2013•2014
FACULTEIT INDUSTRIËLE INGENIEURSWETENSCHAPPEN
master in de industriële wetenschappen: chemie

Masterproef

Applicability evaluation of AspenONE and Dynochem for solubility modeling in the pharmaceutical industry

Promotor :
Prof. dr. ir. Leen BRAEKEN

Promotor :
dr. ir. GEERT VAN DER VORST

Gezamenlijke opleiding Universiteit Hasselt en KU Leuven

Arno Daubies

Proefschrift ingediend tot het behalen van de graad van master in de industriële wetenschappen: chemie

2013•2014

Faculteit Industriële ingenieurswetenschappen

master in de industriële wetenschappen: chemie

Masterproef

Applicability evaluation of AspenONE and Dynochem for solubility modeling in the pharmaceutical industry

Promotor :
Prof. dr. ir. Leen BRAEKEN

Promotor :
dr. ir. GEERT VAN DER VORST

Arno Daubies

Proefschrift ingediend tot het behalen van de graad van master in de industriële wetenschappen: chemie

universiteit hasselt | KU LEUVEN

Foreword

The final work of my education was writing this master's thesis. It was a very difficult task, which was impossible without the help and support of other people. Therefore, I want to show my appreciation and express my gratitude to some people.

First off all, I want to express my thanks to my external promoter dr. ir. Geert Van der Vorst and Elisabeth Schäfer for the support, advice and patience during my internship.

I also want to thank my internal promoter prof dr. ir. Leen Braeken for her support and advice during the internship and for the follow up of the progress of my master's thesis.

Also many thanks for Steven Rusch to give me the opportunity to do my thesis at Janssen Pharmaceutica.

Thanks to all the colleagues and other students for the pleasant work atmosphere at Janssen.

Finally, I want to thank my family for their support and encouragement.

Table of Contents

List of Tables

List of figures

Abstract

Janssen Pharmaceutica produces a large amount of active pharmaceutical ingredients at their plant in Geel. Solubility of these components plays an important role in reaction and in separation steps, as extraction and crystallization as it determines the efficiency and the purity. Solubilities depend on the used solvents meaning that often additional data are required if the process changes or for new processes. To obtain these solubility data samples are usually sent to a laboratory, which is a time consuming and cost increasing activity.

Therefore, predictions of solubility data by software packages AspenONE and Dynochem, which are based on thermodynamic models, are investigated. The goal of this thesis is to select the most appropriate models for prediction of gas-liquid, liquid-liquid and solid-liquid solubilities and to compare the obtained data with experimental data in order to determine the most suitable model for a specific system. Experimental data are collected from literature and from available data within Janssen.

In Aspen Plus the predictions for gas-liquid and liquid-liquid systems deliver good results with an average deviation of ±7% between experimental and simulated data. However, the required model parameters must be available in the desired temperature range. Solid-liquid predictions could not be produced with Aspen due to several errors but Dynochem provides here a good estimation with an average deviation of ±5%. The gas-liquid and liquid-liquid systems however are not extensively supported in Dynochem.

Abstract (Dutch)

Janssen Pharmaceutica produceert grote hoeveelheden actieve farmaceutische ingrediënten in hun productiefabriek in Geel. De oplosbaarheid van componenten speelt een belangrijke rol in de reactie stap en in de scheidingsstappen zoals extractie en kristallisatie omdat het de efficiëntie en zuiverheid bepaald. De oplosbaarheid is afhankelijk van het gebruikte solvent en meestal zijn extra gegevens vereist bij proces aanpassingen of voor nieuwe processen. Om deze oplosbaarheidsgegevens te krijgen zijn labo-analyses nodig die een tijdrovende en kostenverhogende activiteit zijn.

Daarom worden software programma's Dynochem en AspenONE, die gebaseerd zijn op thermodynamische modellen, onderzocht voor het maken van voorspellingen voor de oplosbaarheid. Het doel van deze thesis is het selecteren van de meest passende modellen voor gas-vloeistof, vloeistof-vloesitof en vast-vloeistof oplosbaarheden en de verkregen waardes te vergelijken met experimentele gegevens om zo het meeste bruikbare model voor een specifiek systeem te bepalen. Experimentele oplosbaarheidsgegevens zijn verzameld uit de literatuur en uit beschikbare gegevens binnen Janssen.

In Aspen Plus levert voor gas-vloeistof and vloeistof-vloeistof systemen goede resultaten met een gemiddelde verschil van ±7% tussen experimentele en gesimuleerde data. Maar de benodigde model parameters moeten beschikbaar zijn in het gewenste temperatuursgebied. Vast-vloeistof voorspellingen zijn niet gemaakt met Aspen omdat hier verschillende problemen waren maar Dynochem zorgt hier wel voor een goede schatting met een gemiddelde afwijking van ±5%. Maar gas-vloeistof en vloeistof-vloeistof systemen zijn niet uitgebreid ondersteund in Dynochem.

Introduction

1 Situating

As final year student master in industrial sciences in chemistry at the KULeuven Campus Diepenbeek, I had the opportunity to do my master thesis at Janssen Pharmaceutica. The main goal of this thesis consists in evaluating and comparing existing solubility models and determining the usability of those models for predicting solubilities of different phase systems.

Janssen Pharmaceutica was founded by Dr. Paul Janssen in 1953 with in the beginning a focus on pharmacological research. The objective of Dr. Paul Janssen was to improve the quality of life by developing better medicines.

Today, Janssen Pharmaceutica has establishments all over the world with five plants in Belgium: Geel, Olen, Merksem and two in Beerse. Besides producing medicines, they do also research to develop products for a wide range of disease areas. The main focus lies in pathological areas of neurosciences, oncology, infectious diseases and vaccines.

Different steps are required to go from molecule to medicine. First off all is the basic discovery. In this stage, the researchers try to find a molecule that has an effect on a defect, usually a defective protein in order to treat the disease. Next is pre-clinical development. In this stage the most promising molecules are further developed into potential drugs. These are subjected to further testing in test tubes, on cell cultures (in vitro), or on living test animals (in vivo). This is followed by clinical research. In this stage is tested if the potential drug is safe enough to test on people. Clinical studies are then started with healthy volunteers (phase I) and patients (phases II and III). The safety and effectiveness of the potential drug, and how it behaves in the body (intake, distribution, and excretion) are examined. The following step is the registration of the drug. Only registered drugs may be brought onto the market. Therefore, a file has to be submitted to the government containing all the necessary data relating to the basic, pre-clinical and clinical research. The finally stage is the production and commercialization of the drug. This production consists of the chemical process that produces the active ingredients of a drug and the pharmaceutical process that mixes the active compound with other ingredients which results in a final product. Another action that is taken during the last step is informing the medical profession about the product. The distribution of the medicine to clients starts as soon as approval is received. The safety of the drug is monitored continuously and accurately [1].

During the production of the API or active pharmaceutical ingredients, the solubility is an important parameter because solubility is the limiting factor for the amount of a component that can be present in a solvent. Therefore, it determines the cost and efficiency of different processes such as crystallization, separations and reactions. An example is the separation of unwanted impurities using a liquid-liquid extraction. In this process, it is required that the impurities are readily soluble in the liquid phase that will be disposed of, so that most of these impurities will be removed from the desired liquid phase. Therefore it is also important that the impurities are less or not soluble in the liquid phase that is used in a further process so that the contamination of this stream is limited. If these impurities are too high in the stream that is further used this could result in rejection of this stream and thus an increase of the costs. The determination of the solubility of different solutes in various solvents is thus an important step.

2 Problem statement

At Janssen Pharmaceutica, the solubilities of a solute in different solvents are usually determined by sending samples to an external laboratory for examination. Typically, this laboratory analyses the solute solubility in approximately 20 different solvents. This evaluation costs 1500 euro and takes about one week. Currently new solubility information will always need new analytical work, meaning that new samples are to be sent to the laboratory. For example if the solubility of a solute in an extra solvent is needed and this was not measured during the first time. Also, if the solubility changes due to a temperature change this also needs extra analytical work.

By using software to determine the solubility based on modelling, there is only a little amount of extra data needed. This approach can have the advantage that it is less time consuming because it uses a software package like Dynochem or Aspen. In addition, the data needed for those calculations can be measured in an earlier phase so that they are available when needed. The different software packages that are needed are already available at Janssen Pharmaceutica but not used regularly for solubility problems. Therefore, this approach requires only the cost of measuring the extra data points that are needed for the model. Possible disadvantages are that a certain amount of knowledge is required for using those software packages and the reduced reliability. The required knowledge can easily be solved by making a manual for the use of the software. The problem of reduced reliability will be evaluated during this thesis work. This forms the basis of this thesis.

3 Goal

The main goal consists in evaluating and comparing existing solubility models and determining the possible usability of those models for predicting solubilities.

The first step is to examine the different modelling tools that are available and the different thermodynamic models that can be used with these programs. The collected information will be used to write a literature study. This study is required to understand the theoretical aspects of the different thermodynamic models used in the different tools.

The second step is to compare the different thermodynamic models with each other and with experimental data to determine which one gives the best results and has the best accuracy for specific cases. The division that will be made here for the different cases is as followed. A first division is the use of different solutes namely gasses, liquids and solids. These are divided into two categories namely one liquid or a liquid mixture. One liquid can then be divided in water and other non-water liquids. The cases that will be used for the usability of the different tools will be compared with each other to decide which one is the easiest to use and delivers the best results.

The last step is to write a conclusion about how different solubility models perform and how usable they are for Janssen Pharmaceutica.

4 Materials and methods

To realize the main goal that was explained in the previous part, there are different utilities needed: the software packages and the experimental data needed for the software packages and for comparison.

The first requirement is the availability of the different programs. The two software packages that are used for the evaluation are Dynochem and Aspen Plus. Both are available at Janssen Pharmaceutica.

The second requirement is knowledge of the different software tools. That knowledge is needed for the usage of the software. That information was obtained by research about the two tools in literature and information given by persons from Janssen Pharmaceutica and other experts.

The last requirement is the experimental data that are used in de software and as comparison with the results from the software. The first comparisons are for cases that have a lot data that can be found in articles or databanks. The components in these cases do not have any specific meaning for Janssen but are only for comparison. Afterwards some cases for specific problems that are relevant for the company will be added. The consequence of the first part is that the data that were found in databanks or articles need to be critically evaluated and for the second part that there need to be experimental data available for the comparison.

Part 1: Literature study

1 Introduction

This literature study discusses the theoretical aspects of this thesis starting with a description and definition of the solubility of organic and inorganic components. The general equation for every possible phase equilibrium will be derived as the solubility can be described as an equilibrium between phases. Based on this the general equation for gas-liquid, liquid-liquid and solid-liquid phase equilibrium is derived in order to get a specific and usable equation to predict the solubility. The last part contains more information about different models that can be used to determine behavior of the different systems. The link with the models and the equations is that the models add or adjust a factor that accounts for the non-ideality of the system. There are mainly two types of models that are used to predict the behavior and those two are activity coefficient models and Equation of state models. The focus in this thesis is on the activity coefficient models.

2 Solubility

The solubility of a solute is the maximum amount of that solute that can be dissolved in a solvent. Therefore, it can be described as the equilibrium composition of a saturated solution. The solute can be a gas, a liquid or a solid and the solvent a pure component liquid or mixture of liquids. The solubility is a thermodynamic function depending on the components, temperature and pressure. Another aspect is the amount of solute that is added to a solvent if more than the maximum amount of solute that can dissolve is added in the solvent two different phases are formed. An example is benzene-water. The solubility of benzene in water is 1.8g/L water at 25°C and thus if more than 1.8g of benzene is added to 1L water there will be 2 liquid phases, an organic and an inorganic phase [2] [3].

Solubility can be approached qualitatively and Prof. Martinez made a classification for solutes in specific solvents at room temperature. The first class is the insoluble solutes, meaning that almost nothing of the solute dissolves in the solvent obtaining solute concentrations in the solvent less than 0.01mol/L. The second group consists of slightly soluble compounds with solute concentrations between 0.01mol/L and 0.1mol/L. The third class is the soluble compounds where larger amounts of the solute can be dissolved in the solvent with a solute concentration greater than 0.1mol/L. Finally, miscible solutes are defined when solute and solvent can be mixed in any proportion [3].

2.1 General

In general, solubility in chemical thermodynamics is described as a phase equilibrium. The phase equilibrium can have two phases, three phases or more. In a closed system, the pressure and temperature are assumed constant. The Gibbs free energy is then at a minimum and the differential of the Gibbs free energy for the two phases is equal to zero. The superscript I and II indicated the different phases.

$$dG = dG^{I} - dG^{II} = 0 \qquad 2.1.1$$

Now temperature and pressure are per definition constant.

$$dG = -SdT + Vdp + \sum_{i=1}^{c} \bar{G}_i dN_i \qquad 2.1.2$$

$$dG_{|T,p} = \sum_{i=1}^{c} \bar{G}_i dN_i \qquad 2.1.3$$

In the previous equation $\bar{G}_i$ describes the partial molar Gibbs energy of component i. The total quantity of moles is constant, so that $N_i = N_i^I + N_i^{II}$ or $dN_i^{II} = -dN_i^I$. The equation can then be written as follows.

$$dG = \sum_{i=1}^{c} \bar{G}_i^I dN_i^I + \sum_{i=1}^{c} \bar{G}_i^{II} dN_i^{II} = \sum_{i=1}^{c} (\bar{G}_i^I - \bar{G}_i^{II}) dN_i^I = 0 \qquad 2.1.4$$

The derivative of the Gibbs free energy is set equal to zero, because there is a minimum for the Gibbs free energy, with respect to the molar quantities N. The next equation displays this.

$$\left(\frac{\partial G}{\partial N_i^{II}}\right)_{N_{j\neq i}^I} = 0 = \bar{G}_i^I - \bar{G}_i^{II} \; or \; \bar{G}_i^I = \bar{G}_i^{II} \qquad 2.1.5$$

$\bar{G}_i^I$ And $\bar{G}_i^{II}$ are the partial molar Gibbs energies in phase one and in phase two. These have the same temperature and pressure. They also can be created for different compositions ($x^I \; and \; x^{II}$) as shown in the next equation.

$$\bar{G}_i^I(T, p, x^I) = \bar{G}_i^{II}(T, p, x^{II}) \qquad 2.1.6$$

The partial molar Gibbs energies are sometimes referred to as the chemical potential μ. So that the equation becomes as follows.

$$\mu_i^I(T,p,x^I) = \mu_i^{II}(T,p,x^{II}) \qquad 2.1.7$$

The next equation describes the relation between chemical potential μ and molar Gibbs energies. The γ symbol is the activity coefficient for component i.

$$G = \sum_i x_i \mu_{0i}(p,T) + RT \sum_i x_i \ln x_i + RT \sum_i x_i \ln \gamma_i \qquad 2.1.8$$

And then $\mu_i(T,p)$ is

$$\mu_i(T,p) = \mu_{0i}(T,p) + RT \ln x_i + RT \ln \gamma_i \qquad 2.1.9$$

In this equation, the subscript 0 is an indication that the term is a standard state. The previous equations 2.1.6 and 2.1.7 describe ideal mixtures and because there only exist a few ideal gas-liquid mixtures, the previous equation 2.1.2 is adjusted to be able to estimate the properties of real gas-liquid mixtures. This is done by adding a new concept the fugacity f. the fugacity is a measure of the tendency for a component of a liquid mixture to escape or evaporate from the mixture and has the units of pressure. The equation for fugacity can be derived as follows starting from equation 2.1.1.

$$dG = -SdT + Vdp + \sum_{i=1}^{c} \bar{G}_i dN_i \qquad 2.1.10$$

The second derivatives of the thermodynamic functions are also used.

$$\frac{\partial}{\partial N_i}\Big|_{T,p,N_{j\neq i}} \left(\frac{\partial G}{\partial T}\right)_{p,N_j} = \frac{\partial}{\partial T}\Big|_{p,N_j} \left(\frac{\partial G}{\partial N_i}\right)_{T,p,N_{j\neq i}} \qquad 2.1.11$$

$$\frac{\partial}{\partial N_i}\Big|_{T,p,N_{j\neq i}} \left(\frac{\partial G}{\partial p}\right)_{T,N_j} = \frac{\partial}{\partial p}\Big|_{T,N_j} \left(\frac{\partial G}{\partial N_i}\right)_{T,p,N_{j\neq i}} \qquad 2.1.12$$

Out of the previous equations, the next two equations are obtained.

$$\bar{S}_i = -\left(\frac{\partial \bar{G}_i}{\partial T}\right)_{p,N_j}$$

2.1.13

$$\bar{V}_i = \left(\frac{\partial \bar{G}_i}{\partial p}\right)_{T,N_j}$$

2.1.14

The second equation is more useful than the first and leads to the next relation.

$$\bar{G}_i(\mathrm{T}, p_2, x) - \bar{G}_i(\mathrm{T}, p_1, x) = \int_{p1}^{p2} \bar{V}_i dp$$

2.1.15

The mixture can also be considered as an ideal gas mixture so that $V^{IGM} = RT/p$. The superscript IGM stands for ideal gas mixture properties. The equation becomes then as follows.

$$\overline{\mathrm{G}}_{\mathrm{i}}^{\mathrm{IGM}}(\mathrm{T}, p_2, x) - \overline{\mathrm{G}}_{\mathrm{i}}^{\mathrm{IGM}}(\mathrm{T}, p_1, x) = \int_{p1}^{p2} \frac{\mathrm{RT}}{\mathrm{p}} dp$$

2.1.16

Combining equation 2.1.15 and 2.1.16 gives the next result.

$$\left[\bar{G}_i(\mathrm{T}, p_2, x) - \overline{\mathrm{G}}_{\mathrm{i}}^{\mathrm{IGM}}(\mathrm{T}, p_2, x)\right] - \left[\bar{G}_i(\mathrm{T}, p_1, x) - \overline{\mathrm{G}}_{\mathrm{i}}^{\mathrm{IGM}}(\mathrm{T}, p_1, x)\right] = \int_{p1}^{p2} \bar{V}_i - \frac{\mathrm{RT}}{\mathrm{p}} dp$$

2.1.17

p_1Is set equal to zero $p = 0$, as all fluids are seen as ideal gases and $\bar{G}_i(\mathrm{T}, \mathrm{p} = 0, x) = \overline{\mathrm{G}}_{\mathrm{i}}^{\mathrm{IGM}}(\mathrm{T}, \mathrm{p} = 0, x)$. The equation is than reduced.

$$\bar{G}_i(\mathrm{T}, \mathrm{p}, x) - \overline{\mathrm{G}}_{\mathrm{i}}^{\mathrm{IGM}}(\mathrm{T}, \mathrm{p}, x) = \int_{0}^{p} \mathrm{V} - \frac{\mathrm{RT}}{\mathrm{p}} dp$$

2.1.18

The fugacity can then be written as follows.

$$\bar{f}_i(T,p,x) = x_i p \exp\left(\frac{\bar{G}_i(T,p,x) - \bar{G}_i^{IGM}(T,p,x)}{RT}\right) \qquad 2.1.19$$

For very low pressures ($p \rightarrow 0$) the fugacity equals to $x_i p \equiv p_i$ with p_i the partial pressure of the substance i. This equation of the fugacity is then applied on the equation of the two-phase equilibrium (2.1.6).

$$\bar{G}_i(T,p,x) = \bar{G}_i^{IGM}(T,p,x) + RT \ln\left(\frac{\bar{f}_i(T,p,x)}{x_i p}\right) \qquad 2.1.20$$

$$\bar{G}_i(T,p,x^{II}) - \bar{G}_i(T,p,x^{I}) = \bar{G}_i^{IGM}(T,p,x^{II}) + RT \ln\left(\frac{\bar{f}_i(T,p,x^{II})}{{x_i}^{II} p}\right) - \bar{G}_i^{IGM}(T,p,x^{I}) + RT \ln\left(\frac{\bar{f}_i(T,p,x^{I})}{{x_i}^{I} p}\right) \qquad 2.1.21$$

Then the next relation is used where the ideal gas mixture can be written as ideal gas for the same temperature and pressure and a part for the composition.

$$\bar{G}_i^{IGM}(T,p,\mathrm{x}) = \overline{\mathrm{G}}_\mathrm{i}^{\mathrm{IG}}(\mathrm{T},\mathrm{p}) + \mathrm{RT}\ln \mathrm{x_i} \qquad 2.1.22$$

Using this last equation the equation 2.21 is reduced to the following equation.

$$\bar{G}_i(T,p,x^{II}) - \bar{G}_i(T,p,x^{I}) = RT \ln\left(\frac{\bar{f}_i(T,p,x^{II})}{\bar{f}_i(T,p,x^{I})}\right) \qquad 2.1.23$$

The criterion for equilibrium between two phases is that $\bar{G}_i^I = \bar{G}_i^{II}$ for every type i. Temperature and pressure needs to be constant and equal in both phases. The equation can then be written as follows.

$$\bar{f}_i^I = \bar{f}_i(T,p,x^I) = \bar{f}_i(T,p,x^{II}) = \bar{f}_i^{II} \qquad 2.1.24$$

The final equation is the basis for gas-liquid, liquid-liquid and solid-liquid equilibriums [4] [5].

2.2 Gas-liquid

This part will describe and explain the solubility of a gas in a liquid. As mentioned before the solubility of a solute in a solvent is a two-phase equilibrium. The next equation describes this equilibrium.

$$\bar{f}_i^L(T,p,x) = \bar{f}_i^V(T,p,y) \;\; i = 1,2,\ldots \qquad 2.2.1$$

In equation 2.2.1 the $\bar{f}_i^L$ and $\bar{f}_i^V$ term stands for the fugacity of component I in the liquid and vapor phase. These two fugacities depend on temperature, pressure and composition. The fugacity f_i depends only on temperature and pressure. The x and y are the mole fractions in the liquid and vapor phase. This equation is reduced by using the Lewis-Randall law for the gas phase and the definition of the activity coefficient γ.
The term $\left(\frac{f}{p}\right)_i$ is the fugacity correction for component i.

$$\bar{f}_i^L(T,p,x) = x_i f_i^L(T,p)\gamma_i(T,p,x) \qquad 2.2.2$$

$$\bar{f}_i^V(T,p,y) = y_i f_i^V(T,p) = y_i p\left(\frac{f}{p}\right)_i \qquad 2.2.3$$

$$x_i f_i^L(T,p)\gamma_i(T,p,x) = y_i p\left(\frac{f}{p}\right)_i \qquad 2.2.4$$

Gas that is dissolved in a liquid is in a super-critical state meaning that the gas cannot condense. If the gas is not in the super-critical state, the equilibrium can be seen as a classic liquid-vapor equilibrium. The fugacity needs to be calculated for a dissolved gas in the liquid state under a given pressure and temperature but this leads to problems because this state cannot exist under the given circumstances. The solution is to determine the fugacity for a hypothetical state. There are different methods to make these calculations. The first method is to use a temperature of the mixture is slightly greater than the critical temperature of the solute components ($T \cong T_C$). The critical temperature of a gas is the temperature where the gas is in the super-critical state. The fugacity can then be approximated by its vapor pressure. This calculation is done by extrapolating the vapor pressure curve. Figure 1 gives the vapor pressure curve with on the x-axis, T/T_C and on the y-axis f^L/p_C. T is the temperature, T_C is the critical temperature, f^L is the fugacity in liquid phase and p_C is the critical pressure.

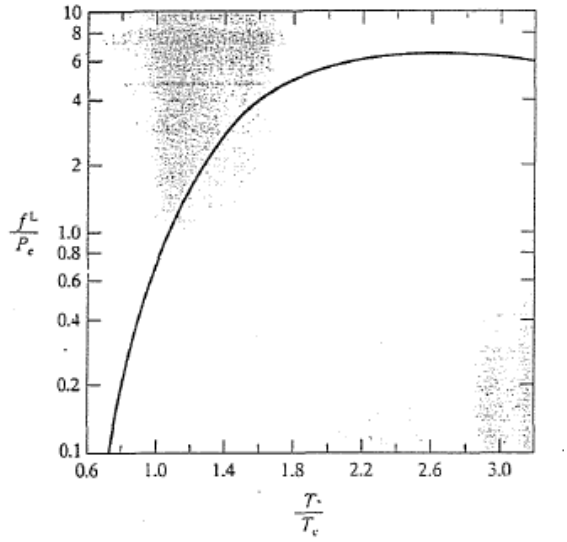

Figure 1: Vapor pressure curve [4].

The second method is to use a temperature of the mixture high above the critical temperature ($T >> T_C$). In this case, the principle of the corresponding-states correlation is used to evaluate the liquid-fugacity at 1atm total pressure. With the Poynting pressure correction the fugacity can be determined for different pressures.

$$f_i^L(T,p) = f_i^L(T,1.013bar)exp\left(\int_{1.013bar}^{p} \frac{v_i^L}{RT}dp\right)$$

$$= f_i^L(T,1.013bar)exp\left(\frac{v_i^L(p-1.013bar)}{RT}\right) \qquad 2.2.5$$

In this equation another unknown and hypothetical term is introduced being the molar volume of the liquid, v_i^L. In the next table (1) a few values that Prausnitz and Shair tabulated can be found.

Table 1: Tabulated values for v^L at 25°C [4].

Gas	v^L(cc/mol)
N_2	32.4
CO	32.1
O_2	33.0
Ar	57.1
CH_4	52
CO_2	55
Kr	65
C_2H_4	65
C_2H_6	70
Cl_2	74

Now for the determination of the gas solubility in a liquid solvent two equations (2.2.4 and 2.2.5) are combined together and this forms the next equation.

$$x_i = \frac{y_i p \left(\frac{f}{p}\right)_i}{\gamma_i(T,p,x)\, f_i^L(T, 1.013bar) exp\left(\frac{v_i^L(p - 1.013bar)}{RT}\right)} \qquad 2.2.6$$

A last method can be found by using Henry's law when the gas is slightly soluble in a liquid. The equation is adjusted as follows.

$$x_i H_i(T,p) = x_i f_i^L(T,p)\gamma_i(T,p,x) = y_i p \left(\frac{f}{p}\right)_i \qquad 2.2.7$$

$$x_i = \frac{y_i p \left(\frac{f}{p}\right)_i}{H_i(T,p)} = \frac{p_i}{H_i} \qquad 2.2.8$$

With the last equation the solubility of the gas can be determined, here $y_i$is considered as one and constant since the solubility of the gas in the liquid is very low and thus the mole fraction of the solute in the gas phase remains constant compared to the mole fraction of the solute in the liquid phase [4]. The association for gas-liquid systems with activity coefficient models is that these models try to calculate the y_i value and the retrieve H_i value from databanks to make a prediction for the solubility.

