UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA INDUSTRIALE

CORSO DI LAUREA MAGISTRALE IN INGEGNERIA CHIMICA E DEI PROCESSI INDUSTRIALI

Tesi di Laurea Magistrale in

Ingegneria Chimica e dei Processi Industriali

USE OF RIGOROUS THERMODYNAMICS IN SOLUBILITY AND KINETIC MODELS

Relatore: Prof. Fabrizio Bezzo

Correlatore: Dr. Sam Wilkinson

Laureanda: Giada Mingozzi

ANNO ACCADEMICO 2017 - 2018

It is what it is :
The form can be shaped
but the essence cannot be blamed

Abstract

Deriving physical and thermodynamic properties for pure compound systems or mixtures in which species, characterised by few or even no data available, are involved, is a common issue in both crystallization process and reactive systems. Furthermore, it is evident that the capability to predict a system behaviour over a wide range of operating conditions leads to limit drastically the expensive and time-consuming experiments required to describe the system. In particular, discrimination among the solvents potentially employable in a process has a significant impact on both crystallization process (solute solubility change) and reactive systems (selectivity and kinetic variations) but also on downstream unit operations. The use of first principle models together with SAFT-γ Mie as advanced EoS, can provide, theoretically, for a reliable description of any molecule and any system characterised by a wide range of operating conditions. It also allows for a significant improvement in analysing and interpreting experimental data. However, this holds true if all the functional groups of the chemical species involved in the system are present in the database and the system interactions can be fully described by quantities as fugacity or activity. The above capabilities (and limitations) have been demonstrated by carrying out three case studies: i) solubility prediction of ibuprofen in seven different solvents, ii) transesterification (homogeneous liquid phase system) of butyl acetate with ethanol employing different amount of heptane as solvent, and iii) hydrogenation of 4-phenyl-2-butanone (multiphase system) employing multiple solvents.

Riassunto

La derivazione delle proprietà fisiche e termodinamiche di composti chimici puri o in miscela, per i quali una scarsa quantità di dati risulta accessibile, presenta un problema comune nei processi di cristallizzazione così come per sistemi reagenti. Inoltre, è evidente che, la possibilità di predire il comportamento di un determinato processo in un ampio intervallo di condizioni operative, permette di limitare il numero di esperimenti necessari alla sua descrizione; questi, infatti, risultano spesso dispendiosi dal punto di vista economico e in termini di tempo richiesto. La possibilità di conoscere il comportamento di un processo a priori, anche se limitatamente all'aspetto qualitativo, dà la possibilità di valutare preliminarmente quali condizioni operative scartare e quali invece portare a un livello di studio di maggior dettaglio. Uno degli aspetti chiave nei processi di cristallizzazione e nei sistemi reagenti è la scelta del solvente, la quale risulta avere un impatto rilevante sulla variazione del grado di solubilizzazione di un determinato composto, sull'accelerazione o ritardo della cinetica di reazione e sulla posizione di equilibrio di reazione. Ad oggi, i modelli utilizzati per la determinazione delle proprietà chimico fisiche di un sistema e per la descrizione di sistemi reagenti sono prevalentemente empirici o semi-empirici. Questi non garantiscono l'affidabilità dei risultati in caso di necessità di estrapolazione e risultano sistema specifici. Questo studio si è concentrato sulla valutazione delle prestazioni relative all'utilizzo di modelli a principi primi e di SAFT-γ Mie come modello termodinamico relativamente alla corretta analisi ed interpretazione dei dati sperimentali e alla capacità predittiva di descrizione del comportamento di processo in un ampio intervallo di condizioni operative. Tale valutazione è stata effettuata attraverso l'analisi di tre casi studio: i) calcolo della solubilità dell'Ibuprofene in un'ampia gamma di solventi, ii) descrizione del comportamento di un sistema di transesterificazione al variare della quantità di eptano come solvente e iii) descrizione del processo di idrogenazione del 4-fenil-2-butanone impiegando diverse tipologie di solvente. Nel primo caso è stata valutata l'affidabilità e l'accuratezza della predizione e i riscontri di un eventuale utilizzo in applicazioni pratiche come il processo di selezione del solvente in un processo di cristallizzazione. Il secondo e terzo caso studio sono interessati alla comprensione della capacità di rappresentazione dei dati sperimentali, relativi a sistemi reagenti, attraverso una maggior accuratezza nella descrizione del sistema reagente. Tale analisi viene eseguita impiegando modelli a principi primi invece di modelli semi-empirici classici come il modello cinetico basato sulla legge di potenza. Nei casi di impossibilità di derivazione di un modello a principi primi, come nel caso di sistemi reagenti complessi, la sostituzione dei termini di

concentrazione con quelli di fugacità è stata impiegata al fine di aumentare la capacità rappresentativa del generico sistema reagente da parte del modello empirico generalmente impiegato per la sua descrizione. Un ulteriore aspetto analizzato in entrambi i casi studio è la capacità di predire il comportamento del sistema reagente al variare della tipologia o della quantità di solvente impiegato. Infine, è stata effettuata una valutazione critica basata sulla stima dei parametri, relativamente all'impatto del modello a principi primi impiegato, rispetto al modello cinetico basato sulla legge di potenza. La stessa analisi è stata svolta per i modelli empirici ai quali sono stati sostituiti i termini di concentrazione con le fugacità.

Il progetto è stato svolto impiegando il software gPROMS sviluppato dall'azienda Process Systems Enterprise, al cui interno è stato svolto il lavoro di tesi. In particolare, gPROMS Formulated Products (gFP) ha fornito l'ambiente di implementazione e i modelli standard per reattori batch mono e multifase impiegati nei casi studio che coinvolgono sistemi reagenti. L'utilizzo di SAFT-γ Mie come modello termodinamico è stato possibile grazie all'implementazione manuale di gSAFT, pacchetto software che permette di utilizzare le equazioni di stato avanzate basate su SAFT (Self-Associating Fluid Theory) in gFP. I modelli a principi primi sono stati derivati a partire dall'equilibrio solido-liquido per il calcolo della solubilità e dalla riformulazione della condizione di equilibrio di reazione per la formulazione del modello cinetico di reazione.

La predizione del grado di solubilità dell'Ibuprofene in diversi solventi è risultata soddisfacentemente accurata da un punto di vista qualitativo ma non sempre quantitativo. La qualità dei risultati è strettamente legata ai parametri relativi alle interazioni tra gruppi funzionali necessari per l'utilizzo di SAFT-γ Mie. Infatti, a seconda che questi siano regrediti da dati sperimentali o approssimati attraverso *combining rules* e che i dati sperimentali siano solo di tipo VLE o anche SLE il risultato finale può variare notevolmente. Per ottenere valori di solubilità affidabili è necessario garantire che i parametri relativi alle interazioni tra gruppi funzionali siano regrediti da dati sperimentali e che questi siano di tipo sia VLE che SLE.

L'analisi sui sistemi reagenti ha dimostrato che, nel caso di sistema omogeneo monofasico, l'utilizzo del modello a principi primi garantisce una descrizione maggiormente accurata rispetto al modello cinetico basato sulla legge di potenza. Inoltre, è stata dimostrata la capacità di predirre la cinetica di reazione dovuta all'influenza della variazione della quantità di solvente. L'aspetto legato all'equilibrio di reazione viene lasciato ad una successiva trattazione dedicata per motivi che sono chiariti nel testo. Sistemi reagenti catalitici multifase non possono essere rappresentati attraverso gli stessi modelli cinetici di sistemi monofasici omogenei. La ragione è dovuta alla presenza di interazioni catalizzatore-solvente, catalizzatore-substrato che non possono essere descritti

attraverso termini come fugacità e attività. I modelli applicati devono quindi essere derivati in modo tale da tenere in considerazione tali interazioni attraverso la struttura del modello stesso e l'utilizzo di parametri dedicati. Per tale motivo è stato analizzato come l'impiego sostitutivo di termini in grado di catturare interazioni solvente-substrato ad esempio, la fugacità al posto delle concentrazioni, abbia migliorato la capacità rappresentativa del sistema. Tale sostituzione ha portato significativi benefici in relazione alla qualità statistica dei parametri regrediti e all'accuratezza della descrizione del sistema in caso di disponibilità di dati sperimentali per il sistema specifico. Invece, non sono stati riscontrati significativi benefici sotto l'aspetto della predizione, per la quale l'utilizzo delle fugacità non risulta produrre alcun miglioramento sensibile rispetto al modello empirico che impiega termini di concentrazione.

Contents

Introduction

Deriving physical and thermodynamic properties for pure compound systems or mixtures in which species, characterised by few or even no data available, are involved, is a common issue in both crystallization process and reactive systems. Furthermore, it is evident that the capability to predict systems behaviour over a wide range of operating conditions leads to drastically limit, or even to avoid, the need of expensive and time-consuming experiments required to characterise the system. This implies the possibility to evaluate, at a preliminary stage, which conditions must be avoided, and for which it worth performing further studies. One of the key choices for both crystallization and reactive systems is the solvent employed, it affects not only the single unit operation in which it is used but also those downstream (e.g. separation process). Moreover, from reactive systems side it must be considered that different solvents and even different amount of the same solvent can affect significantly both reaction kinetics and reaction equilibrium. For these reasons being able to discriminate a priori the most part of solvent candidates would improve considerably the process design procedure reducing drastically the number of expensive and time-consuming experiments otherwise required.
Nowadays, still, empirical and semi-empirical models are applied to overcome the aforementioned issues, but they provide, mainly, results that are specific for the system analysed and that do not allow the extrapolation of values outside the experimental range available. Furthermore, many of the mixtures that are employed in crystallization process and reactive systems are highly non-ideal and involve complex interactions (e.g. solid catalyst and substrate). As a result of recent developments in the advanced equations of state family based on Self-Associating Fluid Theory (SAFT), it is possible to characterise such complex material behaviour with a significant degree of predictive accuracy. In particular, SAFT-γ Mie is the state-of-the-art group contribution approach belonging to SAFT based advanced EoS. It can be employed for the prediction of the thermodynamic properties of systems for which few or no experimental data are available, provided that the functional groups that describe the system are well characterized.

Process System Enterprise is one of the leading suppliers of advanced process modelling software, it develops gPROMS: a modelling and solution platform that provides for several default unit process models but at the same time, gives the possibility to code models from scratch. SAFT-γ Mie together with other SAFT based advanced EoS is implemented at the platform level, thus, it is, potentially, employable in any gPROMS

application. gPROMS framework, therefore, provides the instruments to build and compute any process model employing SAFT-γ Mie as thermodynamic model.
First principle models together with SAFT-γ Mie EoS, that allows for physical and thermodynamic properties computation of any specie, are expected to provide accurate predictions for any process considered.

The objective of this project is, therefore, to examine and quantify the potential impact of the use of rigorous thermodynamics within solubility and kinetic models in terms of the ability to use such models to:

- analyse and correctly interpret experimental data
- predict process behaviour over wide ranges of operating conditions.
- improve the calculation accuracy

In order to make these assessments three case studies have been analysed and for each of them different aspects based on the case study characteristics have been considered. The three case studies consist of: solubility predictions of Ibuprofen in seven different solvents, transesterification reactive system with varying solvent amount and hydrogenation reactive system characterised by different solvents employed. The thesis work presents four chapters, the first one is dedicated to gPROMS framework description with focus on the part of the software specifically employed to carry out the project (gPROMS FormulatedProducts and gSAFT). It is followed by some theoretical background on SAFT with the purpose of introducing SAFT-ϒ Mie advanced equation of state (EoS). The second chapter is entirely dedicated to the solubility prediction case study; the crucial solubility role in crystallization process is described at first; it is followed by the derivation of the models employed for the calculations starting from solid-liquid equilibria definition. The solubility prediction results are reported commented and employed in practical applications (e.g. solvent screening and solvent selection). The third chapter consist mainly of the second case study description and it is structured in such a way to provide, at first, a general overview on the solvent influence in reactive system. Then, a detailed description of the derivation steps followed to obtain thermodynamically rigorous kinetic models based on the thermodynamic reaction equilibrium condition reformulation is presented. This is followed by a pragmatic comparison among the use of concentrations, activities and fugacities as driving forces in reaction process together with the model validation steps followed and the case study objectives. After the mainly theoretical part a detailed description of the transesterification system analysed has been reported. The last chapter is dedicated to assess the degree of improvement attained by the substitution of fugacities terms instead of concentrations in complex kinetic models to describe the solvent effects on catalytic

multiphase systems. The purpose is achieved developing 4-phenyl-2-butanone catalytic hydrogenation that involves four reactions two in series and two in parallel that are strongly affected by the solvent choice.

Chapter 1

gPROMS Framework

In this chapter a general overview of gPROMS structure is firstly presented, it is followed by a more detailed description of the environment (gPROMS FormulatedProducts) and the properties packages (Multiflash, gFPPP and gSAFT) employed in the project. Finally, the objectives and motivations of the thesis are addressed.

1.1 The gPROMS modelling environments

gPROMS® advanced process and material modelling platform is an equation-oriented modelling and optimization framework that provides flowsheeting, first principles custom modelling, parameter estimation, physical properties integration and many other features. Applying advanced process modelling means the use of detailed, high-fidelity mathematical models of process equipment and phenomena, in order to provide accurate predictive information for decision support in process innovation, design and operation. This approach is employed to explore the process decision space to enable better, faster and safer decisions by reducing uncertainty.

The gPROMS platform supports two main environments: gPROCESS and gFORMULATE (recently renamed gPROMS FormulatedProducts) that differ in the types of libraries included, meaning in the subset of specific unit models, that combined, describe the specific processes that characterize industrial sectors such as Food and Dairy, Home and personal care, wastewater treatment and so on. A sketch of gPROMS overall structure is shown in Figure 1.1.

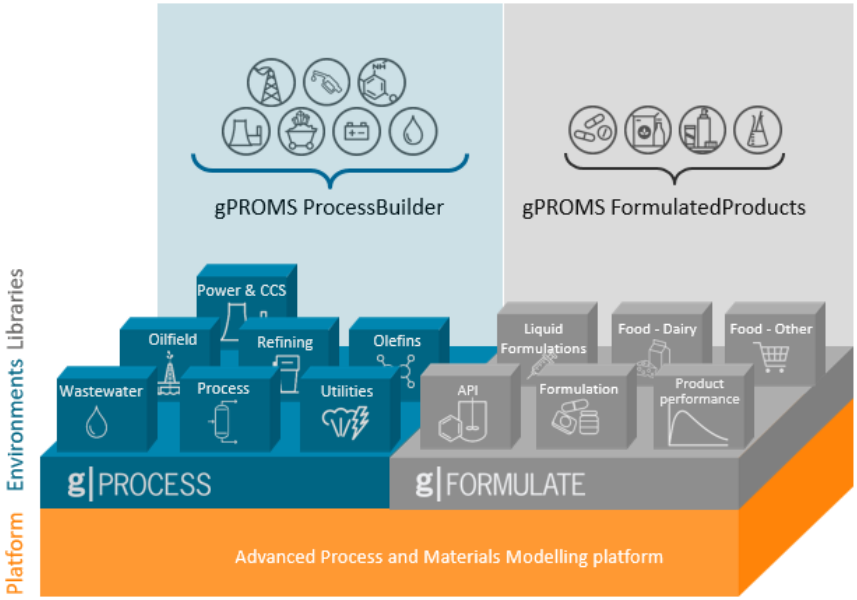


Figure 1.1 *gPROMS current structure (2018): Unique platform (orange) that support two main environments (blue and grey shading) that are characterized by several specific libraries (blue and gray shading squares)Bermingham, (2018).*

The material modelling that comprises SAFT family advanced thermodynamic model implementation (gSAFT) is an integral part of gPROMS platform, this means that by design gSAFT physical properties technology, can be potentially used in any gPROMS product: environments, libraries and models. In order to carry out this thesis project gPROMS FormulatedProducts (gFP) framework with gSAFT implementation has been used.

1.2 gPROMS FormulatedProducts

gPROMS FormulatedProducts is a modelling framework in which it is possible to integrate mechanistic modelling for design and optimization of formulated products and active ingredients manufacturing processes. It is based on libraries present in gCRYSTAL that is related to synthesis and crystallization, gSOLIDS that regards solid processing and gCOAS that deals with product performance; these frameworks have been merged in a unique environment, namely gFP. Other functionalities already present in gPROCESS environment have been specifically added in gFORMULATE in the recent years, such as data import and processing, global sensitivity analysis and external model validation. A complete overview of how gFORMULATE environment is structured is shown in Figure 1.2.

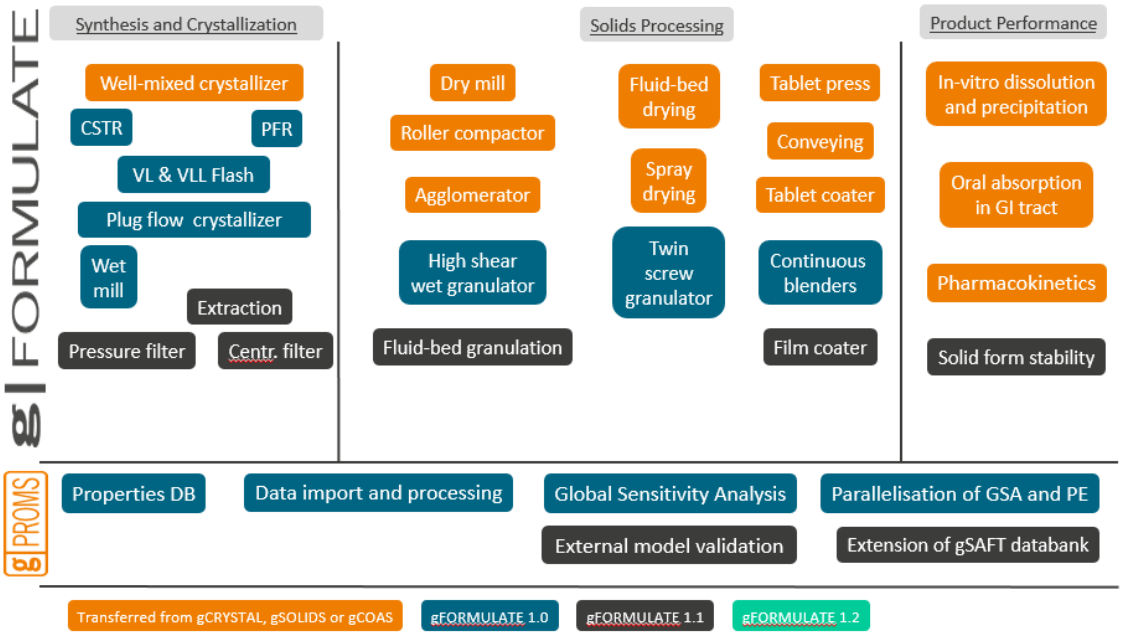


Figure 1.2 *Last gFormulated release contents (2017). Bermingham (2018), '*

Since the focus of this project is on the solubility prediction and kinetic and equilibrium reaction computation of complex molecules, such as API, in a wide range of solvents, the work is focused on the active ingredients manufacture application area. It includes: synthesis, fluid separation, crystallizatio,wet milling, filtration and drying. The focus has been placed on the former because it incorporates both the solvent choice and reaction synthesis process in order to produce the desired molecules in the most efficient way, with the desired purity level and minimizing the formation of chemical by-products.

1.2.1 gFP Custom Modelling

The main drivers for custom modelling are the need of creating novel process units with new configurations that in the standard library models might not exist or improving them in order to include more details that allow the standard library to fit the specific system of interest. Practically what custom modelling is used for can be summed up as:

- Creation of a new model for a full process unit
- Creation of specific elements for an existing model

Pharma and specialty chemical industries are characterized by a product dynamic turnover that usually implies the need of a frequent process modification. For this reason, especially in gFormulatedProducts that operates in these sectors, custom modelling is essential to allow the accurate description of these systems. However, applying custom

modelling is possible in the entire gPROMS platform using gPROMS language, in order to capture corporate knowledge of any system of interest. The models can, then, be published in libraries becoming an integral part of the library itself. The customized models can, also, be validated against laboratory, pilot or operating data to fit empirical constants; in this work, for instance, rate constants values (kinetic parameters) have been regressed using gPROMS state-of-the-art parameter estimation capabilities. Since gFP is the main environment that have been used to carry out the thesis project Figure 1.3 is useful to understand the hierarchy that characterizes custom modelling within gFP architecture.

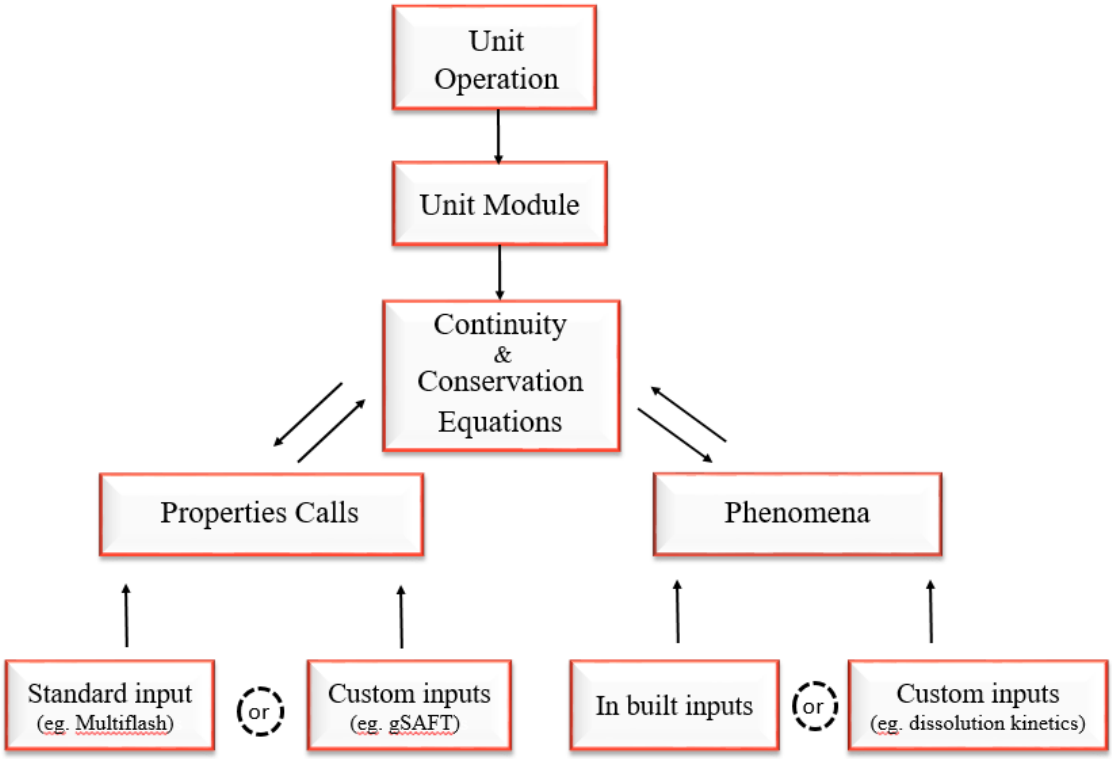


Figure 1.3 *Custom modelling within the gFormulated Product architecture*

A generic unit operation presents a single unit module, few exceptions can be found, for instance in the jacketed CSTR reactor unit operation both the CSTR and the jacket modules are present. Each module is characterized by a system of continuity and conservation equations that arise from properties and phenomena computation. These can be already present in standard libraries, partially customized or coded from scratch. In the last two cases the equations can be implemented through templates that, in turn, can be employed with the default equations or can be filled coding the specific system of equations that best suit the specific application. The settings chosen, then, are used during the computation of the system instead of the pre-defined ones in the specific unit module

model. The custom phenomena comprise mainly reaction kinetic, mass transfer coefficient and solubility models. In the other hand, custom physical properties allow to customize the equations to define how different thermodynamic properties, such as density and enthalpy, should be computed. It also allows for the usage of different property packages for the computation of different physical properties related to pure compound or mixture system. The property calls represent the technical procedure through which the customized computation of physical properties is carried out. It consists of lines of gPROMS code that ask for specific models to compute specific properties and it can refer to different properties packages (Multiflash, gFPPP, gSAFT). Regardless the availability of different property packages that would allow to better compute each property or some phenomena (mass and heat transfer coefficient) it is advisable, from a consistency point of view, to choose a unique property package for the entire system when possible. The models available allow to calculate:

- Constants and pure compound properties (normal boiling point, critical pressure, melting point…)
- Phase properties (Liquid head capacity, vapor density, vapor and liquid fugacity coefficients…)
- Equilibrium flashes (composition, total volume and total enthalpy of all the phases present at equilibrium condition)
- Total properties in mixtures (density, entropy, heat capacity…)
- Phase boundary calculations (bubble and dew temperature, pressure and composition).

