. 26, pp. 1668-1688.

Chen, H., and Heidari, Z., 2016, Pore-Scale Joint Evaluation of Dielectric Permittivity and Electrical Resistivity for Assessment of Hydrocarbon Saturation Using Numerical Simulations, Society of Petroleum Engineers, vol. 21, issue, 6, pp. 1930-1942.

Chung, H., Ku, M-S, 2000, Comparison of Near-Infrared, Infrared, and Raman Spectroscopy for the Analysis of Heavy Petroleum Products, Applied Spectroscopy, vol. 54, no. 2, pp. 239-245.

Coates, G.R., Xiao, L., and Prammer, M.G., 1999, NMR Logging Principles and Applications, Halliburton Energy Services.

Cooper, J.B., Wise, K.L., Welch, W.T., Sumner, M.B., Wilt, B.K., and Bledsoe, R.R., 1997, Comparison of Near-IR, Raman, and Mid-IR Spectroscopies for the Determination of BTEX in Petroleum Fuels, Applied Spectroscopy, vol. 51, no. 11, pp. 1613-1620.

Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A., 1995, A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, Journal of the American Chemical Society, vol. 117, no. 19, pp. 5179-5197.

Cowan, B., 1997, Nuclear Magnetic Resonance and Relaxation, Cambridge University Press, Camgridge.

Cybenko G., 1989, Approximation by Superpositions of a Sigmoidal Function, Mathematics of Control, Signals and Systems, vol. 2, issue 4, pp 303-314.

Derlacki, Z.J., Easteal, A.J., Edge, A.V.J., Woolf, L.A., and Roksandic, Z., 1985, Diffusion Coefficients of Methanol and Water and the Mutual Diffusion Coefficient in Methanol-Water Solutions at 278 and 298 K, The Journal of Physical Chemistry, vol. 89, no. 24, pp. 5318-5322.

Dhanoa M.S., Lister S.J., Sanderson, R., Barnes R.J., 1994, The Link between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) Transformations of NIR spectra, Journal of Near Infrared Spectroscopy, vol. 2, issue 1, pp. 43-47.

Dzida, M., and Kaatze, U., 2015, Compressibility and Dielectric Relaxation of Mixtures of Water with Monohydroxy Alcohols, The Journal of Physical Chemistry, vol 119, no. 38, pp. 12480-12489.

Elliott, J.R. Jr., Suresh, S.J., and Donohue, M.D., 1990, A Simple Equation of State for Nonspherical and Associating Molecules, Industrial & Engineering Chemistry Research, vol. 29, issue 7, pp. 1476-1485.

Ellis, D.V., and Singer, J.M., 2007, Well Logging for Earth Scientists: Springer, Dordrecht, The Netherlands.

Essmann, U., Perera, L., and Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G., 1995, A Smooth Particle Mesh Ewald Method, The Journal of Chemical Physics, vol. 103, issue 19, pp. 8577-8593.

Ernst, R.R., Bodenhausen, G., and Wokaun, A., 2004, Principles of Nuclear Magnetic Resonance in One and Two Dimensions, Oxford University Press, New York.

Falla, F.S., Larini, C., Le Roux, G.A.C., Quina, F.H., Moro, L.F.L., Nascimento, C.A.O., 2006, Characterization of Crude Petroleum by NIR, Journal of Petroleum Science and Engineering, vol. 15, issues 1-2, pp. 127-137.

Filgueiras, P.R., Sad, C.M.S., Loureiro, A.R., Santos, M.F.P., Castro, E.V.R., Dias, J.C.M., Poppi, R.J., 2014, Determination of API Gravity, Kinematic Viscosity and Water Content in Petroleum by ATR-FTIR Spectroscopy and Multivariate Calibration, Fuel, vol. 116, pp. 123-130.

Foloppe, N., MacKerell, A.D., 2000, All-atom empirical force field for nucleic acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data, Journal of Computational Chemistry, vol. 21, pp. 86-104.