2.3 Liquid-liquid

This part discusses the solubility of liquids in liquids and liquid mixtures. In an equilibrium between binary and ternary liquid mixture there will exist different liquid phases with different compositions, rather than a single liquid phase, over a range of temperatures and compositions of the compounds. The reason is that at equilibrium state, the energy of the system is at a minimum and this can be reached if the given mixture splits up in more phases. The equation for a binary liquid is given below. The superscript L1 and L2 indicate the two liquid phases.

$$\bar{f}_i^{L1}(T,p,x^{L1}) = \bar{f}_i^{L2}(T,p,x^{L2}) \;\; i = 1,2,\ldots \qquad 2.3.1$$

The equation can also be written with activity coefficients as follows.

$$x_i{}^{L1} f_i(T,p)\gamma_i{}^{L1}(T,p,x^{L1}) = x_i{}^{L2} f_i(T,p)\gamma_i{}^{L2}(T,p,x^{L2}) \qquad 2.3.2$$

This is reduced into the following equation, because the pure component liquid fugacity on both sides of the previous equation is the same.

$$x_i{}^{L1}\gamma_i{}^{L1}(T,p,x^{L1}) = x_i{}^{L2}\gamma_i{}^{L2}(T,p,x^{L2}) \qquad 2.3.3$$

The compositions of coexisting phases are described as sets of mole fractions.

$$x_1^{L1}, x_2^{L1}, x_3^{L1}, \ldots, x_c^{L1}, x_1^{L2}, x_2^{L2}, x_3^{L2}, \ldots, x_c^{L2} \qquad 2.3.4$$

$$\sum_{i=1}^{c} x_i^{L1} = 1 \; and \; \sum_{i=1}^{c} x_i^{L2} = 1 \qquad 2.3.5$$

The compositions of both coexisting phases can be calculated using the previous equation (2.3.3) and appropriate solution models [4]. In this case, the activity coefficient models will try to approximate the γ_i value in both phases to give a prediction for the solubility.

2.4 Solid-liquid

This section describes the solubility of solids in liquids or liquid mixtures. As the solubility is described as an equilibrium of two-phases, the fugacity in both phases is equal. For solid-liquid systems, two simplifications can be made as the solid is considered as a pure component. The first assumption is that the only equilibrium criterion for the solute counts. The second assumption is that the solid phase fugacity is simplified to that of the pure solid. Next equation is the starting point for describing solid-liquid systems.

$$f_{Sol}^{S}(T,p) = \bar{f}_{Sol}^{L}(T,p,x) \qquad 2.4.1$$

The subscript 'Sol' indicates that it is for the solid solute and superscript 'S' and 'L' indicate the phases solid and liquid. The equation can also be written as follows.

$$f_{Sol}^{S}(T,p) = x_{Sol} f_{Sol}^{L}(T,p)\gamma_{Sol}(T,p,x) \qquad 2.4.2$$

$f_{Sol}^{S}(T,p)$ And $f_{Sol}^{L}(T,p)$ are the fugacity of the pure components for the solid and the liquid at a specified temperature and pressure of the mixture. The x_{Sol} in the equation describes the saturated mole fraction of the solid in the solvent. If the liquid mixture is at the normal melting temperature of the solid (T_m), the fugacity for the solid and liquid phase are equal. This is for pure-component phase equilibrium condition.

$$f_{Sol}^{S}(T_m) = f_{Sol}^{L}(T_m) \qquad 2.4.3$$

$$x_1 = \frac{1}{\gamma_{Sol}(T_m,p,x)} \qquad 2.4.4$$

The solubility for a solid in a liquid at its melting temperature is given by the previous equation and is thus determined by the activity coefficient γ_{Sol} in the solute-solvent mixture. But, usually the temperature is below the melting temperature of the solid thus the fugacity of the liquid phase is higher than the solid phase $f_{Sol}^{L} > f_{Sol}^{S}$ in this case the previous equation is not valid. In this case, the solubility can be described by using an estimation for the f_{Sol}^{L}/f_{Sol}^{S} ratio. A first proposal is using the sublimation pressure for f_{Sol}^{S} and then extrapolating the liquid thermodynamic properties into the solid phase to determine the fugacity for the liquid f_{Sol}^{L}. An alternative is to determine the fugacity ratio using equation 2.4.7 [4]. This estimation is more accurate than the previous one but it requires the parameters heat of fusion $\Delta_{fus}H(T)$ and melting point T_m. The starting point for equation 2.4.7 is given below and the values that are used in this equation are for one component, the solute.

$$\frac{\Delta_{fus}G(T,p)}{RT} = \frac{G^{L}(T,p) - G^{S}(T,p)}{RT} = \ln\frac{f^{L}(T,p)}{f^{S}(T,p)} \qquad 2.4.5$$

And.

$$\Delta_{fus}G(T,p) = \Delta_{fus}H(T) - T\Delta_{fus}S(T) \quad 2.4.6$$

$$f_{Sol}^{L}(T,p) = f_{Sol}^{S}(T,p)\exp\left[\frac{1}{RT}\left[\Delta_{fus}H(T)\left(1-\frac{T}{T_m}\right)+\int_{T_m}^{T}\Delta CpdT - T\int_{T_m}^{T}\frac{\Delta Cp}{T}dt\right]\right] \quad 2.4.7$$

Now equation 2.4.7 is used in equation 2.4.2 resulting into the following.

$$\ln x_{Sol}\gamma_{sol} = \frac{\Delta_{fus}H(T)}{\mathrm{RT}}\left(1-\frac{T}{T_m}\right)+\frac{1}{\mathrm{RT}}\int_{T_m}^{T}\Delta CpdT - \frac{1}{R}\int_{T_m}^{T}\frac{\Delta Cp}{T}dt \quad 2.4.8$$

This is the basic equation for determination of the mole fraction of a solid in a liquid at saturation. Equation 2.4.8 can be simplified by assuming that difference between heat capacities ΔCp is independent of temperature resulting in the next equation.

$$\ln x_{Sol}\gamma_{sol} = -\left[\frac{\Delta_{fus}H(T)}{\mathrm{RT}}\left(1-\frac{T}{T_m}\right)+\frac{\Delta\mathrm{Cp}}{\mathrm{R}}\left(1-\frac{\mathrm{T_m}}{\mathrm{T}}+\ln\left(\frac{\mathrm{T_m}}{\mathrm{T}}\right)\right)\right] \quad 2.4.9$$

If the heat capacity data for either the solid or liquid are not available, they are assumed to be approximately equal to each other giving a $\Delta\mathrm{Cp} = 0$. This results in the next equation.

$$\ln x_{Sol} = -ln\gamma_{sol} - \left[\frac{\Delta_{fus}H(T)}{\mathrm{RT}}\left(1-\frac{T}{T_m}\right)\right] \quad 2.4.10$$

The solubility of a solid in a liquid mixture can be determined by using this last equation. The activity coefficient γ_{sol} can be predicted for non-ideal solutions by using experimental data or a liquid solution model. If the mixture is ideal $\gamma_{sol}$is equal to one [4].

3 Thermodynamic models

Software packages can be used to determine the solubility of various solute-solvent systems. These software packages use different thermodynamic models to predict the behavior of the different systems. The results of those predictions need to be evaluated and compared with experimental data to determine the accuracy of the prediction. To choose the best model for a specific system the following decision tree can be followed, the starting point is figure 2 and the first question that is asked is the polarity of the components in the system. After completing the first question, the tree has to be followed to other questions until a model or a few models are suggested that are usable. In this thesis, the following route was taken: polar, non-electrolytes and below a pressure of 10bar. There are many different models created for different kind of problems. Therefore, always question which model should be used for a certain application. This decision tree is a rough guide that covers popular models that can be used but it does not include every model.

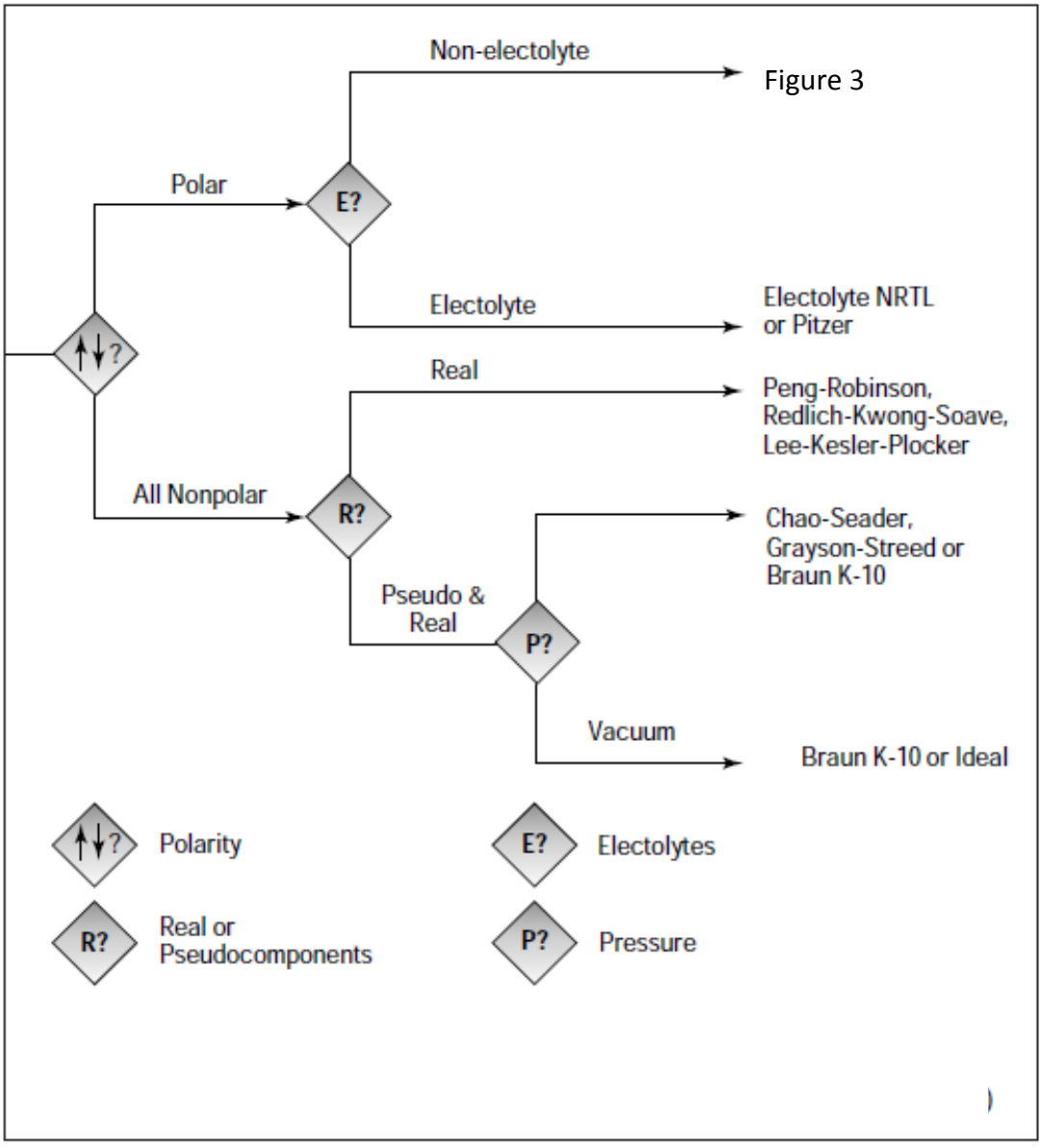

Figure 2: Decision tree (part 1) [6].

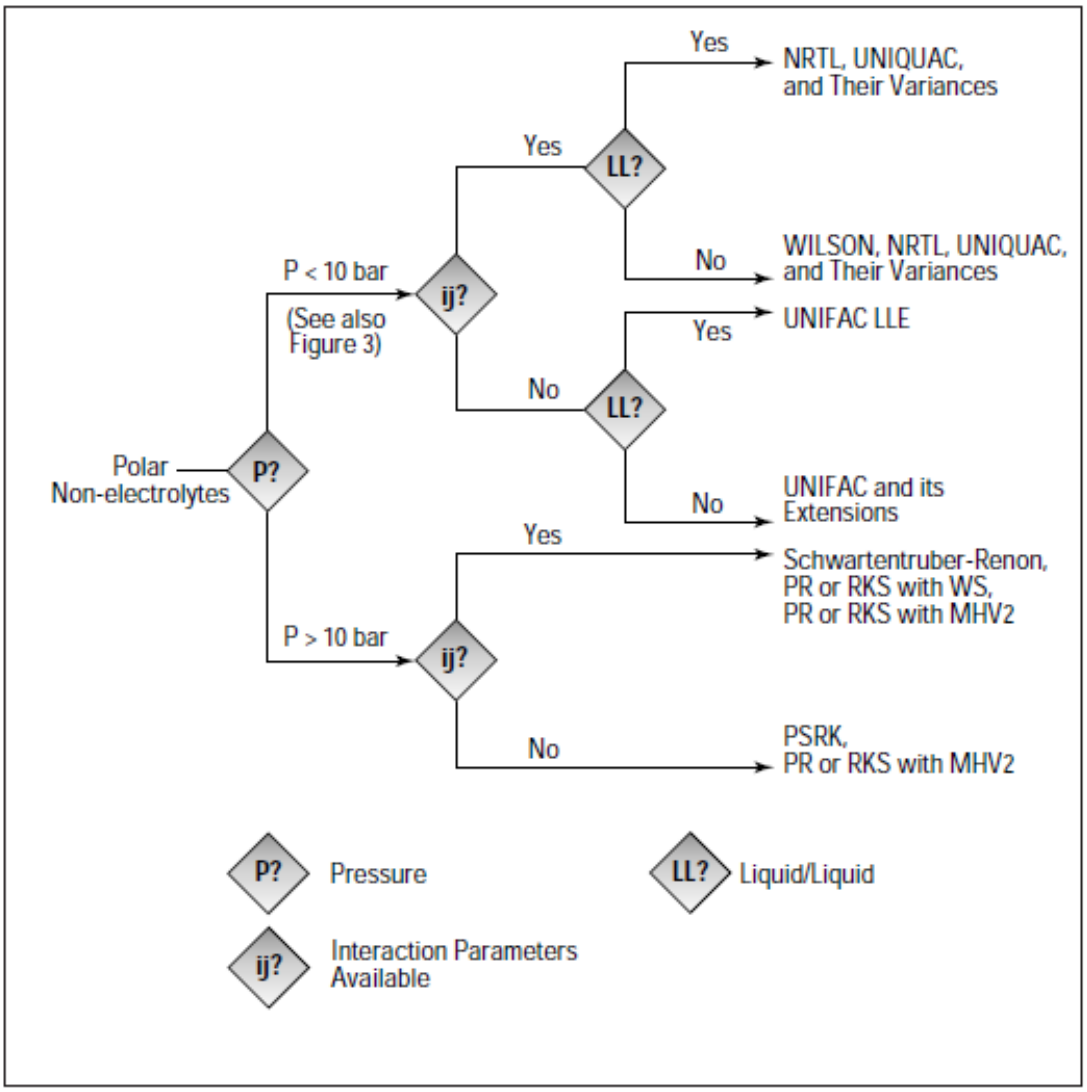

Figure 3: Decision tree (part 2) [6].

In the next part, the general principle of thermodynamic models is explained and different models will be discussed.

3.1 General

Thermodynamic models that describe the behavior of different systems rely on different relations. The main two commonly used techniques to model different systems are the EoS or Equation of state models and the activity coefficient models. In this work, only activity coefficient models such as UNIFAC, UNIQUAC and NRTL will be discussed. These models calculate the activity coefficient γ, which is then used in the equations for gas-liquid, liquid-liquid and solid-liquid systems to determine the solubility. Activity coefficient models as Margules and Van Laar equations use 'average' compositions and are thus based on 'random mixing'. The 'random' mixing rule assumes that mixing molecules is completely random and this means that the ratio of type one too type two molecules in the surrounding of any other molecule is on average the same as the ratio of their mole fractions. However, since mixing molecules is never entirely random due to intermolecular forces those models can be improved and give a better description of the phase behavior by adding a way to account for the non-randomness. The first equation that adds a factor for the non-randomness was the Wilson equation and is described below for two components 1 and 2 (3.1.1). Equations that account the non-randomness of the mixture are called local composition or LC activity models. By using these LC activity models the range of applicability for liquid phase models extended drastically. The LC concept is illustrated in the figure 4.

$$\frac{g^e}{RT} = -x_1 \ln(x_1 + x_2\Lambda_{12}) - x_2 \ln(x_2 + x_1\Lambda_{21}) \qquad 3.1.1$$

In this equation x_i is the mole fraction for component i, $\Lambda_{12}$and $\Lambda_{21}$are the two adjustable parameters and the relation between excess Gibbs energy g^e and the activity coefficient γ is given in the next equation. In this equation i is the component and thus for two components there are two activity coefficients.

$$g^e = RT \sum_{i=1}^{c} x_i \ln \gamma_i \qquad 3.1.2$$

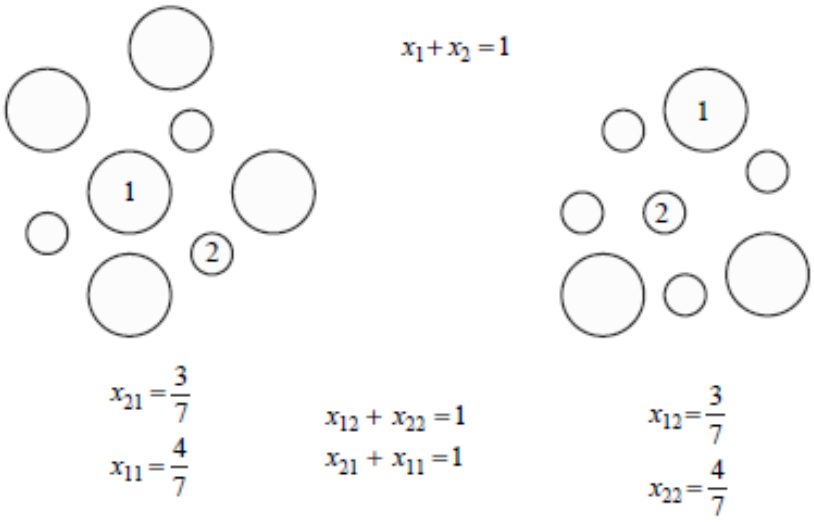

Figure 4: Local composition concept [5].

In this figure x_{ij} is the mole fraction of molecule i around the central molecule j.The total mole fraction of the mixture is one. For example x_{21}, number one is the central molecule and there are seven other molecules surrounding it. Three of the surrounding molecules are number two thus the mole fraction is 3/7.

Examples of these models, which are described further in this thesis, are NRTL, UNIQUAC and UNIFAC. UNIFAC is further developed from UNIQUAC using predictive group contribution and is suitable for an initiating design if no experimental data are available. These LC models have advantages and disadvantages. An advantage is that most of these models only use parameters between two components for binary and ternary systems. An example is the mixture of water, ethanol and cyclohexane. Only parameters between water-ethanol, water-cyclohexane and cyclohexane-ethanol are required. Another advantage is that these models are very usefull for the calculations at low pressure and multi-component vapor-liquid equilibriums and liquid-liquid equilibriums according to Kontogeorgis & Folas [5]. The disadvantage of these models is that they are not suitable for all equilibriums and therefore different variances of these models have been created. Two variances will be included in this work namely NRTL-SAC and Regressed UNIFAC because these models can predict the behavior of solid-liquid equilibriums. The binary parameters are dependent on the temperature because they are fitted to experimental data, from databanks, based on the temperature. This can be seen as a disadvantage because it influences the accuracy of extrapolated result to other temperatures. [5] [7].

3.2 NRTL

The first model is NRTL. The abbreviation NRTL stands for non-random two-liquid model and it is a LC activity coefficient model correlating the activity coefficients γ_i of a component I with its mole fractions x_i in the liquid phase. This model is created out of the Wilson model. The reason for this further adaptation is that the Wilson model cannot predict the behavior for Liquid-liquid or Vapor-liquid-liquid systems. NRTL is thus similar to Wilson and is usable for different phase equilibriums such as multi component vapor-liquid, liquid-liquid and Vapor-liquid-liquid. This models uses in contrast to the Wilson model three adjustable parameters instead of two. [5] The next equation is for the expression of excess Gibbs energy of binary mixture of components 1 and 2.

$$\frac{g^E}{RT} = x_1 x_2 \left(\frac{\tau_{21} G_{21}}{x_1 + x_2 G_{21}} + \frac{\tau_{12} G_{12}}{x_2 + x_1 G_{12}} \right) \qquad 3.2.1$$

And

$$\tau_{ij} = \frac{\Delta g_{ij}}{RT}, \qquad G_{ij} = \exp\left(-\alpha_{ij}\tau_{ij}\right) \; and \; \alpha_{ij} = \alpha_{ji} = \alpha \qquad 3.2.2$$

The three adjustable parameters are α, Δg_{12} and Δg_{21}. α Is the so-called non-randomness factor and has usually a value between 0.2 and 0.47. A few recommendations for α are listed below in the table.

Table 2: α Values [5].

α	Application
0,2	hydrocarbons–polar non-associated compounds
0,3	non-polar compounds, polar mixtures with slight negative deviations or moderate positive deviations from Raoult's law, water–polar components
0,4	hydrocarbons–perfluorocarbons
0,47	alcohols–non-polar, water–butyl alcohols, pyridine, CCl4–acetonitrile and nitro methane

Δg_{12} And Δg_{21} are binary interaction parameters and are independent of the composition. They can be assumed as constants or linear in temperature and they are collected by fitting them to experimental data from databanks.

The activity coefficients can now be derived from equation 3.2.1 as followed.

$$ln\gamma_1 = x_2^2\left(\tau_{21}\frac{\exp(-2\alpha\tau_{21})}{(x_1 + x_2\exp(-\alpha\tau_{21}))^2} + \tau_{12}\frac{\exp(-\alpha\tau_{12})}{(x_2 + x_1\exp(-\alpha\tau_{12}))^2}\right) \quad 3.2.3$$

$$ln\gamma_2 = x_1^2\left(\tau_{12}\frac{\exp(-2\alpha\tau_{12})}{(x_2 + x_1\exp(-\alpha\tau_{12}))^2} + \tau_{12}\frac{\exp(-\alpha\tau_{21})}{(x_1 + x_2\exp(-\alpha\tau_{21}))^2}\right) \quad 3.2.4$$

This can be extended for multicomponent mixtures. An advantage of NRTL is that no multicomponent data is required to obtain the results because all the adjustable parameters are from binary systems [8] [9].

3.2.1 NRTL-SAC

The NRTL model has a few variants including NRTL-SAC, which can be used for the predicting the behavior of solid-liquid phase with a focus on predicting pharmaceutical-solvent phase behavior. This model is developed by ASPEN researchers and is implemented in the Aspen Plus software. The name NRTL-SAC is short for non-random two-liquid segment activity coefficient model. The difference between the normal NRTL is based on the fact that the molecules can be defined as different predefined conceptual segments and that those segments account for the interactions of each molecule in the system. There are four different conceptual segments and they can be defined as follows the first one is for hydrophobicity X, this describes the probability for the molecular surfaces that no hydrogen bond is formed. Then there is hydrophilicity Z, this describes the probability for the molecular surfaces that a hydrogen bond is formed. The last two account for polarity namely repulsive Y- and attractive Y+. These describe the polar molecular surfaces that are “electron pair donor or acceptor”. The expected advantage of this model is a better prediction of the behavior of strongly non-ideal mixtures but the main drawback is that more parameters are required. [10] [11].

3.3 UNIQUAC

Another activity coefficient model that is based on local composition method is UNIQUAC, which is short for UNIversal QUAsiChemical. The basic model can give predictions for vapor-liquid and liquid-liquid equilibriums for binary and multi component mixtures of non-electrolyte components. Only two adjustable parameters per binary mixture are required and it can be extended to a multi component system without need of any ternary parameters. The calculations for the activity coefficient and excess Gibbs energy for this model need of two contributions. The first one is the combinatorial term and it describes the differences of shape and size between the components. The second one is the residual term and it accounts for energy differences between the molecules [12] [5]. The total equation for the activity coefficient of each component i can thus be described in the next equation (3.3.1).

$$\ln \gamma_i = \ln \gamma_i^C + \ln \gamma_i^R \qquad 3.3.1$$

The combinatorial term is $\ln \gamma_i^C$ and the residual term is $\ln \gamma_i^R$. This can also be written for excess Gibbs energy as given below (3.3.2).

$$\frac{g^E}{RT} = \frac{g^E(c)}{RT} + \frac{g^E(r)}{RT} \qquad 3.3.2$$

For a binary mixture the equations are.

$$\frac{g^E(c)}{RT} = x_1 ln \frac{x_1 r_1/r}{x_1} + x_2 ln \frac{x_2 r_2/r}{x_2} - \frac{Z}{2}\left(x_1 q_1 ln \frac{x_1 r_1/r}{x_1 q_1/q} + x_2 q_2 ln \frac{x_2 r_2/r}{x_2 q_2/q} \right) \qquad 3.3.3$$

$$\frac{g^E(r)}{RT} = -x_1 q_1 ln\left(\frac{x_1 q_1}{q} + \frac{x_2 q_2}{q} \tau_{21} \right) - x_2 q_2 ln\left(\frac{x_2 q_2}{q} + \frac{x_1 q_1}{q} \tau_{12} \right) \qquad 3.3.4$$

And

$$\tau_{ij} = \frac{\Delta g_{ij}}{RT}, \qquad r = \sum_{j=1}^{n} x_j r_j \ and \ q = \sum_{j=1}^{n} x_j q_j \qquad 3.3.5$$

In the combinatorial term, the Z value refers to a degree of non-randomness of the mixture and is typically set to 10 and meaning that there are 10 interacting molecules around a central molecule but can be variable depending on the system. The r_i and q_i value are the molecular volume and surface area, which are estimated using group contribution values of Bondi (3.3.5 and 3.3.6) [13] [5] [12].

$$r_i = \sum_k v_k^{(i)} R_k$$

3.3.6

$$q_i = \sum_k v_k^{(i)} Q_k$$

3.3.7

In these equation the parameter $v_k^{(i)}$ stands for the number of functional groups of type k in molecule i and R_k and $Q_k$are the volume and surface area parameters of each functional group k. The only two parameters that are fitted to experimental data for the UNIQUAC model are energy interactions and for a binary system these are Δg_{12} and Δg_{21}. The other parameters that are used in the equations are related to pure components and are usually found in databanks or they have to be determined experimentally. The equations of the activity coefficients can also be derived from the excess Gibbs energy equation as described for the previous model.