It is worth noting that the use of gSAFT in gFP it is possible, so far, only using custom models. For this reason, to carry out the main aim of the project that is assessing the potential impact of SAFT-γ Mie, the most recent SAFT version, on pharmaceutical and specialties chemical applications, the use of custom modelling is essential.

1.2.2 Property packages availability

Since gPROMS has to deal with a wide range of industrial sectors such as oil and gas, chemical and petrochemicals, pharmaceuticals, consumer products and agrochemicals, the requirements in terms of physical and thermodynamic properties cover very different types of molecules and multiphase systems. For instance, in chemical and petrochemical separation the properties required are typically Vapor-Liquid equilibrium (VLE), whereas in crystallization mainly solid-liquid equilibrium (SLE) is employed. So far it has been really challenging for any thermodynamic property model to provide with accurate

properties description/prediction of such a wide range of systems and over broad operating conditions. For this reason, the properties packages available in gPROMS are several and attempt to cater for different needs. The main properties packages currently available are Multiflash, gFormulatedProducts properties package (gFPPP) and gSAFT.

Multiflash is an advanced software package that allows to perform complex equilibrium calculations. The main utility is a multiple phase equilibrium algorithm that is interfaced to Infochem's package of thermodynamic models and a number of physical property data banks (such as Dortmund etc). The models used for fluid phase systems fall mainly into two groups: equation of state and activity coefficient methods. Equations of state provide a consistent framework where vapor and liquid phases are treated on equal footing and a single theory is used for both phases. With an activity coefficient method, instead, the vapor phase properties are derived from an equation of state, whereas the liquid properties are determined from the summation of the pure component properties to which a mixing term or an excess term has been added. Multiflash may also be used to calculate the phase equilibrium of systems containing solid phases, either mixed or pure. These may occur either when a normal fluid freezes or may be due to a particular solid phase such as a hydrate, wax or asphaltene involved in the system; in these cases, specific models are employed. The choice of the thermodynamic model that best represent a specific system usually is taken following the conventional decision tree reported in Figure 1.4. It is sometimes required to change the model and compare the results obtained with a different one since different models may show different accuracies in describing the same system. In case of multiphase systems is even required to choose different models to describe different phases. It is evident, hence, that the choice of the thermodynamic model for a certain system is not always restricted to a single one and may not be optimal.

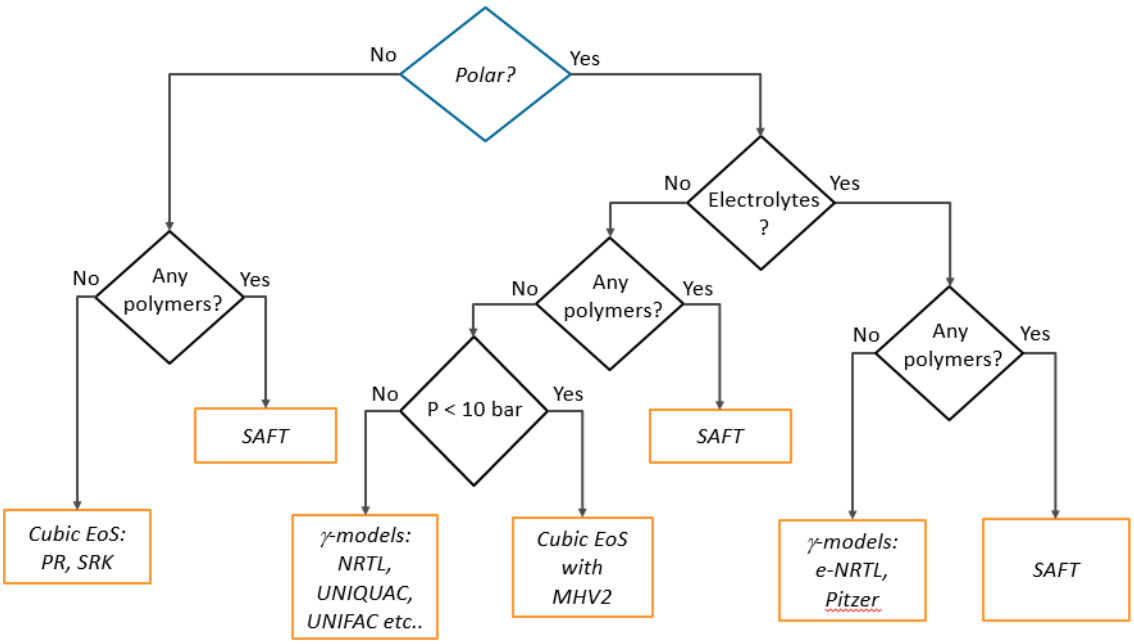


Figure 1.4 *Thermodynamic models conventional decision tree*

gPROMS FormulatedProducts Utilities software, that is technically the environment in which gFPPP can be used, is mainly employed in case of experimental data availability. In this case VLE, LLE and SLE phase equilibrium are not computed through equation of states of activity coefficient methods, instead, gFPPP gives the possibility to customize them through the use of custom template or to use purely empirical equations eligible only in case of experimental data set. gFPPP allows the user to create custom databases, and to define species, phases and reactions. In order to create the material database, the basic information required for each specie is the molecular weight. Thermodynamic property calculations such as density, specific heat capacity and solubility (of a solute in a single solvent) are computed through the same empirical model (1.1) which is a polynomial in temperature that contains up to six parameters that are normally obtained from regression against experimental data.

$$General\ Property = \frac{a}{T^2} + \frac{b}{T} + c + d \cdot T + e \cdot T^2 + f \cdot T^3 \tag{1.1}$$

Different empirical models are applied for anti-solvent systems, sorption isotherm and the solubility product of a salt in a single solvent. It worth pointing out that all the empirical models available can be used only if the parameters in the models are known and this implies to be able to regress them using experimental data. For this reason, the physical properties calculated through empirical models are reliable in a limited range of conditions that correspond to those for which the experimental data have been carried out.

Multiflash and gFPPP provide with two different approach to the physical property calculations but both of them present some issues. First, as aforementioned, the choice of the best thermodynamic model for a specific system is not straightforward and there is the possibility that it does not meet all the requirements needed.

A second aspect to consider, in particular in the pharma and specialty chemical industry, is the lack of data for many species, particularly active pharmaceutical ingredients. In these businesses, improvements and changes of products manufacture and production are common. The chemical species that characterize the system usually are complex molecules for which few experimental data are available and this can be an issue when making changes to manufacturing, particularly if conditions (e.g. temperature) change. The lack of experimental data prevents the use of gFPPP because it is not possible to regress the required parameters. In these cases, it is likely that Multiflash database doesn't provide for those species data as well so it cannot compute any physical properties calculation. Therefore, it is required to carry out ad-hoc measurements that are usually expensive and time consuming; for instance, in order to carry out a solubility measurement the time required is estimated as 6-10 days.

The third issue is related to the limited applicability range of empirical correlations, it is restricted to specific systems and specific range of conditions that correspond to those at which the data used to regress the equation parameters have been carried out. The possibility of extrapolation can be considered but it must be taken into account that the reliability of the result obtained is compromised.

The aforementioned limitations can be theoretically overcome employing a group contribution advanced thermodynamic model such as SAFT-γ Mie that, together with SAFT-VR and PC SAFT, completes the advanced thermodynamic models implemented in gSAFT physical properties package. So far it is not possible to implement gSAFT as property package in any gFP general flowsheet through the default settings, therefore, in order to choose a SAFT EoS as thermodynamic model for a certain system what is required to do is to define, in each specific custom model template, which of the three SAFT versions is to be employed. In order to clarify which are the benefits of employing gSAFT as properties package a general description of SAFT advanced thermodynamic model and how the calculations of physical properties are computed, is required. The second half of this chapter is thereby dedicated to this purpose.

1.3 gSAFT Advanced physical property prediction for process modelling

Progress in thermodynamic modelling has led to the development of accurate and versatile physical property prediction frameworks that can have a significant impact on process modelling reliability. gSAFT is an efficient and robust implementation of three advanced equations of state of general SAFT framework, namely PC SAFT, SAFT-VR SW and SAFT-γ Mie. The latter represents a state-of-the-art group contribution approach, and as such can be employed for the prediction of the thermodynamic properties of systems for which few or no experimental data is available, provided that the functional groups that describe the system are well characterized. For this reason, SAFT-γ Mie has been chosen to be investigated within this thesis.

1.3.1 The Statistical Associating Fluid Theory (SAFT)

The general theory of SAFT is a framework deeply rooted in statistical mechanics that allows for the description of the thermodynamic properties of fluid and fluid mixtures based on a detailed molecular representation (Wertheim, 1984). In general, within the framework of SAFT-type approaches the molecular model employed is the one shown on Figure 1.5. A molecule is assumed to comprise a number of segments as part of a molecular chain, where the segments interact by means of a well-characterized

intermolecular potential, typically describing both repulsive and attractive interactions, similar to a van-der-Waals type of interactions. In addition to these interactions, association sites can be placed on the molecular chains which allow to mimic hydrogen bonding effects, of shorter range and typically increased energy.

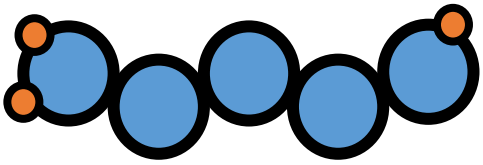


Figure 1.5 *Representation of an associating chain-like molecule in SAFT approach*

The properties of a fluid system are then calculated by summing the appropriate contributions (that arise from the different interactions) to the Helmholtz free energy of the system, as per the below expression:

$$\frac{A}{NkT} = \frac{A^{ideal}}{NkT} + \frac{A^{monomer}}{NkT} + \frac{A^{chain}}{NkT} + \frac{A^{association}}{NkT} \tag{1.2}$$

where:

- A^{ideal} corresponds to the ideal-gas Helmholtz free energy

- $A^{monomer}$ is the residual Helmholtz free energy due to the formation of spherical monomeric segments (includes repulsive and attractive/dispersive interactions);

- A^{chain} refers to the change in Helmholtz free energy associated with the formation of a molecular chain from its constituting segments.

- $A^{association}$ is the Helmholtz free energy due to the association between molecules.

- N is the total number of molecules, k is the Boltzmann constant and T is the absolute temperature.

In Figure 1.6 a visual representation of these individual contributions accounted for in the general framework of SAFT-type EoS is shown.

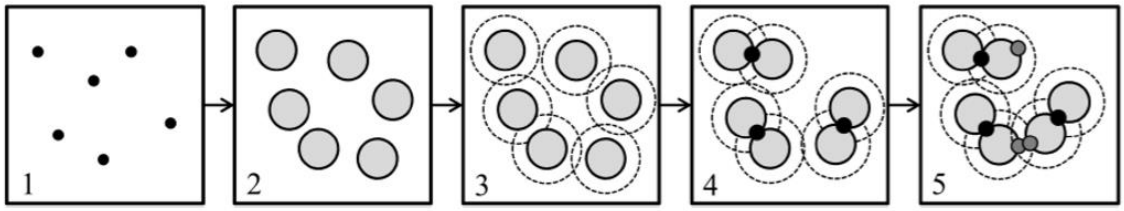


Figure 1.6 *Individual contributions in SAFT EoS*

From the left-hand side what can be seen is the ideal gas contribution (A^{ideal}) followed by the representation of molecules as atomized spherical segments exhibiting repulsive forces (excluded volume, no overlap). The third contribution is characterized by attractive forces between segments (which perturb the fluid of hard-spheres, represented as dashed circles); these repulsive and attractive forces correspond directly to the monomer term ($A^{monomer}$). The fourth and the fifth represent spheres that are fused together to form chain of molecules, and directional forces action respectively. The above describes the characteristics of SAFT as a general framework; the development of an equation of state follows the determination of the intermolecular potential employed to describe interactions between segments. In that respect, several equations of state have been presented over the past 30 years, see a recent review for a more detailed accounting of these (McCabe, 2010). In general, SAFT-based EoSs have been shown successful in the description of the thermodynamic properties of a wide-range of systems and over broad conditions, and this is generally attributed to the detailed description of the structure of the molecule (molecular chains) and molecular interactions.

1.3.2 SAFT-Y Mie: An advanced thermodynamic model

In light of the success of SAFT as a general framework, the development of more predictive SAFT-based equations was attempted, by combining the versatility of SAFT with the predictive capabilities inherent to the group-contribution approach (Papaioannou, 2014). Group contribution (GC) methods are a specific class of predictive models that have drawn increased attention over the past 40 years. In group-contribution approaches molecules are modelled in terms of the functional groups that they comprise, and it is assumed that the properties of any given system can be computed from appropriate contributions of the corresponding groups to the thermodynamic properties of the system. Perhaps the most prominent example of this type of approaches is the universal functional activity coefficient (UNIFAC) approach and its modifications (Fredenslund, 1975). GC-based approaches have also been developed within EoSs, to overcome some of the limitations inherent to activity coefficient approaches, such as the treatment of the liquid phase alone and the inability to compute pure component properties and the phase behaviour near the critical point (as the two phases are inconsistent) (Dufal, 2014).

The group-contribution concept has been applied within a SAFT formalism based on a more detailed heteronuclear molecular model, with different types of monomeric segments that describe the different chemical functional groups representing a given molecule. This heteronuclear molecular model is implemented in the SAFT-γ approach and gives the possibility to predict thermodynamic properties of mixtures based on pure

component data alone. The most recent variant of SAFT-based group-contribution approaches is, therefore, the SAFT-γ Mie EoS where a heteronuclear model is implemented and a Mie (generalized Lennard−Jones) potential of variable repulsive and attractive ranges is used to represent the segment−segment interactions. A brief summary of the SAFT-γ Mie theoretical formalism is first presented, followed by a short discussion on the parameter estimation scheme applied to obtain the specific parameters needed to describe each functional group within the context of the theory.

1.3.2.1 SAFT-γ Mie model and theory

In the SAFT-γ Mie approach the molecules are represented as associating heteronuclear chains of fused spherical segments (Papaioannou, 2014). As a visual example, the representation of Ibuprofen based on SAFT-γ Mie approach is given in Figure 1.7

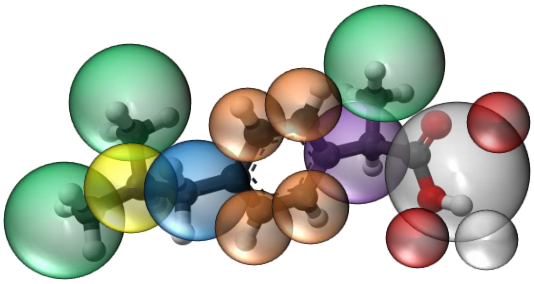


Figure 1.7 *: Example of a molecule decomposition into functional groups: Ibuprofen is composed of three CH3 groups (shaded green),one CH group (shaded in yellow), one aCCH2 group (shaded in blue), one aCCH group (shaded in purple), four aCH groups (shaded in orange) and one COOH group (shaded in grey).*

Each chemical functional group (CH_3, CH, etc) is represented as a fused spherical segment and as a number of segments v_k*. Two segments *k* and *l* are assumed to interact via a Mie intermolecular potential, a generalized Lennard-Jones-type potential, where the attractive and repulsive exponents are not fixed to the values 6 and 12, respectively, but are allowed to vary freely. The form of the pair interaction energy between two monomeric segments is given as a function of the intersegment distance r_{kl} :

$$\Phi_{kl}^{Mie}(r_{kl}) = C_{kl}\epsilon_{kl}\left[\left(\frac{\sigma_{kl}}{r_{kl}}\right)^{\lambda_{kl}^{r}} - \left(\frac{\sigma_{kl}}{r_{kl}}\right)^{\lambda_{kl}^{a}}\right] \tag{1.3}$$

where σ_{kl} the segment diameter, ϵ_{kl} the depth of the potential well, and λ_{kl}^{r} and λ_{kl}^{a} are respectively the repulsive and attractive exponents of the segment−segment interactions. The prefactor C_{kl} is a function of these exponents and ensures that the minimum of the interaction is $-\epsilon_{kl}$:

$$C_{kl} = \frac{\lambda_{kl}^{r}}{\lambda_{kl}^{r}-\lambda_{kl}^{a}}\left(\frac{\lambda_{kl}^{r}}{\lambda_{kl}^{a}}\right)^{\lambda_{kl}^{a}/(\lambda_{kl}^{r}-\lambda_{kl}^{a})} \tag{1.4}$$

The Helmholtz free energy of a generic compound can be obtained from the appropriate contributions of different groups. The implementation of this type of united-atom model of fused segments requires the additional use of a shape factor S_k, which reflects the proportion of a given segment that contributes to the total free energy. Interactions as hydrogen bonding or strongly polar interactions can be treated through the incorporation of a number of additional short-range square-well association sites, which are placed on any given segments as required. The association interaction between two square-well association sites of type *a* in segment *k* and *b* in segment *l* is given by:

$$\Phi_{kl,ab}^{HB}\left(r_{kl,ab}\right) = \begin{cases} -\epsilon_{kl,ab}^{HB} & if \quad r_{kl,ab} \leq r_{kl,ab}^{c} \\ .\ 0 & if \quad r_{kl,ab} \leq r_{kl,ab}^{c} \end{cases} \tag{1.5}$$

where r_{kl} is the distance between the centers of both sites, $-\epsilon_{kl,ab}^{HB}$ the association energy, and r_{kl}^{c} is the cut-off range of both sites that can be equivalently be described in terms of bonding volume $K_{kl,ab}$. In summary, the parameters that describe the *Mie* intermolecular potential are the diameter σ_{kl}, the repulsive and attractive exponents λ_{kl}^{r} and λ_{kl}^{a}, respectively, and the potential depth ϵ_{kl}. The association interactions between pair of sites (denoted, for example, as a and b) are, if present, described by two additional parameters, namely the energy of interaction, $\epsilon_{kl,ab}^{HB}$, and the volume available for bonding, $K_{kl,ab}$. Note that no additional parameters are required to describe the formation of the molecular chains; this contribution is inferred by the knowledge of the distinct types of functional groups that comprise each molecule and their multiplicity on the molecular chains. It worth noting that the number of different association site types and their number of occurrences on a given group needs to be defined a priori. The choice is guided based on chemistry analysis, in particular by understanding how many hydrogen bonding donors and acceptors are present on a given chemical group. The interaction parameters described above relate to all pair interactions. A schematic representation of the physical meaning of each parameter is given in Figure 1.8.

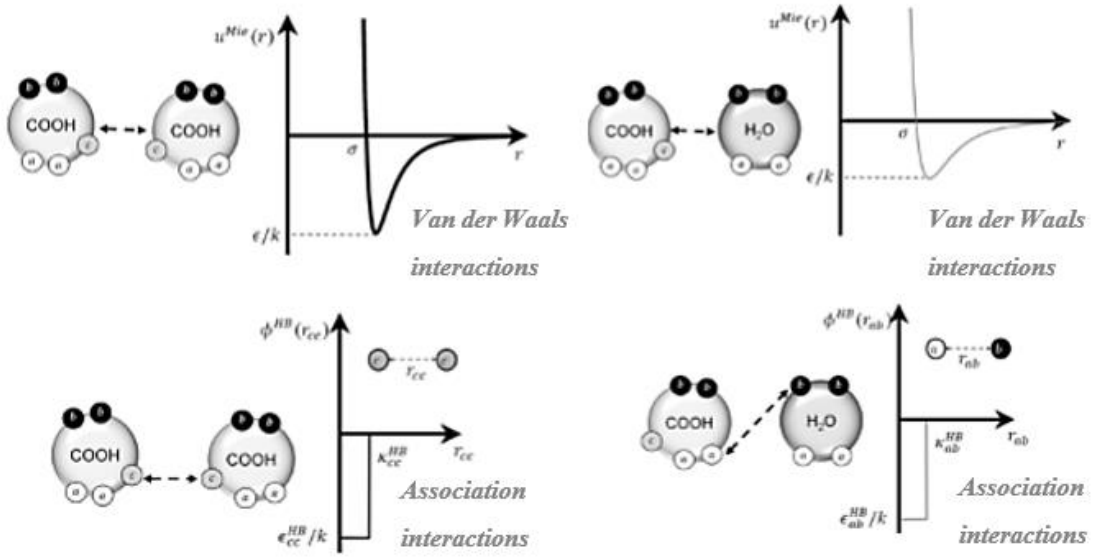


Figure 1.8 *Schematic representation of the SAFT-γ Mie parameters needed to describe the group self- and cross- interactions.*

When developing the group pair interaction parameters, what it is done in practice is separating the treatment of the groups that are of the same type (self-interactions), and the interactions between groups of different types (unlike, or cross-interactions). The self-interaction parameters are usually obtained by regression of experimental data, whereas, the parameters that describe unlike interactions often have to be approximated by combining rules because of experimental data lack. In view of this, it is evident that one of the main challenges for the application of the SAFT-γ Mie approach, for instance in case of pharmaceutical system, relies on the development of a database that contains the maximum number of group interaction parameters for relevant chemical groups typically found in APIs. Those currently available along with the procedure to develop such database are addressed in the next paragraph. Once the relevant parameters are determinate it is possible to evaluate the thermodynamic properties of a pure system or a mixture. The link between the molecular model and the macroscopic thermodynamic properties can be created in terms of the Helmholtz free energy; in this case the expressions are those of the SAFT-γ Mie EoS. The total Helmholtz free energy of a mixture of associating heteronuclear chains of fused spherical segments that interact via Mie potentials can be written in the usual SAFT form as the sum of four separate contributions as shown in equation (1.2). Specific to SAFT-γ Mie EoS within A^{mono} the Mie potential is used to describe the segment−segment interactions (repulsion and dispersion) of the reference monomeric system, and A^{chain} accounts for the change in the free energy due to the formation of molecules from Mie segments. Once the value of the free energy has been obtained, all the other thermodynamic properties can be calculated by taking the appropriate derivative of the Helmholtz free energy function. For example,

the pressure and the fugacity coefficients of a fluid mixture can be obtained with the appropriate derivatives of the free energy of the system, as shown below:

$$\ln \overline{\varphi_i} = \frac{1}{RT}\left(\frac{\partial A^{res}}{\partial n_i}\right)_{T,V,n_{i\neq j}} \tag{1.6}$$

$$P = \left(-\frac{\partial A}{\partial V}\right)_{T,N} \tag{1.7}$$

1.3.2.2 SAFT-γ Mie group parameters estimation

Within the SAFT-γ Mie EoS the group parameters describe the intermolecular interactions between the functional groups that constitute the molecules. They can be determined making use of available experimental data on simple systems where a wide range of different data types can be used, including data on the thermodynamic properties and phase equilibrium behaviour of binary and/or multicomponent systems, but also pure-component data. This results to be a significant benefit over activity coefficient models, since only EoS approaches can be applied to the description of pure components, and hence, benefit from the availability of readily available pure-component data in the development of group parameters. As common in GC-based approaches, and so within SAFT-γ Mie, group parameters are assumed to be transferable so that experimental data are not required to be specific to a given molecule; for the development of the parameters of a given group, data on any molecule containing the group of interest can be used in the regression procedure. As stated above each functional group is described by a set of parameters which describe its size and how it interacts. In order to obtain a good confidence in their values, a commonly used strategy is to include various types of experimental data in the regression procedure, including for instance, vapor pressure, density and caloric data (heat capacity and/or vaporization enthalpy). The details on the exact sequence of development of all the functional group interactions used in this study are beyond the scope of the thesis work but a simple example is used here to illustrate the procedure involving the CH_3, CH_2 and OH groups (Papaioannou, 2014). As typical within GC-based approaches, group parameters are obtained in a sequential manner, so first the methyl and methylene groups are determined by regression to experimental data for a series of n-alkanes, more specifically, from ethane to n-decane. Once the parameters for the CH_3 and CH_2 groups have been determined, the OH group can then be obtained from data on the family of primary alcohols. The use of pure component data allows for both the computation of the self-interactions and the cross-interactions between each group pair, a unique feature of SAFT-γ Mie, due to the underlying heteronuclear molecular model. A similar sequence can then be used to determine any other group, by selecting either mixture or pure component data. It is worth noting that, besides the possibility of defining molecules from functional groups, the SAFT-γ Mie EoS can be used in a

"molecular fashion", in other words, at the limit of a compound comprising a single group. This is usually the case for the first member of each chemical family, e.g., methane, CH_4 group, for the n-alkanes, methanol, CH_3OH group for the alcohols, etc., but can also be employed in the study of more complex molecules. Since the first member of each chemical family behaves in a slightly different way with respect the rest of the compounds that belong to the same family, compound-specific experimental data are required for the development of the single group parameters. Therefore, if methanol is present in the system of interest it must be defined as CH_3OH and not as the combination of CH_3 and OH groups because, otherwise, the results obtained would not be representative of the real system interactions. When the unlike group interaction parameters cannot be regressed from experimental data they are determined using the following combining rules. The unlike segment diameter σ_{kl} is obtained using the Lorentz-like arithmetic mean of the like diameters (Rowlinson, 1982):

$$\sigma_{kl} = \frac{\sigma_{kk}+\sigma_{ll}}{2} \tag{1.8}$$

The unlike dispersion energy ϵ_{kl} between groups k and l is obtained by applying an augmented geometric mean (Berthelot-like rule), which also accounts for asymmetries in size:

$$\epsilon_{kl} = \frac{\sqrt{\sigma_{kk}^3 \cdot \sigma_{ll}^3}}{\sigma_{kl}^3} \cdot \sqrt{\epsilon_{kk} \cdot \epsilon_{ll}} \tag{1.9}$$

The unlike segment-segment interaction exponents λ_{kl}^r and λ_{kl}^a, are obtained from the imposition of the geometric mean of the integrated van der Waals energy for a Sutherland fluid of range λ_{kl} :

$$\lambda_{kl} = 3 + \sqrt{(\lambda_{kk} - 3)(\lambda_{ll} - 3)} \tag{1.10}$$

In case of associating mixture also the association energy $\epsilon_{kl,ab}^{HB}$ and the unlike bonding volume $K_{kl,ab}$ needs to be computed; in the absence of experimental data the former is obtained by applying a simple geometric mean:

$$\epsilon_{kl,ab}^{HB} = \sqrt{\epsilon_{kk,ab}^{HB} \cdot \epsilon_{ll,ab}^{HB}} \tag{1.11}$$

while the latter is obtained as

$$K_{kl,ab} = \left(\frac{\sqrt[3]{K_{kk,aa}} + \sqrt[3]{K_{kk,bb}}}{2} \right)^3 \tag{1.12}$$

The table reported in Figure 1.9 has been used in order to assess the availability of the group cross interactions needed to fully describe the systems involved in the three different case studies analysed. The table presents two types of shading: blue shading indicates the group-cross interaction parameters estimated from experimental data, while the grey shading indicates the unlike interaction parameters approximated by combining rules. This table is updated periodically with the purpose of increasing both the type of functional groups available and the amount of group-cross interaction parameters regressed from experimental data. It is known that in some cases the values of the unlike interactions predicted by combining rules lead to an inaccurate description of the properties of systems. This is more common in cases of systems where one or more polar components are present. In these cases, the regression of the unlike interaction parameters using experimental data is required to achieve a quantitative description of the thermodynamic properties of the system of interest.