Freed, D.E., Burcaw, L., and Song, Y-Q., 2005, Scaling Laws for Diffusion Coefficients in /mixtures of Alkanes: Physical Review Letters, vol. 94, issue 6, pp. 067602.

Freedman, R., New Approach for Solving Inverse Problems Encountered in Well-Logging and Geophysical Applications, 2006, Petrophysics, vol. 47, no. 2, pp. 93-111.

Fuentes-Azcatl, R., and Alejandre, J., 2014, Non-Polarizable Force Field of Water Based on the Dielectric Constant: TIP4P/ε, The Journal of physical chemistry, vol. 118, No. 5, pp. 1263-1272.

Fujisawa, G., Betancourt, S.S., Mullins, O.C., Torgersen, T., O'Keefe, M., Terabayashi, T., Dong, C., and Eriksen, K.O., 2008, Hydrocarbon Compositional Gradient Revealed by In-Situ Optical Spectroscopy, Society of Petroleum Engineers, vol. 11, issue 2, pp. 223-237.

Galicia-Andrés, E., Dominguez, H., Pusztai, L., and Pizio, O., 2015, On the Composition Dependence of Thermodynamic, Dynamic and Dielectric Properties of Water-Methanol Model Mixtures. Molecular Dynamics Simulation Results, Condensed Matter Physics, vol. 18, no. 4, pp. 43602: 1-14.

Garcia, A. P., and Heidari, Z.,2018, Integrated Characterization of Multi-Frequency Dielectric Dispersion Measurements in Mixed-Wet Rocks, SPWLA 59th Annual Logging Symposium, London, UK, June 2-6.

Guevara-Carrion, G., Vrabec, J., and Hasse, H., 2011, Prediction of Self-Diffusion Coefficient and Shear Viscosity of Water and Its Binary Mixtures with Methanol

and Ethanol by Molecular Simulation, Journal of Chemical physics, vol. 134, issue 7, pp. 174508:1-14.

Hansen, J.P., McDonald, I.R., 1990, Theory of Simple Liquids, Academic Press, Second edition, London, UK.

Hansen, P.C., 1994, Regularization Tools: A Matlab Package for Analysis and Solution of Discrete Ill-posed Problems: Numerical Algorithms, vol. 6, no. 1, pp. 1-35.

Heaton, N., LaVigne, J., Bachman, H. N., Decoster, E., Badry, R., and Hemingway, J., 2012, Novel In Situ Characterization Of Heavy Oil Integrating NMR and Dielectric Logs, SPWLA 53rd Annual Logging Symposium, Cartagena, Colombia, June 16-20.

Hegeman, P.S., Dong, C., Varotsis, N., Gaganis, V., 2009, Application of Artificial Neural Networks to Downhole Fluid Analysis, SPE Reservoir Evaluation & Engineering vol. 12, issue 1, pp. 8-13.

Heidari, Z., 2011, Estimation of Static and Dynamic Petrophysical Properties from Well Logs in Multi-Layer Formations: Ph.D. Dissertation, The University of Texas at Austin.

Hess, B., Bekker, H., Berendsen, H.J.C., and Fraaije, J.G.E.M., 1997, LINCS: A Linear Constraint Solver for Molecular Simulations, Journal of Computational Chemistry, vol. 18, no. 12, pp. 1463-1472.

Hizem, M., Budan, H., Deville, B., Faivre, O., Mosse, L., and Simon, M., 2008, Dielectric Dispersion: A New Wireline Petrophysical Measurement. SPE Annual Technical Conference and Exhibition, Denver, Colorado, USA, September 21-24.

Hoffmann, M.M., Conradi, M., 1997, Are There Hydrogen Bonds in Supercritical Water?, Journal of American Chemical Society, vol. 119, No. 16, pp. 3811-3817.

Ijasan, O. and Torres-Verdín, C., 2013, Estimation of Porosity and Fluid Constituents from Neutron and Density Logs using an Interactive Matrix Scale: SPWLA 54th Annual Logging Symposium, New Orleans, Louisiana, June 22-26.