3.4 UNIFAC

The last LC activity coefficient model of this study is the UNIFAC functional-group activity coefficient. UNIFAC is based on UNIQUAC and exist of the sum of a combinatorial activity coefficient, which is identical to UNIQUAC, and a residual activity coefficient, which is different, compared to the residual activity coefficient of UNIQUAC. The residual part is evaluated from interactions between groups. The equations can be described as follows.

$$\ln \gamma_i = \ln \gamma_i^{Comb} + \ln \gamma_i^{Res} \qquad 3.4.1$$

The combinatorial term is then.

$$\ln \gamma_i^{Comb} = \ln\left(\frac{x_i r_i / r}{x_i}\right) + 1 - \left(\frac{x_i r_i / r}{x_i}\right) - \frac{Z}{2} q_i \left(\ln\left(\frac{x_i r_i / r}{x_i q_i / q}\right) + 1 - \left(\frac{x_i r_i / r}{x_i q_i / q}\right)\right) \qquad 3.4.2$$

And r and q are described as.

$$r = \sum_{j=1}^{n} x_j r_j \ \ and \ q = \sum_{j=1}^{n} x_j q_j \qquad 3.4.3$$

The residual term is then.

$$\ln \gamma_i^{Res} = \sum_k v_k^{(i)} \left(\ln \Gamma_k - \ln \Gamma_k^{(i)}\right) \qquad 3.4.4$$

The variables of the residual term are $v_k^{(i)}$, this stands for the number of k groups present in compound i, and $\Gamma_k^{(i)}$ being the residual group contribution of group k in a reference solution containing only molecule i. This last term is needed to obtain an activity coefficient γ_i of one if x_i is going to one (pure component). The next equation is to determine Γ_k the residual contribution of a group k for the mixture and the pure component.

$$\ln \Gamma_k = Q_k \left(1 - \ln\left(\sum_m \Theta_m \Psi_{mk}\right) - \sum_m \frac{\Theta_m \tau_{km}}{\sum_n \Theta_n \tau_{nm}}\right) \qquad 3.4.5$$

In the previous equation the parameter Θ_m is used to define the surface area fraction of group m in the mixture and the equation for Θ_m is:

$$\Theta_m = \frac{X_m Q_m}{\sum_n X_n Q_n}$$

3.4.6

In equation 3.4.6 the parameter X_m is the mole fraction of the group m in the mixture and can be described as follows.

$$X_m = \frac{\sum_j v_m^j x_j}{\sum_j \sum_n (v_n^j x_j)}$$

3.4.7

The parameter τ_{km} in equation 3.4.5 is the function that describes the group interaction parameters between group's m and n. The function is given in the next equation.

$$\tau_{km} = \exp\left(-\frac{U_{mn} - U_{nn}}{T}\right) = \exp\left(-\frac{a_{mn}}{T}\right)$$

3.4.8

Here is U a measure of the energy of interaction between groups and a_{mn} are the group interaction parameters, two parameters per binary mixture of groups that must be retrieved from experimental data. They are considered as temperature independent. The UNIFAC model has also different variants to describe different phase equilibriums. The basic UNIFAC model is good for predicting vapor-liquid phase equilibrium behavior within narrow temperature ranges and extrapolation above 425K should be avoided. Another flaw is that Liquid-liquid phase equilibriums are predicted poorly by UNIFAC. Therefore, there exist many variances of the UNIFAC model to improve the prediction of certain phase equilibriums [14] [5].

3.4.1 Regressed UNIFAC

Another variant of the UNIFAC model is RU or Regressed UNIFAC, which is offered in the software package Dynochem. This model can predict the behavior of solid-liquid phase equilibriums with a focus on 'complex' solids such as pharmaceutical components. The regressed UNIFAC model uses the basis of the standard UNIFAC model that was described in the previous section. This means that the solvent fragments into UNIFAC functional groups and for common solvents, these group interaction parameters are already provided. The assumption that is made for RU is that the solute is seen as one group and thus no fragmentation of the solute in functional groups has to be done. This is also the difference between standard UNIFAC and Regressed UNIFAC. The interaction parameters between the solute and solvent groups are found by regressing them against solubility data. The solubility data must be acquired experimentally and this data depend on the solute-solvent system for which the prediction is required. An example is that the behavior of a solute in an ether is needed. Then the experimental data needed for determining the parameters also need a solvent that has an ether group in it. A specific example is displayed in the next figure [15].

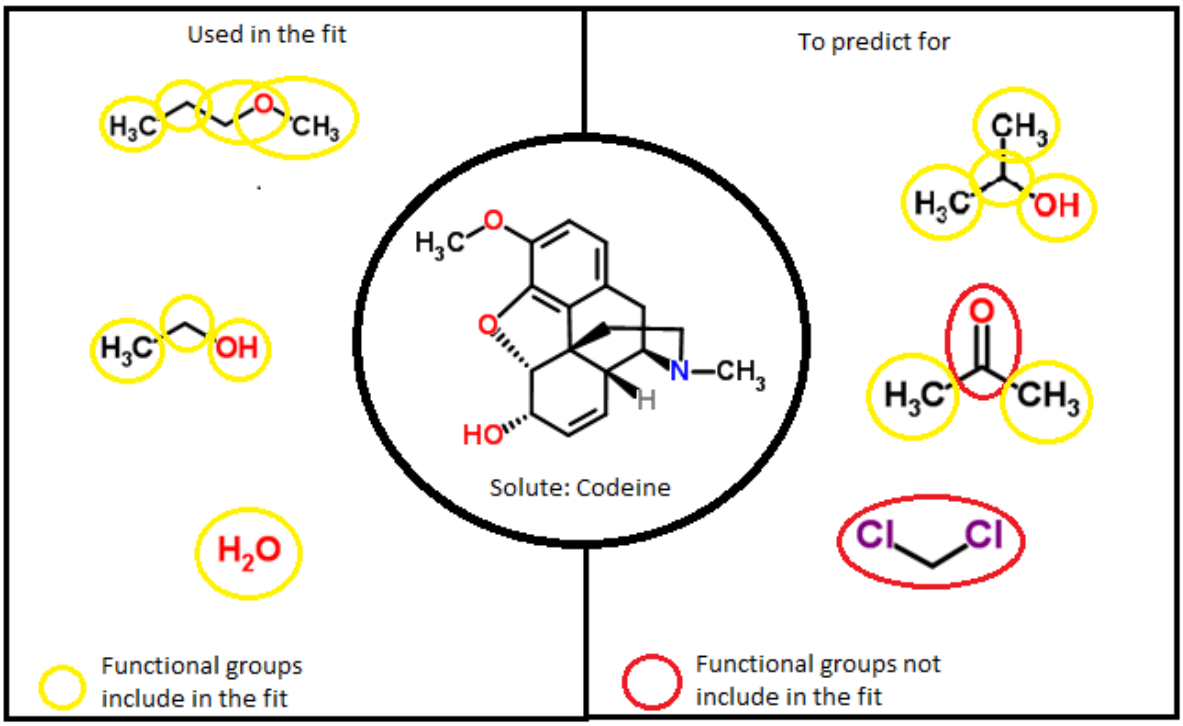

Figure 5: Example Regressed UNIFAC [15] [16].

In this figure, Codeine is the solute and the molecule of codeine is seen as one functional group. In the 'Used in the fit' window are the molecules given that were used for the fit. The structure of the molecules and experimental solubility data of the solute in that molecule is required to created predictions for the solubility of the solute in other molecules with the same functional groups. In the 'To predict for' window are the molecules given where for prediction is needed and because not every functional group was included in the fit not every prediction can be made.

As example the solubility of codeine in the solvent 2-propanol is taken from figure 5Figure 5. The molecule 2-propanol exists out of three carbon groups and one alcohol group. The solute codeine is not divide in groups but is considered as one group and thus only interactions between the codeine and the different groups of the solvent are used. To predict the solubility of codeine in 2-propanol experimental data are needed for minimum two other solvents which contain the same groups as 2-propanol. Examples of experimental data that can be used are ethanol and methyl propyl ether because then the alcohol group is included in one solvent and both solvents have carbon groups. Another possible combination is that two different experimental data for ethanol are used.

Part 2: Materials & methods

1 Introduction

This chapter explains the materials and methods that are used to collect data, generate results and formulate conclusions. The first part discusses the software tools that are used. Then a complete overview of the different systems with the software packages is given and the procedure for each system is explained. Finally, the evaluation method to determine the accuracy and usefulness of the results and software tools is explained.

2 Materials

The two software packages used to generate solubility predictions are Dynochem and AspenONE. Dynochem is created by Scale-up Systems, this is a privately held technology company, with headquarters in Dublin, Ireland. The focus of this company is to produce the best software for the pharmaceutical and related industries in order to accelerate and facilitate the process development, scale-up and continuous improvement in API manufacturing. Dynochem 2011 is an add-in for excel that uses different worksheets created by the company, available on their website [15]. The worksheets used are 'Early phase solvent selection: solubility prediction' for solid-liquid systems using the thermodynamic model Regressed UNIFAC and 'Vapor-Liquid and Liquid-Liquid equilibrium' for liquid-liquid systems using NRTL or UNIFAC. The gas-liquid systems were not investigated in this thesis with Dynochem because no direct method is available.

AspenONE is created by the company AspenTech. Headquarters is situated in Burlington, Massachusetts. The focus in this company is to create software that optimizes process manufacturing for different industries such as oil & gas, chemicals, engineering & construction, pharmaceutical and others. The software package AspenONE includes specific programs for different applications. For solubility prediction of gas-liquid and liquid-liquid system Aspen Plus V8.0 is used. The thermodynamic models used in this program are NRTL, UNIFAC and UNIQUAC. For solid-liquid systems an excel add-on is used namely the Aspen solubility modeler which is created for solid-liquid systems with the focus on API's-liquid systems usingthe thermodynamic model NRTL-SAC.

The experimental data to compare with the predicted results are collected from literature and from available data within Janssen.

3 Methods

3.1 Overview general approach

In table 3 a complete overview is shown explaining the practical side of the thesis. The more detailed explanation of each system is described in of the next sections. The general approach for each system such as gas-liquid, liquid-liquid and solid-liquid is as follows. In each system, different cases are studied to make sure the solubility predictions are case independent. Then if possible, different thermodynamic models and programs are used to predict the solubility of the compounds and these data are then compared with each other and with the experimental data. The different programs are evaluated to determine whether they are user-friendly.

Table 3: Overview practical approach.

Solubility prediction			
Systems / Programs →	Gas-liquid	Liquid-Liquid	Solid-Liquid
Aspen ONE	Model: NRTL with Henry's law	Model: NRTL, UNIQUAC or UNIFAC	Model: NRTL-SAC
	Cases	Cases	Cases
	methyl chloride & water methyl bromide & water methyl bromide & ethanol methyl bromide & water & ethanol	toluene & water benzene & water dichloromethane & water 1-butanol & 3-methyl-1-butanol & water 1-propanol & benzene & water n-hexane & methanol & cyclohexane ethanol & water & cyclohexane	thebaine & dipropyl ethers codeine & dipropyl ethers oripavine & dipropyl ethers morphine & dipropyl ethers thebaine & dimethyl ether codeine & dimethyl ether oripavine & dimethyl ether morphine & dimethyl ether
Dynochem	Model: /	Model: NRTL or UNIFAC	Model: Regressed UNIFAC
	Cases	Cases	Cases
		1-butanol & 3-methyl-1-butanol & water 1- propanol & benzene & water n-hexane & methanol & cyclohexane ethanol & water & cyclohexane	thebaine & different ethers codeine & different ethers oripavine & different ethers morphine & different ethers thebaine & dimethyl ether codeine & dimethyl ether oripavine & dimethyl ether morphine & dimethyl ether

3.2 Gas-liquid systems

The method used to predict the solubilities for gas-liquid systems is explained in this section. An overview of the gas-liquid systems is given with an explanation on the method that was chosen to predict solubilities.

3.2.1 Overview gas-liquid systems

The gas-liquid systems are divided into two categories depending on the number of solvents One solvent systems and a mixture of two solvents have been investigated. The cases are listed in table 4

Table 4: Investigated cases for gas-liquid.

Case	Solute	Solvent
1	methyl chloride	water
2	methyl bromide	water
3	methyl bromide	ethanol
4	methyl bromide	water and ethanol

Janssen Pharmaceutica recommended methyl bromide and methyl chloride because they are part of some production processes. Only Aspen Plus in AspenONE was used because Dynochem has no direct method to generate predictions for gas-liquid solubilities.

3.2.2 Generation of gas-liquid solubility curves

In this section, the method is described for Aspen Plus to create solubility predictions for gas-liquid systems. In the following steps, the procedure to calculate gas-liquid solubilities with Aspen Plus is introduced based on the example of methyl chloride in water. At first, the different compounds are added as shown in the next figure (Figure 6).

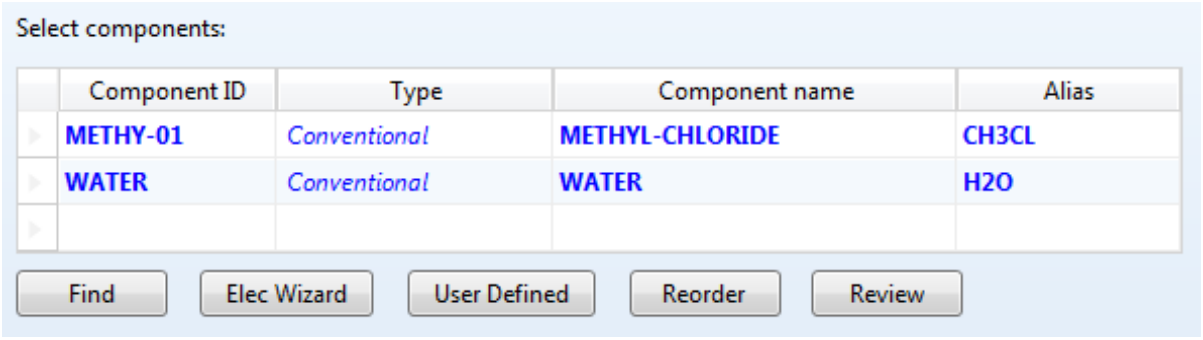

Figure 6: Adding components for gas-liquid systems in Aspen Plus.

The next and most important step is to choose the desired model and in this case NRTL/UNIQUAC with Henry's law or Henry components was chosen. This model was chosen because it was used before in similar case about an absorber in Janssen Pharmaceutica and delivered good results.

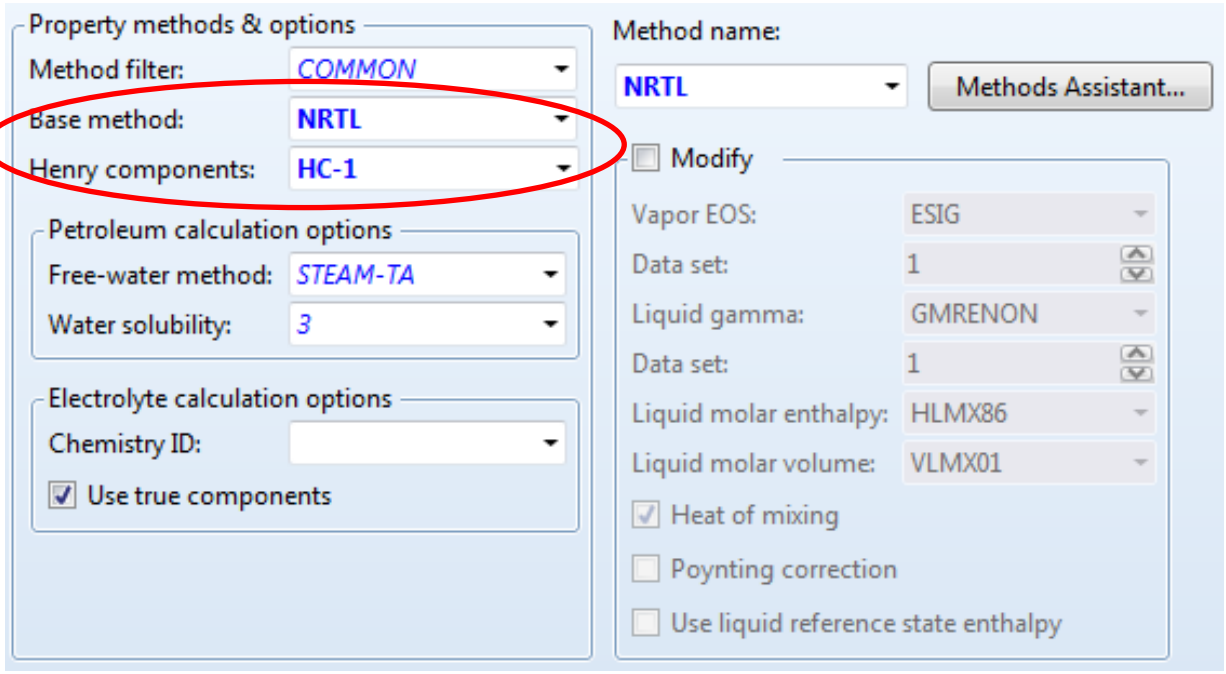

Figure 7: Choosing thermodynamic model for gas-liquid systems in Aspen Plus.

The Henry components are added because they are specifically used for gas-liquid systems as mentioned in the literature study. The third step is to create a flow sheet with a flash separator as shown in the next figure (8)

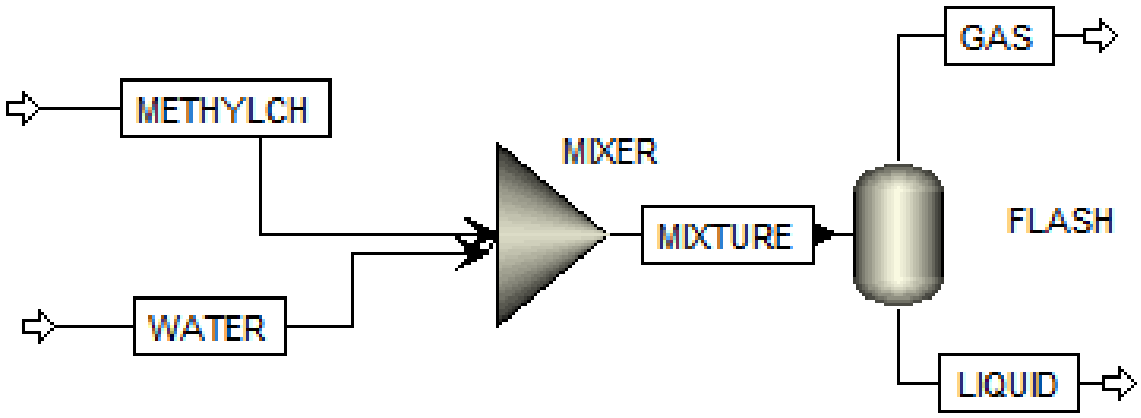

Figure 8: Flow sheet gas-liquid system in Aspen Plus.

Figure 8 an example is given with 'METHYLCH' as the gas or solute and 'WATER' as the liquid or solvent, thus creating a binary system. For ternary gas-liquid systems the liquid will be a mixture of two liquids. The gas and the liquid go to a mixer to create a homogeneous mixture, which is separated using a two-outlet isothermal flash with a gas and liquid phase. The fourth step is to do a sensitivity analysis for the flash temperature. This means that the temperature of the flash separator changes between two values and thereby the changes of the compositions of liquid outlet are registered. Only the outlet of the liquid phase is used because only the solubility of a gas in a liquid is needed. An example of the results is given in table 5.

Table 5: The results for a gas-liquid system in Aspen Plus.

Temperature	Methyl chloride	Water
°C	kg/h	kg/h
0	0.131	9.98
5	0.105	9.97
10	0.0854	9.96
15	0.0703	9.94
20	0.0585	9.91

The final step is recalculated these values in g/L solvent, the time is set at 1h and then the liquid is converted to liter by using the density and the gasfraction to grams. To obtain the solubility the resulting values will be divided by each other. An example of these calculations is given for 5°C. The density of water changes with the temperature but this neglected because the average error is less than 1.5%. Thus the density is assumed constant at 1.000kg/L

$$Solvent(Water) \rightarrow {}^{9.97\text{kg}}/_{1.000kg/L} = 9.97\text{L}$$

$$Solute(Methylchloride) \rightarrow 0.105\text{kg} = 105\text{g}$$

$$Solubility\ {}^{g}/_{L} = {}^{105g}/_{9.97L} = 10.5\ {}^{g}/_{L}\ for\ 5°C$$

In case of a ternary system, two liquids will be present and the total volume of these two liquids is used for the final calculation. The calculations are then created for every temperature and plotted in a temperature-solubility curve. This curve is then compared with experimental data to determine the accuracy. This is further explained in the part about the evaluation method.

3.3 Liquid-liquid systems

This section discusses about the method to create solubility curves for binary liquid-liquid systems and liquid-liquid phase diagrams for ternary liquid-liquid systems. First, an overview of the liquid-liquid system is given and then the methods are explained in detail to generate the desired results.

3.3.1 Overview liquid-liquid systems

The liquid-liquid systems are divided into two categories namely usage of one liquid solvent and of a mixture of two liquids solvents. In case of one solvent the used cases are listed in the table below

Table 6: Investigated cases for liquid-liquid.

Case	Solute	Solvent
1	toluene	water
2	benzene	water
3	dichloromethane	water
4	1-butanol	3-methyl-1-butanol & water
5	1-propanol	benzene & water
6	n-hexane	methanol & cyclohexane
7	ethanol	water & cyclohexane

Aspen Plus in AspenONE was used for binary and ternary systems. Dynochem was used only for ternary systems using the worksheet 'Vapor-Liquid and Liquid-Liquid equilibrium'.

3.3.2 Generation of liquid-liquid solubility data

This section describes the approach that was followed with each software package to create solubility predictions for liquid-liquid systems.

3.3.2.1 Dynochem

In Dynochem, only the ternary liquid-liquid systems were investigated. At first, desired input data e.g. temperature, components, feed composition and model were added as shown in figure 9. NRTL or UNIFAC were selected as thermodynamic models.

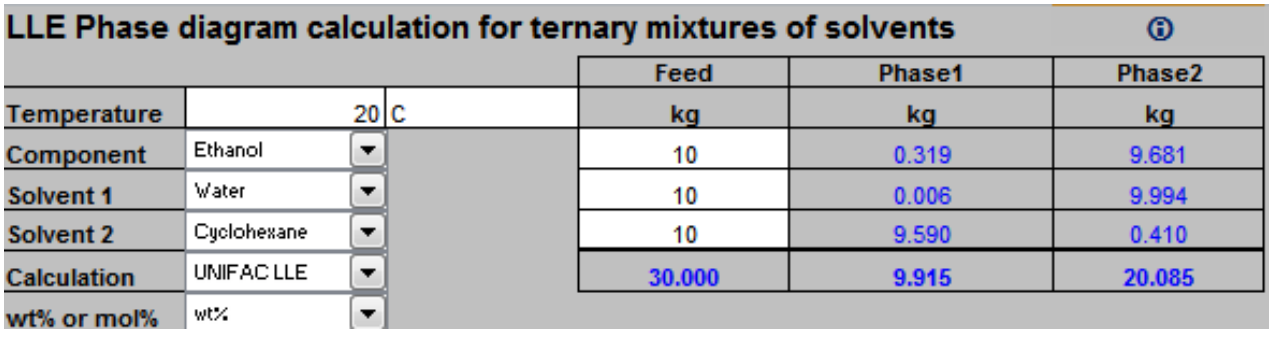

Figure 9: Input for a ternary liquid-liquid system in Dynochem.

After the input data are filled in, the results are shown in a table and are plotted in a ternary diagram as displayed in table 7 and figure 10.

Table 7: The calculated compositions of the two phase system for ethanol-cyclohexane-water at 20°C in weight precent

Phase1			Phase2		
Ethanol	Water	Cyclohexane	Ethanol	Water	Cyclohexane
wt%	wt%	wt%	wt%	wt%	wt%
0.000	0.011	99.989	0.000	99.955	0.045
0.907	0.020	99.072	8.117	91.791	0.093
1.581	0.029	98.390	16.038	83.782	0.180
2.061	0.037	97.902	23.487	76.186	0.327
2.613	0.046	97.342	34.708	64.527	0.765
2.685	0.047	97.268	36.369	62.766	0.864
2.681	0.047	97.273	36.257	62.886	0.857
2.563	0.045	97.392	33.589	65.707	0.704

In table 7 the equilibrium weight percentages of the two-liquid phases for the ternary system ethanol-water-cyclohexane are calculated by Dynochem at 20°C beginning with the binary system water-cyclohexane. Then ethanol is added until the two-liquid phases disappear, meaning that the components are fully miscible. These values are plotted in figure 10. The pink square in the middle is the feed that is defined in figure 9 and it splits in two-liquid phases (other pink squares). The blue points are the different compositions of both phases and the blue lines are tie lines that connect corresponding compositions of both phases with each other.

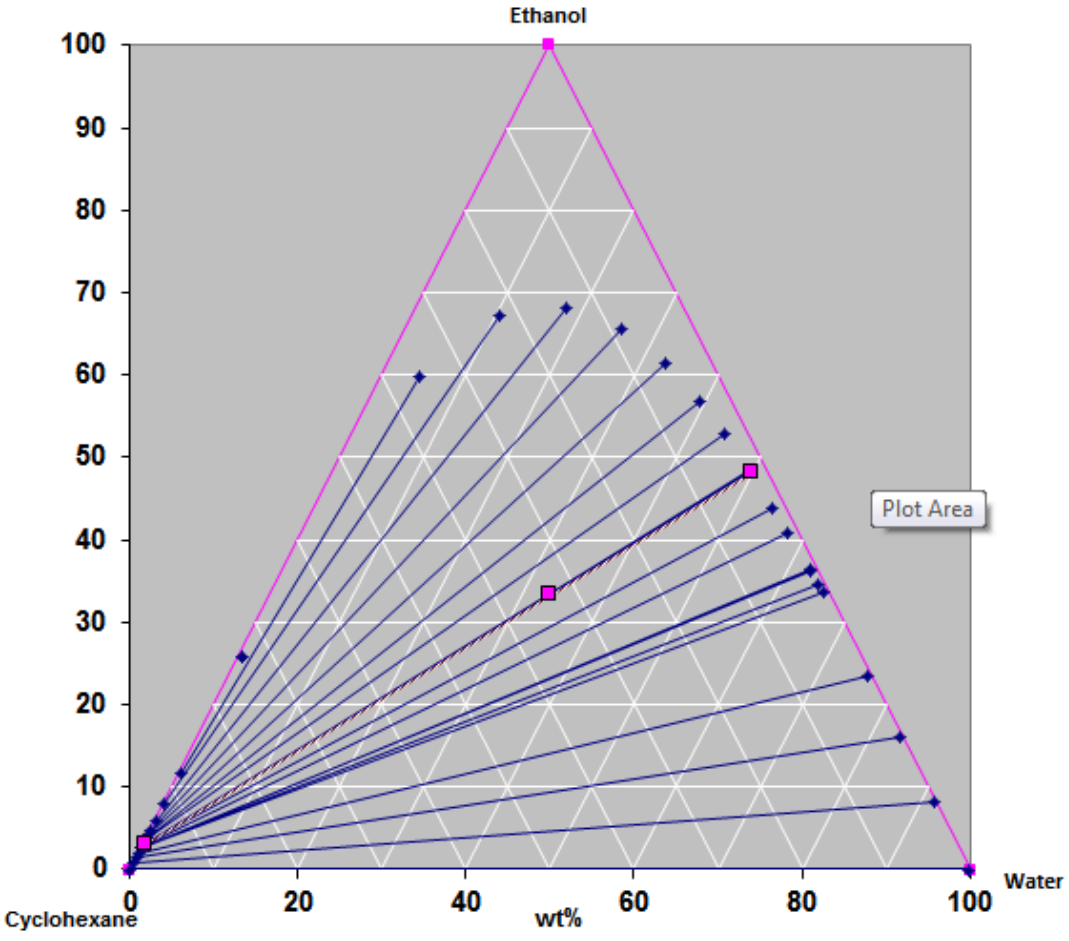

Figure 10: The plotted compositions for ethanol-cyclohexane-water system at 20°C in weight precent.