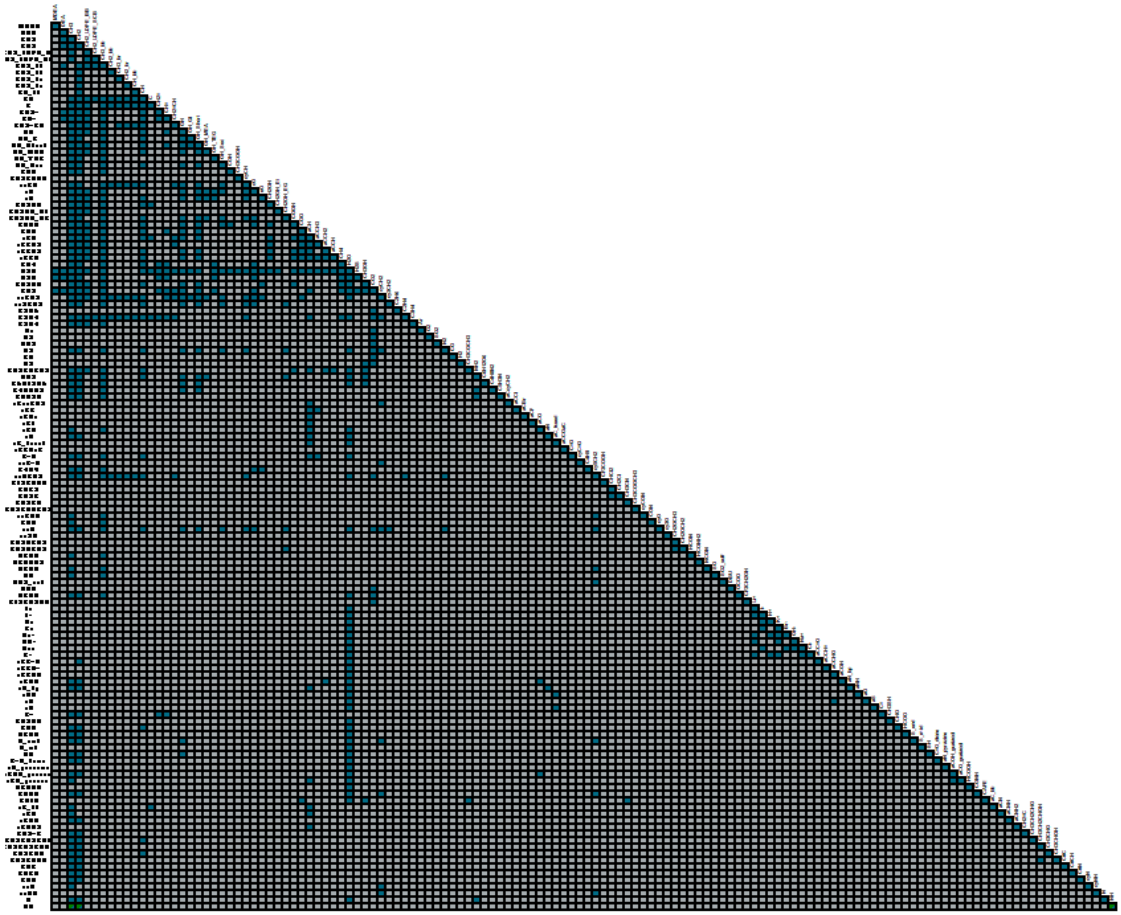


Figure 1.9 *Group parameter matrix developed for use within SAFT-γ Mie approach*

1.3.3 Physical properties computation

Equations of state can be used over wide ranges of temperature and pressure, including the sub-critical and supercritical regions. They are frequently used for ideal or slightly non-ideal systems but they don't necessarily represent highly non-ideal chemical systems for which an activity coefficient approach is preferred. The system description accuracy increases when binary interaction parameters (BIPs) have been regressed from experimental data. This brief description shows the limitations of cubic EoS in terms of handling highly polar system and their dependence of experimental data to ensure an accurate description of the system. A non-cubic group contribution based EoS as SAFT-γ Mie can satisfactorily describe pure component systems and complex mixture over a wide range of operating condition requiring limited experimental data. Therefore, there is no need to choose the most suitable thermodynamic model for each system depending on the operating conditions and the species involved anymore.

In order to describe real systems an equation of state must be able to describe the difference between the properties of the real system compared to the ideal gas at the same thermodynamic conditions. This difference is the so called residual term and its role in thermodynamic properties for the computation of enthalpy and entropy in mixture is shown in equations 1.11 and 1.13. The relations reported refer to mixtures, but the same relations are valid for pure compounds provided that the sum over all the species and the composition dependence are removed.

$$H(T,P,n) = \sum_i n_i \left[\Delta H_i^{F,ig} + \int_{T\theta}^{T} c_{p,i}^{ig}(T')dT'\right] + H^{res} \tag{1.13}$$

$$S(T,P,n) = \sum_i n_i \left[\Delta S_i^{F,ig} + \int_{T\theta}^{T} \frac{c_{p,i}^{ig}(T\prime)}{T\prime} dT' - R\ln\left(\frac{P}{P\theta}\right)\right] - NR\ln(x_i) + S^{res} \tag{1.14}$$

Where :

- H [J] and S [J/K] are the total enthalpy and entropy of the system respectively

- $\Delta H_i^{F,ig}$ [J/mol] and $\Delta S_i^{F,ig}$ [J/K mol] the ideal gas enthalpy and entropy of formation of the specie *i* respectively

- $c_{p,i}^{ig}$ [J/K mol] the specific heat capacity of specie *i* in the ideal gas state

gSAFT computes the departure from the ideal gas system this means that it is responsible of describing all the non-ideal interactions between molecules. It worth noting that 1.13 and 1.14 are written in term of formation properties (element reference) instead of applying the pure ideal gas reference (pure compound reference). This choice is due to the fact that the pure ideal gas reference doesn't provide an invariant reference point in

case of reactive systems. When chemical reactions occur, different possible states of the system will generally involve different amounts of each chemical species; as a result, the property value cannot be compared if the pure compound reference is used. In order to move from element reference to pure ideal gas the formation properties are needed, hence, they must be present in the database for accurate calculations. Sometimes formation properties in the ideal gas phase are present neither in the database nor in the literature. What gSAFT models allow to do is to use not only the formation properties in the ideal gas state but also in the pure liquid phase state and in the infinite dilution state. gSAFT automatically converts these quantities in such a way to make them usable to compute the ideal gas state. Whereas if an activity coefficient model is employed only one among the three types of formation properties mentioned can be used. In the absence of any kind of formations properties what gSAFT does is to approximate their value using Joback method (Joback, 1987) that according to the author of the original paper doesn't provide for high accuracy formations properties values. For this reason, the latter option should be avoided when possible. In Figure 1.10 is shown the thermodynamic path from element reference state to real mixture, in the sketch it is highlighted when gSAFT comes into play, namely in the conversion from element reference to pure ideal gas state and from ideal gas mixture state to real mixture one. In summary, using SAFT-γ Mie as thermodynamic model it is possible to compute every thermodynamic property provided that: the formation properties are set (if not present in the database) the ideal gas heat capacity of each compound as pure is available (readily available and it is a function only of temperature) and the interactions between the functional groups that characterize the system species are well defined.

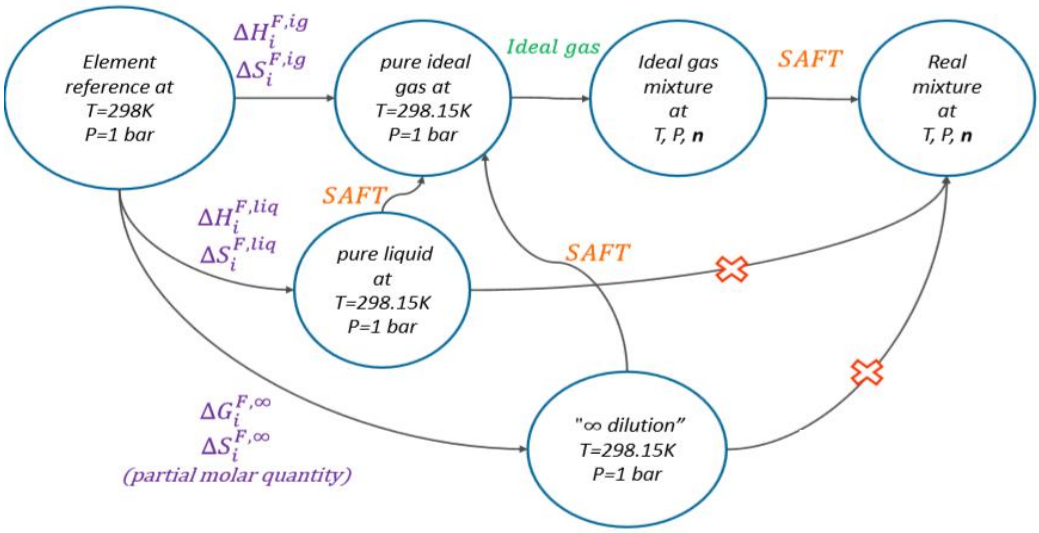


Figure 1.10 *Thermodynamic path to compute chemical potential in a generic system*

1.4 Motivation and objective of the project

The most part of chemical processes involve reactive systems and separation sections. These two types of process unit usually require the use of a solvent that can be either highly beneficial or detrimental to the system considered. For this reason, being able to predict, even from a qualitative point of view, the effect of different solvents in a specific system, would improve considerably the process design procedure reducing drastically the number of experiments required saving money and time. Moreover, pharma and specialty chemical industries are characterized by a dynamic product turnover, therefore new molecules are continuously introduced to the market for which few or no experimental data are available. For this reason, the capability to calculate thermodynamic properties and phase equilibria condition for these molecules as pure or in mixtures is highly beneficial because it drastically reduces the experimental data required. The improvement on the reactive and separation systems representation can be theoretically achieved employing first principle models, that together with a group contribution method as thermodynamic model, should be able to handle even mixtures for which few or no experimental data are available. The objective of this project is, therefore, to examine and quantify the potential impact of the use of rigorous thermodynamics and SAFT-γ Mie as thermodynamic model within solubility and kinetic models in terms of the ability to use such models to:

- analyse and correctly interpret experimental data
- predict process behaviour over wide ranges of operating conditions.
- improve the calculation accuracy

In order to make these assessments three case studies have been analysed and for each of them different aspects based on the case study characteristics have been considered. The first case study is related to solubility calculations of a complex molecule as solute while the other two deal with common reaction systems in the pharma and specialty chemical industry. For what regards the former the following aspects are considered:

- quantitative and qualitative assessment of solubility prediction
- evaluation of the possibility to employ the prediction obtained in solvent selection and solvent screening process
- possibility to find a quicker alternative to brute force experimentation

In the reaction unit operation study what has been analysed is the capability to capture the solvent effect on reaction kinetics and reaction equilibrium for system characterised by an increasing level of complexity. The first aspect that is evaluated is the accuracy improvement in the calculation with respect empirical/semi-empirical model

concentration-based typically used when experimental data are available. The second and main one is the assessment of the prediction capabilities in case of experimental data lack.

Chapter 2

Solubility prediction

This chapter deals with the first case study faced during the development of the project namely the solubility prediction of ibuprofen in several solvents. Firstly, the main industrial drivers and the resulting objectives are described. Solid-Liquid equilibrium expressed through classical thermodynamics is presented and it is followed by the derivation of the models employed based in gSAFT capabilities. The results obtained applying two of the models described are reported and discussed from both the quantitative and the qualitative point of view. Then, the results carried out are used to perform both solvent screening and solvent selection process bringing to the most suitable solvents for the specific system. Finally, an alternative approach is separately presented because regardless the high potential is not yet applicable.

2.1 Solubility role in crystallization process

Crystallization is a complex, multi-phase unit operation used in a wide range of manufacturing industries to achieve separation and purification of products. For the pharmaceutical industry, active pharmaceutical ingredient (API) crystallization may be regarded as the first step in the formulation process with molecules stabilized within the crystal lattice throughout the subsequent processing steps until the crystal dissolves upon administration to the patient allowing the molecular form of the drug to be absorbed. As a consequence, crystallization is a critical process step. The pharmaceutical industry is also placing increasing demands on crystallization, for example, as drug structures become more complex they can be more challenging to crystallize. The main reason is related to the poor solubility that may characterize new API and drug.

In Figure 2.1 is shown the general framework for designing a crystallization process. This work covers the first three stages; in particular the focus is on how to obtain the maximum number of the chemical compound physic and thermodynamic properties through predictions, namely knowing only the functional groups that characterized the chemical specie analyzed. The same approach has been used, then, to compute solubility calculations that have been carried out without any experimental data on the system of

interest. This is made possible by the use of SAFT-γ Mie that by design, as described extensively in chapter 1, allows, as an advanced thermodynamic model, to predict the properties of interest.

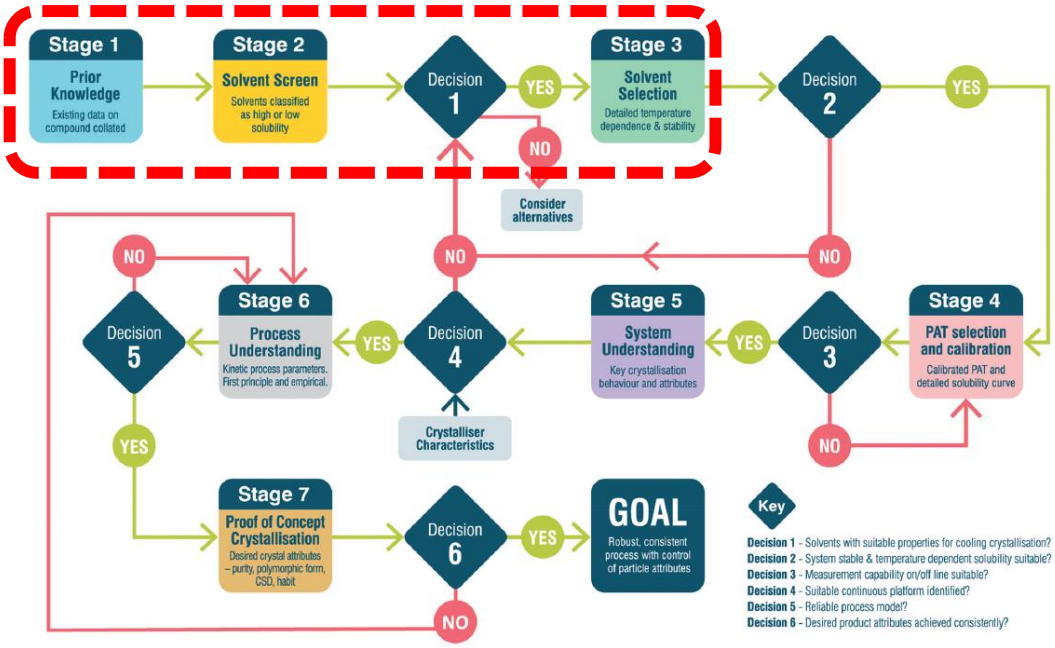


Figure 2. 1 *General workflow for crystallization process (J.Brown et al. 2018)*

The first stage requires the collection of all the information available on the API of interest. This can include molecular design, synthetic route development, solvate propensity, reactivity and impurity profiles. The second and the third stages regard solvent screening and selection, namely discrimination among the wide range of solvents available. The criteria applied to discriminate among different solvents can vary depending, for instance, on the operating conditions and on the characteristics of the API we want to purify. The criteria taken into account in this thesis are solubility magnitude, solubility temperature dependence, absolute and relative yield. All of these will be computed through SAFT-γ Mie as thermodynamic model.

In pharmaceutical manufacturing, even if predictive models have been applied, they don't currently provide sufficiently accurate quantitative prediction for a wide range of compounds. However, they can be used to give qualitative rankings of solvent/solute solubility giving an essential contribution in moving from stage two to stage four. Thus, solubility is one of the most important properties that needs to be understood for the production and purification of pharmaceuticals. The most common approaches used to estimate the solubility of drug substances are empirical/semi-empirical correlations.

These methods require a large amount of experimental solubility data to fit the respective model coefficients. Hence their reliability mostly depends on the amount of experimental data available to correlate the coefficients. Moreover, it's not possible extrapolate values outside the experimental range available. The experiments required are usually carried out preparing solutions of known concentration, *C* (molar or mass base), and subjecting them to temperature cycles to determine the point of dissolution. This procedure can be both time consuming and expensive, hence a model that can compute solubility calculations without the need of such experimental data would be beneficial under multiple aspects. A third method that overcomes the weaknesses of the empirical and semi-empirical approaches is based on the application of rigorous thermodynamics, namely the equality of chemical potential of the solute in all the phases in which it is present (solid and liquid).

2.2 Objectives

The main limitations in solubility models is the need of experimental solubility data and the inability of extrapolate outside the operating conditions in which they are carried out. This is especially the case with new pharmaceutical drugs whose structure becomes more and more complex. Due to these issues there is a need to provide a model that is capable of computing solubility calculations without requiring experimental solubility data and one that can handle solutes and even solvents for which few or no physical properties are available. Examples include complex pharmaceutical molecules such as fenofibrate, ibuprofen, ketoprofen, paracetamol etc). In order to fulfill the former requirement a model derived from solid-liquid phase equilibrium conditions has been employed. This model requires only the knowledge of the heat of fusion of the specific solute and the temperature at which the heat of fusion is observed. The latter requirement is fulfilled using SAFT-γ Mie advanced thermodynamic model that allows to compute all the physical and thermodynamic properties of pure species or mixtures setting only the functional groups that formed each component in the system. The main objective of the study is, therefore, to assess the qualitative and quantitative solubility predictions using two different models described in the next section. What they have in common is the need of few experimental data to be carried out, hence the aim is to minimize time, resources required and costs. The assessment of the model performance has been followed by a practical application: solvent screening and solvent selection processes. The system

analyzed uses ibuprofen as a solute in seven different solvents. The objective will be to assess if SAFT-γ Mie can be used to predict and correctly rank the solubility of ibuprofen in these solvents.

2.3 Solid-Liquid equilibria calculations

The general phase equilibrium conditions for a neutral system comprising NC chemical compounds i = 1,…,NC distributed over a number NP of phases k = 1,..,NP at a given temperature T and pressure P are such that the total Gibbs free energy of the system is minimized. This leads to the following expression:

$$\mu_i^{[k]}(T,P,n_k) = \mu_i^{[1]}(T,P,n_1) \qquad k = 2,\dots,NP \tag{2.1}$$

For solubility calculations in a pharmaceutical context, our interest is focused primarily on systems comprising only solid and liquid phases. Hence for compounds i that exist both in solid and liquid phases (e.g. non-dissociating or weakly dissociating APIs), equation (2.1) is modified to remove the composition dependence of the chemical potential in all solid phases. The assumptions made are the following: no appreciable solubility of the liquid solvent in the solid phase and pressure effect neglected. Under these consideration (2.1) becomes:

$$\mu_i^{[k]}(T,P) = \mu_i^{[1]}(T,P,n_1) \tag{2.2}$$

For any solid phase k comprising compound i. For simplicity, phase 1 is considered always a liquid. Equation (2.1) remains unchanged for other liquid phases k. Therefore, the only thermodynamic property required for solubility calculations is the chemical potential of each compound present in the system. In particular, there is no need for introducing concepts such as “reaction equilibrium constants (or related quantities such as pKa) or “solubility products”. Since in general it is needed to manage chemical transformations (e.g. dissociation of composite compounds from solid to liquid phases; liquid-phase reactions), it is needed to adopt a reference datum of chemical elements at a reference temperature, T^{θ} and pressure, P^{θ}. Thus, the chemical potential of compound i in a liquid phase l can be computed by:

$$\mu_i^{[l]}(T,P,n) = \mu_i^{ig}(T,P,n) + \mu_i^{res}(T,P,n) \tag{2.3}$$

Here μ_i^{ig} is the ideal gas chemical potential of compound i given by:

$$\mu_i^{ig}(T,P,n) = \Delta H_i^{F,ig} - T\,\Delta S_i^{F,ig} + \left[\int_{T^\theta}^{T} cp_i^{ig}\,(T')\,dT' - T\int_{T^\theta}^{T}\frac{cp_{p,i}^{ig}(T')}{T'}\,dT'\right] \quad (2.4)$$

$$+ RT\ln\left(\frac{P}{P^\theta}\right) + RT\ln(x_i)$$

The residual term can be computed in terms of fugacity, symmetric activity and asymmetric activity but in this work only the fugacity approach is reported and applied for SLE purpose:

$$RT\ln(f_i(T,P,n)) \equiv RT\ln(P) + RT\ln(x_i) + \mu^{res}(T,P,n) \quad (2.5)$$

where :

- T^θ [K] and P^θ [Pa] are, respectively, the standard temperature and pressure
- $\Delta H_i^{F,ig}\ \left[\frac{\mathrm{J}}{\mathrm{mol}}\right]$ and $\Delta S_i^{F,ig}\ \left[\frac{\mathrm{J}}{\mathrm{mol\ K}}\right]$ are respectively, the enthalpy and entropy of formation of compound i in the ideal gas standard state at T^θ and P^θ
- $cp_i^{ig}\ \left[\frac{\mathrm{J}}{\mathrm{mol\ K}}\right]$ is the ideal gas specific heat capacity of compound i
- x_i is the molar fraction of compound i

For what regards the solid-state chemical potentials, two alternative approaches that differ in the requirements for solid-state properties have been reported. In particular, the first one makes use of enthalpy and entropy of formation, while the second relies on properties related to the melting point of the solid phase.