Indo, K., Hsu, K., Pop, J., 2015, Estimation of Fluid Composition From Downhole Optical Spectrometry, SPE Journal, vol. 20, issue 6, pp. 1326-1338.

Jia, G-Z, Huang, K-M, Yang L-J, and Yang X-Q, 2009, Composition-Dependent Dielectric Properties of DMF-Water Mixtures by Molecular Dynamics Simulations, International Journal of Molecular Sciences, vol 10, no. 4, pp. 1590-1600.

Jiang, J., Berry, R.J., Siesler, H.W., Ozaki, Y., 2002, Wavelength Interval Selection in Multicomponent Spectral Analysis by Moving Window Partial Least-Squares Regression with Applications to Mid-Infrared and Near-Infrared Spectroscopic Data, Analytical Chemistry, vol. 74, no. 14, pp. 3555-3565.

Jorgensen, W.L., Maxwell, D.S., Tirado-Rives, J., 1996, Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids, Journal of American chemical society, vol. 118, no. 45, pp. 11225-11236.

Jossi, J.A., Stiel, L.I., and Thodos, G., 1962, The Viscosity of Pure Substances in the Dense Gaseous and Liquid Phases, Journal of AICHE, vol. 8, No. 1, pp. 59-63.

Kalinichev, A.G., and Bass, J.D., 1997, Hydrogen Bonding in Supercritical Water. 2. Computer Simulations, The Journal of Physical Chemistry, vol. 101, No. 50, pp. 9720-9727.

Kaatze, U., Behrends, R., and Pottel, R., 2002, Hydrogen Network Fluctuations and Dielectric Spectrometry of Liquids, Journal of Non-Crystalline Solids, vol. 305, issues 1-3, pp. 19-28.

Killner, M.H.M., Rohwedder, J.J.R., Pasquini, C., 2011, A PLS Regression Model using NIR Spectroscopy for On-Line Monitoring of the Biodiesel Production Reaction, Fuel, vol. 90, issue 11, pp. 3268-3273.

Kobayashi, K., Liang, Y., Amano, K-I., Murata, S., Matsuoka, T., Takahashi, S., Nishi, N., and Sakka, T., 2016, Molecular Dynamics Simulation of Atomic Force Microscopy at the Water-muscovite Interface: Hydration Layer Structure and Force Analysis, Langmuir, vol. 32, issue 15, pp. 3608-3616.

Krynicki, K., 1996, Proton Spin-Lattice Relaxation in Pure Water between 0°C and 100°C, Physica, vol. 32, Issue 1, pp. 167-178.

Lama, R.F., Lu, B.C.-Y., 1965, Excess Thermodynamic Properties of Aqueous Alcohol Solutions, Journal of Chemical & Engineering Data, vol. 10, No. 3, pp. 216-219.

Lemmon, E.W., Huber, M.L., and McLinden, M.O., 2013, NIST Standard Reference Database 23: Reference Fluid Thermodynamic and Transport Properties-REFPROP, Version 9.1, National Institute of Standards and Technology, Standard Reference Data Program, Gaithersburg.

Lindahl, E., and Edholm, O., 2001, Molecular Dynamics Simulation of NMR Relaxation Rates and Slow Dynamics in Lipid Bilayers, The Journal of Chemical Physics, vol. 115, no. 10, pp. 4938-4950.

Lo, S-W., 1999, Correlations of NMR Relaxation Time with Viscosity/Temperature, Diffusion Coefficient and Gas/Oil Ratio of Methane-Hydrocarbon Mixtures: Ph.D. Dissertation, Rice University, Houston.

Luinge, H.J., Van der Maas, J.H., Visser, T., 1995, Partial Least Squares Regression as a Multivariate Tool for the Interpretation of Infrared Spectra, Chemometrics and Intelligent Laboratory Systems, vol. 28, issue 1, pp. 129-138.

Luzar, A., and Chandler, D., 1996, Effect of Environment on Hydrogen Bond Dynamics in Liquid Water, Physical Review Letters, vol. 76, No. 6, pp. 928-931.