3.3.2.2 Aspen Plus

In Aspen Plus predictions are generated for binary and ternary liquid-liquid systems. The first step is adding the desired components. An example for a binary system is shown in figure 11.

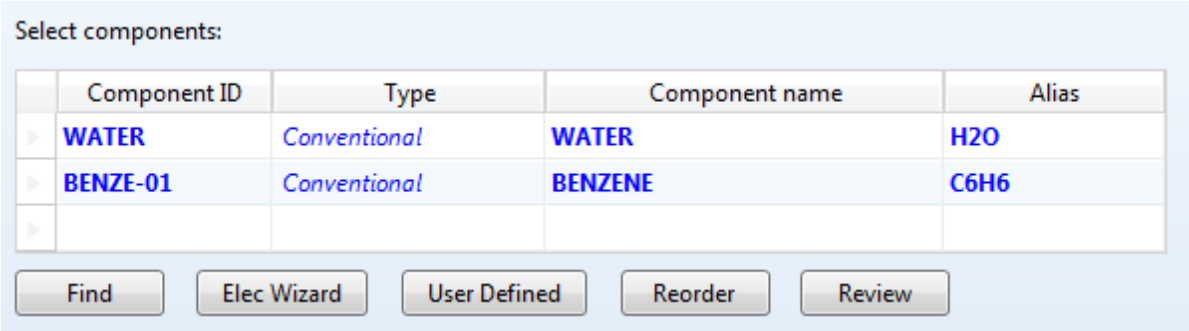

Figure 11: Adding components for a liquid-liquid system in Aspen Plus.

The second step is the choice of an appropriate model which means for liquid-liquid systems NRTL, UNIQUAC or UNIFAC (Figure 12).

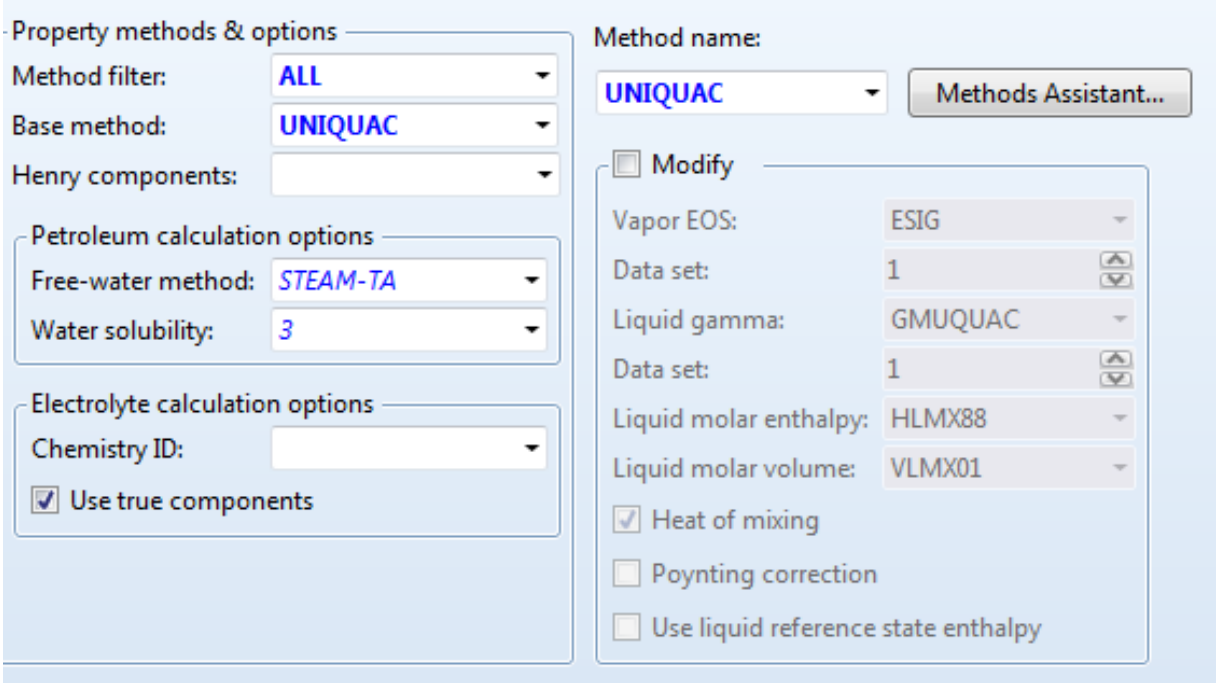

Figure 12: Choosing the model for a liquid-liquid system in Aspen Plus.

The third step is creating a flow sheet with an isothermal flash separator as shown in figure 13 for realisation of the solution of liquid benzene in water. The reason that an isothermal flash is taken is that the temperature needs to be constant to determine the solubility.

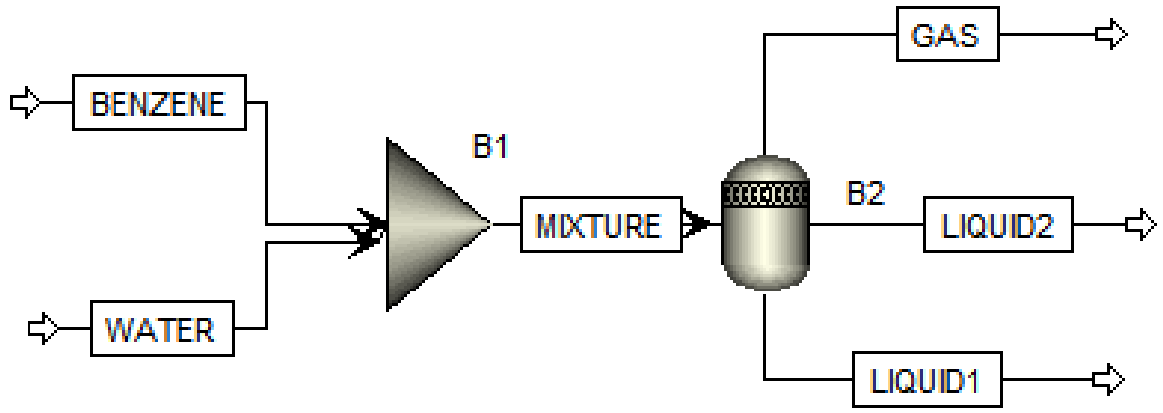

Figure 13: Flow sheet for liquid-liquid systems in Aspen Plus.

Both liquids go in a mixer to create a homogeneous mixture which is then separated using a three-outlet flash. The solvent rich liquid phase, in this case liquid phase 1, is used for solubility prediction. Finally, to create a temperature-solubility curve, a sensitivity analysis for the flash temperature is performed. This means that the temperature will be change between two limiting values with a given step size of 5°C and thereby the changes of the compositions of the liquid outlets are registered (Table 8)

Table 8: The results for a liquid-liquid system in Aspen Plus.

	Liquid phase 1		Liquid phase 2	
Temperature	Solute	Solvent	Solute	Solvent
°C	**kg/h**	**kg/h**	**kg/h**	**kg/h**
0	0.013	9.99	0.005	9.99
5	0.013	9.99	0.005	9.99
10	0.014	9.99	0.005	9.99
15	0.015	9.99	0.006	9.99
20	0.016	9.99	0.006	9.98
25	0.017	9.99	0.007	9.98

The obtained data are recalculated for all temperatures to g/L solvent as described in '3.2.2 Generation of gas-liquid solubility curves'.

In case of a ternary system the ternary phase diagrams are created by using a ternary analysis on the selected components (1, 2 and 3) and thermodynamic model (property model) as shown in figure 14.

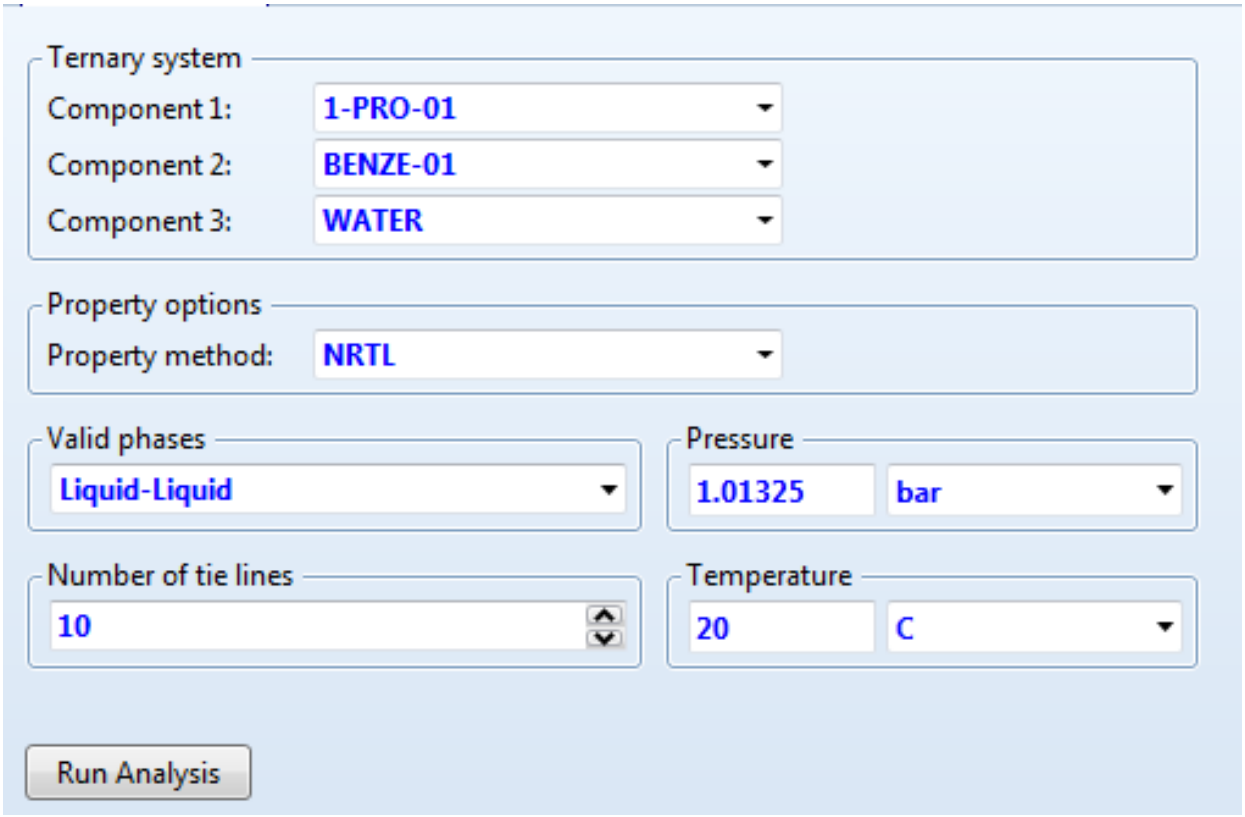

Figure 14: Input data for a ternary liquid-liquid system in Aspen Plus.

The pressure and temperature are set to a desired constant value. In this example the pressure is 1atm or 1.01325bar and the temperature is 20°C. An example of the results is given in figure 15.

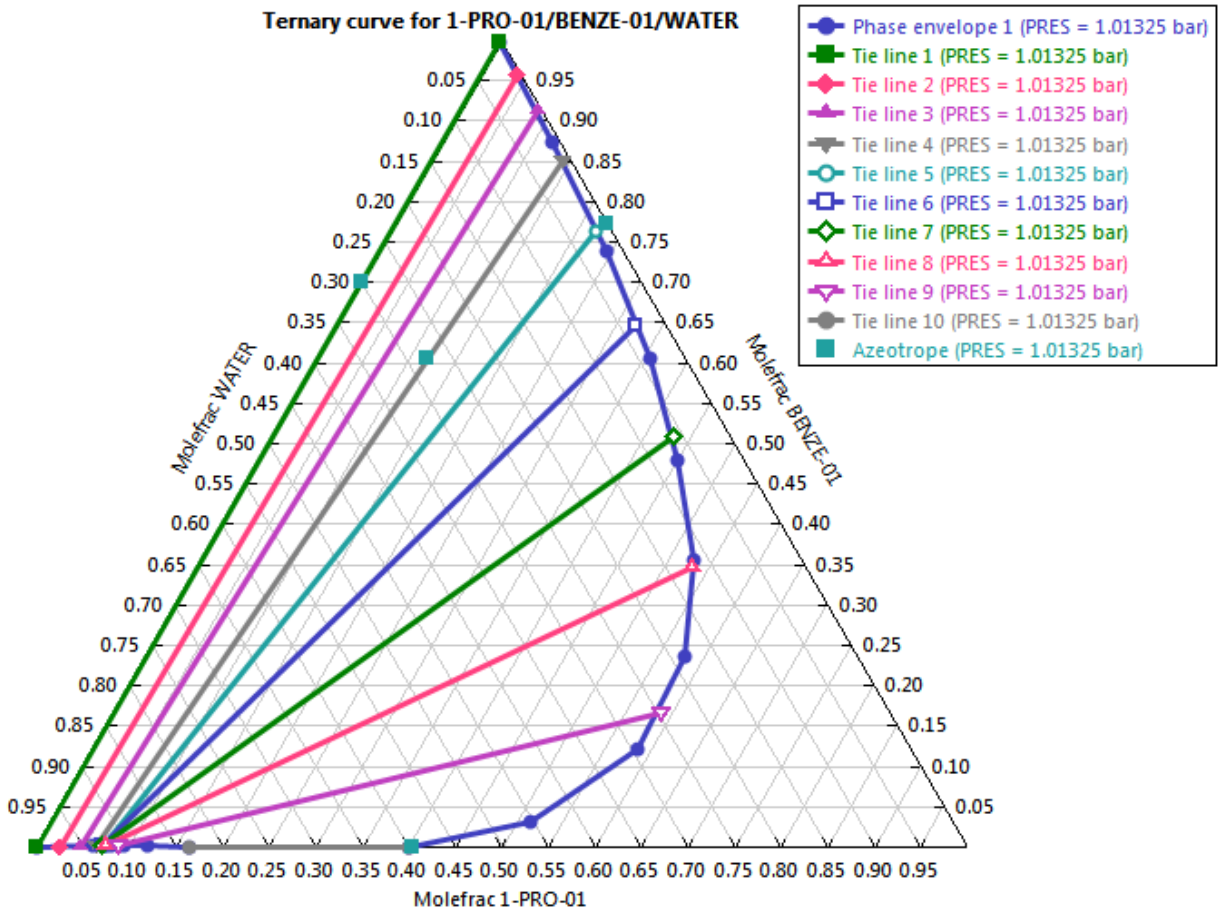

Figure 15: The plotted compositions for 1-propanol-benzene-water system at 20°C in mole fraction.

The equilibrium mole fractions of the two-liquid phases for the ternary system 1-propanol-benzene-water are calculated and plotted by Aspen Plus beginning with the binary system water-benzene (tie line 1) by 20°C. The other tie lines are two opposite liquid phases that are connected and the phase envelope line connects al the mole fractions to define the boundary between miscibility and immiscibility of the components.

3.4 Solid-liquid systems

This section discusses about the method to generate solubility data for solid-liquid systems. First, an overview of the solid-liquid systems is shown and then the methods are explained in detail.

3.4.1 Overview solid-liquid systems

The solid-liquid systems have only one category namely usage of one liquid solvent. The cases that are investigated are listed below (Table 9):

Table 9: Investigated cases for solid-liquid.

Case	Solute	Solvent
1	thebaine	dipropyl ether
2	codeine	dipropyl ether
3	oripavine	dipropyl ether
4	morphine	dipropyl ether
5	thebaine	dimethyl ether
6	codeine	dimethyl ether
7	oripavine	dimethyl ether
8	morphine	dimethyl ether

These cases are investigated because they were given by Janssen Pharmaceutica. In AspenONE, 'the solubility modeler' was used and for Dynochem the worksheet 'early phase solvent selection: solubility prediction'.

3.4.2 Generation of solid-liquid solubility curves

3.4.2.1 Dynochem

At first the required information e.g. the chemical formula, amount of aromatic rings, amount of non-aromatic rings, the heat of fusion, the normal melting point and the solute density are specified. The chemical formula is used to calculate the molecular weight, the amount of aromatic and non-aromatic rings is used to estimate the R and Q values, the heat of fusion and the normal melting point are used in the equation for the solubility calculation (equation 2.4.10) and the solute density and molecular weight is used for unit conversions. The name and the structure are optional and not used further. The R and Q values are use in the formula of UNIFAC and are explained in the literature study (section 3.4 UNIFAC). An example is given for thebaine in the next table (Table 10)

Table 10: Input for solid-liquid systems for Dynochem.

Solute data		Solid form
Name	TheBaine	
Chemical Formula	C19H21NO3	
# aromatic rings	1	
# non-aromatic rings	4	
Heat of fusion	23060.01174	J/mol
Normal melting point	465	K
Solute density	1305	g/L
Quick solute R and Q estimates		
R	11.51	-
Q	8.13	-
Molecular weight	311.37	g/mol
Optional: Paste molecular structure		

The second step is to define the solvents and add measured data for those solvents. The reason for this is explained in the literature study (section 3.4.1 Regressed UNIFAC) An example is shown in table 11. The experimental data is collected from available data within Janssen, which will be added in the results for solid-liquid systems.

Table 11: Input of experimental data for solid-liquid systems in Dynochem.

Define solvents and enter measured data		**Vol frac 1**	**Temp**	**Measured**	**Fit to data**
Solvent1	Solvent2	-	C	g/L solvent	
1_2_dimethoxyethane	<select solvent>	1	20	22.6	No
Diglyme	<select solvent>	1	20	57.3	No
Dipropyl_ether	<select solvent>	1	20	2.3	No
Dibutyl_ether	<select solvent>	1	20	2.1	No
Isopropyl_ether	<select solvent>	1	20	1.6	No
Methyl_propyl_ether	<select solvent>	1	20	6.8	No
Diethyl_ether	<select solvent>	1	20	4.2	No

At least two different measured data need to be included.

The third step is then to determine if the heat of fusion, the normal melting temperature, R-value and Q value need to be included in the fit. In this example (Table 12) the heat of fusion and normal melting temperature are not included in the fit because the experimental values were available. The R and Q values are included to create a better estimation of these values and thus created predictions that are more accurate.

Table 12: Fit additional parameters in Dynochem.

Additional parameters:	
Fit ln (DHf)	No
Fit Tm	No
Fit ln R	Yes
Fit Q	Yes

The fourth step is then to fit the experimental values and additional parameters using the Regressed UNIFAC model. An example for the results of a fit is given in the next table (13)

Table 13: The results for a solid-liquid system in Dynochem.

Solubility in single solvent		Refresh		
Solvent	Methyl_propyl_ether			
Method	RU			
Covered	TRUE			
Temperature	Start	Increment		
C	20	1		
xsolvent1	T	x1	x2	C1
solute free	C	molfrac	molfrac	g/L solv
1	20	0.0014	0.9986	4.27
1	21	0.0015	0.9985	4.47
1	22	0.0015	0.9985	4.67
1	23	0.0016	0.9984	4.88
1	24	0.0017	0.9983	5.10

The solvent where prediction is needed for is selected, in this case methyl propyl ether. The predictions are only valid for solvents that are covered as explained in the literature study, which is indicated by the value 'True' as shown in table 13

3.4.2.2 AspenONE

Predictions are based on two different worksheets namely the regression and the calculations file. The first step is to add all the required information in the regression file which includes the molecular weight, the melting point, enthalpy of fusion or entropy of fusion. The melting point and the enthalpy of fusion are used in the equation for the solubility calculation (equation 2.4.10). The NRTL-SAC parameter can be filled in manually or regressed using experimental data.

Table 14: Input for solid-liquid systems in the regression file.

NRTL-SAC Parameter Regression For Pure Solvent System

Step 1. Open Aspen Properties File

C:\Users\adaubies\Desktop\ArnoDaubies\Aspen solubility modeler\NRTL-SAC_130_Solvents_43_Excipients.aprbkp — Execute Step 1

Step 2. Define Drug

Name	MW (Kg/Kgmole)	Melting Point (K)	Enthalpy of Fusion (KJ/Kmole)	Entropy of Fusion (J/Kmole-K)
TheBaine	311.37	465	23060.01174	49591.4231

Execute Step 2

Either the entropy or the enthalpy can be used here. If both are entered the entropy will be used.

Step 3. Enter NRTL SAC Model Parameters for Drug
(For parameter regression, go to Step 4)

Parameter X	Parameter Y-	Parameter Y+	Parameter Z	Ksp A	Ksp B	Ksp C
1	1	1	1	5.964522166	-2773.502807	0

The values of Ksp A and Ksp B can be either calculated by the programe or designated by the user.

Default value of Ksp C is 0, which can be overrided by the user

Calculate Ksp A & Ksp B

Execute Step 3

The second step is to add at least four experimental values to calculate the NRTL-SAC parameters through regression. An example is shown in the next table (Table 15). In this case, also a selection can be made which experimental data is included in the regression by changing the last row. The same experimental data as for the predictions in Dynochem was used.

Table 15: Input experimental data in AspenONE.

Step 4. Perform Data Regression to Compute Model Parameters
(At least four data points are required)

Parameter X	Parameter Y-	Parameter Y+	Parameter Z	Ksp A	Ksp B	Ksp C
REGRESS	REGRESS	REGRESS	REGRESS	REGRESS	REGRESS	EXCLUDE

Solvents	Temp (Exp) (C)	Solubility (Exp.) (g drug/g solvents)	Std-Dev Solubility(%)	Data to Regress
1,2-DIMETHOXYET	25	0.026027871	10	Yes
METHYL-N-PROPY	25	0.009244154	10	Yes
BUTYL-ETHER	25	0.002730819	10	Yes
DIETHYL-ETHER	25	0.005726752	10	Yes
DIISOPROPYL-ETH	25	0.002206897	10	Yes
DI-N-PROPYL-ETH	25	0.002206897	10	Yes

The third step is to do the regression and calculate the NRTL-SAC parameters based on the input data. Finally, a calculation sheet is created by Aspen as shown in table 16. The solvent, the temperature and the pressure can still be changed in this worksheet. The solute is the solid that was specified in the regression file. The solubility predictions are given in the last column of table 16 given in g solute/100g solvent for the specified temperature and pressure.

Table 16: Results for a solid-liquid system in AspenONE.

Comparison for Solubility in Solvents									
Temperature	(deg C)	15							
Pressure	(atm)	1							
			Input x		Calculated x		Calculated w		Solubility (g
Solvent Name	Solute	Solvent ID	Solute	Solvent	Solute	Solvent	Solute	Solvent	Calculated
WATER	**DRUG**	WATER	0.5	0.5	8.59E-05	0.999914	0.001483	0.99851707	0.148513249
1,2-DIMETHOXYETHANE	**DRUG**	12DIMEC2	0.5	0.5	2.12E-05	0.999979	7.33E-05	0.99992673	0.007327439
DIMETHYL-ETHER	**DRUG**	DIME	0.5	0.5	0.5	0.5	0.871114	0.12888642	675.8769013
DIETHYL-ETHER	**DRUG**	DEETHER	0.5	0.5	3.66E-06	0.999996	1.54E-05	0.99998461	0.001539108

3.5 Evaluation method

This section discusses about the evaluation and comparison of the predicted results and the experimental values. The criteria for the evaluation are divided in two main categories e.g. the usability and the comparison. These two categories are then divided in subcategories as shown in table 17.

Table 17: Evaluation criteria.

Usability	Data evaluation
Input	Requirements
Predictions	Accuracy of the experiments
Knowledge	Accuracy of the predictions
Interface	

The first criterion under the category of the usability is 'Input' and discusses the time needed to fill in and to create the necessary input with a specific model. The second is 'Predictions' containing the required time needed for calculation for one specific system. The 'Knowledge' criterion evaluates how much the user needs to know about the software to be able to use it. The final criterion evaluates the interface by its clarity and its user-friendliness.

The first criterion under the category of the data evaluation is 'Requirements' evaluating the amount of input that is needed to create the desired predictions in the software packages. The following is 'Accuracy of the experiments' and this evaluates the deviations and errors of the used experimental data. The final criterion is 'Accuracy of the predictions' and this contains an evaluation about deviation between the predicted and the experimental results. Another aspect for this criterion is the comparison between the trend of the experimental and the predicted results. This means that the predicted results can deviate from the experimental results but still follow the same trend. An example is that the solubility decreases with the temperature according to the experimental values and the predicted values deviate from them but they also decrease in the same trend.

Part 3: Results & discussions

1 Introduction

This chapter discusses the results of the performed case studies for both the gas-liquid and liquid-liquid systems. Every part starts with an overview of compounds in each case study. Then for each case, the input and the results are displayed and finally the conclusion for each system is explained.

2 Gas-liquid systems

2.1 Cases gas-liquid

The different cases used for the evaluation of the gas-liquid systems with the models are listed below (Table 18).

Table 18: List of different cases with the models used for gas-liquid systems.

			Model with henry		
Case	**Solute**	**Solvent**	**NRTL**	**UNIQUAC**	**UNIFAC**
1	methyl chloride	water	x	x	0
2	methyl bromide	water	x	x	0
3	methyl bromide	ethanol	0	0	0
4	methyl bromide	water and ethanol	0	0	0

Results are available for the cases marked with an 'x'. The UNIFAC model had for every case the same error, saying that there is a missing structural parameter for methyl chloride or methyl bromide. Case 3 and 4 failed for NRTL and UNIQUAC because the binary parameters Henry are not available in Aspen Plus or found in literature. The binary parameters are for supercritical components for which Henry's law is used.

2.1.1 Input gas-liquid cases

The practical method for the cases is explained in Part 2: Material & methods in section 3.2.2 Explanation gas-liquid systems. The inputs that are specific for each case are the solute and the solvent or mixture of solvents and based on these Aspen searches the correct binary parameters.

2.1.2 Result case 1: methyl chloride & water

The obtained results for the solubility of methyl chloride in water are summarized in the following graph (Figure 16). The exact data are presented appendix A.