2.3.1 Solid-phase chemical potential in terms of formation properties

Although equation (2.3) holds always true, there is currently no generally applicable way of computing the residual term, μ_i^{res}in case of solid phase. In particular, both equations of state and activity coefficient models are restricted to phases exhibiting liquid-like structures only, and cannot describe the behavior of phases where the underlying microscopic structure follows a well-defined geometric pattern such as crystalline solid phases. It worth noting that amorphous phases often exhibit liquid-like structures and can be described by fluid-phase Equation of State. Therefore the chemical potential of a compound i in a solid phase corresponding to a particular crystal structure as:

$$\mu_i^{s}(T) = \Delta H_i^{F,s} - T\,\Delta S_i^{F,s} + \left[\int_{T^\theta}^{T} cp_{p,i}^{s}\,(T')\,dT' - T\int_{T^\theta}^{T}\frac{cp_{p,i}^{s}(T')}{T'}\,dT'\right] \quad (2.6)$$

Where:

- $\Delta H_i^{F,s}\ \left[\frac{\mathrm{J}}{\mathrm{mol}}\right]$ and $\Delta S_i^{F,s}\ \left[\frac{\mathrm{J}}{\mathrm{mol\ K}}\right]$ are, respectively, the enthalpy and entropy formation of the solid phase of the same crystal structure at T^{θ} and P^{θ}.
- $cp_{p,i}^{s}\ \left[\frac{\mathrm{J}}{\mathrm{mol\ K}}\right]$ is the specific heat capacity of the solid at temperature T and pressure P^{θ}

The above equation omits the pressure dependence of $\mu_i^s(T)$ as this is typically very small at the pressures of interest to pharmaceutical applications. Pure compound parameters involved:

- Enthalpies and entropies of formation of crystalline phases at standard temperature T^{θ} and pressure P^{θ}.
- Parameters describing the temperature dependence of solid-phase specific heat capacities.

2.3.2 Solid-phase chemical potential in terms of melting properties

As aforementioned, since an equation such as SAFT-γ Mie cannot compute the chemical potentials in the solid phase directly, the idea behind this method is to relate the chemical potential of component i in the solid phase to the chemical potential of component i in a pure liquid phase, quantity that SAFT-γ Mie can compute. Therefore (2.5) can be re-write as:

$$\mu_i^s(T) = \mu_i^l(T,P) + \Delta G^{l\to s}(T,P) \tag{2.7}$$

Where $\mu_i^L(T,P)$is defined as :

$$\mu_i^L(T,P) = \Delta H_i^{F,ig} - T\,\Delta S_i^{F,ig} + \left[\int_{T^{\theta}}^{T} cp_{p,i}^{ig}\,(T')\,dT' - T\int_{T^{\theta}}^{T}\frac{cp_{p,i}^{ig}(T')}{T'}\,dT'\right] + \tag{2.8}$$

$$+ RT\ln\left(\frac{P}{P^{\theta}}\right) + \mu_i^{res}(T,P)$$

It worth noting that equation (2.7) is not a phase equilibrium relation, but a simple integration between two different thermodynamic states:

Liquid at T → Liquid at T_m → Solid at T_m → Solid at T

One possibility to express $\Delta G^{l\to s}(T,P)$ is in terms of melting enthalpy and melting temperature. The relation can be obtained by integrating the enthalpy and entropy of both

the solid and liquid state from the melting temperature T_m(measured at pressure P_m) to the temperature and pressure of interest, $T$and P.

$$\Delta G^{l\to s}(T,P) = \Delta H^{l\to s}(T,P) - T\Delta S^{l\to s}(T,P) \tag{2.9}$$

$$\Delta G^{l\to s}(T,P) = H^{s}(T,P) - H^{l}(T,P) - T\Big(S^{s}(T,P) - S^{l}(T,P)\Big) \tag{2.10}$$

The thermodynamic integration for the enthalpy and entropy of the liquid is given by:

$$H^{s}(T,P) = H^{s}(T_m,P_m) + \int_{T_m}^{T} \mathrm{C_p^s}(T',P)dT' + \int_{P_m}^{P} \mathrm{d}H^{s}dP' \tag{2.11}$$

$$S^{s}(T,P) = S^{s}(T_m,P_m) + \int_{T_m}^{T} \frac{\mathrm{C_p^s}(T',P)}{T'}dT' + \int_{P_m}^{P} dS^{s}dP' \tag{2.12}$$

where $\mathrm{C_p^s}(T,P)$ denotes the heat capacity of the solid phase.

Equivalent expressions can be written for the metastable liquid phase:

$$H^{l}(T,P) = H^{l}(T_m,P_m) + \int_{T_m}^{T} \mathrm{C_p^l}(T',P)dT' + \int_{P_m}^{P} \mathrm{d}H^{\mathrm{l}}dP' \tag{2.13}$$

$$S^{l}(T,P) = S^{l}(T_m,P_m) + \int_{T_m}^{T} \frac{\mathrm{C_p^l}(T',P)}{T'}dT' + \int_{P_m}^{P} dH^{l}dP' \tag{2.14}$$

Therefore, the term $\Delta G^{l\to s}(T,P)$ in equation (7) can be written as:

$$\Delta G^{l\to s}(T,P) = \Delta H^{l\to s}(T_m,P_m) - \mathrm{T}\Delta S^{l\to s}(T_m,P_m) + \int_{T_m}^{T} \Delta C_p^{l\to s}(T',P)dT' - \tag{2.15}$$

$$-T\int_{T_m}^{T} \frac{\Delta C_p^{l\to s}(T',P)}{T'}dT' + \int_{P_m}^{P} \Delta^{\mathrm{l}\to\mathrm{s}}(\mathrm{d}G)dP'$$

where

$$\Delta C_p^{l\to s}(T,P) = \mathrm{C_p^s}(T,P) - \mathrm{C_p^l}(T,P) \tag{2.16}$$

and

$$\int_{P_m}^{P} \Delta^{\mathrm{l}\to\mathrm{s}}(\mathrm{dG})dP' = \int_{P_m}^{P} \mathrm{dG^s}dP' - \int_{P_m}^{P} \mathrm{dG^l}dP' \tag{2.17}$$

The expression used to compute $\Delta G^{l\to s}(T)$ term can be simplified by noticing that, at the melting temperature, $\Delta G^{l\to s}(T_m,P_m) = 0$, so that:

$$\Delta H^{l\to s}(T_m,P_m) = T_m\Delta S^{l\to s}(T_m,P_m) \tag{2.18}$$

Therefore, neglecting the pressure dependence we can write the following relation:

$$\Delta G^{l\to s}(T) = \Delta H^{l\to s}(T_m)\left(1-\frac{T}{T_m}\right) + \int_{T_m}^{T}\Delta C_p^{l\to s}(T')dT' - T\int_{T_m}^{T}\frac{\Delta C_p^{l\to s}(T')}{T'}dT' \qquad (2.19)$$

where:

- T_m is the melting temperature of the solute.
- $\Delta H^{l\to s}(T_m)$ is the enthalpy of fusion of the solute at the melting temperature.
- $\Delta C_p^{l\to s}$ is the difference between the heat capacity in the solid phase and the in the liquid one of the solute.

Usually it is possible to assume that ΔC_p is independent on temperature and to bring it outside the integral. This approximation leads to negligible errors in solubility calculations only if the system temperature is not too far from the melting temperature of the solute. The melting properties can derive from measurements (e.g. Differential Scanning Calorimetry (DSC)) or can be regressed from experimental data if available. In few cases triple point properties can be available instead, in this circumstance it is still possible to use this approach provided that the triple point temperature and the enthalpy of sublimation are substituted to the melting properties in (2.19).

2.4 Solubility predictions of Ibuprofen in seven different solvents

Solubility of Ibuprofen in pure solvents and in solvent mixtures plays a key role in the crystallization process. It determines, at the condition of the process, the production rate and the yield. The solubility also determines the method by which supersaturation is generated in the process and how the supersaturation varies during the course of operation. The methods usually applied in crystallization to obtain supersaturation conditions are the following: solvent evaporation (evaporative crystallization), solvent cooling (cooling crystallization) and solvent dilution (antisolvent crystallization). Only few solubility data for ibuprofen in the solvents taken into consideration are available in the literature. What has been carried out is, therefore, a test of the solubility prediction of Ibuprofen in different solvents versus experimental data taken from the literature using a custom test-solubility flowsheet in gFP environment.

2.4.1 System description

The solubility of ibuprofen in a temperature range of 10 °C to 35 – 45 °C and using seven different solvents has been performed. The solvents considered in the study are: 1-propanol, 2-propanol, toluene, acetone, methanol, 4-methyl-2-pentanone and ethanol. The experimental solubility data have been taken for 1-propanol, 2-propanol, acetone and ethanol from Wang et al. (2010) and for toluene, methanol and 4-methyl-2-pentanone from Garcin et al. (2002).

2.4.2 Model employed

The models employed are both derived from the phase equilibrium conditions, thus from a combination of the approaches described above.

2.4.2.1 Rigorous solubility calculation: melting properties approach model (MPAM)

The rigorous solubility calculation model employed has been obtained substituting equation (2.8) and (2.19) in relation (2.7) and substituting equations (2.7) and (2.3) in (2.2). After these two steps the solubility model results as follows:

$$\ln({x_i}^L) = \frac{\Delta G^{l\to s}(T)}{RT} - \ln\left(\frac{\overline{\varphi_i^L}(T,P,n)}{\varphi_i^L(T,P)}\right) \tag{2.20}$$

Where $\overline{\varphi_i^L}(T,P,n)$ is the fugacity coefficient of the solute in the solution and $\varphi_i^L(T,P)$ is the fugacity coefficient of the solute as the pure liquid. The experimental values needed for caring out the solubility prediction are the Ibuprofen melting temperature 347.15 [K] and the Ibuprofen enthalpy of fusion at the same temperature 25500 [J/mol].

2.4.2.2 Semi-empirical model (SEM)

The second model employed has been obtained substituting the terms in (2.2) with those in (2.3),(2.4) and (2.6), the equation obtained can be directly used to calculate the solubility of compound *i*, provided that, the terms of equation (2.6) are known and the residual chemical potential is computed by a thermodynamic model capable of calculate it. The latter constraint is satisfied because SAFT-γ Mie is capable of computing it but $\Delta H_i^{F,S}$ and $\Delta S_i^{F,S}$, within equation 2.6, are not always available. The same model can be rearranged in such a way to require solubility experimental data instead of formation properties. Defining the function $\Delta g_i(T)$ as:

$$\Delta g_i(T) \equiv \mu_i^s(T) - \Delta h_i^{F,ig} + T\,\Delta s_i^{F,ig} - \left[\int_{T^\theta}^{T} cp_{p,i}^{ig}\,(T')\,dT' - T\int_{T^\theta}^{T}\frac{cp_{p,i}^{ig}(T')}{T'}\,dT'\right] \quad (2.21)$$

It is possible to come up with the relations reported below.

$$\Delta g_i(T) = RT\ln\left(\frac{P}{P^\theta}\right) + RT\ln({x_i}^L) + \mu_i^{res}(T,P,n) \quad (2.22)$$

Applying some rearrangements, the explicit expression for solubility calculations can be found:

$${x_i}^L = e^{\frac{\Delta g_i(T)}{RT}} - \frac{\overline{\varphi_i^L}(T,P,n)\cdot P}{P^\theta} \quad (2.23)$$

It worth noting that $\Delta g_i(T)$ is purely a characteristic of the solute, it doesn't depend on any other compounds in the system; it depends only on temperature. Thus, the value of $\Delta g_i(T)$ for a given solute at a given temperature will be the same regardless the system in which the solute can be found. Using equation (2.22) it is possible to compute $\Delta g_i(T)$ from any available experimental solubility data for a specific solute by using the measured liquid-phase composition to replace the variables ${x_i}^L$ and n. Since equation (2.1) holds true also in multiphase systems, provided that they are at equilibrium, the solubility data can be obtained also from system characterized by the existence of other solid and/or liquid phases. What is needed is the composition of one of the liquid phases that contains the solute of interest.

It is fundamental to have an equation of state able to compute the term $\mu_i^{res}(T,P,n)$ namely the fugacity coefficient of solute in the solution. In this work the model used to compute the residual term is SAFT-γ Mie. The expression of $\Delta g_i(T)$ for Ibuprofen has been taken from Lafitte at al. (2018) that has used experimental solubility data for a wide range of solvents over a wide range of temperatures to compute the right-hand side of (2.22). The use of multiple experimental data is related to the will of demonstrate that all the points, actually, define a single function that can be well approximated by a straight line. The resulting expression is:

$$\Delta g_{ibuprofen}(T) = 238.78\,T - 117460 \quad (2.24)$$

Substituting equation (2.24) in (2.23) it is possible to calculate the solubility of ibuprofen in any other solvent for any temperature. The values that appear on the right-side of (2.21)

are not required in this case because $\Delta g_i(T)$ has been found as an empirical relation derived by fitting experimental data.

2.4.3 Solubility predictions

The solubility prediction computed using the two models described above are reported in Figure 2.2. It can be noticed that the two models provide very close solubility predictions but nevertheless, from the quantitative point of view both of them result to be poor for half of the analyzed systems.

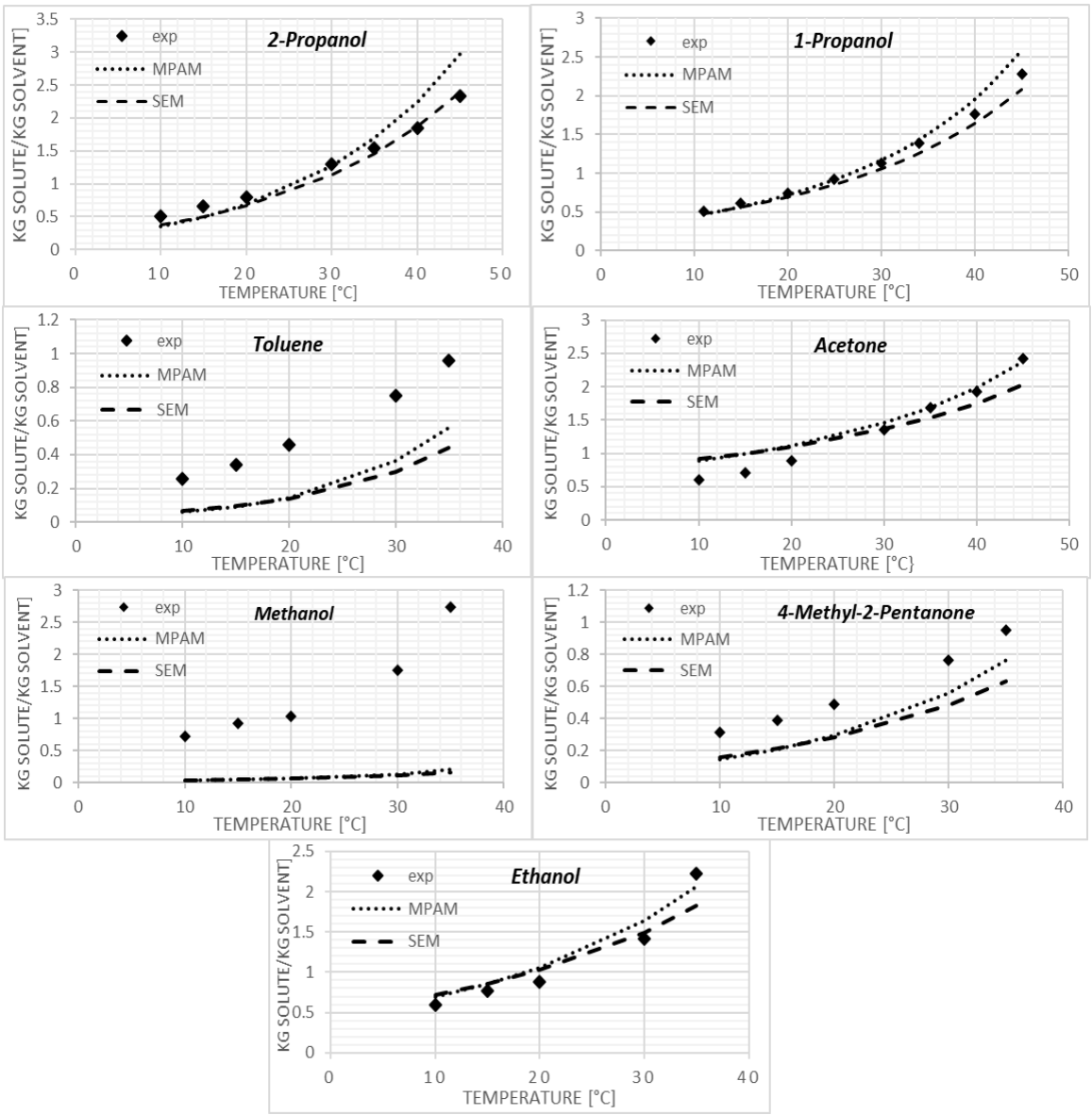


Figure 2.2 *Experimental and predicted solubility values computed applying melting properties approach model (MPAM) and semi-empirical model.(SEM). Solubility is solvent base [kg_solute] / [kg_solvent]*

From the calculations carried out is not possible stating which of the two models used give better quantitative predicted results. However, what can be stated for sure is that in both cases at least the qualitative results represent in a satisfactory way all the experimental trends (i.e there is clearly a bias, but the behavior is quite consistent), except for methanol. In order to point out the gap between the experimental solubility values and the predicted ones Figure 2.3 and Figure 2.4 have been reported, they show the relative error between the predicted values and the experimental ones for both solubility models considered.

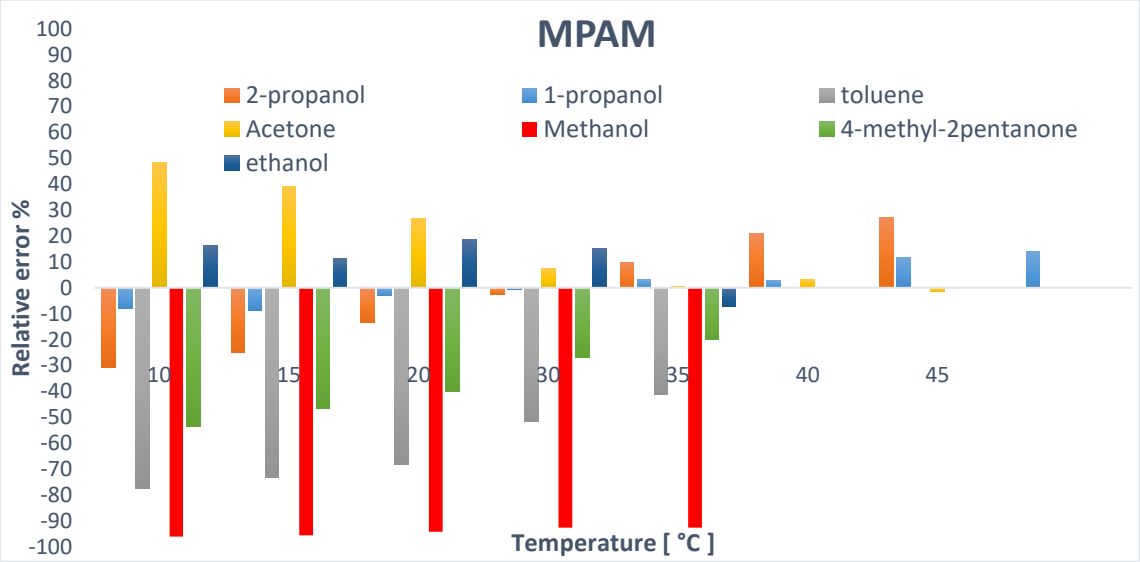


Figure 2.3 *Relative error of solubility predictions using melting properties approach model (MPAM), with respect to the experimental data available for the range of temperature 10 °C to 45 °C*

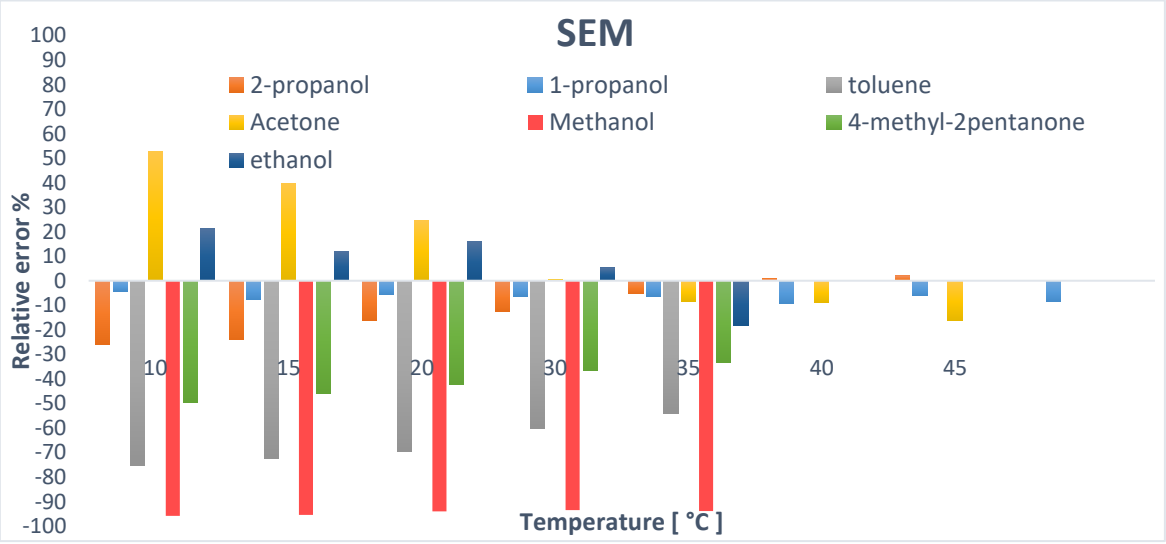


Figure 2.4 *Relative error of solubility predictions using the semi-empirical model (SEM) with respect to the experimental data available for the range of temperature 10 °C to 45 °C*

As it can be seen from the bar chart the relative error is less than 50 % except for methanol and toluene. Thus, the order of magnitude of most of the predicted solubility values is the same order of magnitude of the experiments, this gives good prospective on the reliability of the models. The accuracy of predictions depends on whether the majority of the group cross-interactions available in gSAFT database have been regressed using experimental data, instead of being approximated by combining rules. Indeed, for both methanol and toluene three group cross-interactions are approximated by combining rules and this clearly has a strong impact on the accuracy of the results. Other solvents as 2-propanol and 4-methyl-2-pentanone are characterized by some approximated group cross-interactions but the impact of such approximations in these cases doesn't affect dramatically the results. This is due to the different interaction strength between different groups. The group cross interactions approximated by combining rules employed are reported below.

Table 2.1 *Group cross interactions that characterize the system which have been approximated by combining rules.*

2-Propanol		**Toluene**		**Methanol**		**4-methyl-2-pentanone**	
Sec_OH	aCCH2	aCCH3	CH	CH3OH	aCCH2	C=O	COOH
Sec_OH	aCCH	aCCH3	aCCH2	CH3OH	aCCH	C=O	aCCH2
Sec_OH	aCH	aCCH3	COOH	CH3OH	aCH	C=O	aCH

In order to improve the accuracy of the predictions the availability of experimental data that allows the regression of such group cross interactions is required. The functional group pairs highlighted in yellow are the most critical ones because of their stronger interactions, hence they should be considered as high priority. The values predicted through the rigorous solubility calculation approach have been used to carry out both the solvent screening and the solvent selection.

2.4.3.1 Solvent Screening

The aim of this stage is to take a broad library, that in this case study is composed of seven solvents, and rapidly identify only those that show desirable solubility. The choice of solvents for this screen can be narrowed by considering, for instance, only those included in the International Conference on Harmonisation (ICH) classification for

residual solvents as class 2 (solvents that are limited in their use) or class 3 (solvents with low toxic potential).