Markelov, D.A., Falkovich, S.G., Neelov, I.M., Ilyash, M.Y., Matveev, V.V., Lahderanta, E., Ingman, P., and Darinskii, A.A., 2015, Molecular Dynamics Simulation of Spin-Lattice NMR Relaxation in Poly-L-Lysine Dendrimers: Manifestation of the Semiflexibility Effect, Physical Chemistry Chemical Physics, vol 17, issue 5, pp. 3214-3226.

Martens H, Nielsen J.P., Engelsen S.B., 2003, Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to Near-Infrared Transmission Analysis of Powder Mixtures, Analytical Chemistry, vol. 75, no. 3, pp. 394-404.

Martínez-Jiménez, M., and Saint-Martin, H., 2018, A Four-site Molecular Model for Simulations of Liquid Methanol and Water-Methanol Mixtures: MeOH-4P, Journal of Chemical Theory and Computation, vol. 14, No. 5, pp. 2526-2537.

McConnel, J., 1987, The Theory of Nuclear Magnetic Relaxation in Liquids, Cambridge University Press, Camgridge.

McKeon, D.C. and Scott, H.D., 1989, SNUPAR - A Nuclear Parameter Code for Nuclear Geophysics Applications: IEEE Transactions on Nuclear Science, vol. 36, issue 1, pp. 1215-1219.

McQuarrie, D.A., Statistical Mechanics, 1976, Harper and Row, New York

Medellin, D., Ramachandran, V.R., Torres-Verdín, C., Pore-Size Dependent Fluid Substitution Method for Magnetic Resonance Measurements, Geophysics, vol. 18, issue 1, pp. 1-79.

Meiboom, S., and Gill, D., 1958, Modified Spin Echo Method for Measuring Nuclear Relaxation Times, Review of Scientific Instruments, vol. 29, pp. 688-691.

Meksiarun, P., Ishigaki, M., Huck-Pezzei, V.A.C., Huck, C.W., Wongravee, K., Sato, H., Ozaki, Y., 2017, Comparison of Multivariate Analysis Methods for Extracting the Paraffin Component from the Paraffin-Embedded Cancer Tissue Spectra for Raman Imaging, Scientific Report, vol. 7, pp. 44890.

Mendoza, A., Torres-Verdín, C., and Preeg, W.E., 2007, Rapid Simulation of Borehole Nuclear Measurements based on Spatial Flux-Scattering Functions, SPWLA 48th Annual Logging Symposium, Austin, Texas, June 3-6.

Moore, P.B., 1980, Small-angle Scattering. Information Content and Error Analysis, Journal of Applied Crystallography, vol. 13, pp. 168-175.

Murdock, J.W., 1993, Fundamental Fluid Mechanics for the Practicing Engineer, CRC Press, pp. 25-27.

Mullins, O. C., Hashem, M., Elshahawi, H., Fujisawa, G., Dong, C., Betancourt, S., and Terabayashi, T., 2004, Hydrocarbon Compositional Analysis In-Situ in Openhole Wireline Logging: SPWLA 45th Annual Logging Symposium, Noordwijk, Netherlands, June 6-9.

Nadaraya, E.A., 1964, On Estimating Regression, Theory of Probability and its Applications, vol 9, issue 1, pp. 141-142.

Neumann, M., 1983, Dipole Moment Fluctuation Formulas in Computer Simulations of Polar Systems, Molecular Physics, vol. 50, issue 4, pp. 841-858.

Noda, K., Ohashi, M., and Ishida, K., 1982, Viscosities and Densities at 298.15 K for Mixtures of Methanol, Acetone, and Water, Journal of Chemical & Engineering Data, vol. 27, No. 3, pp. 326-328.

Ollila, O.H.S., Heikkinen, H.A., and Iwai, H., 2018, Rotational Dynamics of Proteins from Spin Relaxation Times and Molecular Dynamics Simulations, The Journal of Physical Chemistry, vol. 122, no. 25, pp. 6559-6569.

Ortega, E., Torres-Verdín, C., Preeg, W. E., and Miles, J., 2013, Multidetector LWD Sigma Logging Principles, Petrophysical Applications, and Environmental Effects: SPWLA 54th Annual Logging Symposium, New Orleans, Louisiana, June 22-26.