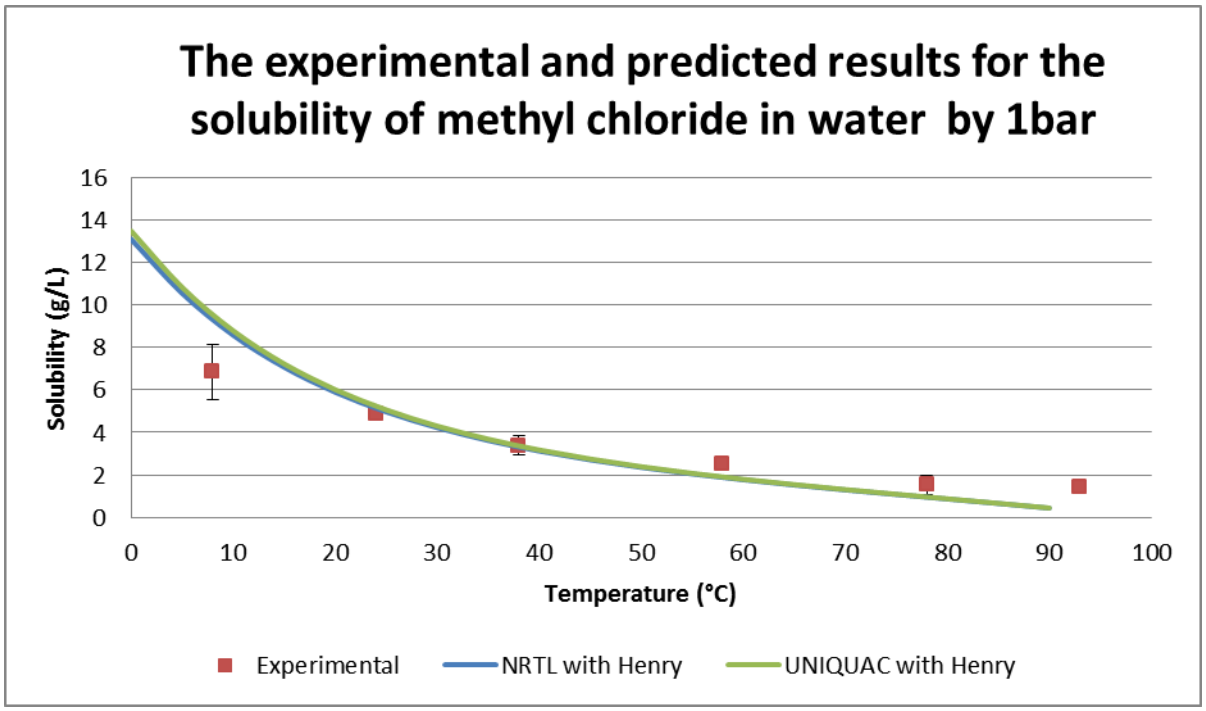

Figure 16: The solubility of methyl chloride in water for different temperatures.

The blue and green line describes the prediction using the NRTL or UNIQUAC model with Henry and the red points are experimental values collected from experiments done by Fei Chen et al [17]. The results seem less accurate around a temperature of 10°C with a deviation of 25% between predictions and experimental values. For temperatures between 20°C and 40°C, the simulated values show an average deviation of 5%. For temperatures higher than 60°C, accuracy starts to decrease again with an average deviation of 36%. The deviation are calculated as follows. First, the difference between the predicted and experimental result is made as shown for 8°C.

$$|experimental: 6.85 - predicted: 8.58| = 1.73$$

This is the divided by the experimental result and multiplied by 100 for the percentage.

$$\frac{1.73}{experimental: 6.85} * 100\% = \pm 25\%$$

The deviation for 8°C is than ±25% and for the average deviation the mean is taken from the deviations.

2.1.3 Result case 2: methyl bromide & water

The predicted and experimental results for the solubility of methyl bromide in water are plotted in the next graph (Figure 17) with the detailed values summarized in appendix A.

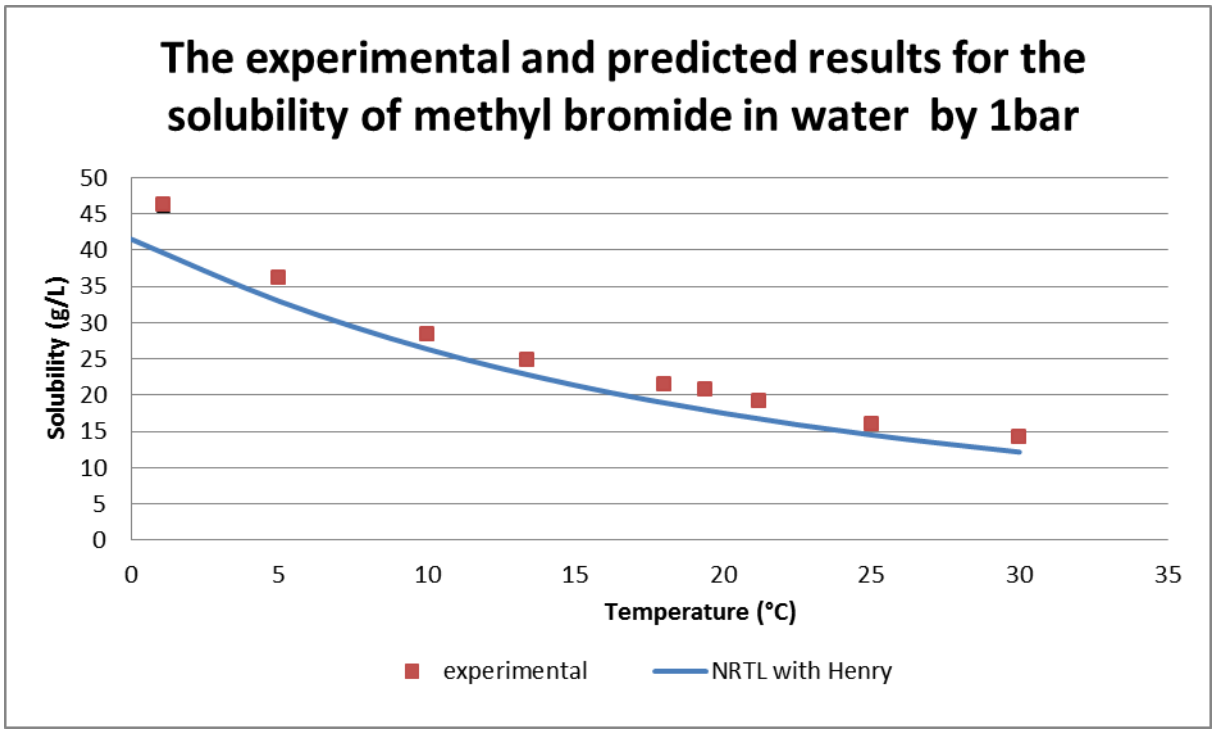

Figure 17: The solubility of methyl bromide in water for different temperatures.

The blue line are the prediction using the NRTL model with Henry and the red points are experimental values collected from experiments done by Warren De Bruyn and Eric Saltzman [18]. In this case, no large deviation exists and the average deviation between the experimental and predicted values is 11%. The method for calculating the deviations is explained in 'case 1: methyl chloride & water'.

2.2 Discussion gas-liquid

For the discussion the category ‘data evaluation’ of 3.5 the evaluation method in the materials & methods is now used. The criterions are ‘Requirements’, ‘accuracy of the predictions’ and accuracy of the experiments.

The requirements for these predictions are the different components and the models with their parameters. The components and models are available and easy to add. Both models UNIQUAC and NRTL deliver similar results with an average difference between the two models of less than 2% as shown in figure 16. This applies for case one and two therefore only NRTL was displayed in case 2. The availability of the binary parameters for Henry is limited in Aspen Plus to binary systems with water. A possible solution for this is adding these manually if they are found in literature or created by doing experiments.

The last two criterions are discussed together for each case that gave results. In case one, relatively large variations exist on the deviation between the predicted and experimental results. The reason for this can be the accuracy of the experimental data with reported experimental standard deviations of 18.9% for the temperature of 8°C and 29.4% for 78°C. Another possible cause is that the parameters for Henry are valid for a specific temperature range and for this case; it is between 3.85°C and 79.85°C according to Aspen Plus. Thus by extrapolating an error is created because the Henry parameter are temperature depended as explained in the literature study in section 2.2 Gas-liquid.

In case two, the deviations between the experimental and predicted results are lower. Two reasons can explain the deviations the first is that the error on the experimental results is only 2% and the second is that the temperatures are within the limits of temperature range of the Henry parameters. The temperature range of the binary parameters in case two are between 4.85°C and 78.85°C according to Aspen Plus.

Finally, the trends followed by the predictions are similar to the trends of the experimental results even if there are deviations between the predicted and experimental results. The reason is that the precision of the results depends on the type and precision of the experimental data that aspen uses for the regression of the binary parameter.

3 Liquid-liquid systems

3.1 Cases liquid-liquid

Table 19 gives a list of the evaluated cases for the liquid-liquid systems.

Table 19: List of different cases dived in solute and solvent used for liquid-liquid systems.

Case	Solute	Solvent
1	toluene	water
2	benzene	water
3	dichloromethane	water
4	1-butanol	3-methyl-1-butanol & water
5	1-propanol	benzene & water
6	n-hexane	methanol & cyclohexane
7	ethanol	water & cyclohexane

3.1.1 Input liquid-liquid cases

The practical method for the cases is explained in Part 2: Material & methods in section 3.3.2 Explanation liquid-liquid systems.

3.1.1.1 Aspen plus

The models NRTL, UNIQUAC and UNIFAC-LL are used to create predictions. Sometimes Aspen Plus offers different sets of binary parameters from various databanks for the models NRTL and UNIQUAC. These parameters describe the interactions between two components. NRTL and UNIQUAC have different parameters for the same system and if the system has different phase equilibriums depending on the temperature, there are also different sets to describe them. An example is that dichloromethane and water can have vapor-liquid and liquid-liquid equilibrium parameters. The explanation of these parameters is given in the literature study for NRTL in section 3.2 NRTL and for UNIQUAC in section 3.3 UNIQUAC. Thus for examining liquid-liquid systems the liquid-liquid binary parameters are needed to created predictions. The inputs that are specific for each case are the solute and the solvent or mixture of solvents and based on these inputs Aspen Plus searches for the binary parameters. For ternary systems the ternary diagram is created and evaluated because found experimental data is given in ternary diagrams and these are also more available.

3.1.1.2 Dynochem

The models NRTL and UNIFAC-LL are chosen for predictions using ternary systems. In Dynochem also the results are displayed in ternary diagrams with the same reason as for Aspen Plus. Only Aspen Plus is used and the reason for this is explained in the discussion part.

3.1.2 Results case 1: toluene & water

Figure 18 displays the obtained results for the solubility of toluene in water with the calculated values reported in appendix B.

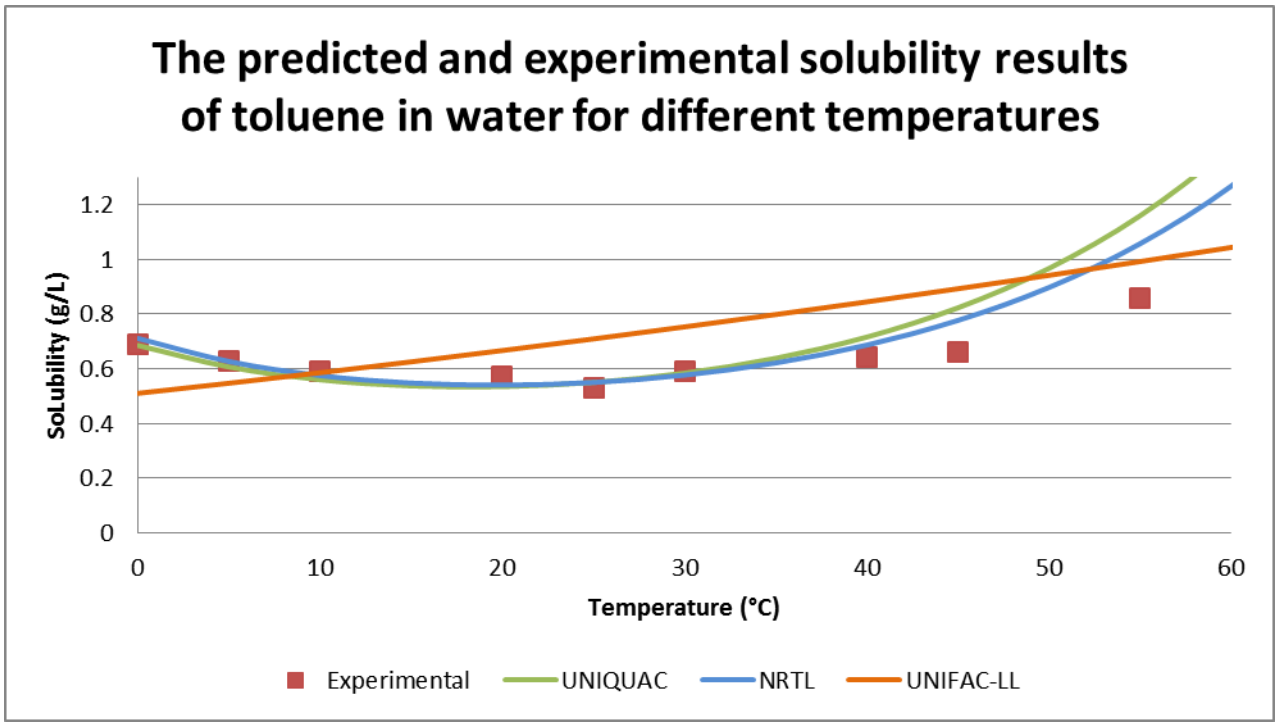

Figure 18: The predicted and experimental solubility results of toluene in water.

The experimental results are collected from experiments performed by D.G. Shaw [19]. The NRTL and UNIQUAC model both resulted in a good prediction for temperatures below 40°C. The average deviation between experimental and predicted results is for NRTL 3.5% and for UNIQUAC 4.6%. For temperatures above 40°C, the average deviation is 20% for NRTL and 30% for UNIQUAC. The UNIFAC-LL model results has an average deviation of 22% for the given temperature range between 0°C and 55°C. The obtained curve with this model is a straight line, which does not follow the experimental trend. The calculations for the deviations are based on the same principle used in the gas-liquid systems.

3.1.3 Results case 2: benzene & water

Figure 19 shows the obtained solubility results of benzene in water and the real values are added in appendix B.

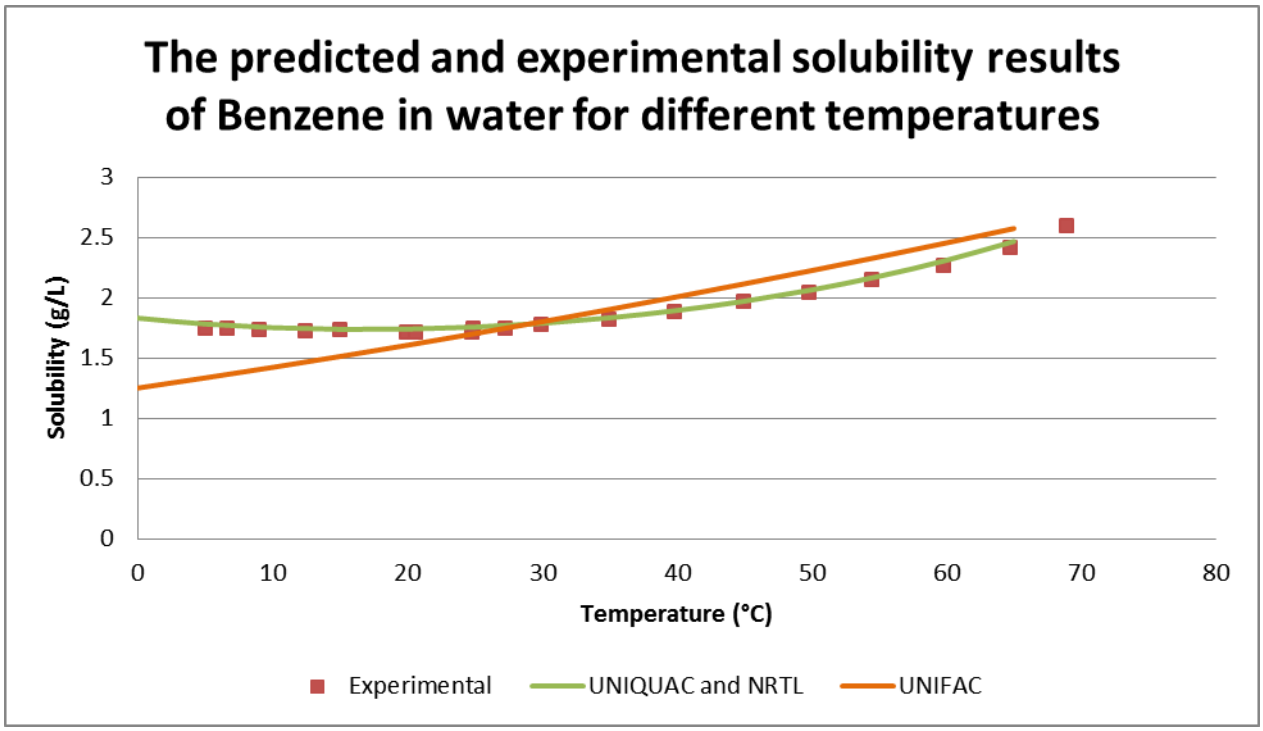

Figure 19: The predicted and experimental solubility results of benzene in water.

The experimental results are collected from experiments performed by D. S. Arnold et al [20]. In this figure, the UNIQUAC and NRTL results are identical to each other and have an average deviation between the experimental and predicted results of 1.3%. UNIFAC LL shows for temperatures below 25°Ca large difference of 18% exists between the predicted and experimental results. This is reduced to 6% for the higher temperature range.

3.1.4 Results case 3: dichloromethane & water

Figure 20 gives the solubility results of dichloromethane in water and the actual values are given in appendix B.

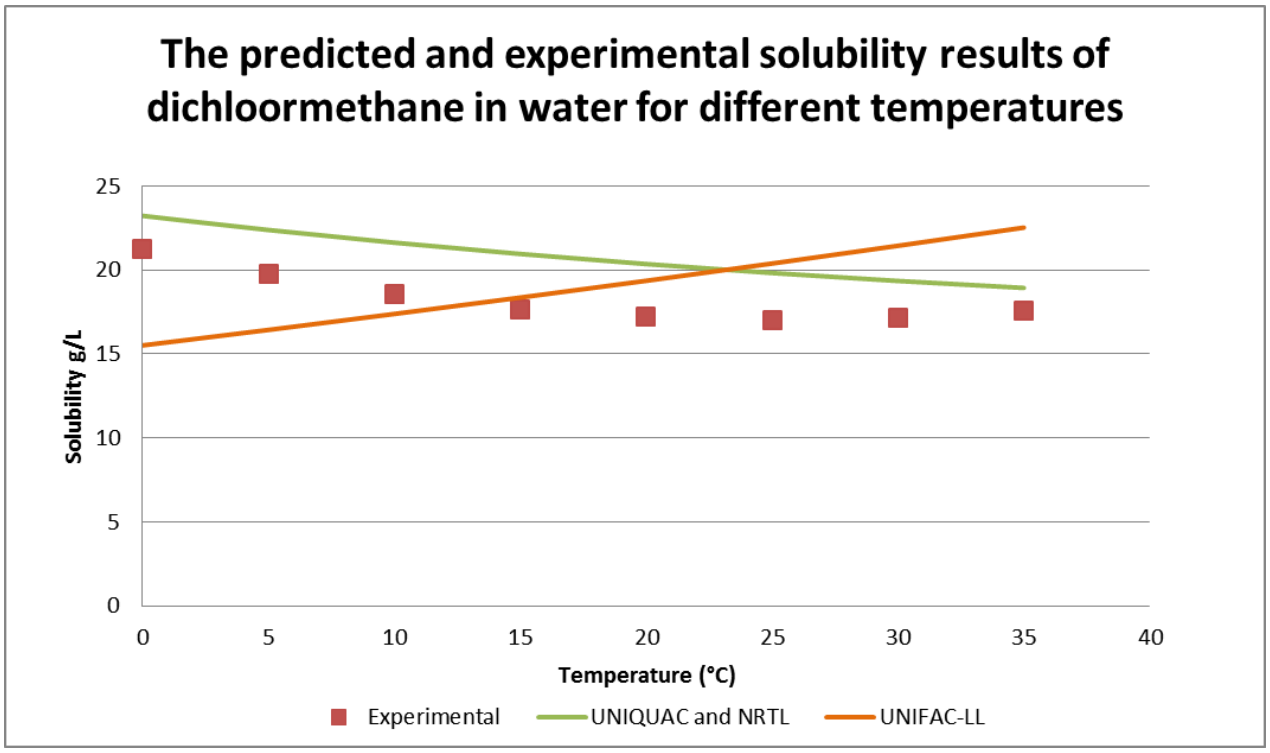

Figure 20: the predicted and experimental solubility results of dichloromethane in water.

The experimental results are collected from IUPAC-NIST Solubility Database [21]. The results for the NRTL and UNIQUAC model are equal to each other therefore only one line is visible. The average difference between the experimental results and NRTL or UNIQUAC model is 14%. The average deviation for UNIFAC LL model is 18% and the obtained curve does not follow the experimental trend.

3.1.5 Results case 5: 1-butanol & 3-methyl-1-butanol & water

Figure 21 displays the predicted and experimental results of the ternary diagram of 1-butanol & 3-methyl-1-butanol & water at 25°C in percentage mole fractions.

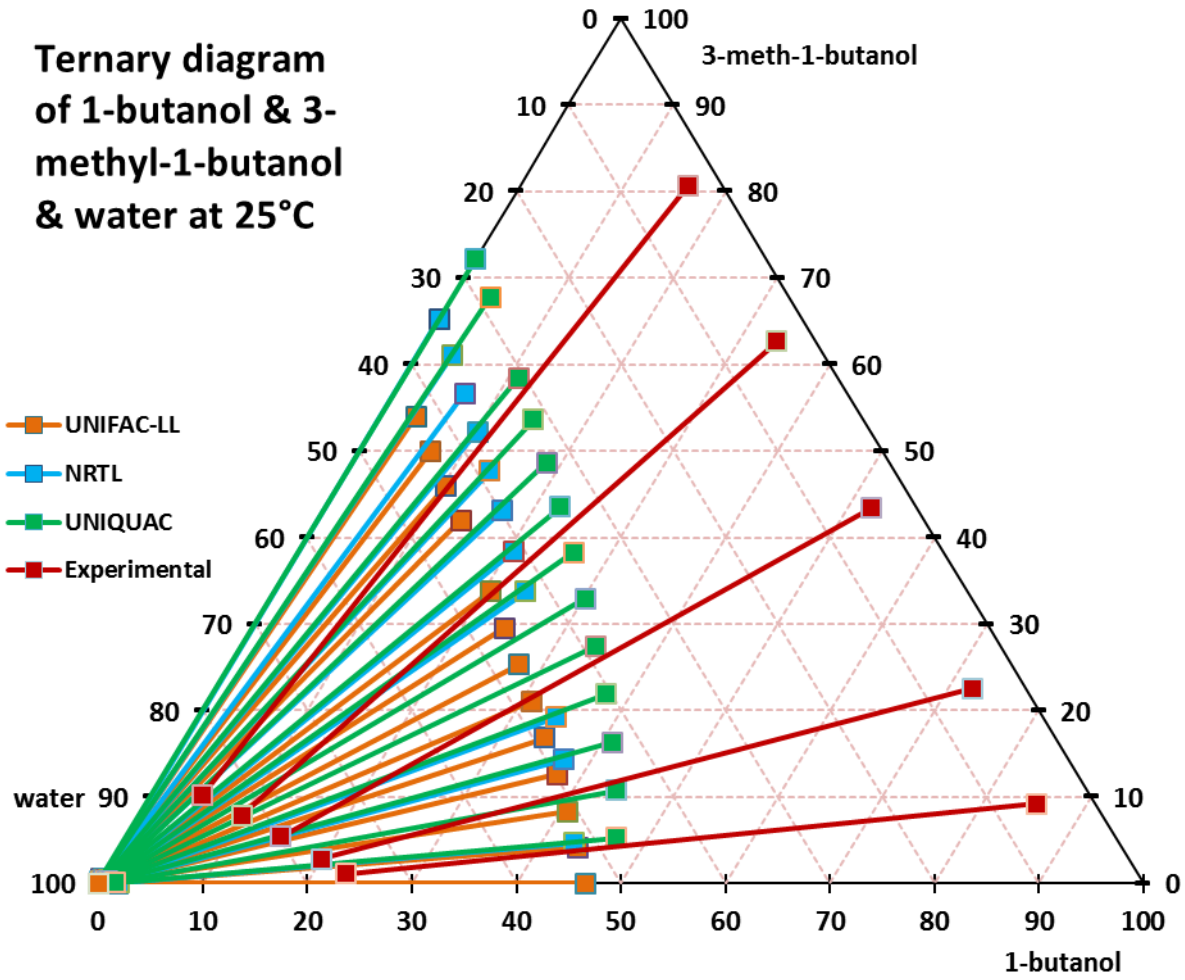

Figure 21: Ternary diagram of 1-butanol & 3-methyl-1-butanol & water for predicted and experimental results in percentage mole fractions.

The experimental results are collected from experiments performed by Zhaoyou Zhu et al [22]. The temperature for this ternary diagram is 25°C. The different compositions of the two existing phases are plotted and connected with tie lines. In this case there is a large difference between the experimental and predicted results. Predictions indicate that pure water is obtained as aqueous phase whereas the experimental results show compositions for the aqueous phase with 1-butanol between 10 to 23% and 3-methyl-1-butanol between 2 to 10%. The organic phase contains less water according to, the experimental data with water fraction around 5% where as predicted results gave water fractions between 30 to 50%.

3.1.6 Results case 6: 1-propanol & benzene & water

Figure 22 shows the predicted and experimental results for the ternary system of 1-propanol & benzene & water in percentage mole fractions.