ICH classification for residual solvents	
Class 2	**Class 3**
Methanol	Acetone
MIBK	Ethanol
Toluene	1-propanol
	2-propanol

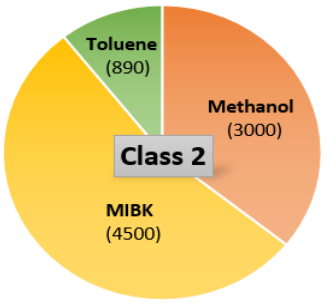


Figure 2.5 *The table on the left-hand side shows the solvents, which have been considered in the study, listed according to the ICH classification of residual solvents. The Pie chart on the right-hand side shows the concentration limit acceptable in ppm of class 2*

From Figure 2.5 it is clear that there are four solvents that are preferable. In case a solvent of class 2 is needed MIBK will be consider first, then Methanol and Toluene. Another evaluation criteria applied is the solubility value at the highest feasible temperature, for this specific case study has been considered 35 ºC.

Table 2.2 *Solubility values obtained from experiments (exp) Garcin et al. and calculated (CV) using the melting property model approach*

Solubility at 35 °C (CV)	Solubility at 35 °C (exp)
ethanol	methanol
2-propanol	ethanol
acetone	acetone
1-propanol	2-propanol
4-methyl-2-pentanone	1-propanol
toluene	toluene
methanol	4-methyl-2-pentanone

Since the results for methanol, as it is evident from Figure 2.2, are not reliable, methanol is not taken into consideration for the solvent choice procedure. From Table 2.2 the solvents that show the highest solubility result to be those belonging to class 3 so all of them have been taken to the solvent selection stage.

2.4.3.2 Solvent Selection

This stage is geared towards further narrowing the solvents selected using different criteria[1] among which there is the evaluation of the absolute (Y_i) and relative ($Y_i^{rel}\%$) yield defined as follows:

$$Y_i = \omega_i(35\ ℃) - \omega_i(10\ ℃) \tag{2.25}$$

$$Y_i^{rel}\% = \frac{Y_i}{\omega_i(35\ ℃)} \cdot 100 \tag{2.26}$$

Where ω_i is the solubility expressed on a solute free basis $\left[\frac{\mathrm{kg_{solute}}}{\mathrm{kg_{solvent}}}\right]$. Since the difference in the amount of solid dissolved in a defined temperature range is one of the key properties considered in crystallization process, the absolute yield has been considered first. Only in case of very close values of absolute yield the relative ones are taken into account. The latter can be seen as a sort of solubility "efficiency" related to the specific pair solvent-solute. Since the yield is based on solubility difference, if the qualitative trend of calculated solubility values is close enough to the experimental data one, the solvent selection should come out to be the same using either the experimental data or the predicted ones.

Table 2.3 *Absolute yield evaluated for the four solvents taken to the solvent selection stage.*

	Ethanol	2-Propanol	Acetone	1-Propanol
Absolute Yield (CV)	1.376	1.34	0.804	0.969
Absolute Yield (exp)	1.6374	1.035	1.083	0.89
Relative error %	-15.96	29.51	-25.73	8.93

From the absolute yield reported in Table 2.3 it is clear that the two solvents with higher yield are ethanol and 2-propanol. Thus, as the solvents with the highest absolute yield, they have been chosen as the solvents for which it worth looking at the relative yield.

Table 2.4 *Relative yield evaluated for the four solvents taken to the solvent selection stage*

	Ethanol	2-Propanol
Relative Yield (CV)	66.6	79.29
Relative Yield (exp)	73.42	67.19

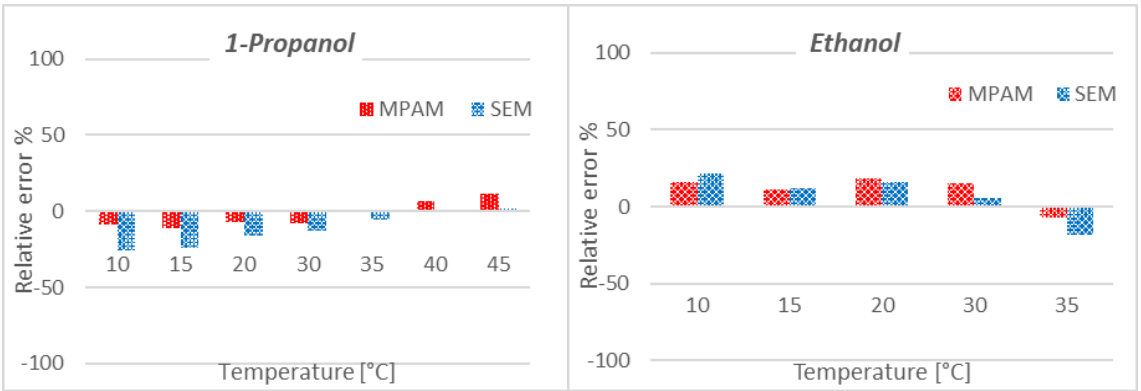
.

[1] Listed in Table 4 of C.J.Brown et al. (2018)

From Table 2.4 it is not clear which of the two solvents, effectively, has the highest relative yield. It worth notice that the error of predicted solubility value with respect the experimental one for 2-propanol is positive at 35 ºC and negative at 10 ºC; while for ethanol results to be exactly the opposite. Thus, since the yield calculations are based on the difference between these two values, they result to be overestimated in case of 2-propanol and underestimated in case of ethanol. For this reason, it cannot be stated, with confidence, which of the two solvents is the most suitable one. At this stage, both solvents could potentially be utilized for the cooling crystallization process to be designed. Further selection criteria for the desirable solvent(s) could be applied in the solvent selection process. A key additional selection criterion for crystallization process design, would be the solid form produced of the API (solute) from the acceptable solvents, assessed via a solid form screen. This is outside of the scope or capabilities of gSAFT.

2.4.4 Results and discussion

The solubility values computed by the two models employed result to be in good agreement from both quantitative and qualitative sides for 1-propanol and ethanol as it can be seen from Figure 2.6*a* and Figure 2.6*b* which show that the relative error is always lower than 20%. In both cases all the group cross interactions needed are present in the databank as regressed from experimental data.



(a) *(b)*

Figure 2.6 *Relative error of Ibuprofen solubility predictions employing MPAM and SEM with respect experimental data in Ibuprofen-Ethanol (b) and Ibuprofen-1-Propanol (a) systems.*

The relative error for Ibuprofen-Acetone system is shown in Figure 2.7. As it can be seen, although all the group cross interactions have been regressed from experimental data the

℃

results are not accurate as the previous cases. Applying the rigorous solubility calculation model, as it can be seen also in Figure 2.2, the predictions are accurate both quantitatively and qualitatively, but they worsen for low temperature. While, using the semi-empirical model the results are inaccurate both at high and low temperature.

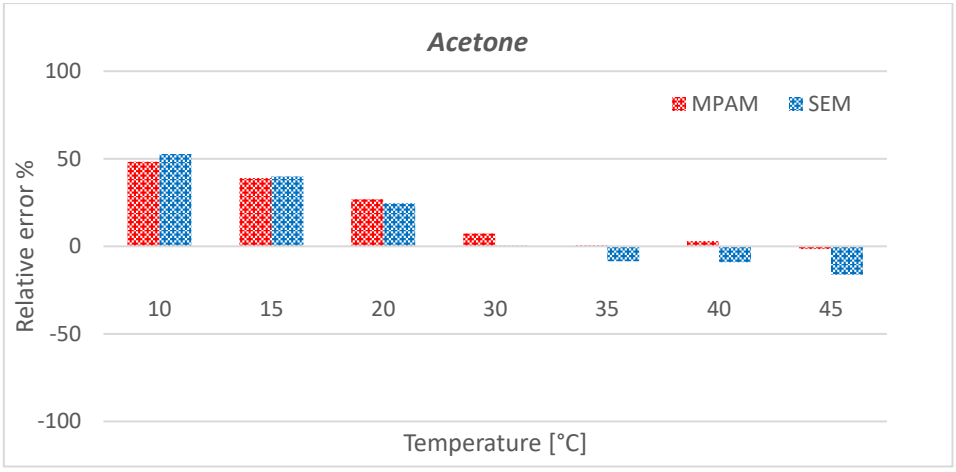


Figure 2.7 *Relative error of Ibuprofen solubility prediction in Acetone employing MPAM and SEM with respect experimental data*

For what regards the systems involving Toluene and 4-Methyl-2-Pentanone as it can be seen from Figure 2.8, what is in good agreement with the experimental solubility values it is only the qualitative trend of the prediction. The quality of prediction of absolute solubility values is poor. As shown in Table 2.1, some group cross interactions needed to correctly define these systems are approximated by combining rules. It these specific cases seem that the lack of group cross interactions regressed using experimental data have led to a shift in the predicted values.

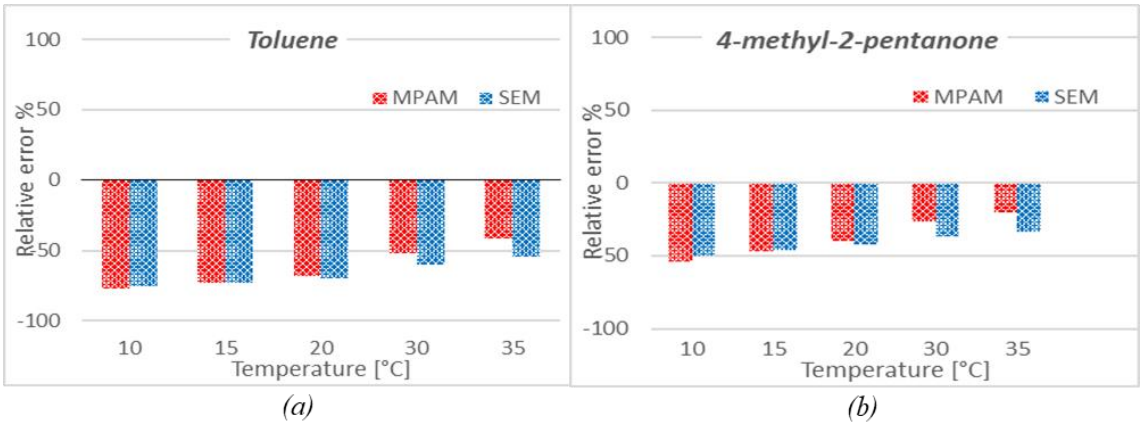


Figure 2.8 *Relative error of Ibuprofen solubility predictions employing MPAM and SEM with respect experimental data in Ibuprofen-Toluene (a) and Ibuprofen-4-Methyl-2-Pentanone (b) systems*

The last system considered is Ibuprofen-Methanol, this is the only one that is characterized by both quantitative and qualitative poor predicted solubility values. As it can be seen from Figure 2.9 the relative error is greater than 90% for each experimental point.

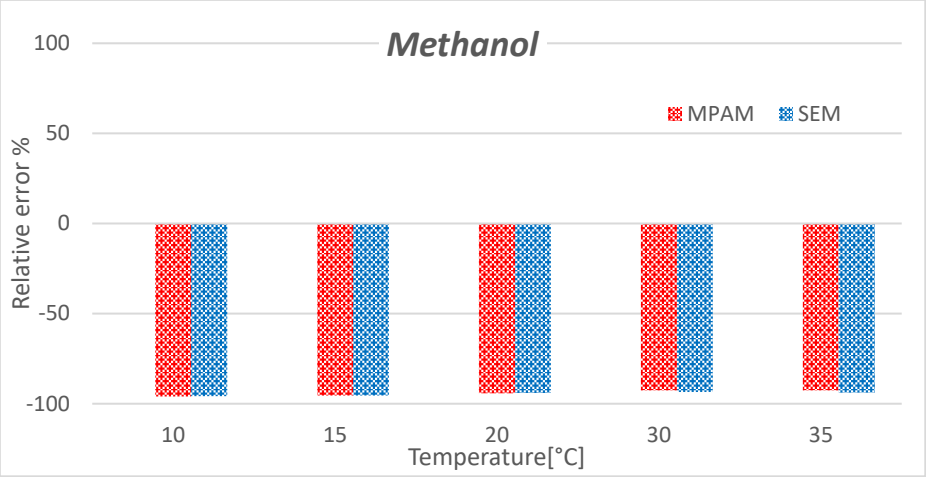


Figure 2.9 *Relative error of Ibuprofen solubility predictions in Methanol employing MPAM and SEM with respect experimental data*

It worth noting that for this system three group cross interactions are approximated through combining rules but the group cross interaction for the functional groups pair that shows the strongest interaction (CH3OH – COOH) has been regressed from experimental data.

2.4.5 Conclusion and further improvements

From the study carried out, it has been possible to understand when it is worth using the rigorous solubility calculation approach that implies the use of a thermodynamic model for the residual term computation. The assessment is also related to the degree of accuracy required. The main steps to take into considerations are the following:

- Evaluate if all the group cross interactions needed to characterize the system of interest are present as regressed from experimental data.
- In the case that all of them are in the database, what is expected is that both qualitative and quantitative predictions be in good agreement with the experimental data.

If some of group cross interactions are approximated by combining rules, it is important to assess which type of functional groups involved:

- In case of group cross interaction between non-polar groups the predictions will likely result in being qualitatively accurate but quantitatively inaccurate.
- On the other hand, if the functional groups are both polar (e.g. COOH and Sec_OH) certainly, the predictions expected would be both qualitatively and quantitatively inaccurate. Since methanol is a common solvent in the pharmaceutical industry and is the one for which the worst results have been obtained, it is worth analyzing the system behavior in more details.

The group cross interactions are obtained from experimental VLE data, this means that the interactions considered are representative of the fluid phase. Since this study deals with SLE systems, the interactions between the same functional groups may need to be refined. In order to understand if this is the case for the methanol and Ibuprofen system, the functional groups with stronger interactions have been considered. The pair methanol – carboxyl group is the only one characterized by both polar functional groups so it is the one with the greatest impact on solubility prediction computation. Once identified the strongest pair both VLE and SLE calculations have been carried out for binary mixture characterized by species which contain methanol and carboxyl group. The binary mixture for which the VLE has been computed is the butyric acid and methanol pair; the results are shown in Figure 2.10.

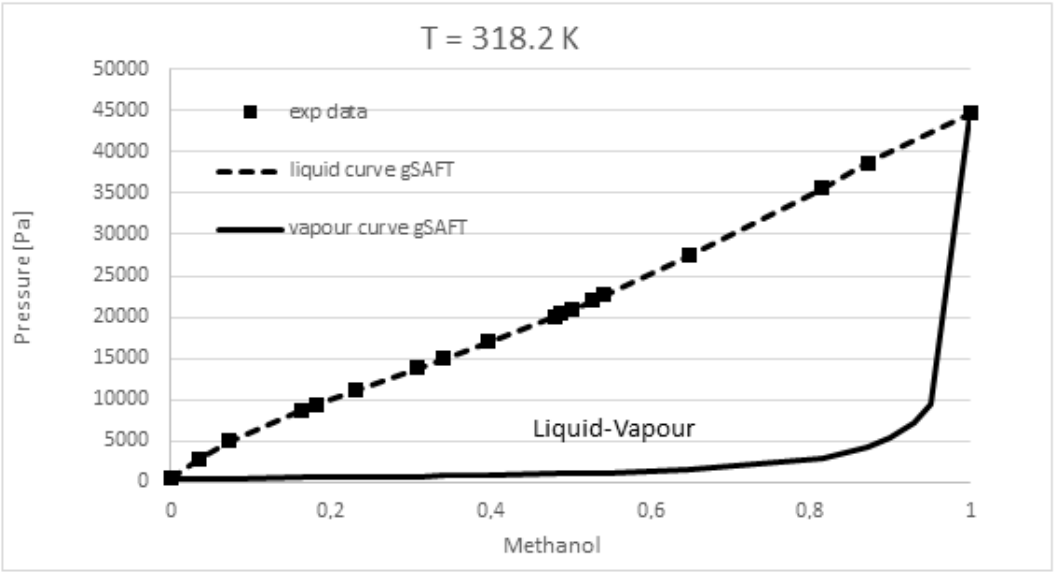


Figure 2.10 *Vapor-Liquid equilibria for Methanol-Butyric Acid binary mixture.*

As it can be seen from Figure 2.10, the predictions are in perfect agreement with the experimental data. This means that the interactions between the functional group of interest, in the fluid phase, are very well described. The results for lauric acid-methanol SLE are reported in Figure 2.11.

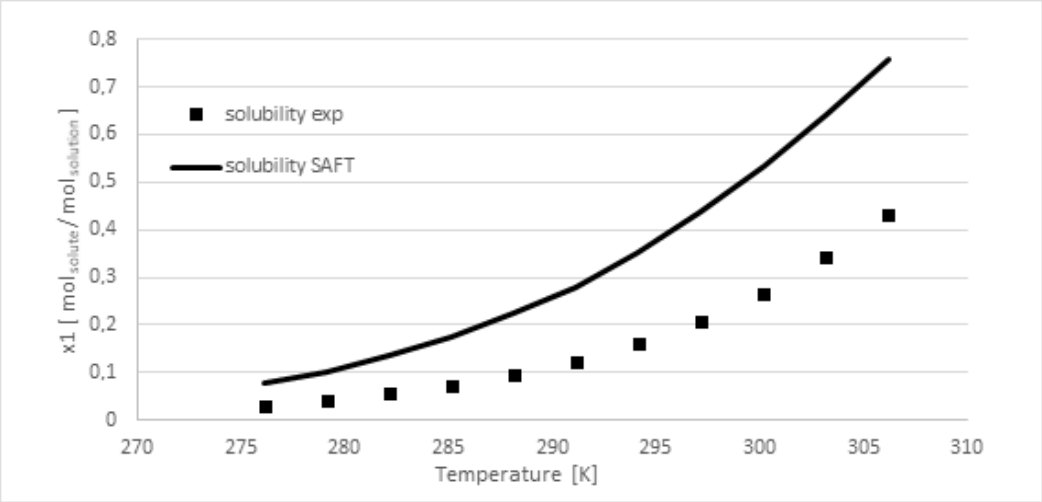


Figure 2.11 *Solubility of lauric acid in methanol for a range of temperature 276.17-306.17 K at 1 bar.*

The solubility x_1 is defined as:

$$x_1 = \frac{m_1/M_1}{\frac{m_1}{M_1}+\frac{m_2}{M_2}} \tag{2.27}$$

Where m_1 represents the mass of solute, m_2 the mass of the solvent and M_i the molecular weight. From the computation of VLE and SLE it is clear that the latter is not in agreement with the experimental value as the former one. The different interactions nature between the same functional groups in different phases cannot be neglected. For this reason, in order to rely on solubility predictions carried out using SAFT-ϒ Mie, it is needed to regress the group cross interactions from both SLE and VLE data. The new parameters obtained would theoretically keep the VLE predictions accurate while significant improvements in the SLE computation are expected.

2.5 Application of TPFlash as further solubility computation method

So far it has been presented the solubility predictions resulted from the use of two models derived from rigorous thermodynamic relations. gPROMS will be able to provide an alternative approach to carry out the solubility computation as rigorous as the two models aforementioned. This is possible through the use of "flash properties calls" (e.g. TPFlash) that is capable to provide the composition in total amount of moles, the total enthalpy and volume of each phase, at equilibrium condition, present in the system. What is required from the function is just the temperature, pressure and composition of the system of interest. So far TPFlash has been used only for VLE and LLE, but it has been recently taken into account the possibility of creating an enhanced version that will be able to handle solid phases as well.

The thermodynamic and mathematical approach behind the phase equilibrium computation is described as follows. The phase equilibrium conditions applied for solubility calculations assume that both the number of phases (*NP*) that are present in the system and their nature are known. In general, this information is not always available. One approach, to establish the complete set of phases that are present at the temperature T and the pressure P of interest, is to start with a single liquid phase comprising all of the specified amounts, and then test its stability with respect to all solid phases that may potentially form given the set of compounds under consideration. In general, such solids may include both pure API phases and API salts, solvates, hydrates and/or co-crystals, some of which may exist in polymorphic forms. A stability criterion based on the Gibbs surface tangent plane (Michelsen, 2007) can be applied to reach this aim. Given a fluid phase of composition **n**, it tests whether the formation of an infinitesimally small amount of a new "trial" phase would result in an overall reduction of the Gibbs free energy. If no such trial phase can be found, the original fluid phase is deemed to be stable. Mathematically, this is equivalent to requiring that the following quantity:

$$\Omega_G(\boldsymbol{w}) = \sum_{i=1}^{NC} w_i(\mu_i^{trial}(T, P, \boldsymbol{w}) - \mu_i(T, P, \boldsymbol{n})) \tag{2.28}$$

is nonnegative for any is the trial composition **w**. Testing the stability of a liquid phase with respect to solid trial phases comprising a single compound that exist in both the solid and the liquid phase lead to a simplified form of criterion (2.28):

$$\mu_i^{[s]}(T,P) \geq \mu_i^{[l]}(T,P,\boldsymbol{n}) \tag{2.29}$$

A slightly different relation is obtained for compounds that dissociate in the liquid phase. Given a liquid phase of composition, $\boldsymbol{n}$, the above stability criteria can be checked for each and every solid phase s under consideration using the values of the liquid and solid phase chemical potentials. If the criterion is found to be violated for any phase s, then that phase may be present in the system, and therefore must be added to the set of phases included in the formulation of the phase equilibrium conditions. The solution of the equations resulted from this procedure will then yield a new liquid-phase composition $\boldsymbol{n}$ which again can be tested for stability against potential existence of more solid phases. The procedure is repeated until a stable liquid phase is reached. In some cases, more than one trial solid phase, for a given liquid-phase composition, may be found to violate the stability criterion. In these situations, the approach adopted is tentatively add each such solid phase s separately to the system, perform the associated solubility calculation, and then use the resulting equilibrium compositions to compute the total Gibbs free energy of the system:

$$G_S = \sum_{k=1}^{NP} \sum_i n_i^{[k]} \mu_i^{[k]}(T,P,n_k) \tag{2.30}$$

The trial solid phase s that results in the lowest free energy G_S is selected to be added to the system, and the corresponding liquid-phase composition is subjected to further stability. Being able to apply this approach would lead to a much faster and simpler way to compute solubility calculations.

The required specifications are as follows:

- If a specie can be found, eventually, in the solid phase
- If the species that for their nature can only be found in the solid phase and therefore how these dissociate in the fluid phase.

The user doesn't need to know a priory neither the number of phases (*NP*) that are present in the system nor their nature. When the solubility is analyzed, information about composition enthalpy and volume of all the phases present at equilibrium are provided as well. So far, the use of TPFlash for the described purpose has not been assessed because the final implementation is not ready to be released yet. For this reason, it hasn't been possible extending the solubility prediction accuracy evaluation also to this approach.

Chapter 3

Prediction of solvents impact on reactive systems

This chapter is structured in such a way to combine the theoretical description of solvent influence in generic reactive system, the derivation of first principle models to describe reactive systems and a practical application on a system of industrial interest: transesterification reaction. Therefore, firstly, the solvent effects in liquid and multiphase reactive systems is defined; then, starting from the thermodynamic definition of reaction equilibria, a rigorous thermodynamic kinetic model has been derived and compared to the currently most employed one, namely the power law concentration-based kinetic model. Once the critical role of solvents in reactive systems is known and the theoretical background is well understood, the transesterification case study (pure liquid phase system) is described.

3.1 Solvent influence in reactive systems

The organic chemists usually work with compounds, which possess labile covalent bonds and are relatively involatile, thereby often rendering the liquid-phase a more suitable reaction medium than the gas one. Of the thousands of reactions known to occur in solution only few have been studied in the gas-phase, even though a description of reaction mechanisms would be much simpler. In a liquid-phase system, indeed, the reaction depends on a larger number of parameters than in the gas-phase. Consequently, the experimental results can often be only qualitatively interpreted because the state of aggregation in the liquid phase has so far been insufficiently studied. On the other hand, the fact that the interaction forces in solution are much stronger and more varied than in the gas-phase, permits to affect the properties and reactivity of the solute in manifold modes. Therefore, to carry out a chemical reaction it is not sufficient to take into consideration the proper reaction vessels and the appropriate reaction temperature. One of the most important features for the success of a reactive process is the selection of a suitable solvent. Solvents create the reaction environment for catalytic liquid-phase reactions (e.g. transesterification), multiphase reactions (e.g. hydrogenations) and

crystallization processes. The choice of the solvent is never arbitrary because it can be highly beneficial or detrimental to the system considered. With a focus on heterogeneous catalytic reactions, the solvent can also significantly affect both kinetic and selectivity, due to interactions with the catalyst. As well as influencing reactants and products behaviour the solvent may also interact with the metal and/or support of the catalyst employed in the system. These kinds of interactions are not embedded in quantities dedicated to system interaction description as activity or fugacity. The same holds true for the interactions between the substrate and the catalyst (metal plus support). A sketch that represents the interactions described is reported in Figure 3.1.