Pang, F-M., Seng, C-E, Teng, T-T, and Ibrahim, M.H., 2007, Density and Viscosities of Aqueous Solutions of 1-propanol and 2-propanol at Temperatures from 293.15 K to 333.15 K, Journal of Molecular Liquids, vol. 136, issues 1-2, pp. 71-78.

Pascal, T.A., and Goddard, W.A., 2012, Hydrophobic Segregation, Phase Transitions and the Anomalous Thermodynamics of Water/Methanol Mixtures, The Journal of Physical Chemistry, vol. 116, no. 47, pp. 13905-13912.

Pearson, K., 1901, On Lines and Planes of Closest Fit to Systems of Points in Space, Philosophical Magazine, vol. 2, issue 6, pp. 559-572.

Phillips, J.C., Braun R., Wang W., Gumbart J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L., and Schulten K., 2005 Scalable Molecular Dynamics with NAMD, Journal of Computational Chemistry, vol. 26, pp. 1781-1802.

Plimpton, S., 1995, Fast Parallel Algorithms for Short-range Molecular Dynamics, Journal of Computational Physics, vol. 117, pp. 1-19.

Pratt, K.C., and Wakeham, W.A., 1975, The Mutual Diffusion Coefficient for Binary Mixtures of Water and the Isomers of Propanol, Proceedings of the Royal Society of London. Series A, vol. 342, no. 1630, pp. 401-419.

Robinson, D., and Peng, D., 1978, The Characterization of the Heptanes and Heavier Fractions for the GPA Peng-Robinson Programs, Gas Processors Association Research Report, RR-28.

Rønne, C., Thrane, L., Astrand, P-O, Wallqvist, A., Mikkelsen, K.V., and Keiding, S.R., 1997, Investigation of the Temperature Dependence of Dielectric Relaxation in Liquid Water by THz Reflection Spectroscopy and Molecular Dynamics Simulation, The Journal of Chemical Physics, vol. 107, issue 14, pp. 5319-5331.

Rumelhart, D.E., McClelland, J.L., 1986, Parallel Distributed Processing: Explorations in the Microstructure of Cognition. vol. 1, MIT Press, Cambridge, MA, USA.

Salas, F.J., Méndez-Maldonado, A., Núñez-Rojas, E., Aguilar-Pineda, G.E., Dominguez, H., and Alejandre, J., 2014, Systematic Procedure To Parametrize Force Fields for Molecular Fluids, Journal of Chemical Theory and Computation, vol. 11, No. 2, pp. 683-693.

Santos, V.O., Oliveira, F.C.C., Lima, D.G., Petry, A.C., Garcia, E., Suarez, P.A.Z., Rubim, J.C., 2005, A Comparative Study of Diesel Analysis by FTIR, FTNIR and FT-Raman Spectroscopy using PLS and Artificial Neural Network Analysis, Analytica Chimica Acta, vol. 547, issue 2, pp. 188-196.

Sato, T., 1994, Application of Principal-Component Analysis on Near-Infrared Spectroscopic Data of Vegetable Oils for Their Classification, Journal of the American Oil Chemists' Society, vol. 71, issue 3, pp 293-298.

Sato, T., Chiba, A., and Nozaki, R., 1999, Hydrophobic Hydration and Molecular Association in Methanol-Water Mixtures Studied by Microwave Dielectric Analysis, The Journal of Chemical Physics, vol. 112, issue 6, pp. 2924-2932.

Sato, T., and Buchner, R., 2003, Dielectric relaxation spectroscopy of 2-propanol-water mixtures, The Journal of Chemical Physics, vol. 118, issue 10, pp. 4606-4613.

Sato, T., and Buchner, R., 2005, Cooperative and Molecular Dynamics of Alcohol/Water Mixtures: The View of Dielectric Spectroscopy, Journal of Molecular Liquids, vol. 117, issues 1-3, pp. 23-31.