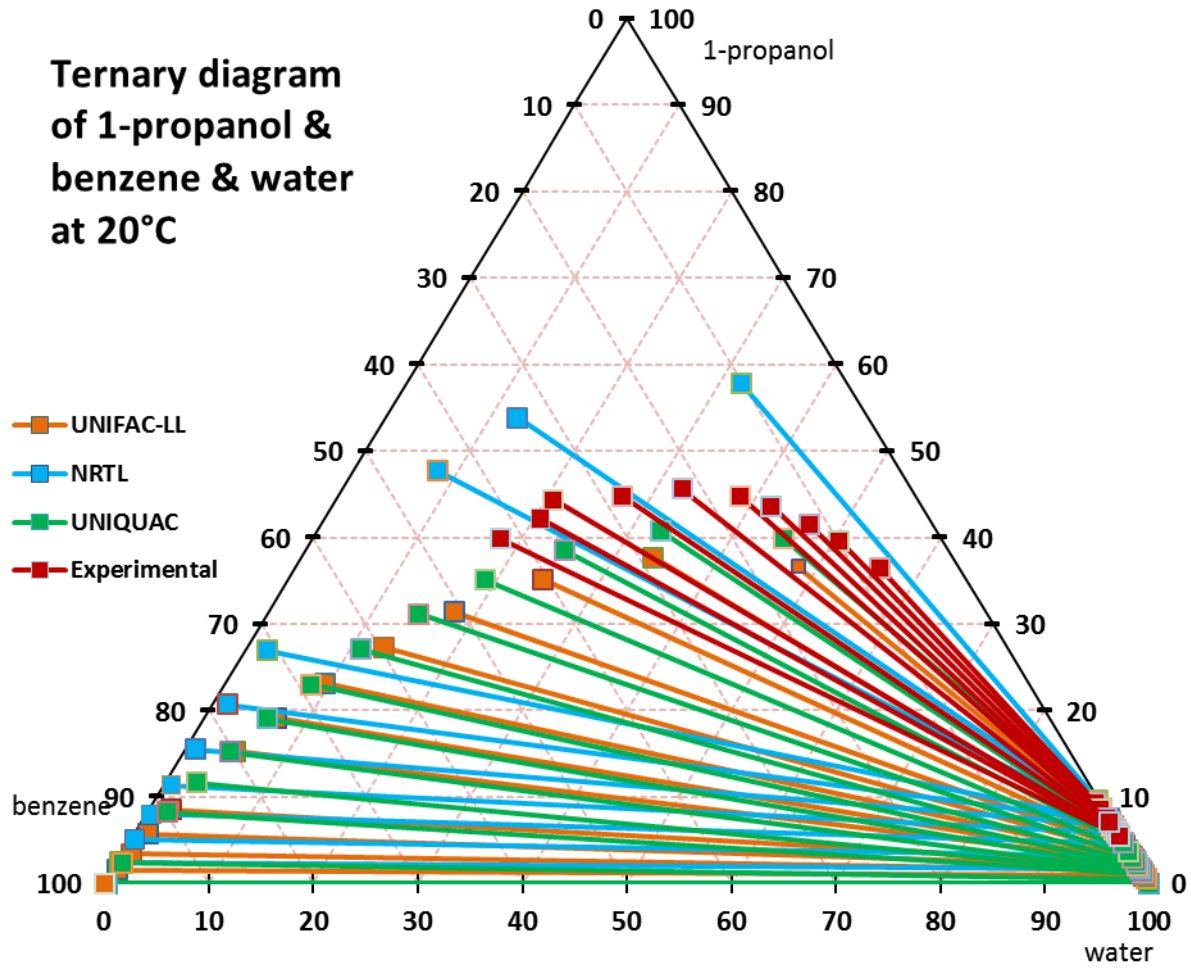

Figure 22: Ternary diagram of 1-propanol & benzene & water for predicted and experimental results in percentage mole fractions.

The experimental results are acquired from experiments performed by C. G. Denzler [23]. The temperature for this ternary diagram is 20°C because the experimental values found for this case are at 20°C. In this case, the deviations between experimental and predicted results are high for NRTL with an average deviation of 13%. The results for the UNIQUAC and UNIFAC-LL model are very similar; the average deviation between the two models is around 2%. The UNIQUAC model deviates the least from the experimental results with an average of ±5%. The calculations of the deviations are based on the difference between two values above each other . An example is the difference between the data point of NRTL with 57% of 1-propanol and the experimental data point with 45% of 1-propanol. The X coordinates of these points are both around 60.90 meaning that they are above each other and the difference between the Y coordinates is then the deviation.

3.1.7 Results case 6: n-hexane & methanol & cyclohexane

Figure 23 displays the obtained ternary diagram for n-hexane & methanol & cyclohexane at 20°C in percentage mole fractions.

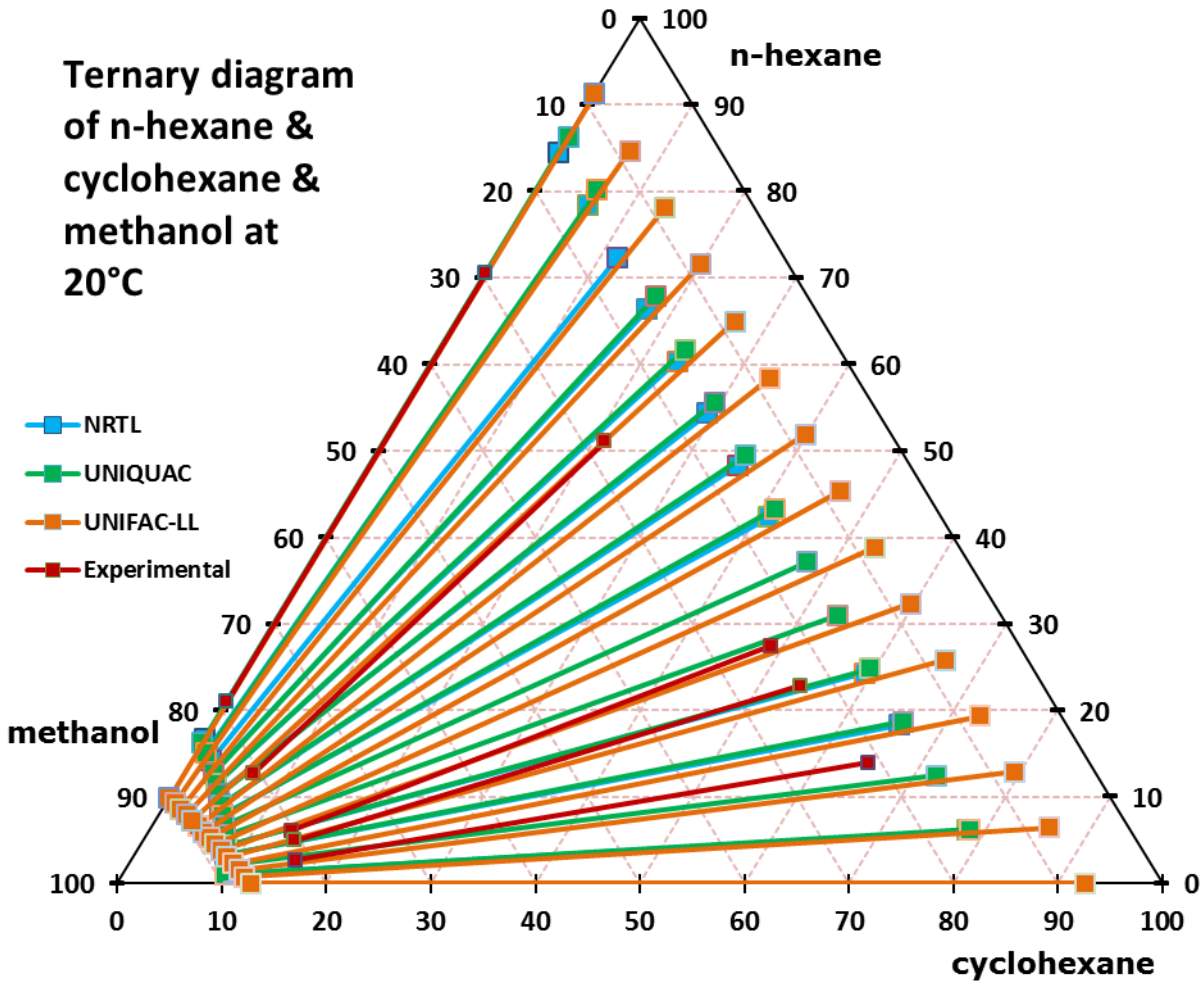

Figure 23: Ternary diagram of n-hexane & methanol & cyclohexane for predicted and experimental results in percentage mole fractions.

The experimental results are acquired from experiments performed by Paolo Alessi et al [24]. The temperature is kept constant at 20°C. The two liquid phases in this ternary diagram are a methanol rich phase and an organic phase containing more cyclohexane and n-hexane. The results for the NRTL and UNIQUAC model are similar and the deviation between the two is approximately 2%. The average deviation for the methanol rich phases between the experimental results and the NRTL or UNIQUAC model is 5%. UNIFAC-LL shows a bigger deviation of 9%. For the organic phase the average deviation from the experimental results for UNIFAC-LL (±15%) is higher than for the other two models (±7%).

3.1.8 Results case 7: ethanol & water & cyclohexane

Figure 24 shows the predicted and experimental ternary diagram for the ternary system of ethanol & water & cyclohexane at 25°C in percentage mole fractions.

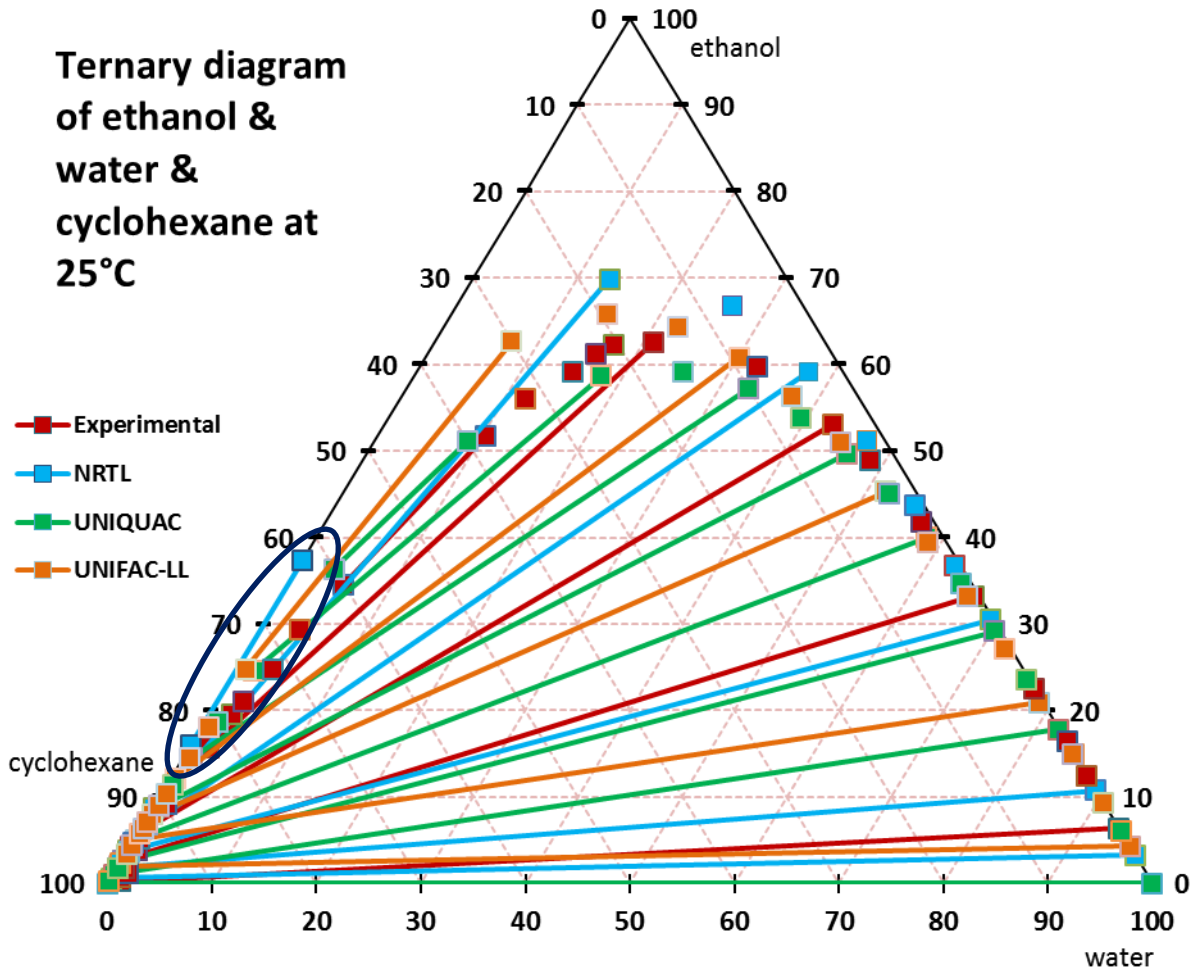

Figure 24: Ternary diagram of ethanol & water & cyclohexane for predicted and experimental results in percentage mole fractions.

The experimental results are acquired from experiments performed by Takashi Moriyoshi et al [25]. In this case, the average deviations between the predicted and experimental results are below 8% mole fraction for every model. The NRTL model assumes that there is an open miscibility gap for the components ethanol and cyclohexane, which not true according to the experimental results. An open miscibility gap is a miscibility gap where two binary systems form two separate phases when mixed. Normally a miscibility gap has only one binary system that forms two phases and by adding the third component, the two phases will disappear if a certain amount is added. However, for open miscibility gaps the third component forms also separate phases with one of the other components. The NRTL model assumes this for cyclohexane and ethanol as shown on figure 24 with the dark blue oval. The experimental results do not give tis for the cyclohexane and ethanol.

3.2 Discussion liquid-liquid

The discussion is built on the category 'data evaluation' of evaluation method described in the materials & method section 3.5. The requirements for the solubility predictions for liquid-liquid systems are the presence of the different substances in the database and the availability of models and their parameters related to the used compounds. Two software packages are available and usable but only Aspen plus was used as Dynochem uses exactly the same models e.g. the NRTL and UNIFAC-LL model but with a very limited binary parameter databank for NRTL (83 binary systems) compared to Aspen Plus (3000 binary systems). Finally, the components databank in Dynochem (44) is also limited compared to Aspen Plus (2500). The three models used for the investigation are NRTL, UNIFAC-LL and UNIQUAC. NRTL and UNIQUAC rely on the availability of the liquid-liquid binary parameters, UNIQUAC has two available databanks whereas NRTL has only one databank for finding parameters. The difference between the two data banks is that one is based on regressions of experimental data done by Aspen and the other one is based parameters found in the literature.

The accuracy of the predictions results from the deviations between the experimental and predicted results. In table 20 the deviations between the experimental and predicted results are given and the error on the experimental data. For the first four cases, the average difference is expressed in percentage and these calculations are based on the same method as used in the results of the gas-liquid systems Part 3 section 2.1.2. For the other case the maximum difference expressed in percentage mole fraction. The available experimental errors are given with their units.

Table 20: Listed differences between experimental and model data and the experimental error

Case	NRTL	UNIQUAC	UNIFAC-LL	Difference in	Experimental error	Units
1	3.5	4.6	22	%	/	g/L
2	1.3	1.3	6 to 18	%	0.03	g/L
3	14	14	18	%	3.7	g/L
4	40	35	45	%mole frac	0.1	% mole frac
5	13	5	13	%mole frac	0.3	% mole frac
6	7	7	15	%mole frac	0.05	% mole frac
7	8	8	8	%mole frac	0.1	% mole frac

Comparing the results of the seven cases, the UNIFAC-LL model shows the highest overall deviation between experimental and predicted results and UNIQUAC has the lowest. The explanation is that predictions with the UNIFAC-LL model have a poor accuracy for systems that contains water [5]. The prediction of the UNIFAC model are based on group interactions and these interactions are found through general evaluation of experimental data. Comparing NRTL and UNIQUAC the deviations are very similar for cases one, two, three, six and seven but for the remaining cases the deviations are more different from each other and in the advantage of the UNIQUAC model. This means that UNIQUAC is more suitable and the

reason is the limited availability of the required parameters for NRTL because of the availability of only one databank.

The first three cases are binary systems and for case three, the deviation is higher compared to the others. The reason is the experimental error, which is high as displayed in Table 20. Another aspect that can be evaluated for the binary systems is the trend followed by the predicted values. This is good for UNIQUAC and NRTL for every case but the UNIFAC model is not good for cases one, two and three because it creates only predictions following a straight line with a slope. The reason for this is explained in the discussion of the gas-liquid systems, Part 3 section 2.2, and is thus based on the used binary parameters. The explanation for UNIFAC-LL is the same as for the poor accuracy.

The last four cases are ternary systems and for case five, the deviation is high for every model as listed in table 20. The reason for this high deviation is that not all liquid-liquid parameter were used for every binary system since they were not available in Aspen. A solution to this problem is to fill out manually the correct parameters if they are found in the literature or be doing experiments. In this case, the parameters are available in the document with the experimental data [22] and they were filled in to achieve the same results as the paper but the results did not improve. This means that the parameters mentioned in the paper are not the correct and the reason is possibly the numbers were mixed when reported in the paper.

4 Solid-liquid systems

4.1 Cases solid-liquid

The different test cases are displayed in the table 21 beneath. For the first four cases, experimental data are available within Janssen Pharmaceutica and for the other cases no experimental results have been reported.

Table 21: List of different cases dived in solute and solvent used for solid-liquid systems.

Case	Solute	Solvent
1	thebaine	dipropyl ether
2	codeine	dipropyl ether
3	oripavine	dipropyl ether
4	morphine	dipropyl ether
5	thebaine	dimethyl ether
6	codeine	dimethyl ether
7	oripavine	dimethyl ether
8	morphine	dimethyl ether

4.1.1 Input solid-liquid cases

The practical method for the cases is explained in Part 2: Material & methods in section 3.4.2 Explanation solid-liquid systems. Tables 33 to 37 list the experimental input data that were used for AspenONE and Dynochem and can be found in appendix C. This consists of the structures and properties of the solid compounds (Table 33), the solubility of the solutes in the different solvents in g/L at 20°C (Table 34), and finally the solubility of specific solutes in a selective number of solvents at different temperatures (Tables 35 to 37). The data in these tables are experimental data acquired for Janssen Pharmaceutica. Solvents chosen in table 34 can be explain that the Regressed UNIFAC model needs experimental solubility data with groups related to the solvent where predictions is needed for as explained in the literature study in section 3.4.1.

The predicted results for solid-liquid systems are only created with Dynochem because the solubility modeler in AspenONE is not usable due to different problems. Another aspect is that Dynochem is much easier to use then the solubility modeler and more user-friendly.

4.1.2 Results case 1: thebaine & dipropyl ether

Figure 25 displays the obtained results for the solubility of thebaine in dipropyl ether and table 38 gives a list of the different combinations of experimental solubility data used for the predictions, which is added in appendix C

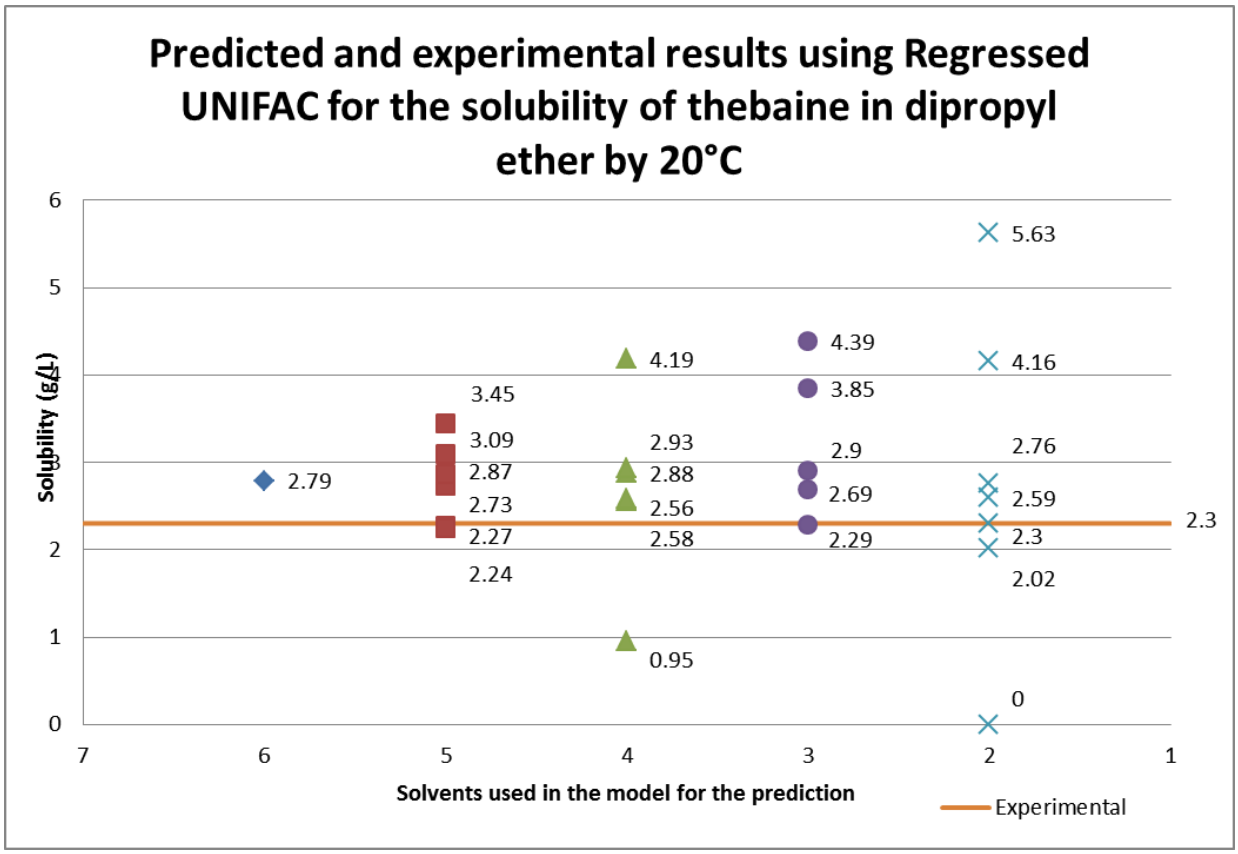

Figure 25: The predicted and experimental results for the solubility of thebaine in dipropyl ether at 20°C.

In figure 25 the results of the predictions for the solubility of thebaine in dipropyl ether are given with the different combinations of experimental data that was used in the Regressed UNIFAC for the prediction. This started with including every available experimental solvent data for thebaine with exception of the data for water and dipropyl ether. This means that there are six experimental points left to use for the model and the result of this prediction was 2.79g/L. Then only five solvents are taken to create the predictions this means that there are different combinations possible with the six usable experimental data points and thus also different results for the prediction. Next step is to include only four solvents in the creation of the prediction and this continued until only two solvents are used. The reason for this approach is to see if also accurate prediction can be created with the usage of less experimental data in the model. The usage of two solvents is the minimum that is needed to be used in the Regressed UNIFAC model as explained in the literature study section 3.4.1 Regressed UNIFAC. The deviation between experimental and predicted results is calculated again based on the method used in the results of gas-liquid systems. For 6 solvents the deviation is ±20% and for 5, 4, 3 and 2 the deviations depends on the combination of the

experimental data used for the prediction but by using two solvents a deviations can be achieved of less than 12% with a specific combinations.

4.1.3 Results case 2: codeine & dipropyl ether

Figure 26 shows the created results for the solubility of codeine in dipropyl ether using six experimental data points and different combinations of only two points in the Regressed UNIFAC model.

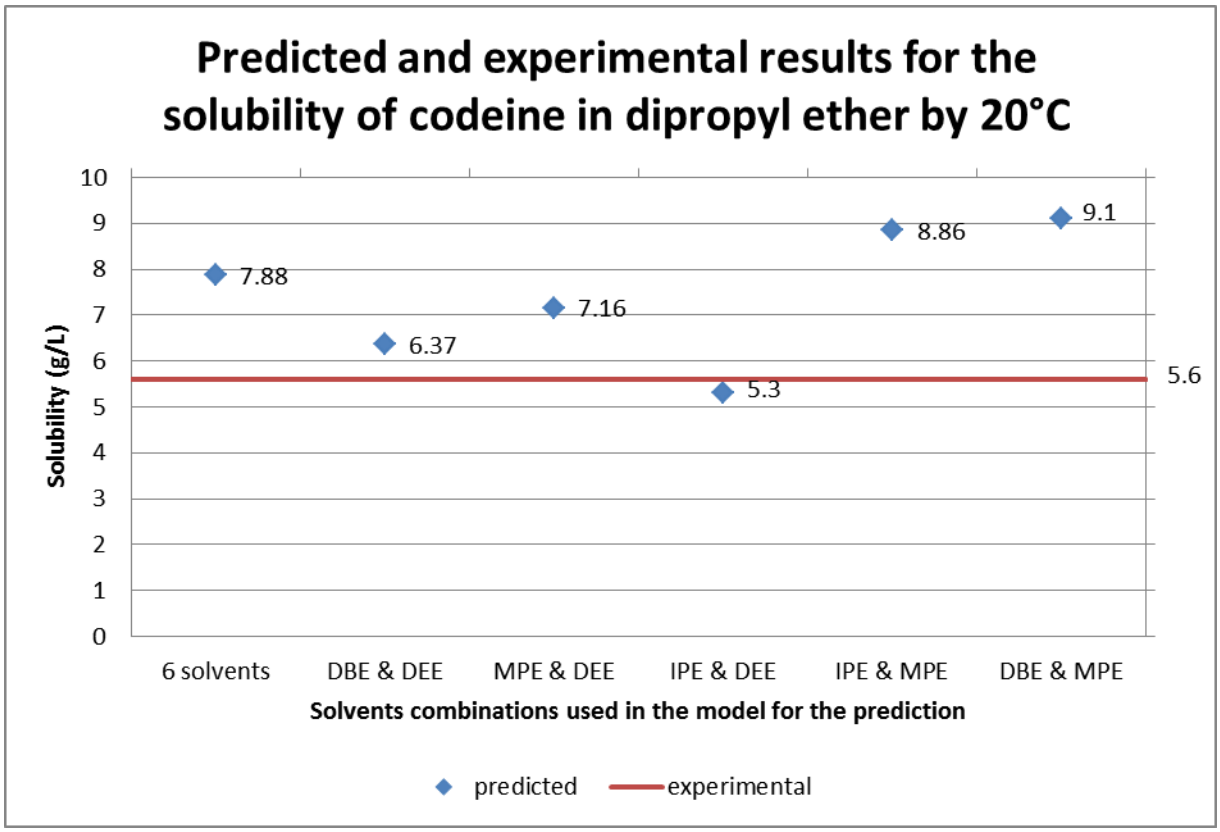

Figure 26: The predicted and experimental results for solubility of codeine in dipropyl ether at 20°C and the abbreviations are DBE=dibutyl ether; DEE=diethyl ether; MPE=methyl propyl ether and IPE= isopropyl ether.

Only the results for the usage of six and different combinations of two solvents in the model are displayed because in the previous case accurate predictions were created with only two experimental data points and the results of with the usage of six points is include as comparison. The used combinations for two solvents are related to the structure of dipropyl ether. The deviations between experimental and predicted values is for six solvents is ±40%, which is high compared to specific combinations of two solvents such as dibutyl ether & diethyl ether = ±14%, methyl propyl ether & diethyl ether = ±28% and isopropyl ether & diethyl ether = ±5%.

4.1.4 Results case 3: oripavine & dipropyl ether

Figure 27 displays the created results for the solubility of oripavine in dipropyl ether using six experimental data points and different combinations of only two points in the Regressed UNIFAC model.