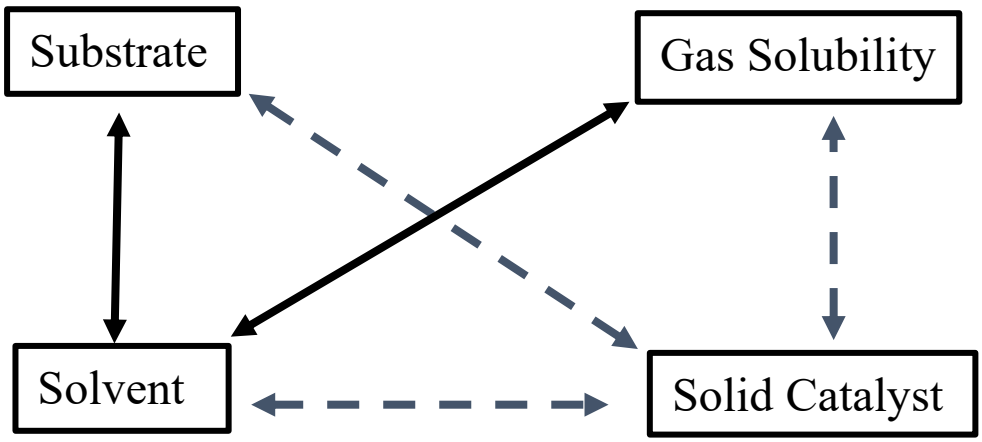


Figure 3.12 *A general representation of the interactions that characterise a multiphase system. The dashed arrows represent the interactions that need parameters to be described, the continuous arrows represent the interactions that can be captured by thermodynamics through fugacity or activity terms.*

What is needed to take into account such interactions, is to employ a certain number of parameters that are regressed using experimental data derived from each specific system. Numerous effects arise also from the solvent interactions with the substrate and ultimately from the variation of the system composition (formed mainly by the solvent) and the solubility of the hydrogen in case of hydrogenation.
The selection of the solvent also affects the succeeding purification steps and, therefore, is quite important for the efficiency of a reaction process and the separation units afterwards. In addition to the application of pure solvents, solvent mixtures are also used to facilitate catalyst recycling or to provide specific properties to reaction media. Solvent effects can support or suppress reaction kinetics, reaction equilibria and consequently, yields.

Today, there are about three hundred common pure solvents available (Reichardt and Welton, 2011), and several solvent mixtures. For this reason, the need to understand and

predict the solvent effects is critical and this field can be explored through methodologies that are able to probe reactions behaviour from a fundamental physical and chemical stand point but are pragmatic at the same time. Such characteristics are required in order to cover a broad spectrum of application.

3.2 Theoretical background: from reaction equilibrium to reaction kinetics

In this subsection it is recalled the concept of chemical potential and it is shown how it is used to derive the thermodynamic rigorous kinetic models. Firstly, the fundamental reaction equilibria form is derived starting from the Gibbs free energy (*G*) role in reactive systems, then how thermodynamic rigorous kinetic models are derived as an alternative way of writing the fundamental reaction equilibria form is described.

The Gibbs free energy of reaction, ΔGr, is defined as the slope of the graph of the Gibbs energy plotted against the extent of reaction, ξ, that is a quantity that measures the extent in which the reaction proceeds. This definition holds true in a case when temperature, *T*, and pressure, *P*, are constant:

$$\Delta G_R = \left(\frac{\partial G}{\partial \xi}\right)_{P,T} \tag{3.1}$$

It is worth noting that $\Delta G_{r,}$ can also be interpreted as the difference between the chemical potentials, $\mu_i(T, P, n)$, of the reactants and products at the composition of the reaction mixture. In particular, considering the reaction with stoichiometric coefficients ν_i when the reaction advances by $d\xi$ (infinitesimal of reaction extent) the amounts of reactants and products change by

$$dn_i = \nu_i d\xi \tag{3.2}$$

where *i* refers to a generic reactive specie.

The resulting infinitesimal change in the Gibbs energy at constant temperature and pressure is

$$dG = \sum_i \mu_i \, dn_i = \sum_i \mu_i \nu_i \, d\xi = (\sum_i \nu_i \mu_i) d\xi \tag{3.3}$$

It follows that

$$\Delta G_R = \left(\frac{\partial G}{\partial \xi}\right)_{P,T} = \sum_i \nu_i \mu_i \tag{3.4}$$

Since the chemical potential varies with composition, the slope of the plot of Gibbs energy against extent of reaction changes as the reaction proceeds as it can be seen in Figure 3.2.

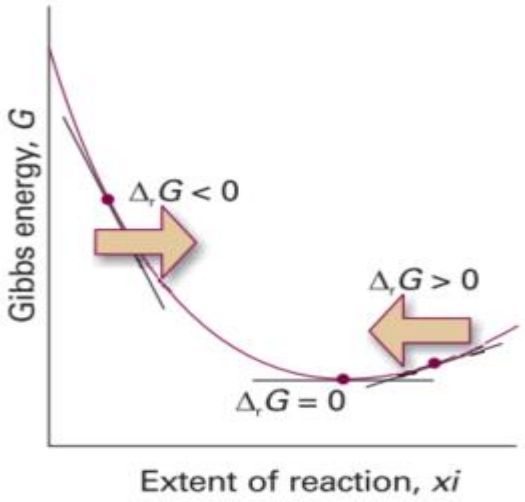


Figure 3.13 *Variation of Gibbs free energy of reaction as a function of the extent of reaction*

From Figure 3.2 it can be seen that, since the reaction runs in the direction of decreasing *G* (that is, down the slope of *G* plotted against ξ), the direct reaction proceeds spontaneously when the chemical potential of the reactants is greater with respect the one of the products, whereas the reverse reaction proceeds spontaneously in the exactly opposite conditions. When the slope is zero ($\Delta G_r = 0$) the reaction is spontaneous in neither direction, thus ΔG_r is equal zero implies that the sum of the chemical potential of the reactive species multiplied by the corresponding stoichiometry coefficients (negative in sign for products and positive in sign for reactants) is null. This condition result to be the fundamental reaction equilibria form:

$$\sum_{i=1}^{NC} \nu_{ij} \mu_i = 0 \tag{3.5}$$

Where *NC* refers to the number of reactive species in the system and subscript *j* refers to the specific reaction.

3.2.1 Alternative reaction equilibria form

Equation 3.5 can be written in alternative forms as it is shown in the work of Lafitte (2017) and reported as follows:

$$K = \prod_{i=1}^{NC} (W_i)^{\nu_i} \tag{3.6}$$

where W_i is a measure of the chemical potential's dependence on the composition of the system and K is the thermodynamic equilibrium constant defined as follows:

$$K \equiv exp\left(-\frac{\sum_{i=1}^{NC} \nu_i Z_i}{RT}\right) \tag{3.7}$$

where Z_i is the chemical potential part that does not depend on composition.
The alternative forms of reaction equilibrium obtained equating equations 3.6 and 3.7 vary depending on how the chemical potential in the mixture is expressed. Therefore, it is possible to define mainly three general ways to reformulate equation 3.5, namely fugacity reformulation, symmetric activities reformulation and asymmetric activity reformulation. For the derivation of all the reformulations, the procedure is analogous. The first step is to write the chemical potential of specie *i ;* the expressions can be slightly different depending on the reference state chosen and the definition of the quantity that is required to use, namely fugacity, symmetric activity or asymmetric activity. The second step consist of writing equation 3.5 using the chemical potential obtained in step one. In the third step the expression obtained in the second step is rearranged as described in equations 3.6 and 3.7. Equating the two terms found in the third step the reformulation of interest has been derived. In this way it is possible to obtain the corresponding definition of thermodynamic equilibrium constant for any combination of chemical potential expression chosen for any chemical specie in a reactive system.

3.2.1.1 Fugacity reformulation

Fugacity reformulation is characterised by the use of fugacity terms in the chemical potential relation and the use of ideal gas state for formation properties. The chemical potential of a specie *i* in a mixture written in term of formation properties (element reference) in the ideal gas state at standard pressure and temperature ($ig\theta$) has the following form taken from the work of Lafitte (2017):

$$\mu_i(T,P,n) = \Delta H_i^{F,ig\theta} - T\Delta S_i^{F,ig\theta} + \left[\int_{T^\theta}^{T} c_{p,i}^{ig}(T')dT' - T\int_{T^\theta}^{T} \frac{c_{p,i}^{ig}(T')}{T'}dT'\right] +$$
$$RT\ln\left(\frac{P}{P^\theta}\right) + RT\ln(x_i) + \mu^{res}(T,P,n) \tag{3.8}$$

Fugacity of component *i* in a generic mixture is defined as:

$$RT\ln\big(f_i(T,P,n)\big) \equiv RT\ln(P) + RT\ln(x_i) + \mu^{res}(T,P,n) \tag{3.9}$$

Employing 3.8 and 3.9 to compute 3.5 the equilibrium condition results equivalent to:

$$\sum_{i=1}^{NC} -\nu_{ij}\left\{\Delta H_i^{F,ig\theta} - T\Delta S_i^{F,ig\theta} + \left[\int_{T^\theta}^{T} c_{p,i}^{ig}(T')dT' - T\int_{T^\theta}^{T}\frac{c_{p,i}^{ig}(T')}{T'}dT'\right] + RT\ln\left(\frac{1}{P^\theta}\right)\right\} =$$

$$\sum_{i=1}^{NC}\nu_{ij}\, RT\ln(f_i(T,P,n)) \tag{3.10}$$

Equation 3.10 is rearranged in such a way that all the terms on the left-hand side are independent on composition while on the right-hand side only the composition dependent term are present. Equations 3.6 and 3.7 become respectively:

$$K(T) = \prod_{i=1}^{NC}\left(\frac{f_i(T,P,n)}{P^\theta}\right)^{\nu_i} \tag{3.11}$$

$$K(T) \equiv \exp\left(-\frac{\sum_{i=1}^{NC}\nu_{ij}\left\{\Delta H_i^{F,ig\theta} - T\Delta S_i^{F,ig\theta} + \left[\int_{T^\theta}^{T} c_{p,i}^{ig}(T')dT' - T\int_{T^\theta}^{T}\frac{c_{p,i}^{ig}(T')}{T'}dT'\right]\right\}}{RT}\right) \tag{3.12}$$

It worth noting that the thermodynamic equilibrium constant computation depends only on the temperature of the system and on the type of species involved in the specific reaction.

3.2.1.2 Symmetric activity reformulation

Symmetric activity reformulation is characterised by the use of symmetric activity terms in the chemical potential relation and the use of pure liquid state for formation properties. The chemical potential of specie *i* in a mixture written in term of formation properties (element reference) in the pure liquid state at standard pressure and temperature ($l\theta$) has the following form taken from the work of Lafitte (2017):

$$\mu_i(T,P,n) = \Delta H_i^{F,l\theta} - T\Delta S_i^{F,l\theta} + \left[\int_{T^\theta}^{T} c_{p,i}^{ig}(T')dT' - T\int_{T^\theta}^{T}\frac{c_{p,i}^{ig}(T')}{T'}dT'\right] +$$

$$+ RT\ln\left(\frac{f_i(T,P)}{P^\theta}\right) + RT\ln(\gamma_i(T,P,n)x_i) \tag{3.13}$$

Where $f_i(T,P)$ is the fugacity of component *i* as pure at the temperature and pressure of the system. The symmetric activity of component *i* in a mixture, a_i, is dimensionless and defined as:

$$a_i(T,P,n) = \gamma_i(T,P,n)\cdot x_i \tag{3.14}$$

$$RT\ln(\gamma_i(T,P,n)x_i) \equiv \mu_i(T,P,n) - \mu_i^0(T,P) \tag{3.15}$$

$$\mu_i^0(T,P) = \Delta H_i^{F,l\theta} - T\Delta S_i^{F,l\theta} + \left[\int_{T\theta}^{T} c_{p,i}^{ig}(T')dT' - T\int_{T\theta}^{T}\frac{c_{p,i}^{ig}(T')}{T'}dT'\right] \quad (3.16)$$

The terms $\gamma_i(T,P,n)$ and $\mu_i^0(T,P)$ are respectively the symmetric activity coefficient of a specie *i* in the mixture and the chemical potential of pure specie *i* considered at the same temperature and pressure of the system. Equation 3.5 results equivalent to:

$$\sum_{i=1}^{NC} -\nu_{ij}\left\{\Delta H_i^{F,l\theta} - T\Delta S_i^{F,l\theta} + \left[\int_{T\theta}^{T} c_{p,i}^{ig}(T')dT' - T\int_{T\theta}^{T}\frac{c_{p,i}^{ig}(T')}{T'}dT'\right] + \right.$$

$$\left. + RT\ln\left(\frac{f_i(T,P)}{P^\theta}\right)\right\} = \sum_{i=1}^{NC}\nu_{ij}\{RT\ln(\gamma_i(T,P,n) + RT\ln(x_i)\} \quad (3.17)$$

The comments on relation 3.10 hold true also for 3.17. In this case equations 3.6 and 3.7 become respectively:

$$K'(T,P) = \prod_{i=1}^{NC}(\gamma_i(T,P,n)x_i)^{\nu_i} \quad (3.18)$$

$$K'(T,P) \equiv \exp\left(-\frac{\sum_{i=1}^{NC}\nu_{ij}\left\{\Delta H_i^{F,l\theta} - T\Delta S_i^{F,l\theta} + \left[\int_{T\theta}^{T} c_{p,i}^{ig}(T')dT' - T\int_{T\theta}^{T}\frac{c_{p,i}^{ig}(T')}{T'}dT'\right] + RT\ln\left(\frac{f_i(T,P)}{P^\theta}\right)\right\}}{RT}\right) \quad (3.19)$$

It worth noting that K' represents theoretically the same quantity as K found in 3.12 but in this case, it depends not only on the system temperature and the species involved in the reaction but also on the pressure of the system.

3.2.1.3 Asymmetric activity reformulation

Asymmetric activity reformulation is characterised by the use of asymmetric activity terms in the chemical potential relation and the use of infinite dilution state for formation properties. The chemical potential of a specie *i* in a mixture written in term of formation properties (element reference) in the infinite dilution state (∞) has the following form taken from the work of Lafitte (2017):

$$\mu_i(T,P,n) = \hat{\mu}_i^{w,\infty}(T,P) + RT\ln(\hat{a}_i) \quad (3.20)$$

Where the chemical potential of specie *i* in the infinite dilution state $\hat{\mu}_i^{w,\infty}$ and the asymmetric activity, $\hat{a}_i$ are defined as follows:

$$\hat{\mu}_i^{w,\infty}(T,P) = \Delta\hat{H}_i^{F,\infty} - T\Delta\hat{S}_i^{F,\infty} - RT\ln\left(\varphi_i\left(T^\theta,P^\theta\right)\right) + \left[\int_{T^\theta}^{T} c_{p,i}^{ig}(T')dT' - \right.$$
$$\left. - T\int_{T^\theta}^{T}\frac{c_{p,i}^{ig}(T')}{T'}dT'\right] + RT\ln\left(\frac{P}{P^\theta}\right) + RT\ln\left(\varphi_i(T,P)\right) \quad (3.21)$$

$$RT\ln\hat{a}_i = RT\ln(\hat{\gamma}_i x_i) \equiv \mu_i(T,P,n) - \hat{\mu}_i^{w,\infty}(T,P) \quad (3.22)$$

Where φ_i is the fugacity coefficient of specie i and $\hat{\gamma}_i$ the asymmetric activity coefficient of the same chemical compound. Both quantities are dimensionless.
In this case equations 3.6 and 3.7 become respectively:

$$\widehat{K'}(T,P) = \prod_{i=1}^{NC}(\hat{\gamma}_i(T,P,n)x_i)^{\nu_i} \quad (3.23)$$

$$\widehat{K}'(T,P) \equiv \exp\left(-\frac{\sum_{i=1}^{NC}\nu_{ij}\hat{\mu}_i^{w,\infty}(T,P)}{RT}\right) \quad (3.24)$$

As K', $\widehat{K}'$ depends on both system temperature and pressure along with the reactive specie involved in the system. Asymmetric activity can be employed in different basis, namely mole fraction basis, molality basis and concentration basis. The above derivation is on a mole fraction basis, but it can be converted in any other of the other two through a slight change in the chemical potential formulation to ensure dimension consistency.

3.2.1.4 General comments

The possibility to choose among fugacity, symmetric and asymmetric activity reformulation of the chemical potential allows to choose the most suitable form for each chemical species. Some example to clarify the concept are reported: ions cannot be found in pure liquid state, so it would be impossible to use the symmetric activity reformulation, nevertheless they can be found both in ideal gas state and as infinitely diluted so both asymmetric activity coefficients and fugacity reformulation can be employed, A second example deals with incondensable gasses that neither at the standard conditions nor at usual operating condition at which the formation properties are obtained are found as pure liquid or at infinite dilution state. This implies that neither symmetric nor asymmetric activity reformulation can be used, while fugacity reformulation can always be applied with gases. The two examples reported show that for different species different formation properties reference state are required or preferred but the reformulation of the chemical potential expression allow to use all of them. The definition of the thermodynamic equilibrium constant as it can be seen in the previous section is different for each approach. The definitions obtained are based on the assumption that for all the species that take part in the reaction the chemical potential is written with the same approach. This is not always the case, for instance the dissociation of water that forms hydrogen

ions (H3O+) and hydroxide (OH-) ions required symmetric activity approach for water chemical potential and asymmetric activity approach for the product (ion) chemical potential. In this case the final form of the thermodynamic equilibrium constant is a combination of those derived in the previous subchapters as explained in §3.2.1. Ideally, what would be preferred is to use the same approach for all the chemical species, but this requires that all chemical compounds exist in the reference state at which the formation properties are calculated that is specific of the approach chosen. Among the three approaches presented, fugacity, which is characterised by ideal gas reference state, is the only one that would be capable of fulfil the requirement.

3.2.2 Thermodynamically rigorous kinetic model

In this project the main focus is on the use of fugacities for several reasons that will be addressed later on in this chapter. Therefore, the thermodynamically rigorous kinetic model employed in the case study analysed is the one based on fugacity reformulation; however, the rationale behind the derivation procedure is completely general. The model is derived starting from the fundamental form of reaction equilibrium (3.5) and the chemical potential written in terms of fugacity (3.8, 3.9). Due to this approach the rigorous thermodynamic equilibrium constant ($K(T)$) has been defined as unique function of temperature as it can be seen from 3.12. In order to obtain a thermodynamically consistent treatment of a reacting system, the definition of $K(T)$ theoretically derived must holds true when the reaction equilibrium is reached and in this case the model is said to be based on thermodynamic consistency. The fugacity-based kinetic model (FKM) shows therefore, the following structure:

$$r_j = k_{1,j}\left[\prod_{i=1}^{NC}\left(\frac{f_i(T,P,n)}{P^{\theta}}\right)^{\max(0,-\nu_{i,j})} - \frac{1}{K_j(T)}\prod_{i=1}^{NC}\left(\frac{f_i(T,P,n)}{P^{\theta}}\right)^{\max(0,\nu_{i,j})}\right] \tag{3.25}$$

$$K_j(T) = \exp\left(\frac{-\sum_{i=1}^{NC}\nu_{ij}\Delta G_i^F(T,P^{\theta})}{RT}\right) = \frac{k_{1,j}}{k_{-1,j}} \tag{3.26}$$

$$f_i(T,P,n) = P \cdot x_i \cdot \overline{\varphi_i}(T,P,n) \tag{3.27}$$

Where

- r_j is the rate of reaction j [mol/m^3 s]
- $k_{1,j}$ is the intrinsic forward rate constant [mol/m^3 s]
- K_j is the thermodynamic equilibrium constant for reaction j [-]
- ΔG_i^F Gibbs free energy of formation [J/mol]

As the progression of reaction kinetics leads to the reaction equilibrium at infinite time (where $r_j = 0$), $k_{1,j}$ and $k_{-1,j}$ are connected via the thermodynamic equilibrium constant,

according to 3.26. This means that once the thermodynamic equilibrium constant is known, only one kinetic parameter out of two needs to be regressed for modelling the reaction kinetics. It is worth pointing out that once the operating conditions and the reactive species involved in the reactive system are set the thermodynamic equilibrium constant can be calculated without the need of any experimental data. On the other hand, the intrinsic rate constant needs to be regressed from experimental data and it is expressed as a single parameter for isothermal systems and through the Arrhenius equation for non-isothermal
ones.
As a matter of comparison, the model derived based of symmetric activity approach is reported:

$$r_j = k_{1,j}\left[\prod_{i=1}^{NC}(x_i\gamma_i)^{\max(0,-\nu_{i,j})} - \frac{1}{K_j(T)}\prod_{i=1}^{NC}(x_i\gamma_i)^{\max(0,\nu_{i,j})}\right] \tag{3.28}$$

$$K_j(T,P) = \exp\left(\frac{-\sum_{i=1}^{NC}\nu_{ij}\left[\Delta G_i^F(T,P^\theta)+RT\ln\left(\frac{f_i(T,P)}{P^\theta}\right)\right]}{RT}\right) = \frac{k_{1,j}}{k_{-1,j}} \tag{3.29}$$

As it can be seen from 3.28 the model structure is analogous to the fugacity approach one, what changes is the driving force terms used, namely activities instead of fugacities. Any comments provided for the previous model described hold true also for 3.28.
The last model reported is the widespread power law concentration based kinetic model (PLCKM) whose structure is the following:

$$r_j = {k'}_{1,j}\left[\prod_{i=1}^{NC} {C_i}^{\mathrm{nr}_{i,j}} - \frac{1}{{K_j}''(T,P,C_i,\mathrm{Solvent})}\prod_{i=1}^{NC} {C_i}^{\mathrm{np}_{i,j}}\right] \tag{3.30}$$

Where ${k'}_{1,j}$ is the rate constant whose unit of measurement depends on those of the term in brackets, the product between the latter and the rate constant should give [mol/m^3 s]. The terms ${C_i}^{\mathrm{nr}_{i,j}}$ and ${C_i}^{\mathrm{np}_{i,j}}$ are concentration [mol/m^3] of the reactants and products respectively at the power of two different parameters usually derived by regressing experimental data for the specific reactive system or assuming a certain order of reaction verified a posteriori. The apparent equilibrium constant, ${K_j}''$, depends on temperature, pressure, concentration of the reactive species in the system and the type of solvent as well. It is usually obtained from experimental concentration data at equilibrium conditions as the ratio between the product of the products concentration and the product of the reactant concentration at equilibrium. Also, in this case the ratio between the forward and the backward rate constants results to be equal to the apparent equilibrium constant.

3.2.3 Theoretical comparison among concentrations, activities and fugacities as reaction driving forces

All the models reported in §3.2.2 present a very similar structure, what is different is the driving force they are based on and the equilibrium constant that in fugacity and activity based model is called thermodynamic equilibrium constant and in the concentration based model is called apparent equilibrium constant. Regardless the physical meaning that is the same for the three models, the word "thermodynamic" in thermodynamic equilibrium constant indicates the rigorous derivation of the relations found and the independence on concentrations and other factors except temperature (both activity and fugacity approach) and pressure (only activity approach). The adjective "apparent" used in apparent equilibrium constant, on the other hand, is due to the need of using experimental data to calculate it, since the model is not thermodynamically consistent, and it must capture the influence of different operating conditions on the equilibrium position. For instance, the presence of different solvents, that macroscopically are inert, may cause variation in the selectivity at equilibrium conditions and the apparent equilibrium constant must be able to capture it. This instance brings to the analysis of the driving force used in the different models.

Starting from the concentration-based model the first issue to notice is that the reaction is driven by the difference in concentrations, thus, every kind of interactions among different species and the reaction environment is neglected. For this reason, the apparent equilibrium constant should take care of it. Therefore, it has to be observed experimentally for a given mixture each time, rather than being predictive.