Schuler, L.D., Daura, X., van Gunsteren, W.F., 2001, An Improved GROMOS96 Force Field for Aliphatic Hydrocarbons in the Condensed Phase, Journal of Computational Chemistry, vol. 22, pp. 1205-1218.

Schulz, H., Schrader, B., Quilizsch, R., and Steuer, B., 2002, Quantitative Analysis of Various Citrus Oils by ATR/FT-IR and NIR-FT Raman Spectroscopy, Applied Spectroscopy, vol. 56, no. 1, pp. 117-124.

Sciortino, F., Geiger, A., and Stanley, H.E., 1992, Network Defects and Molecular Mobility in Liquid Water, The Journal of Chemical Physics, vol. 96, issue 5, pp. 3857-3865.

Senthilkumar, P., Ganesh, T., Vinoth, K., Sylvester, M.M., Karunakaran, D.J.S.A., Hudge, P.G., and Kumbharkhane, A.C., 2018, Dielectric Dispersion, Relaxation and Molecular Interaction of Pyrazine Binary Mixtures, Journal of Physics Communications, vol. 2, pp. 035042:1-11.

Singer, P.M., Asthagiri, D., Chapman, W.G., and Hirasaki, G.J., 2017a, Molecular Dynamics Simulations of NMR Relaxation and Diffusion of Bulk Hydrocarbons and Water, The Journal of Magnetic Resonance, vol. 277, pp. 15-24.

Singer, P.M., Chen, Z., Alemany, L.B., Hirasaki, G.J., Zhu, K., Xie, Z.H., and Vo, T.D., 2017b, NMR Relaxation of Polymer – Alkane Mixes, A Model System for Crude Oils, Transactions of the Society of Petrophysicists and Well Log Analysts 58th Annual Logging Symposium, Oklahoma City, Oklahoma, June 17-21.

Shikhov, I., Arns, C.H., 2016, Temperature-Dependent Oxygen Effect on NMR D-T_2 Relaxation-Diffusion Correlation of n-Alkanes, Applied Magnetic Resonance, vol. 47, issue 12, pp. 1391-1408.

Singer, P.M., Asthagiri, D., Chen, Z., Parambathu, A.V., Hirasaki, G.J., and Chapman, W.G., 2018, Role of Internal Motions and Molecular Geometry on the NMR Relaxation of Hydrocarbons, The Journal of Chemical Physics, vol 148, issue 16, pp. 164507: 1-10.

Smith, W., Yong, C.W., and Rodger, P.M., 2010, DL_POLY: Application to Molecular Simulation, Molecular Simulation, vol. 28, issue 5, pp. 385-471.

Soper, A.K., 1997, Site-site Pair Correlation Functions of Water from 25 to 400 °C: Revised Analysis of New and Old Diffraction Data, The Journal of Chemical Physics, vol. 106, issue 1, pp. 247-254.

Smith, J.D., Cappa, C.D., Wilson, K.R., Messer, B.M., Cohen, R.C., and Saykally, R.J., 2004, Energetics of Hydrogen Bond Network Rearrangements in Liquid Water, Science, vol. 306, issue 5697, pp. 851-853.

Soave, G., 1972, Equilibrium Constants from a Modified Redlich-Kwong Equation of State, Chemical Engineering Science, vol. 27 issue 6, pp. 1197-1203.

Starling, K.E., 1973, Fluid Properties for Light Petroleum Systems, Gulf Publishing Company, pp. 270.

Steene, M.V., Ardila, M., Nelson, R., Fekry, A., and Farghaly, A., 2012, Fluid Identification in Light Hydrocarbons using NMR and Downhole Fluid Analyzers: SPE North Africa Technical Conference and Exhibition, Cairo, Egypt, February 14-17.

Straley, C., 2006, Reassessment of Correlations of Between Viscosity and NMR Measurements: SPWLA 47th Annual Logging Symposium, Veracruz, Mexico, June 4-7.