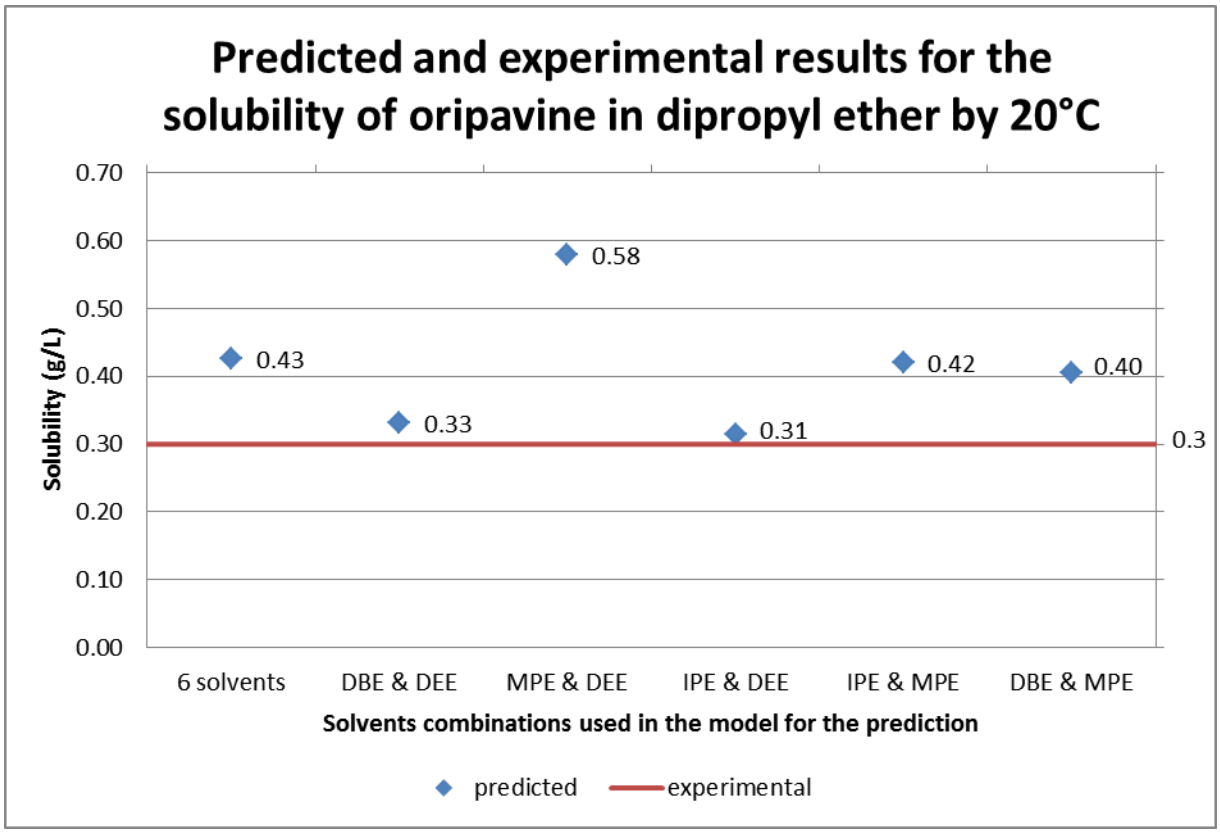

Figure 27: The predicted and experimental results for solubility of oripavine in dipropyl ether at 20°C and the abbreviations are DBE=dibutyl ether; DEE=diethyl ether; MPE=methyl propyl ether and IPE= isopropyl ether.

The same method is used for this case as for case two.The deviation between experimental and predicted results is for six solvents, approximately 42%, higher than specific combinations with two solvents such as dibutyl ether & Diethyl ether = ±10% and isopropyl & diethyl ether = ±5%.

4.1.5 Results case 4: morphine & dipropyl ether

Figure 28 shows the obtained results for the solubility of morphine in dipropyl ether using six experimental data points and different combinations of only two points in the Regressed UNIFAC model.

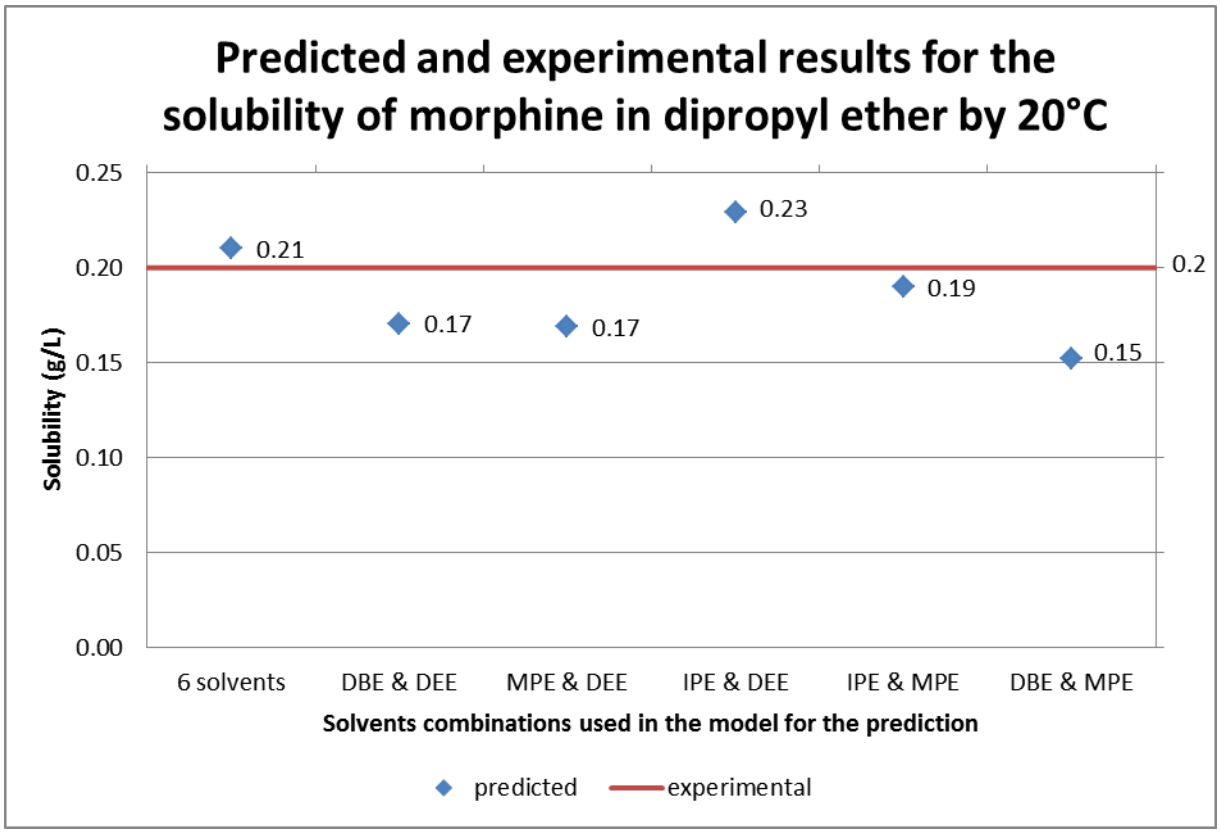

Figure 28: The predicted and experimental results for solubility of morphine in dipropyl ether at 20°C and the abbreviations are DBE=dibutyl ether; DEE=diethyl ether; MPE=methyl propyl ether and IPE= isopropyl ether.

For this case, also the same method was used as explained in case one. The deviations between the experimental and predicted results are lower with the usage of six solvents (±5%) in the model than for most combinations of two solvents, an average of 20%, with an exception for isopropyl ether & methyl propyl ether, which is ±5%.

4.1.6 Combined results of the last four cases

Cases five to eight have no experimental results because the measurements are difficult to perform and there is no information in the literature. The reason for adding this is that when solubility data is required there is no information is available and then the results of predictions need to be evaluated based on other criteria. The evaluation can then be based on previous predictions were comparison was made with experimental data. There are two different approaches executed for this problem. The first method is creating a prediction with the combination of experimental data from two other solvents. The two solubility data chosen for this prediction are these of water and diethyl ether because the structure of dimethyl ether lies between water and diethyl ether. The second method is by looking at other linear ether such as diethyl ether, dipropyl ether and dibutyl ether and doing a regression with a power function based on the molecular weight for the four linear ethers. The experimental data for these methods is available and listed in tables 34 , 35 and 36 and are added in appendix C.

Figure 29 presents the gathered results for the solubility of different narcotics in dimethyl ether using method one and the actual values are listed in table 39 in appendix C.

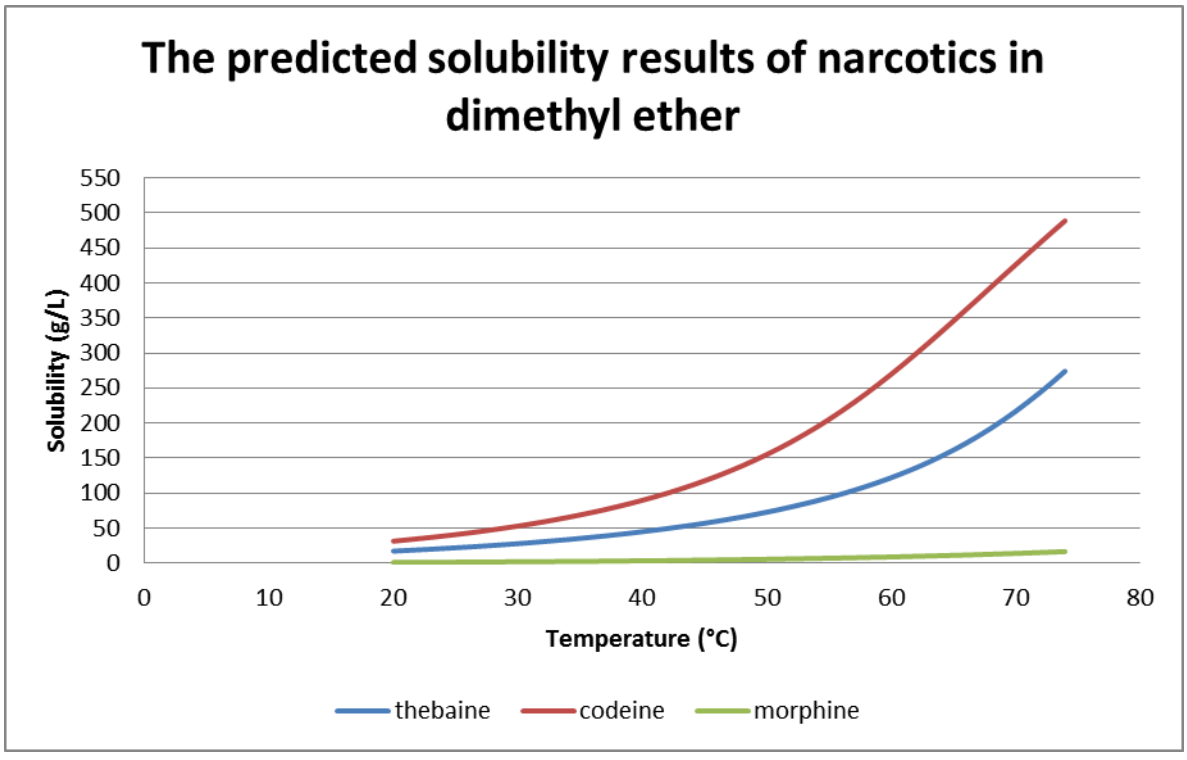

Figure 29: The predicted solubility results of different narcotics in dimethyl ether with method 1.

The predictions are created based on only the experimental solubility data of the narcotics in water and diethyl ether by 20°C. oripavine is not included because no solubility data was available for water. The predicted values for the narcotics start below 35g/L for 20°C and for codeine en thebaine the solubility increases fast to high values compared to morphine. The predictions for higher temperatures is not recommended because the experimental data

used in the prediction was only for 20°C and extrapolating the results could lead into bigger deviations if the model cannot handle it. This is further explained in the discussion.

Figure 30 displays the collected results using method two for the solubility of the different narcotics in dimethyl ether and the actual values are given in table 40 in appendix C.

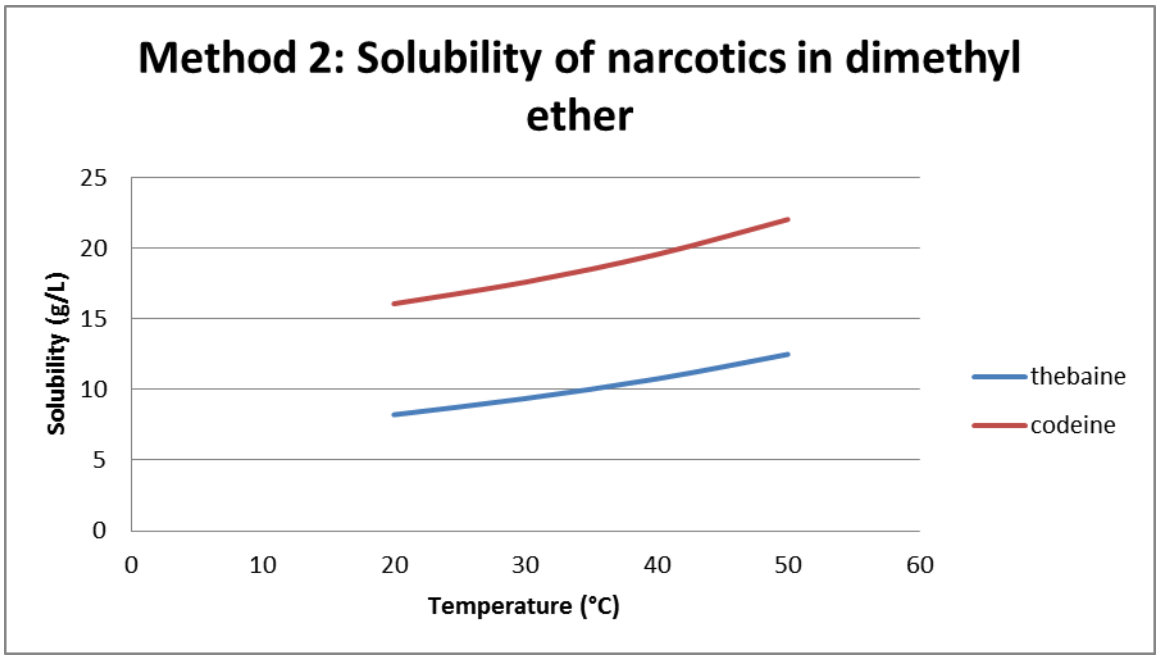

Figure 30: The predicted solubility results of different narcotics in dimethyl ether with method 2.

This method is only applied on thebaine and codeine because there is missing experimental data for the morphine and oripavine. Comparing both methods, the predicted results are for codeine again higher than for thebaine but the results are much lower than in method one. Using method one the result for thebaine is 17.5g/L compared to 8g/L with method two and for codeine this is 31.6g/L with method one and 16g/L for method two. Only the comparison for 20°C is made because extrapolating is not recommended for the Regressed UNIFAC model.

4.2 Discussion solid-liquid

The discussion is based on the category 'data evaluation' of evaluation method described in the materials & method section 3.5. The different criteria in these category are 'the requirements', 'the accuracy of the experiments' and 'the accuracy of the predictions'.

The requirements for the Regressed UNIFAC model of Dynochem are the different solute specifications and a minimum of two experimental solubility data points. These need to be available to create predictions with the model. In this case, the necessary input for the model was available in the literature and is listed in the tables33 to 37. Some of the solute specification can be include in the fit of the model as explained in 'Part 3 materials & methods section 3.4.2' but this means also that more experimental solubility data are required.

The models use experimental data to create predictions. Therefore it is important that these data are accurate for good predictions. In this case, the error on experimental data was not available, which needs to be taken into account.

The accuracy of the predictions depends thus on the accuracy of the experiments but also on the combination of the different experimental solvent data used in the predictions. The first case was to determine the minimum of required experimental data used in the prediction and taking in account the accuracy of the predictions. The results are that a combination of two data points can also deliver accurate predictions. Using more than two data points can deliver also accurate predictions with specific combinations but the goal was to limit the used experimental data points .The following cases two, three and four are then to determine which combination delivers the best results with the least deviation between the experimental and predicted results. Table 22 gives a list with the different combinations used for the predictions and also the deviation between the experimental and the predicted results.

Table 22: The deviation (%) between experimental and predicted values for the different narcotics with the used combinations for the prediction.

Used in the prediction		Thebaine	Codeine	Oripavine	Morphine
Dibutyl ether	Diethyl ether	20	14	10	15
Methyl propyl ether	Diethyl ether	0	28	93	16
Isopropyl ether	Diethyl ether	12	5	5	15
Isopropyl ether	Methyl propyl ether	13	58	40	5
Dibutyl ether	Methyl propyl ether	43	63	35	24

The combination of isopropyl ether and diethyl ether delivers the best predictions in general if every narcotic is taken into account. The second best is the combination dibutyl ether and diethyl ether. The reason that diethyl ether & isopropyl ether give generally the best prediction give is that isopropyl and diethyl ether are the closest relatives of dipropyl ether. The second best can be explained because in the homologue series of ethers with the same groups dipropyl ether lies between diethyl and dibutyl ether. The solubility of the narcotics

increases with decreasing size of the groups on the ether. For example in thebaine the solubility of diethyl ether is 4.2, dibutyl ether is 2.1 and dipropyl ether is 2.3 thus the solubility of thebaine in dipropyl ether lies between the two other values. This is also valid for the other narcotics.

The results of case five, six, seven and eight are evaluated based on the two methods because no experimental results are available. Comparing method one with method two leads to the conclusions that the results of both methods deviate a lot from each other. The deviation between the two methods is ±50% for a temperature of 20°C. The first reason to explain this is the usage of water in the first method. Water is described in Dynochem as one group and is thus not divided in the different groups such as dimethyl and diethyl ether. Another reason that can explain this difference is the reliability of the second method. In the second method, predictions for different temperatures are created with Dynochem for the dipropyl ether, diethyl ether and dibutyl ether based on two experimental values at 20°C and 30°C. The important aspect is inter- and extrapolation capabilities of the model. This is tested based on the experimental data in table 35 by using only two values and predicting the others. The results are shown in the next figures (Figure 31 and Figure 32).

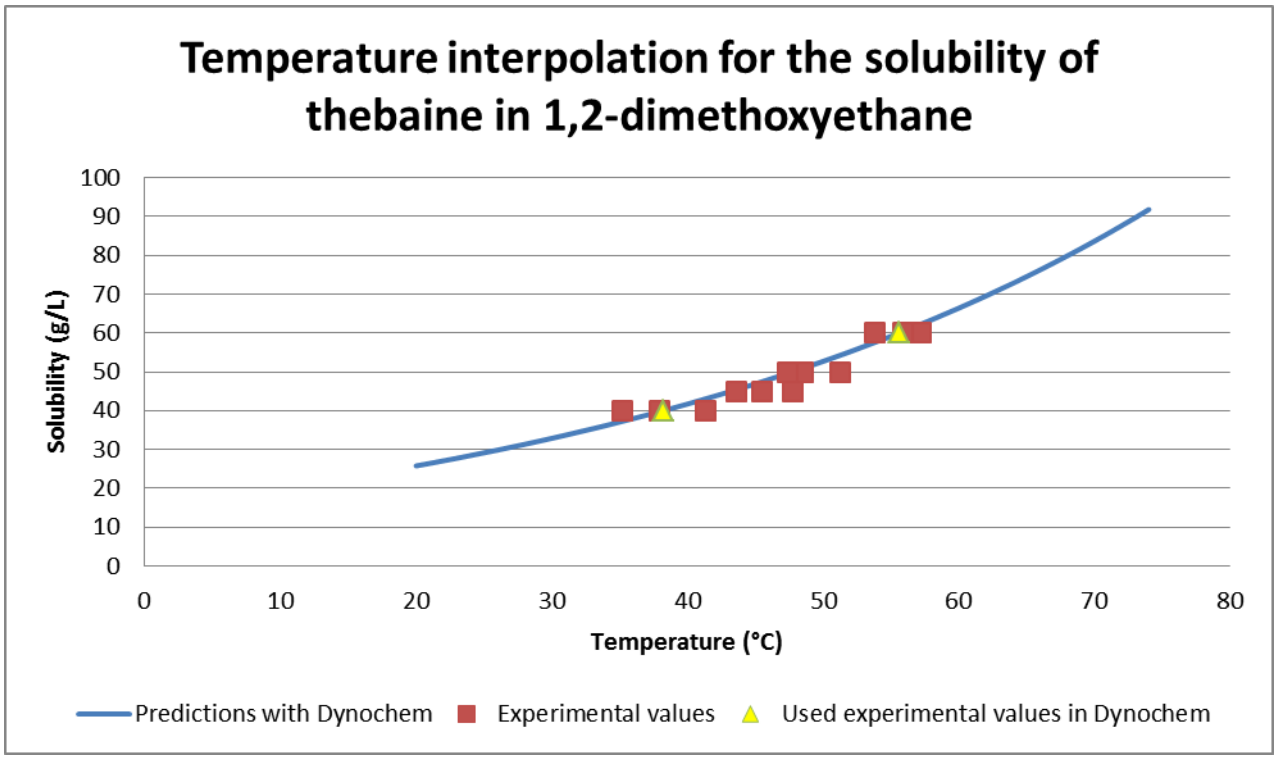

Figure 31: Temperature interpolation for the solubility of thebaine in 1,2-dimethoxyethane with the Regressed UNIFAC model.

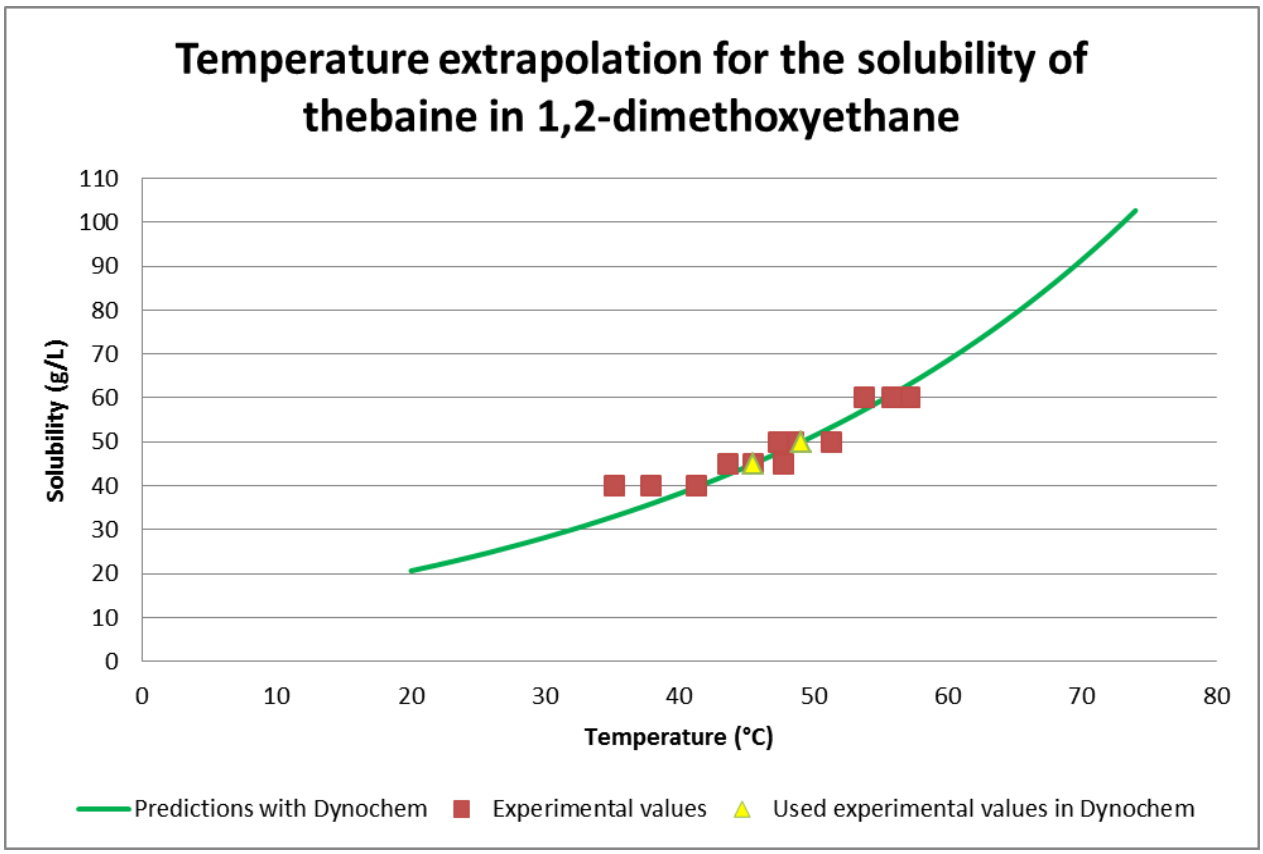

Figure 32: Temperature interpolation for the solubility of thebaine in 1,2-dimethoxyethane with the Regressed UNIFAC model

Based on the previous figures (Figure 31 and Figure 32) it is concluded that interpolation is usable and that extrapolation results have to be evaluated with care before using them.

The problem of the second method is that no second experimental data points were available for diethyl and dibutyl ether. This was solved by creating artificial second points by assuming that they have the same slope as between the two points of dipropyl ether. This is an assumption and if the slope change for example with 10% the predicted results for dimethyl ether will also, change as shown in the next figure (33). The conclusion is that the other experimental data are needed to avoid this error.

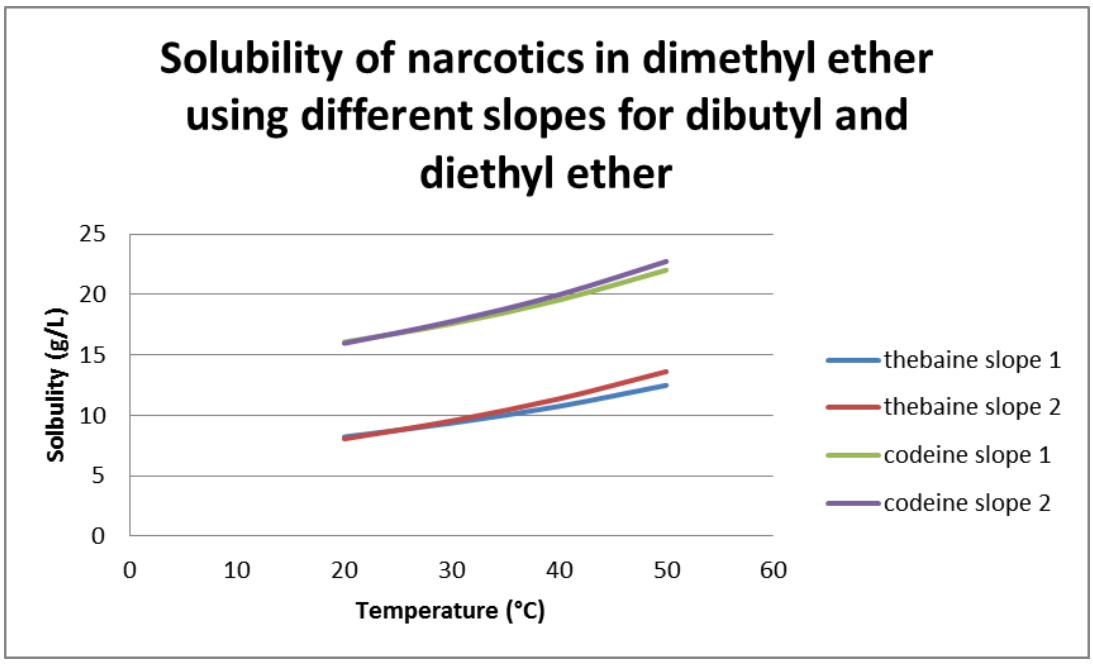

Figure 33: The different slopes between the experimental and artificial data point in method 2

The conclusion that can be made based on the differences between the two methods is that method one uses less experimental data points than the second method. But problems with the first method are the limited temperature range and the usage of the water component in this specific prediction. The limited temperature range can be solved by using more experimental data points from literature data for different temperatures in the model. The second method is possibly better for this specific problem because it requires no experimental values for water in the model but then the other data needs to be available.