When fugacities or activities are employed as driving force in a reactive system the interactions between reactive species themselves but also those between them and the reaction environment, are take into account. Therefore, both ideal and highly non-ideal system, theoretically, can be described with the same degree of accuracy. For this reason, the impact of different solvents employed for and the impact of a variation in the amount of the same solvent concentration in the same reactive system should be captured in terms of interactions from fugacity and activity driving force.

From a physical point of view the structure of the models presented (Equations 3.25, 3.28, 3.30) can be seen as a combination of two contributions: the probability that the reactant species have to collide and the probability that the molecules that collides actually react. The former contribution is given by the terms inside the brackets and includes the driving force considered and the equilibrium constant, factors already discussed. The latter is represented by the rate constant; from its physical meaning it is not supposed to vary with a change in the reaction environment if the pressure and the temperature remain constant.

Therefore, variation in the type of solvent or in the amount of the same solvent for a given reactive system, theoretically does not cause any change in the rate constant value. This holds true if the interactions and non-idealities present in the reactive species are captured by the driving force term. Thereby, since the concentration driving force is not capable to account for interactions, and it only accounts for the dilution of the reactants through the solvent, it is not possible to expect, in principle, a constant value for the rate constant. On the other hand, since symmetric activity and fugacity are driving forces based on the interactions among chemical species and not only on the composition of the system, it is expected, according to the theory, a constant value for the intrinsic rate constant. This result implies that in case of power law concentration-based model the rate constant should be regressed from experimental data every time a different solvent is employed in the same reactive system, the same holds true even if there is a variation in the amount of the same solvent. This does not hold true for symmetric activity and fugacity based models for which once the intrinsic rate constant is regressed it should be, in principle, the same for every other solvent employed in the reactive system considered.

The limitations of the concentrations-based kinetic model raised make evident the benefit derived from the use of symmetric activity and fugacity based models regarding the capability of described a reactive system. A fugacity based kinetic model should be used instead of the symmetric activity based one; the reasons are the following:

- The thermodynamic equilibrium constant depends only on temperature; therefore, its computation only depends on formation properties.
- For any chemical species is possible to measure/calculate the formation properties in the ideal gas state, thus it is always possible use 3.25.
- From the computation point of view fugacities require less computational effort than activities .

The second and the third bullets require clarifications: it is always possible to calculate formation properties in the ideal gas state using Joback method (Joback, 1987) and the improvements in the computational effort are related specifically to SAFT-γ Mie implementation. Since it is an advanced equation of state, the computation of the fugacity coefficients is “direct”, while the calculation of activity coefficients requires additional steps. This is the reason why the computational effort is said to be less in case of fugacity approach.

3.3 Objectives

Reactive systems are extremely important in chemical industries, therefore the need to accurately describe them and even being able to predict their behaviour is still a very

active research field. So far purely empirical models have been applied to describe these complex processes because the several interactions involved are rarely captured by the most widespread kinetic model employed. The aim of the case study analysed is therefore to assess the capability of the fugacity based kinetic model with fugacities computed by SAFT-γ Mie advanced equation of state to accurately described the system through:

- the correct interpretation of experimental data
- the capability of predict the process behaviour over a wide range of conditions

Particular attention is given to solvent effects on reactive systems, since they are known to have a significant influence on both reaction equilibrium and reaction kinetics. The objectives set have been attained in the transesterification system through the evaluation of:

- The impact on the accuracy of the results applying the fugacity based kinetic model (FKM) instead of the power law concentration based one (PLCKM)
- Comparison of prediction performance between FKM and PLCKM employing a constant rate constant for different solvent amount in the same reactive system

3.4 Model validation

A process model is constructed from equations describing the physical phenomena that take place in the system. For what regards the case study faced the model of the process was already given by the system of equation embedded in the “CSTR one phase liquid” unit, which is one of the process units available in gPROMS FormulatedProduct. What has been modified is the rate of reaction expression for which the most suitable kinetic model has been deployed.

Once a model has been developed to describe a certain phenomenon, in this specific case a reactive system, its validation is obtained applying the following procedure:

- Identify the unknown parameters in kinetic models and related measurements that are influenced by them.
- Perform a sensitivity analysis using the model to ensure that the measurements selected are affected by the values of the parameters.
- Perform parameter estimation to find values for the unknown parameters that minimise the difference between simulated and measured values. This usually includes the following:
  - Checking the meaningfulness of initial guesses for model parameters before running a full parameter estimation.
  - Applying an appropriate variance model to the regressed experimental data.

  - Parameter estimation of rate constants using data at a single operating temperature. This is the single step required for isothermal reactive systems. For adiabatic systems, instead, this first step is applied to ensure a satisfactory description of the reaction kinetic model, then multitemperature data are regressed. In this thesis work only isothermal reactive system are considered.
- Assessment of the parameters estimated quality through the evaluation of confidence intervals, t-statistics and standard deviation.
- If necessary, improvements to experiment and/or model should be made and then the procedure should be repeated. In the case of reaction kinetics, the power law model initially employed may not be sufficient to describe the reaction kinetics and custom kinetics model expressions may be required.

After the model validation, the model is ready to be employed in design and optimization applications. Model validation in gPROMS is based on the Maximum Likelihood formulation, which provides simultaneous estimation of parameters in both the physical model of the process and the variance model of the process. The latter has not be carried out because a fixed relation for the variance model has been set. When solving a Maximum Likelihood validation problem, the attempts is to determine values for the uncertain physical parameters, θ, that maximize the probability that the mathematical model will predict the measurement values obtained from the experiments, namely the parameters value that minimizes the overall maximum likelihood objective function. Assuming independent, normally distributed measurement errors, ε_{ijk}, with zero means and standard deviation, σ_{ijk}, this maximum likelihood goal can be captured through the following objective function:

$$\Phi = \frac{N_m}{2}\ln(2\pi) + \frac{1}{2} min_{\theta} \left\{ \sum_{i=1}^{NE} \sum_{j=1}^{NV_i} \sum_{k=1}^{NM_{ij}} \left[\ln\left(\sigma^2{}_{ijk}\right) + \frac{(\tilde{z}_{ijk} - z_{ijk})^2}{\sigma^2{}_{ijk}} \right] \right\} \qquad (3.31)$$

Where:

Φ is the Maximum likelihood objective function

N_m is the total number of measurements taken during all experiments

θ is the set of parameters to be estimated

NE is the number of experiment performed

NV_i is the number of variables measured in the i^{th} experiment

NM_{ij} is the number of measurement of the j^{th} variable in the i^{th} experiment

σ_{ijk} is the variance of the k^{th} measurement of variable j in experiment i

$\tilde{z}_{ijk}$ is the k^{th} measured value of variable j in experiment i

z_{ijk} is the k^{th} predicted value of variable j in experiment i

The contribution of a measured variable to the overall objective function is split into three terms: the constant term (the first term of the right-hand side of eq. 3.31), the variance term (the first half of the summation in eq. 3.31) and the residual term (the reminder of the second term in eq. 3.31). Both the constant and the variance terms remain fixed during parameter estimation and the optimal point is achieved by only minimizing the contribution of the residual term of all the measured variables to the objective function. The number of data points for a given measurement and the variance model applied can therefore significantly impact the minimization of the objective function during parameter estimation. The variance models available are: constant variance, constant relative variance and linear variance. For measurements that may take very small values or span a large range (reactant and product concentration or mass fraction in this case), it is recommended to use the linear variance model rather than the constant or constant relative one. This is because the linear variance model has a constant component which will ensure that measurements with a small value are not given a high weight. Thus, it will remove the bias in the results of the parameter estimation towards measurements with a small value. For this reason, a linear variance model has been selected for the case study analysed.

As discussed, confidence intervals are one of the tools used to assess the parameter estimated. It is a type of interval estimate, computed from the statistics of the observed data, that might contain the true value of an unknown population parameter. The interval has an associated confidence level that represents the frequency of possible confidence intervals that contain the true value of the unknown population parameter. Most commonly, the 95% confidence level is used (Zar, 1984); however, other confidence levels can be used, for instance, 90% and 99%. As a rule of thumb applied in this work if the value corresponding to 95% of confidence is lower than the 10% of the absolute value of the parameter then the regressed parameter value is considered to be accurate.

To assess the statistical quality of the parameters, it is necessary to compare the value of each parameter estimated with its confidence range. In other words, what is required to do is to understand how large the confidence region is assigned to the parameter with respect to the absolute value of the parameter itself. This can be achieved assessing a t-test with 95% of significance and comparing the t-value of each estimated parameter with the reference t-value of a Student distribution. If the t-value results much lower than the reference one that parameter is considered statistically meaningless meaning that a change in the absolute value does not affect the model computation. For this reason, in this project the parameters that result statistically meaningless are removed from the model.

3.5 Transesterification case study

Transesterification is a classic organic reaction that has enjoyed numerous laboratories use and industrial applications. This reaction is often employed as a convenient means to prepare esters, especially in cases where the carboxylic acids, that reacts to alcohol to give ester through esterification, are labile, difficult to isolate and have poor solubility in organic solvents that usually characterise the reaction environment. Other industrial applications of transesterification are: production of PET (polyethylene terephthalate), formation of lactones or macrocycles through intramolecular transesterification, transesterification of triglycerides to mixture of fatty acid esters (eg. FAME biodiesel component) and glycerol (cosmetics, toothpastes, pharmaceuticals, food applications), SPE (sugar polyesters) production (Schuchardt, 1997). Therefore, the prediction of the reaction environment (eg. solvent) effects on transesterification is of interest for several industrial application. The case study is focused on one specific transesterification reaction, and it has been chosen to represent this class of reactions. Thus, the study carried out should be seen as a general analysis.

3.5.1 System description and model employed

The transesterification of butyl acetate with ethanol to butanol and ethyl acetate in heptane has been investigated.

$$C_2H_4O + C_6H_{12}O_2 \leftrightarrow C_4H_{10}O + C_4H_8O_2 \tag{3.32}$$

The reaction is catalysed by $K^+(CH_3)_3CO^-$. The concentration of this catalyst is three order of magnitude less than the other species present in the system, and so has been neglected. The process operating conditions are:

- Homogeneous liquid phase
- Isothermal system at 293 K
- Isobaric system at 1 bar
- Variable amount of heptane (0%vol, 32%vol, 51%vol, 70%vol, 90%vol) that leads to consider five different systems
- Equimolar Reactants

A stirred tank reactor is used. gPROMS Formulated Products provides a unit operation called “CSTR one phase liquid”, which contains the system of equation that describes the process. Template custom reaction kinetics modelling approach is implemented in this unit operation using a custom reaction kinetics template. The assumptions made in deriving the “CSTR one phase liquid” model are:

- The reactor is perfectly mixed at all times and so there are no special variations in the intensive properties.
- The phase is in thermal equilibrium.
- Only liquid phase chemical reactions take place.
- The pressure is constant.
- The reaction liquid mixture is incompressible.

The experimental data employed are those published in 1999 by Schmidt et al. (1999) and regard the initial concentration [mol/kg] of the reactants for different amount of heptane volume percentage and the concentration [mol/kg] of ethyl acetate at different time (from 0 to 240 minutes) and for different amount of solvents. In Schmidt et al. (1999), the authors investigated experimentally both the reaction equilibrium and the reaction kinetics. They found that varying the amount of solvent heptane does not affect the reaction equilibrium but significantly influences the reaction kinetics. They describe the solvent effects on the reaction kinetics by empirical correlations that contain the experimentally-observed rate constants with the dielectric constants of the different reaction mixtures. In this case study the experimental data are re-evaluated in order to present a thermodynamic approach to consistently predict the solvent effect on both, the reaction equilibrium and the reaction kinetics. The models employed in this study are FKM (eq. 3.25) and PLCKM (eq. 3.30), that for this specific system take respectively the following form:

$$\mathrm{r} = k_1 \left[\frac{f_{ethanol} \cdot f_{butylacetate}}{(P^{\theta})^2} - \frac{1}{\mathrm{K}} \frac{f_{butanol} \cdot f_{ethylacetate}}{(P^{\theta})^2} \right] \tag{3.33}$$

$$\mathrm{r} = k'_1 \left[C_{ethanol} \cdot C_{butylacetate} - \frac{1}{\mathrm{K}''} C_{butanol} \cdot C_{ethylacetate} \right] \tag{3.34}$$

The former is characterised by the computation of fugacity coefficients that are performed by the thermodynamic model chosen that, in this study, is SAFT-γ Mie. The main concern is therefore to identify the group cross interactions missing in the description of the system. In this case all the group cross interactions result to be present in the database. Accounting for the fugacity coefficients of the reactants/products, the FKM allows for considering the interactions of the reactants/products among themselves and also with the solvent heptane. This type of interactions, indeed, can be completely described by thermodynamics, therefore, theoretically, they do not require any parameters to capture their effect.

3.5.2 Results and discussions

Since the two aspects in which the solvent can play a role are reaction equilibrium and kinetics, they have been treated firstly individually and then combined. In both cases the model validation as described in §3.4 is carried out and comments on statistics are present. The set of experimental data used to regress the parameters are batch measurements at different time of ethyl acetate in mass fraction basis provided for each volume percentage of heptane in the system. The experimental values are obtained from infrared spectroscopy analysis and are employed for the regression of parameter in mass fraction basis. Therefore, the variance model set is the linear one comprising a fixed variance of 0.05 that accounts for typical 5% error in data collection and a constant relative term of 10^{-4} to account for base accuracy of infrared spectroscopy analysis.

3.5.2.1 Reaction Equilibrium

The reaction equilibrium is defined through the condition in equation 3.5. It is taken into account in kinetic models through the thermodynamic equilibrium constant. The equilibrium constant is defined, theoretically, both for FKM and PLCKM through eq. 3.12. The quantities required to calculate it are the formations properties in the ideal gas state provided by Lemberg et al. (2017), and the system temperature. From Table 3.1 it can be seen that the Gibbs free energy of reaction is lower than 10 kJ/mol, this implies that the reaction does not promote neither the forward nor the backward reaction and a small change in its absolute value can have a significant impact on the thermodynamic equilibrium constant. As a matter of consistency, the formation properties values have been taken also from the Perry's Chemical Engineers' Handbook 7th edition and calculated using Joback method (Joback, 1987) to be compared to those from Lemberg's paper, the results are reported in Table 3.1. It worth underline that the experimental equilibrium constant value is approximately 0.93 for all the systems analysed namely for five different amounts of solvents.

Table 3.5 *Gibbs free energy of formation [J/mol] taken from different sources (Perry's Chemical Engineers' Handbook, Lemberg et al., 2017, Joback method calculation) for the transesterification reactive species used to calculate the Gibbs free energy of reaction and then the thermodynamic equilibrium constant*

	ΔG_F [J/mol] Perry's	ΔG_F [J/mol] Lemberg's	ΔG_F [J/mol] Joback's
Ethanol	-167850	-169686	-179280
Butyl acetate	-312600	-315766	-361250
Butanol	-150300	-153201	-145600
EthylAcetate	-324200	-329696	-327570
ΔG_R [J/mol]	5950	2555	67360
K (293 K)	0.08705	0.35038	9.90E-13
1/*K*	11.49	2.85	1.01E+12

Since the results obtained for the thermodynamic equilibrium constant are very different from each other what can be stated for sure is that the approximated Joback method is not reliable because very far from both the other two sources and the experimental value. The two literature sources provide values with one order of magnitude difference. In order to quantify the effect of the departure between the two values a Global System Analysis (GSA) has been carried out. GSA is a tool that allows to analyse "what-if" scenarios, through a comprehensive exploration of the behaviour of the system over domains of any subset of input and output variables. The input that has made varying is the inverse of the thermodynamic equilibrium constant in a range that lays between the two values highlighted in Table 3.1, while the output tracked is the mass fraction of ethyl acetate. The GSA result has been reported through the plot in Figure 3.3.

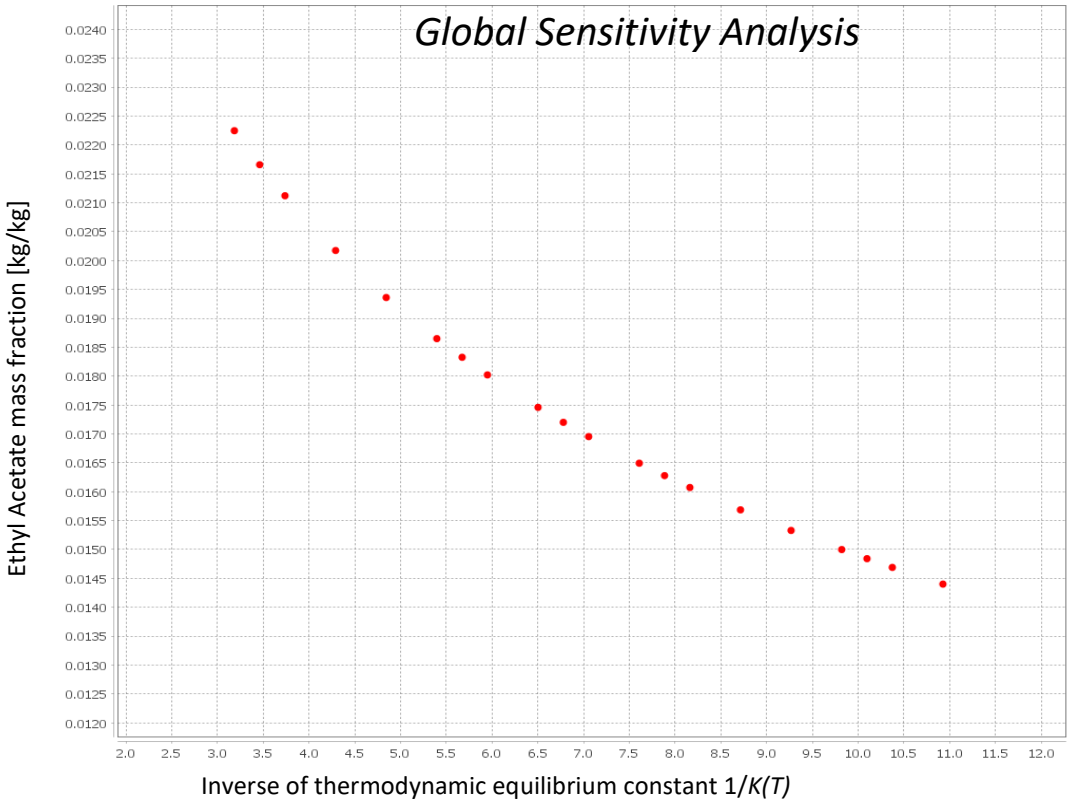


Figure 3.14 *Global sensitivity analysis carried out consider as input the inverse of the thermodynamic equilibrium constant and as output the ethyl acetate mass fraction*

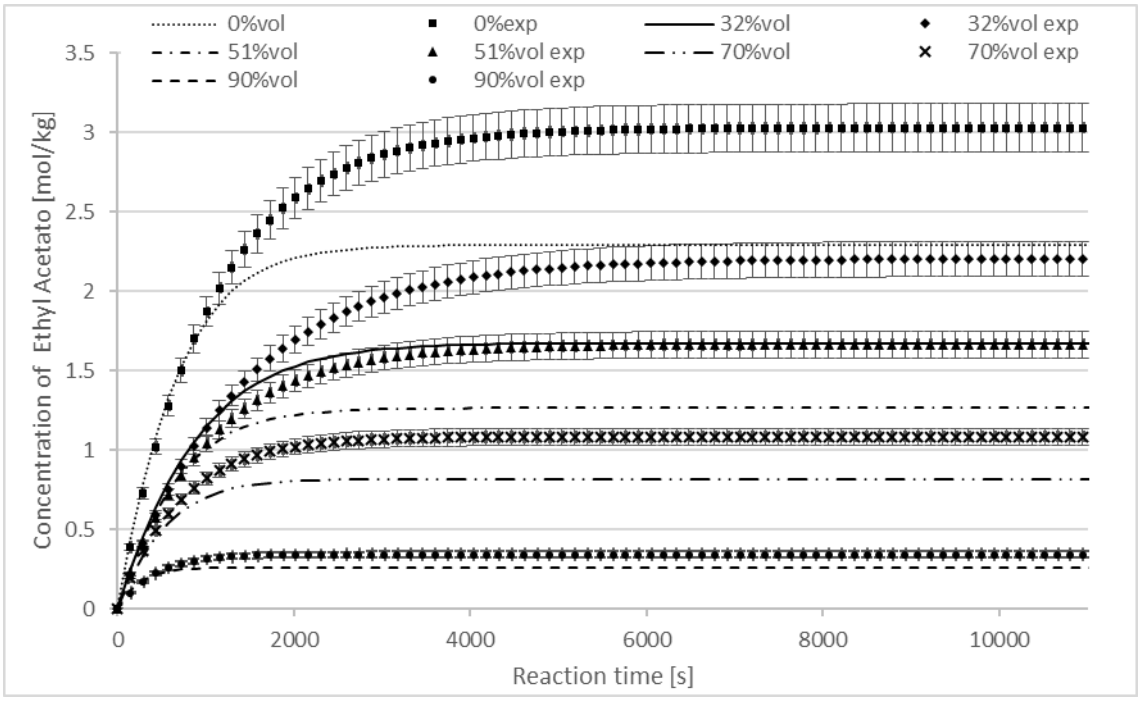
From Figure 3.3 it can be seen that the mass fraction of ethyl acetate varies from a value of 0.0145 to 0.0223, therefore the source from which the values of formation properties are taken can have a significant impact on the system computation when the Gibbs free energy of formation is low. It is worth noting that in PLCKM the concentration terms are not able to take into account the interactions among species in the system because by

design they are capable of consider only the "dilution" effect of different compounds. In particular PLCKM is not capable to capture the interactions effect on equilibrium condition because while the impact on kinetics can be partially included in the rate constant parameter, the equilibrium is characterised by the reaction rate equal zero, so the rate constant has no effect on equilibrium computation. Therefore, a parameter dedicated to capture the solvent effect only on the equilibrium condition is not present. What is done in practice is to calculate the value of the apparent equilibrium constant from the experimental equilibrium data for each specific system. Because of these reasons, the reaction equilibrium described by rigorous thermodynamics expression (e.g. equation 3.12) is not reliable for this system and can lead to very poor reaction equilibrium results. It can affect the kinetic computation as well.

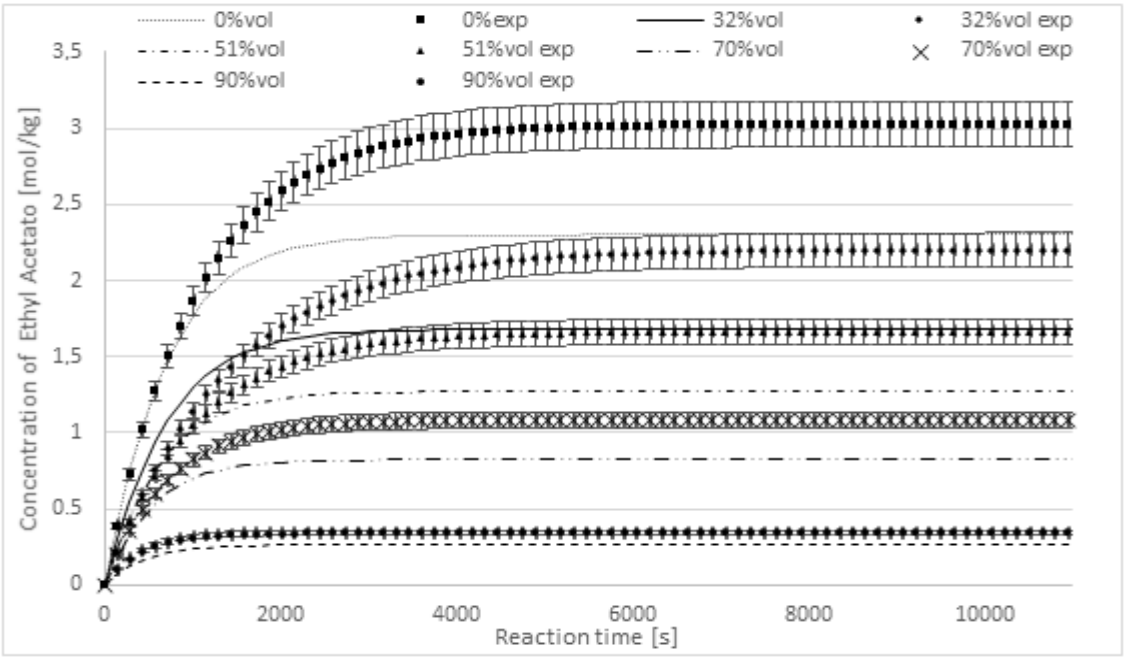
Table 3.6 *Rate constant values regressed from single dataset for each system. The different systems are individuated as a different percentage of C7 (heptane) solvent employed. The equilibrium constant in set at 0.35.*

	0%vol C7	32%vol C7	51%vol C7	70%vol C7	90%vol C7
$k_{,1}$ [mol/m^3 s]	10749.2	8020.94	10050.1	12954.2	15000
$k'_{,1}$ [m^3/mol s]	9.21E-08	1.00E-07	1.89E-07	4.57E-07	2.71E-06

In Figure 3.4 the batch curves describing the system computed by both FKM and PLCKM employing the thermodynamic equilibrium constant calculated based on Lemberg (2017) formation properties have been reported. The rate constants for both kinetic models corresponding to the five different systems regressed from each single data set are reported in Table 3.2.