Su, J.T., Duncan, P.B., Momaya, A., Jutila, A., and Needham, D., 2010, The Effect of Hydrogen Bonding on the Diffusion of Water in n-Alkanes and n-Alcohols Measured with a Novel Single Microdroplet Method, The Journal of Chemical Physics, vol. 132, issue 4, pp. 044506:1-8.

Svozil, D., Kvasnicka, V., Pospichal, J., 1997, Introduction to Multi-Layer Feed-Forward Neural Networks, Chemometrics and Intelligent Laboratory Systems, vol. 39, issue 1, pp 43-62.

Swope, W.C., Andersen, H.C., Berens, P.H., Wilson, K.R., 1982, A Computer-simulation Method for the Calculation of Equilibrium-constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters, The Journal of Chemical Physics, vol. 76, issue 1, pp. 637-649.

Tandon, S., Heidari, Z., and Daigle, H., 2017, Pore-Scale Evaluation of Nuclear Magnetic Resonance Measurements in Organic-Rich Mudrocks Using Numerical Modeling, Unconventional Resources Technology Conference, Austin, Texas, USA, July 24-26.

Taylor, E.L., Bertrand. G.L., 1973, Thermochemical Investigations of Nearly Ideal Binary Solvents. I. Standard Heats and Volume Changes of Solution in Methanol-Isopropanol Mixtures at 25°C, The Journal of solution chemistry, vol. 3, issue 6, pp. 479-491.

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., Mackerell, A.D. Jr., 2010, CHARMM general force field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-atom Additive Biological Force Fields, Journal of computational chemistry, vol. 31, no. 4, pp. 671-690.

Villa, A., and Stock, G., 2006, What NMR Relaxation can Tell Us about the Internal Motion of an RNA Hairpin: A Molecular Dynamics Simulation Study, Journal of Chemical Theory and Computation, vol. 2, no. 5, pp. 1228-1236.

Voss, B., Torres-Verdín, C., Gandhi, A., Alabi, G., and Lemkecher, M., 2009, Common Stratigraphic Framework to Simulate Well Logs and to Cross-Validate Static and Dynamic Petrophysical Interpretation, SPWLA 50th Annual Logging Symposium, The Woodlands, Texas, June 21-24.

Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., and Case, D.A., 2004, Development and Testing of a General Amber Force Field, Journal of computational chemistry, vol. 25, no. 9, pp. 61-74.

Wei, I.C., and Rowley, R.L., 1984, Binary Liquid Mixture Viscosities and Densities, Journal of Chemical & Engineering Data, vol. 29, No. 3, pp. 332-335.

Wensink, E.J., Hoffmann, A.C., van Maaren, P.J., and van der Spoel, D., 2003, Dynamic Properties of Water/Alcohol Mixtures Studied by Computer Simulation, The Journal of Chemical Physics, vol. 119, issue 14, pp. 7308-7317.

Wold, S., Sjöström, M., Eriksson, L., 2001, PLS-Regression: A Basic Tool of Chemometrics, Chemometrics and Intelligent Laboratory Systems, vol. 58, issue 2, pp. 109-130.

X-5 Monte Carlo Team, 2005, MCNP - A general Monte Carlo N-Particle Transport Code: Overview and Theory, version 5, vol. 1: Los Alamos National Laboratory.

Zhang, B., and Daigle, H., 2017, Nuclear Magnetic Resonance Surface Relaxation Mechanisms of Kerogen, Geophysics, vol. 82, issue 6, pp. JM15-22

Vita

Hyungjoo Lee received a Bachelor's degree (2006) in electrical engineering and computer science from Kyungpook National University, South Korea, and a Master's degree (2013) in petroleum engineering from the University of Texas at Austin. He was an intern with Schlumberger Formation Evaluation group and BHP Billiton Petrophysics group during the summers of 2013 and 2014, respectively. From 2001 to 2004, he had been working at Korean Air Force, and from 2007 to 2010, he was an offshore facility design engineer with Daewoo Shipbuilding and Marine Engineering Co. Ltd. He is a member of SPE and SPWLA.

Permanent e-mail address: hyungjoo.lee@utexas.edu

This dissertation was typed by Hyungjoo Lee.