Part 4: General conclusion

In this part, a general conclusion is made based on the evaluation method and the discussion of gas-liquid, liquid-liquid and solid-liquid systems.

1 Gas-liquid

The predictions for gas-liquid systems are all done by using AspenPlus and the overall conclusion is that predictions for gasses in water are satisfying if the binary parameters for henry are available, which is the case for most gasses in water. The average error between the predicted and experimental results is ±15%, which is high .Predictions for gasses in other solvents than water are possible if the Henry parameters are added in manually. The most time consuming step is to fill out all necessary input e.g. defining of compounds, selecting the model, creating a flow sheet and doing the sensitivity analyses. This depends on the user's knowledge and the availability of a manual because Aspen Plus is an extensive program with a broad range of applications possibilities. The time that the model uses actually to calculate and predict is negligible. The wide varieties of application possibilities makes the interface of Aspen Plus a bit overwhelming for an unexperienced user.

2 Liquid-liquid

In case of liquid-liquid systems, both programs are capable of creating predictions. The difference between the programs is that Dynochem has a less extensive database to its disposal compared to Aspen Plus but Dynochem is easier and faster to use because it uses separate files for different applications meaning that it is more obvious to use even for an unexperienced user. Another difference is that Dynochem has a limited amount of models available for liquid-liquid systems only NRTL and UNIFAC. Therefore, only the results of Aspen Plus are included and the overall conclusion is that the predictions for binary and ternary systems are good for NRTL and UNIQUAC if the liquid-liquid binary parameters are available. The average deviation between the predicted and experimental results is ±5%, which is acceptable. The UNIQUAC model has more available parameters therefore this is model is preferred. The most time consuming steps are again the required inputs.

3 Solid-liquid

Theoretical both AspenONE and Dynochem are capable of creating predictions for solid-liquid systems but only Dynochem was used. Dynochem is again faster and easier than the solubility modeler of AspenONE. There are different reasons for this. The first is that Dynochem uses only one file whereas AspenONE requires two files. The second reason is that Dynochem requires a minimum of two experimental solubility data compared to four for AspenONE. The third is that there occurred problems with the solubility modeler that working with it was impossible and Dynochem had no problems. The required time to calculate the prediction is negligible. Predictions are created with two experimental solubility data points for two different solvents that are closely related to the solvent that needs prediction. Through the usage of a combination of two solvents that are closely related the average deviation of the predicted and experimental results is ±9%.

4 Decision tree

Figure 34 shows the decision tree to decide when modeling is possible and when other options need to be considered.

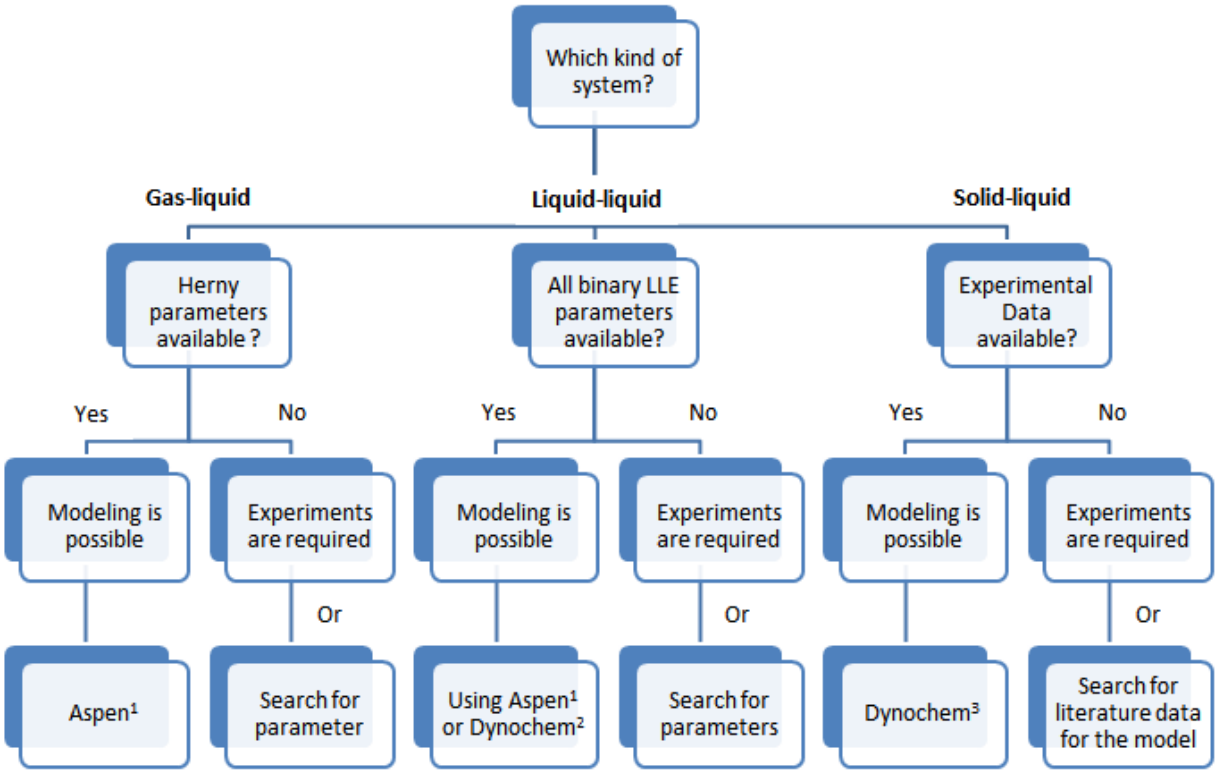

Figure 34: The modeling decision tree

First an important comment is that if the system is chosen also is checked if the desired components are available in the software package. Aspen has a very large databank with different components which means it is unlikely that the wanted components are unavailable. Dynochem has a smaller databank of available components thus here it can be possible that they are not available. In both programs components can be added but then more information about the compound is needed.

Secondly the numbers used in the decision tree are explained

- Number 1: refers to the fact that using Aspen requires some knowledge about how to get the desired predictions for the solubility, which can be solved by following a manual;
- Number 2: refers to the fact that Dynochem has a limited data bank for liquid-liquid binary parameters for NRTL;
- Number 3: refers to the fact that the combination of the solvents is important to create accurate predictions. The combination used for the prediction needs to be closely related to the solvent wherefore prediction is needed.

An example for 3 is that prediction is needed for dipropyl ether with experimental solubility data available for two closely related solvents e.g. isopropyl ether and diethyl ether. In that case it is recommended to use those two for the Regressed UNIFAC model. Another option to obtain related data is to look for solvents that lie in the same homologue series as the solvent wherefore prediction is needed. An example for dipropyl ether is the usage of dibutyl ether and diethyl ether.

Thirdly is when the second question leads to the answer ‘no’ then there are always two options. The first option is doing the required experiments or sending samples to a laboratory. The second is to search for the parameters or data for the model in the literature. In this last option the user needs to know how to fill in and where to possibly find them and this can be time consuming and lead to no information if nothing was found.

The information needed for the gas-liquid and liquid-liquid systems can be found in documents where the same system and same model is used. This is found in scientific literature which can be found using scientific search engines such as Google scholar. The experimental data for the solid-liquid systems can be found in compound databases such as Chemspider [16] and technical factsheets of the compounds.

Finally, the requested accuracy and specific temperature range of the data may also influence the decision between modelling and experiments as models will give in most cases only an approximation with a certain deviation which often increases drastically outside the models temperature range.

References

[1] Janssen Pharamaceutica, "Our Story: Janssen Pharmaceutica," 6 12 2013. [Online]. Available: http://en.janssenpharmaceutica.be/.

[2] United States Environmental Protection Agency, "Technical Factsheet on: BENZENE," 4 Januari 2010. [Online]. Available: http://water.epa.gov/drink/contaminants/basicinformation/historical/upload/Archived-Technical-Fact-Sheet-on-Benzene.pdf. [Accessed April 2014].

[3] I. Martinez, "THERMODYNAMICS OF SOLUTIONS," 1995-2014. [Online]. Available: http://webserver.dmt.upm.es/~isidoro/. [Accessed April 2014].

[4] S. I. Sandler, Chemical, Biochemical, and Engineering Thermodynamics Fourth edition, John Wiley & Sons, 2006.

[5] G. M. Kontogeorgis and G. K. Folas, Thermodynamic Models for Industrial Applications, Chichester: John Wiley & Sons, 2010.

[6] E. C. Carlson, "Don't Gamble With Physical Properties For Simulations," *CHEMICAL ENGINEERING PROGRESS,* pp. 35-46, 1996.

[7] S. K. Ghosh and S. J. Chopra, "Activity Coefficients from the Wilson Equation," *Ind. Eng. Chem.,Process Des. Dev.,Vol 14,* pp. 304-307, 1975.

[8] H. Renon and J. M. Prausnitz, "Local Compositions in Thermodynamic Excess Functions for Liquid Mixtures," *AlChE Journal,* pp. 135-144, 1968.

[9] J. P. O'Connell and J. M. Haile, Thermodynamics, New York: Cambridge University Press, 2005.

[10] H.-H. Tung, J. Tabora, N. Variankaval, D. Bakken and C.-C. Chen, "PHARMACEUTICS, PREFORMULATION AND DRUG DELIVERY," *JOURNAL OF PHARMACEUTICAL SCIENCES,* pp. 1813-1820, 2008.

[11] D.-B. Ren, Z.-H. Yang, Y.-Z. Liang, Q. Ding, C. Chen and M.-L. Ouyang, "Correlation and prediction of partition coefficient using nonrandom two-liquid segment activity coefficient model for solvent system selection in counter-current chromatography separation," *Journal of Chromatography A,* pp. 10-18, 2013.

[12] D. S. Abrams and J. M. Prausnitz, "Statistical Thermodynamics of Liquid Mixtures: A New P txpression for the Excess Gibbs Energy of Partly or Completely Miscible Systems," *AlChE Journal,* pp. 116-127, 1975.

[13] A. A. Bondi, Physical Properties of Molecular Crystals, Liquids, and Glasses, New York: John Wiley & Sons, 1968.

[14] A. Fredenslund, R. L. Jones and J. . M. Prausnitz, "Group-Contri bution Estimation of Activity Coefficients in Nonideal Liquid Mixtures," *AlChE Journal,* pp. 1086-1099, 1975.

[15] Dynochem, "Dynochem Resources," 7 August 2012. [Online]. Available: http://www.scale-up.com/.

[16] Royal Society of Chemistry, "Chemspider: Search and share chemistry," 2014. [Online]. Available: http://www.chemspider.com/.

[17] F. Chen, D. L. Freedman, R. W. Falta and L. C. Murdoch, "Henry's law constants of chlorinated solvents at elevated temperatures," *Chemosphere,* p. 156–165, 2012.

[18] W. J. De Bruyn and E. S. Saltzman, "The solubility of methyl bromide in pure water, 35% sodiumchloride and seawater," *Marine Chemistry,* pp. 51-57, 1997.

[19] D. Mackay, W. Y. Shiu, K.-C. Ma and S. C. Lee, Physical-Chemical Properties and Environmental Fate for Organic Chemicals, Boca raton: CRC Press, 2006.

[20] D. S. ARNOLD, C. A. PLANK, E. E. ERICKSON and F. P. PIKE, "Solubility of Benzene in Water," *Industrial and engineering chemistry,* pp. 253-256, 1958.

[21] Measurement Services Division of the National Institute of Standards and Technology, "Solubility System: Dichloromethane with Water," 2007. [Online]. Available: http://srdata.nist.gov/solubility/sol_detail.aspx?goBack=Y&sysID=60_242.

[22] Z. Zhu, Y. Liu and Y. Wang, "Liquid–liquid equilibrium for the ternary system of 1-butanol + 3-methyl-1-butanol + water at different temperatures," *Fluid Phase Equilibria,* pp. 14-19, 2012.

[23] C. G. Denzler, "Liquid-liquid equilibrium data," *Chemical engineering,* pp. 358-365, 1945.

[24] P. Alessi, M. Fermeglla and I. Kikic, "Liquid-liquid equilibrium of cyclohexane-n-hexane-methanol mixtures: Effect of water content," *J. Chem. Eng. Data,* pp. 236-240, 1989.

[25] T. Moriyoshi, Y. Uosaki, K. Takahashi and T. Yamakawa, "(Liquid + liquid) equilibria of (water + ethanol + cyclohexane) at the temperatures 298.15 K and 323.15 K," *j. Chem. Thermodynamics,* pp. 37-42, 1991.

List of appendixes

Appendix A: Overview of experimental results for the gas-liquid systems

Table 23: The predicted solubility values for methyl chloride in water.

NRTL with Henry			
Temperature	**Methyl chloride**	**Water**	**Solubility**
°C	**kg/h**	**kg/h**	**g/L**
0	0.131	9.98	13.1
5	0.105	9.97	10.6
10	0.0855	9.96	8.58
15	0.0703	9.94	7.08
20	0.0585	9.91	5.90
25	0.0491	9.88	4.97
30	0.0416	9.84	4.23
35	0.0355	9.79	3.63
40	0.0304	9.71	3.13
45	0.0261	9.62	2.71
50	0.0224	9.49	2.36
55	0.0191	9.33	2.05
60	0.0161	9.10	1.77
65	0.0134	8.80	1.53
70	0.0108	8.37	1.30
75	0.0083	7.74	1.08
80	0.0059	6.76	0.867
85	0.0033	5.07	0.654
90	0.0007	1.57	0.436
95	0	0	/
100	0	0	/

Table 24: The experimental solubility values for methyl chloride in water with the standard deviation.

Experimental		
Temperature	**Solubility**	**SD**
°C	**g/L**	**%**
8	0.942	18.9
24	0.240	4.82
38	0.673	13.5
58	0.189	3.80
78	1.46	29.4
93	0.792	15.9

Table 25: The predicted solubility values for methyl bromide in water.

NRTL with Henry			
Temperature	**Methyl bromide**	**Water**	**Solubility**
°C	**kg/h**	**kg/h**	**g/L**
0	0.415	9.99	41.5
5	0.329	9.98	33.0
10	0.263	9.98	26.4
15	0.213	9.97	21.4
20	0.175	9.95	17.5
25	0.145	9.94	14.5
30	0.121	9.92	12.2

Table 26: The experimental solubility values for methyl bromide in water.

Experimental	
Temperature	**Solubility**
°C	**g/L**
1.1	46.2
5	36.3
10	28.4
13.4	24.9
18	21.6
19.4	20.7
21.2	19.2
25	15.9
30	14.2

Appendix B: Overview of experimental results for the liquid-liquid systems

Table 27: The predicted solubility values for toluene in water.

	NRTL			UNIQUAC			UNIFAC-LL		
T	Toluene	Water	Solubility	Toluene	Water	Solubility	Toluene	Water	Solubility
°C	kg/h	kg/h	g/L	kg/h	kg/h	g/L	kg/h	kg/h	g/L
0	0.00711	10.0	0.711	0.00685	10.0	0.685	0.0051	10.0	0.510
5	0.00627	10.0	0.627	0.00606	10.0	0.606	0.00547	10.0	0.547
10	0.00575	10.0	0.575	0.00559	10.0	0.560	0.00585	10.0	0.586
15	0.00547	10.0	0.548	0.00537	10.0	0.537	0.00625	10.0	0.626
20	0.00540	10.0	0.540	0.00534	10.0	0.534	0.00666	10.0	0.667
25	0.00549	10.0	0.550	0.0055	10.0	0.550	0.00709	10.0	0.709
30	0.00576	10.0	0.577	0.00584	9.99	0.584	0.00753	10.0	0.753
35	0.00621	10.0	0.621	0.00638	9.99	0.638	0.00798	9.99	0.799
40	0.00686	10.0	0.687	0.00716	9.99	0.716	0.00845	9.99	0.845
45	0.00776	10.0	0.777	0.00823	9.99	0.823	0.00893	9.99	0.893
50	0.00897	10.0	0.898	0.00967	9.99	0.968	0.00942	9.99	0.942
55	0.0106	10.0	1.06	0.0116	9.99	1.16	0.00992	9.99	0.993
60	0.0127	10.0	1.27	0.0142	9.98	1.42	0.0104	9.99	1.04
65	0.0155	10.0	1.55	0.0176	9.98	1.76	0.0110	9.99	1.10

Table 28: The experimental solubility values for toluene in water.

Experimental	
T	Solubility
°C	g/L
0	0.69
5	0.63
10	0.59
20	0.57
25	0.53
30	0.59
40	0.64
45	0.66
55	0.86

Table 29: The predicted solubility values for benzene in water.

	NRTL			UNIQUAC			UNIFAC-LL		
T	Benzene	Water	Solubility	Benzene	Water	Solubility	Benzene	Water	Solubility
°C	kg/h	kg/h	g/L	kg/h	kg/h	g/L	kg/h	kg/h	g/L
0	0.0183	10.0	1.83	0.0128	10.0	1.28	0.0125	10.0	1.25
5	0.0178	10.0	1.78	0.0137	10.0	1.37	0.0134	10.0	1.34
10	0.0175	10.0	1.75	0.0146	9.99	1.46	0.0142	9.99	1.42
15	0.0174	10.0	1.74	0.0155	9.99	1.55	0.0151	9.99	1.52
20	0.0174	9.99	1.74	0.0164	9.99	1.64	0.0161	9.99	1.61
25	0.0176	9.99	1.76	0.0174	9.99	1.74	0.0170	9.99	1.70
30	0.0179	9.99	1.79	0.0184	9.99	1.84	0.0180	9.99	1.80
35	0.0183	9.99	1.84	0.0194	9.99	1.95	0.0190	9.99	1.91
40	0.0190	9.99	1.90	0.0205	9.99	2.05	0.0201	9.99	2.01
45	0.0197	9.99	1.97	0.0216	9.99	2.16	0.0212	9.99	2.12
50	0.0207	9.99	2.07	0.0227	9.99	2.27	0.0223	9.99	2.23
55	0.0218	9.98	2.18	0.0238	9.99	2.39	0.0234	9.99	2.34
60	0.0231	9.98	2.31	0.0250	9.99	2.50	0.0246	9.99	2.46
65	0.0246	9.98	2.47	0.0262	9.99	2.62	0.0257	9.99	2.58

Table 30: The experimental solubility values for benzene in water with the given standard deviations.

Experimental					
T	Solubility	SD	T	Solubility	SD
°C	g/L		°C	g/L	
5.0	1.74	0.024	30.0	1.77	0.019
6.7	1.74	5.190	35.0	1.82	0.023
9.0	1.73	0.031	39.9	1.88	0.015
12.5	1.72	0.002	45.0	1.97	0.002
15.0	1.73	0.026	49.8	2.04	0.040
20.0	1.71	0.021	54.5	2.15	0.027
20.6	1.71	0.020	59.8	2.26	0.051
24.8	1.71	0.034	64.8	2.41	0.021
24.9	1.74	0.024	69.0	2.60	0.023
27.3	1.74	0.023			

Table 31: The predicted solubility values for dichloromethane in water.

	NRTL			UNIQUAC			UNIFAC-LL		
T	Dichloro	Water	Solubility	Dichloro	Water	Solubility	Dichloro	Water	Solubility
°C	kg/h	kg/h	g/L	kg/h	kg/h	g/L	kg/h	kg/h	g/L
0	0.233	9.99	23.3	0.232	9.99	23.2	0.155	9.99	15.5
5	0.224	9.99	22.4	0.224	9.99	22.4	0.164	9.99	16.4
10	0.216	9.99	21.6	0.216	9.99	21.6	0.174	9.99	17.4
15	0.209	9.99	20.9	0.209	9.99	21.0	0.183	9.99	18.4
20	0.203	9.99	20.3	0.203	9.99	20.4	0.193	9.99	19.4
25	0.198	9.98	19.8	0.198	9.98	19.8	0.204	9.98	20.4
30	0.194	9.98	19.4	0.193	9.98	19.4	0.214	9.98	21.4
35	0.190	9.98	19.0	0.189	9.98	18.9	0.225	9.98	22.5

Table 32: The experimental solubility values for dichloromethane in water.

Experimental	
T	Solubility
°C	g/L
0	21.3
5	19.8
10	18.6
15	17.7
20	17.2
25	17.0
30	17.1
35	17.5

Appendix C: Overview of experimental data and results for the solid-liquid systems

Table 33: Specifications of the different solute compounds.

Solute	Molecular weight (g/mol)	# Aromatic rings	# Non-aromatic rings	Heat of fusion (J/mol)	Normal melting point (K)	Solute density (g/L)
thebaine	311.37	1	4	23060	465.41	1305
codeine	299.36	1	4	26661	429.71	1300
oripavine	297.35	1	4	94304	473.93	1400
morphine	285.34	1	4	31764	529.63	1444

Table 34: Experimental solubilities for 20°C in different solvents.

Solvents	Thebaine (g/L)	Codeine (g/L)	Oripavine (g/L)	Morphine (g/L)
1,2-dimethoxyethane	22.6	82.6	7.4	6.7
diglyme	57.3	105.4	38	0.07
dipropyl ether	2.3	5.6	0.3	0.2
dibutyl ether	2.1	4.9	0.2	0.1
isopropyl ether	1.6	4.8	0.2	0.2
methyl propyl ether	6.8	25	1.2	0.3
diethyl ether	4.2	9.5	0.7	0.4
water	0.7	9	/	0.149

Table 35: Solubility of thebaine at different temperatures.

Thebaine		
Solvent	**Temperature (°C)**	**Solubility (g/L)**
1,2-dimethoxyethane	38.2	40
1,2-dimethoxyethane	45.5	45
1,2-dimethoxyethane	49.1	50
1,2-dimethoxyethane	55.6	60
dipropyl ether	39.5	5
dipropyl ether	51	7.5
dipropyl ether	58.8	10
dipropyl ether	73.6	15

Table 36: Solubility of codeine at different temperatures.

Codeine		
Solvent	**Temperature (°C)**	**Solubility (g/L)**
1,2-dimethoxyethane	36.4	125
1,2-dimethoxyethane	40.4	150
1,2-dimethoxyethane	49.5	175
1,2-dimethoxyethane	53.7	200
1,2-dimethoxyethane	59	250
dipropyl ether	45.1	15
dipropyl ether	52.3	20
dipropyl ether	59.2	25

Table 37: Solubility of oripavine and morphine at different temperatures.

	Oripavine		Morphine	
Solvent	**Temperature (°C)**	**Solubility (g/L)**	**Temperature (°C)**	**Solubility (g/L)**
1,2-dimethoxyethane	45.3	10	46.9	10
1,2-dimethoxyethane	52.9	12.5	55.2	12.5
1,2-dimethoxyethane	56.7	15	61.1	15
1,2-dimethoxyethane	65.8	17.5	68.9	17.5
1,2-dimethoxyethane	70.5	20	75	20

Table 38: The different combinations of solvents used for the prediction of the solubility of thebaine in dipropyl ether and the abbreviations are: DME=1,2-dimethoxyethane; DG=diglyme; DBE=dibutyl ether; IPE=isopropyl ether; MPE=methyl propyl ether and DEE=diethyl ether.

	6 Solvents	5 Solvents	4 Solvents	3 Solvents	2 Solvents
2.79	All				
3.45		DME+DG+DBE+IPE+MPE			
3.09		DME+DG+DBE+IPE+DEE			
2.27		DME+DG+DBE+MPE+DEE			
2.24		DME+DG+IPE+MPE+DEE			
2.73		DME+DBE+IPE+MPE+DEE			
2.87		DG+DBE+IPE+MPE+DEE			
4.19			DME+DG+DBE+IPE		
0.95			DME+DG+MPE+DEE		
2.56			DBE+IPE+MPE+DEE		
2.58			DME+DG+DBE+DEE		
2.88			DME+DBE+IPE+DEE		
2.93			DME+DG+IPE+MPE		
4.39				DME+DG+DBE	
2.29				IPE+MPE+DEE	
2.69				DME+DBE+DEE	
3.85				DME+DG+IPE	
2.9				DBE+IPE+MPE	
2.3					MPE+DEE
5.63					DBE+IPE
0					DME+DG
2.59					IPE+MPE
2.76					DBE+DEE
4.16					DME+DBE
2.02					IPE+DEE

Table 39: The values for the combined results for method 1.

	Thebaine	Codeine	Morphine
T(°C)	**Solubility (g/L)**	**Solubility (g/L)**	**Solubility (g/L)**
20	17.5	31.7	1.28
21	18.4	33.4	1.35
22	19.3	35.1	1.43
23	20.2	37.0	1.51
24	21.2	39.0	1.59
25	22.2	41.0	1.68
26	23.3	43.2	1.77
27	24.4	45.5	1.87
28	25.6	47.9	1.97
29	26.8	50.5	2.08
30	28.1	53.2	2.19
31	29.5	56.0	2.30
32	30.9	59.0	2.43
33	32.4	62.2	2.55
34	34.0	65.5	2.69
35	35.6	69.0	2.82
36	37.3	72.7	2.97
37	39.1	76.6	3.12
38	41.0	80.8	3.28
39	43.0	85.2	3.45
40	45.1	89.8	3.62
41	47.3	94.8	3.80
42	49.6	100	3.99
43	52.0	106	4.18
44	54.5	111	4.39
45	57.2	118	4.60
46	60.0	124	4.83
47	63.0	131	5.06
48	66.1	139	5.30
49	69.4	147	5.56
50	72.9	155	5.82
51	76.6	164	6.10
52	80.6	173	6.38
53	84.7	183	6.68
54	89.1	194	6.99
55	93.8	205	7.32
56	98.7	217	7.66
57	104	229	8.01

58	110	242	8.37
59	116	255	8.75
60	122	269	9.15
61	129	284	9.56
62	136	299	9.99
63	144	314	10.4
64	152	329	10.9
65	161	345	11.4
66	171	361	11.9
67	181	377	12.4
68	192	393	13.0
69	204	409	13.5
70	216	425	14.1
71	230	441	14.7
72	244	457	15.4
73	259	473	16.0
74	274	489	16.7

Table 40: The values for the combined results for method 2.

	Thebaine	Codeine
T(°C)	**Solubility (g/L)**	**Solubility (g/L)**
20	8.06	16.0
30	9.55	17.8
40	11.4	20.0
50	13.6	22.7

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Applicability evaluation of AspenONE and Dynochem for solubility modeling in the pharmaceutical industry

Richting: **master in de industriële wetenschappen: chemie**
Jaar: **2014**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Daubies, Arno

Datum: **10/06/2014**