(a)



(b)

Figure 3.15 *Batch concentration-reaction time plots for ethyl acetate. Lines denote model predictions for the same reactive system characterised by variable amount of solvent. Symbols with error bars (linear error interval with coefficient0.05 and bias 10^{-4}) represent denotes experimental reading. The kinetic models employed for the calculation are: (a) FKM and (b) PLCKM. The parameter used are 0.35 as equilibrium constant and the values reported in Table 3.2 as rate constant*

From Figure 3.4 it is evident that the equilibrium results are very poor for both models. While for PLCKM the results obtained were expected, the FKM was supposed to compensate for the discrepancies between the rigorous thermodynamic equilibrium constant value and the apparent equilibrium constant thanks to the fugacity terms. The reason why the result differ from the expectations is probably related to the wide range of values that, for this specific system, the thermodynamic equilibrium constant can assume depending of the literature source chosen. For cases in which the Gibbs free energy results to be lower than 10^4 J/mol it is advisable setting the thermodynamic equilibrium constant equal to the value calculated using experimental data at equilibrium or regressing the thermodynamic equilibrium constant as a parameter. Both options have been carried out in this study and the results are presented in the following section with focus on the reaction kinetics.

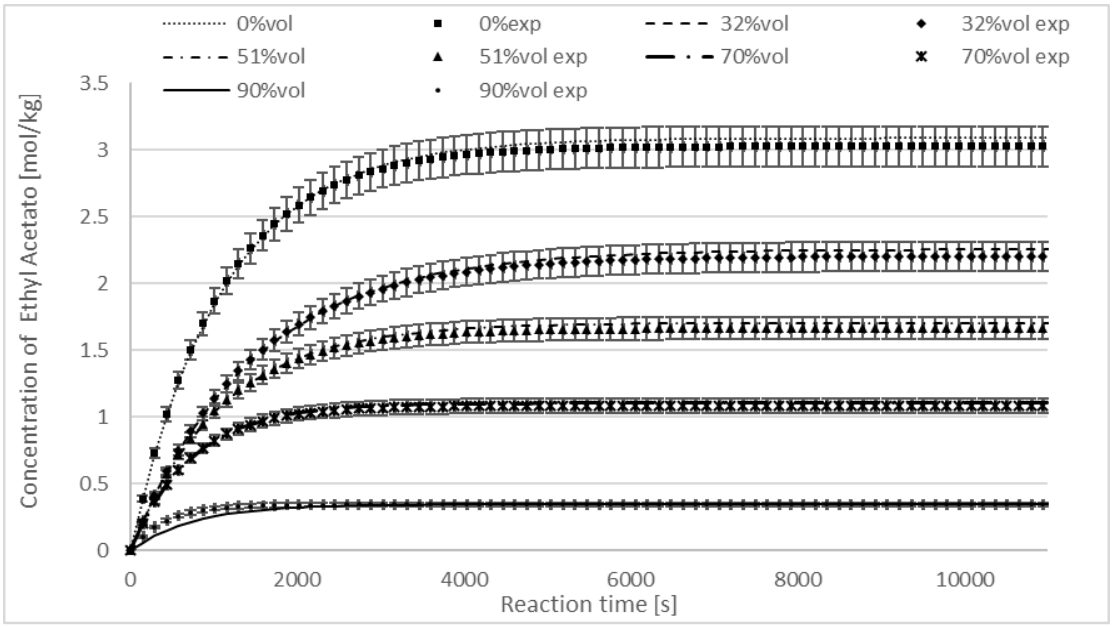
3.5.2.2 Reaction Kinetics

In order to assess the potential accuracy improvements derived by using FKM over PLCKM when the experimental data are available, batch curves have been performed. The thermodynamic equilibrium constant and the apparent equilibrium constant are set to the experimental value 0.93 (Schmidt et al., 1999) obtained as the ratio between the product of the products concentration and the product of the reactants concentration at the equilibrium condition.

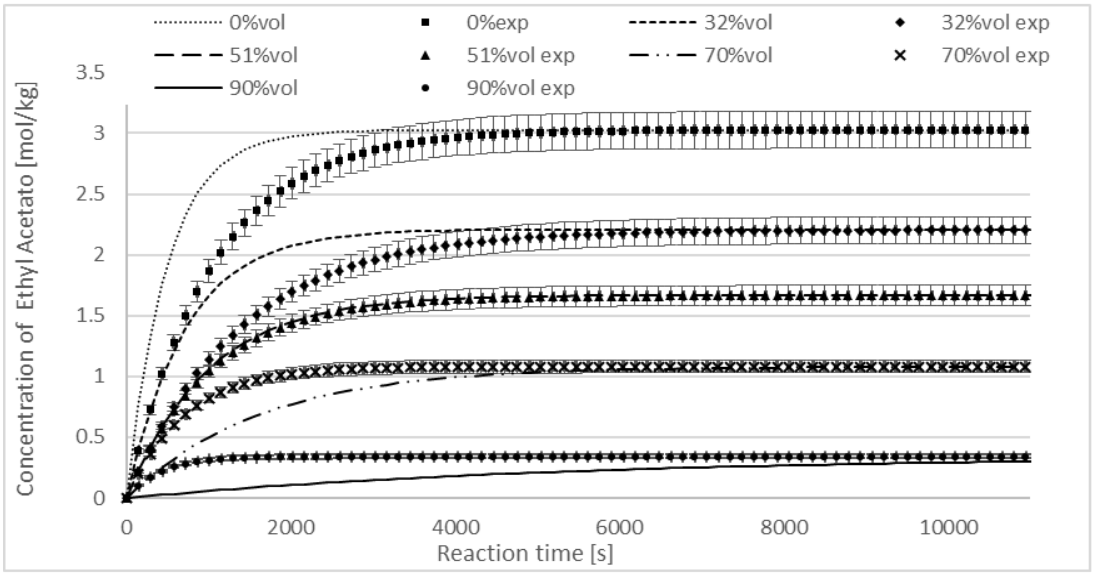
Table 3.3 *Rate constant values regressed from single dataset for each system. The different systems are individuated as a different percentage of C7 (heptane) solvent employed. The equilibrium constant in set at 0.97.*

	0%vol C7	32%vol C7	51%vol C7	70%vol C7	90%vol C7
$k_{,1}$ [mol/m^3 s]	10195	7717.43	9516.88	11926.6	16506.3
$k'_{,1}$ [m^3/mol s]	8.85139E-08	9.77299E-08	1.81242E-07	4.2576E-07	2.39E-06

Since accuracy rather than predictive capability is being compared, the rate constant has been regressed for every set of experiment, namely for different amount of solvent. The rate constants for both kinetic models corresponding to the five different systems regressed from each single data set are reported in Table 3.3.



(a)



(b)

Figure 3.5 *Batch concentration-reaction time plots for ethyl acetate. Lines denote model predictions for the same reactive system characterised by variable amount of solvent. Symbols with error bars (linear error interval with coefficient0.05 and bias 10^{-4}) represent denotes experimental reading. The kinetic models employed for the calculation are: (a) FKM and (b) PLCKM. The parameter used are 0.93 as equilibrium constant and the values reported in Table 3.3 as rate constant*

As can be seen in Figure 3.5*b*, PLCKM underestimates the concentrations of ethyl acetate and the deviation increases with increasing heptane content, indeed, what is captured is only the dilution effect of the solvent that increase with solvent content. The greater the solvent amount, the lower the concentration of the reactant, the slower the reaction described by PLCKM. On the other hand, Figure 3.5*a* shows that once the equilibrium condition is fixed properly the kinetic profile is accurately captured by the model,

meaning that the fugacity terms are able to allow the interactions among reactive species and solvent prevail over the simple dilution effect. The system trend may seem counter-intuitive: the more solvent in the system the more difficult should be for the reactant to collide and react, especially if the reactants result to be poorly soluble in the solvent. For this reason, an analysis of the fugacity coefficients has been made. They result to increase as the amount of solvent increase. What can be concluded is that the propensity of both reactants to "escape" from the solvent makes their probability of collision higher and this increase with the solvent amount.

In order to assess the prediction capabilities of the models, batch curves have been computed, but in this instance the thermodynamic equilibrium constant and the apparent equilibrium constant are set to the theoretical value found applying equation 3.26. Since that value causes a sensible departure from experimental equilibrium condition the focus is only on the reaction kinetics. According to the physical meaning of the rate constant, namely the probability that the molecule that collide actually react it should be independent on the type of solvent and on the different amount of the same solvent. Therefore, conversely to the accuracy analysis, the prediction assessment employs a single value of the rate constant for each of the five systems characterised by different solvent amount. Firstly, an analysis on the rate constant values is reported in which the single rate constants, obtained from the regression of each experiment dataset, are normalized from the rate constant regressed to the entire dataset available at once (global rate constant). The result obtained are shown in Figure 3.6.

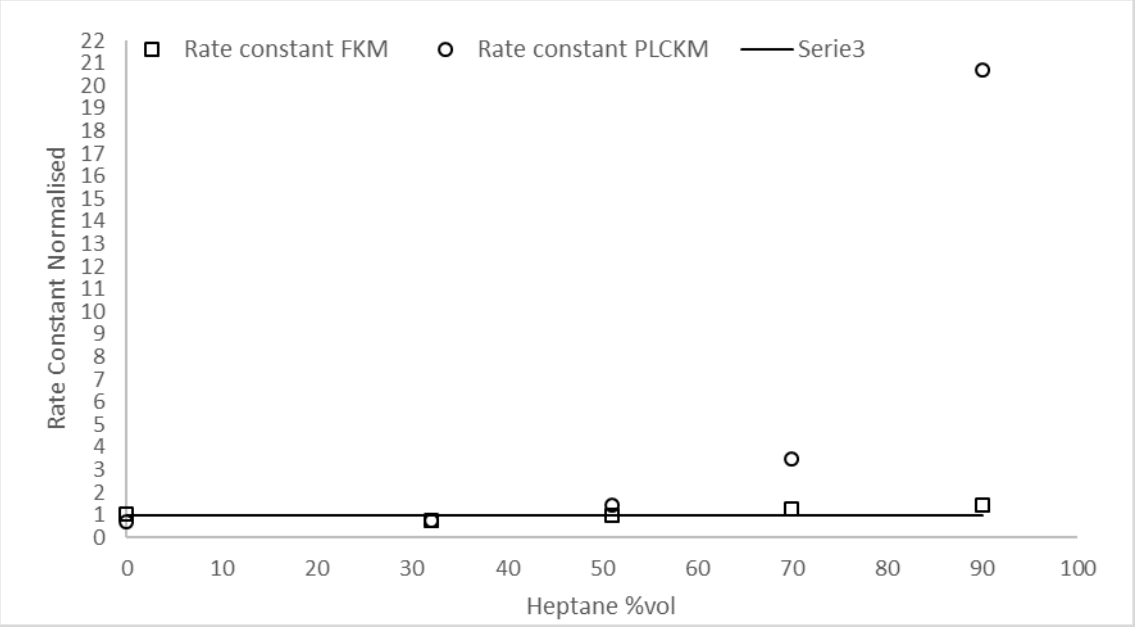


Figure 3.6 *FKM and PLCKM rate constant values for different amount of solvent (%vol) normalised over the global rate constant.*

As it can be seen from Figure 3.6 the value of the normalized rate constants is very close to one for FKM model, meaning that the results are in good agreement with what was expected from the theory. On the other hand, PLCKM is characterised by a variation of the rate constant that reaches more than twenty times the global one. This is due to the different driving force taken into account in the two models. As aforementioned the FKM employs as driving force the fugacity terms that can describe accurately the probability of the reactant molecules to collide thanks to the capability to consider all the interactions present in the system. Since the driving force term is in good agreement with what is expected from its physical meaning, the rate constant should show the same behaviour. Conversely the PLCKM employing as driving force concentration terms is not capable to accurately describe the probability of the molecules to collide, therefore the rate constant must also compensate for this discrepancy and loose the independence from the solvent adopted.

The reaction kinetics predictions are obtained employing the global rate constant and setting the thermodynamic equilibrium constant and the apparent equilibrium constant to 0.35 (value theoretically calculated). In Table 3.4 the numerical value of the global rate constant and the statistics related to the quality of parameter estimates are reported.

Table 3.4 *Numerical value of the global rate constant and the statistics related to the quality of parameter estimates for FKM and PLCKM*

	Global rate constant [mol/m^3 s]	*95% confidence*	*reference t-value (95%)*	*95% t-value*
FKM	10322.3	271.79	1.69	38.26
PLCKM	1.31E-07	3.51E-09	1.69	36.97

From Table 3.4 it can be seen that the values of the global rate constant for the two models have thirteen orders of magnitude of difference, this is due to the different driving force employed in the kinetic model used, namely fugacity (equation 3.33) or concentration (equation 3.34) terms. For what regards the confidence intervals, the “10% rule of thumb”, which requires that the 95% confidence interval is less than the 10% of the absolute value of the parameter, it is satisfied by both models. Since the 95% t-value is more than twenty times the refence one for both models, the global rate constant is a statistic significant parameter. The batch curves obtained have been reported in Figure 3.7 only for the reaction kinetic range.

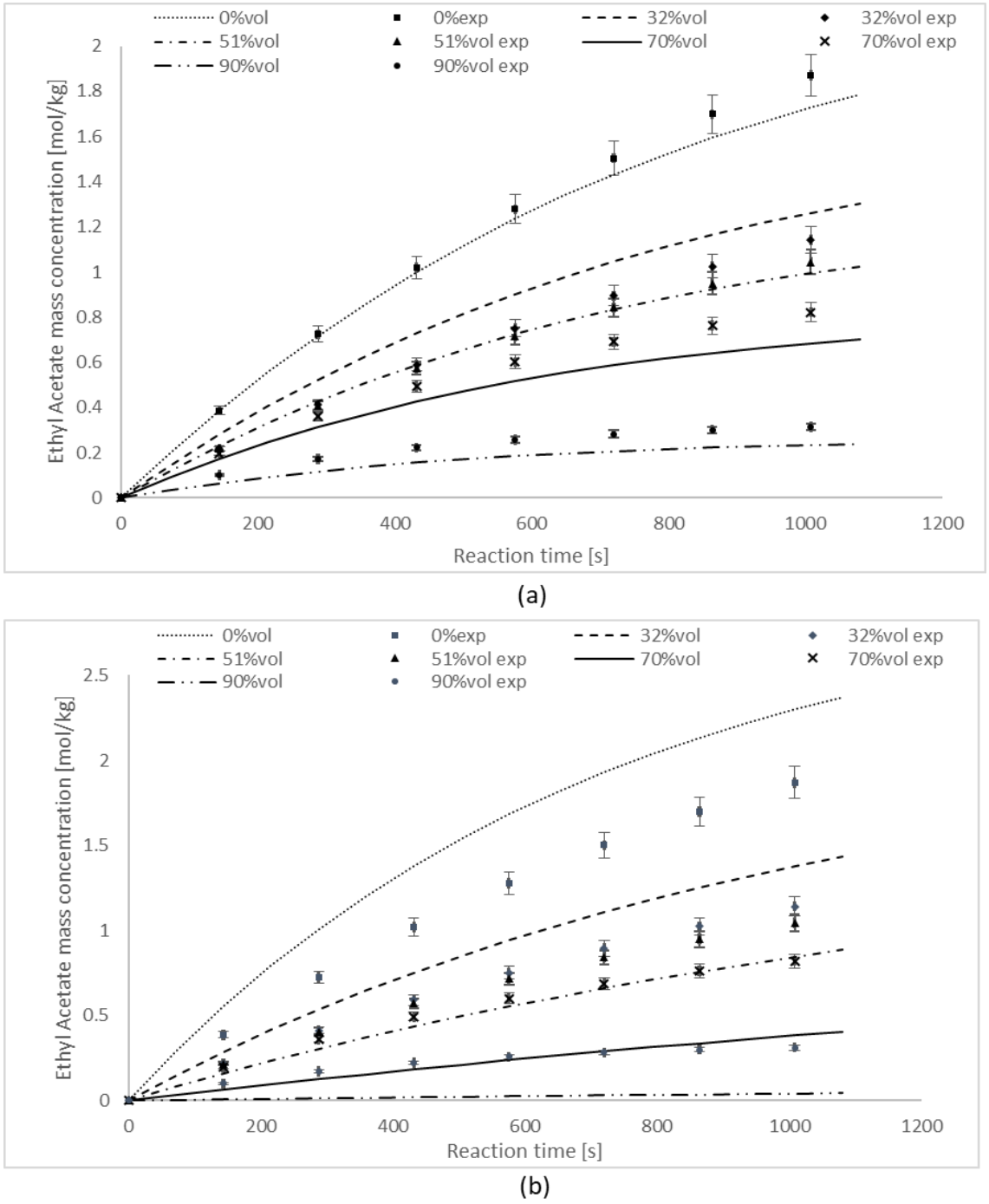


Figure 3.7 *Batch concentration-reaction time plots for ethyl acetate. Lines denote model predictions for the same reactive system characterised by variable amount of solvent. Symbols with error bars (linear error interval with coefficient0.05 and bias 10^{-4}) represent denotes experimental reading. The kinetic models employed for the calculation are: (a) FKM and (b) PLCKM. The parameter used are 0.35 as equilibrium constant and the values reported in Table 3.4 as rate constant*

Comparing Figure 3.7*a* and Figure 3.7*b* it is clear that FKM provides for much more accurate predictions of reaction kinetics with respect PLCKM. This result is in agreement with the theoretical expectations. What is worth noting is that even if the equilibrium condition computation provides for non-reliable results as shown in Figure 3.4, the

computation of the kinetic part keeping the same value of the thermodynamic equilibrium constant results reliable. Therefore, in case of $\Delta G_R(T)$ lower than 10^4 J/mol even if the equilibrium calculation cannot be trusted, it is possible rely on the reaction kinetic computation if it is performed with FKM. In order to show the departure from experimental data applying FKM and PLCKM, parity plots for each different amount of solvent are reported in Figure 3.8. In no case, prediction results obtained from PLCKM are better than those performed through FKM. The latter in turn, provides calculated Ethyl Acetate concentration that are in very good agreement with the experiments: almost in every system reported it is very close to the parity line.

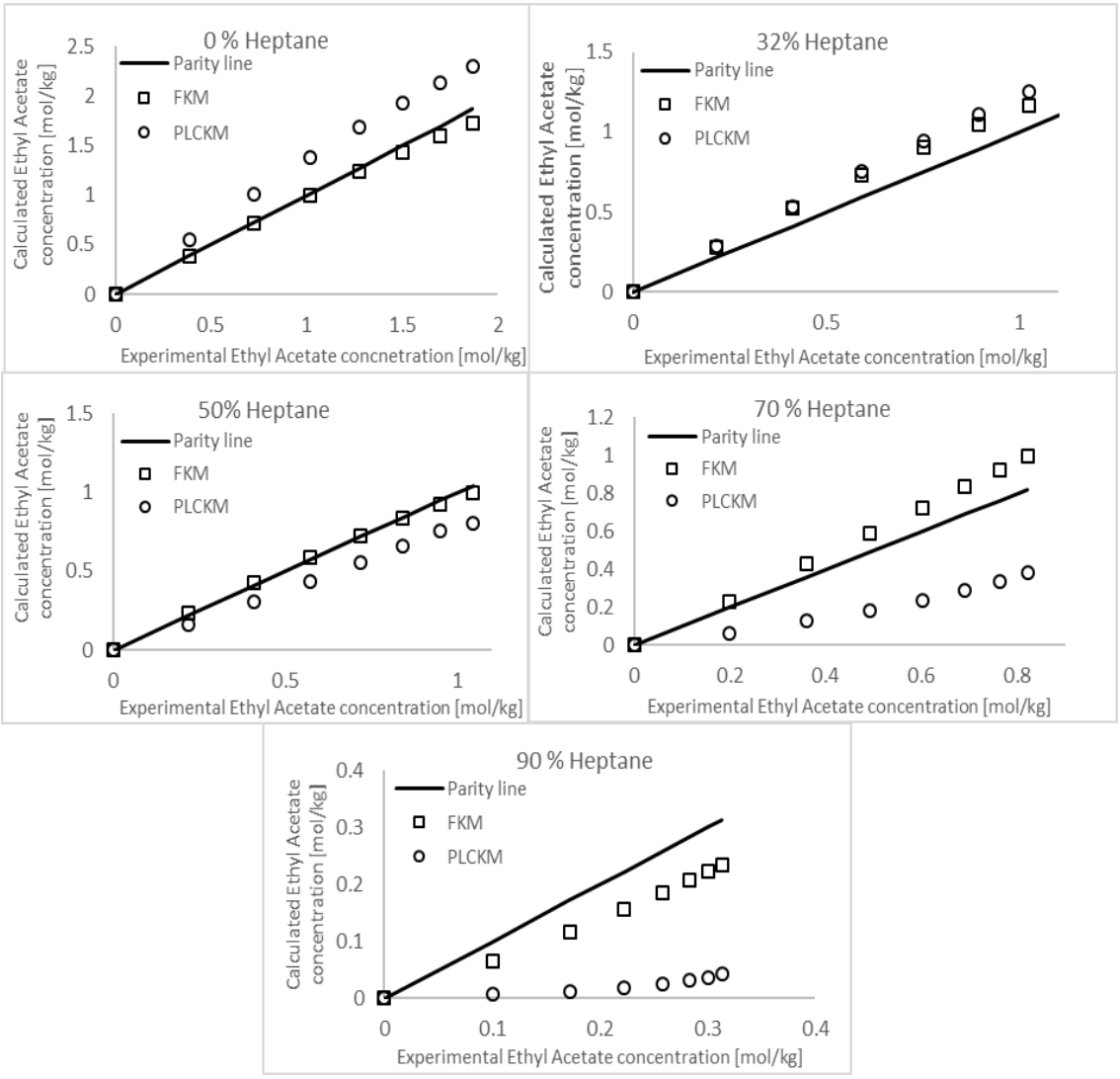


Figure 3.8 *Parity plots that illustrate the ethyl acetate concentration predicted applying FKM and PLCKM for each ethyl acetate concentration measurements. Each parity plot refers to the same reactive system where different amount (%vol) of heptane is employed.*

Comparing Figure 3.4 to Figure 3.5 it is evident that the use of the experimental equilibrium constant instead of the thermodynamic one affects not only the reaction equilibrium calculations but also the reaction kinetics description. Therefore, what has been done is to consider the equilibrium constant as a parameter to regress together with the rate constant in order to give more flexibility to the model. The calculations have been carried out for both FKM and PLCKM; the initial guess for the parameters are set as the values found in the previous calculations. Table 3.5 shows the numerical values of the global rate constant, the inverse of the equilibrium constant and the statistics related to the quality of parameters estimated. The reference 95% t-value depends on the experimental measurements and on the number of parameters regressed, therefore is the same for both systems and it is equal to 1.648.

Table 3.5 *Numerical value of the optimal estimate value of regressed parameters: k (rate constant) and 1/K (inverse of the equilibrium constant) and statistics related to the quality of parameter estimates for FKM and PLCKM*

	Optimal estimate	95% confidence intervals	95% t-value
FKM k [mol/m^3 s]	9852.12	198.73	49.87
FKM 1/K	1.09	0.022	49.89
PLCKM k [mol/m^3 s]	1.87E-07	3.53E-09	53.17
PLCKM 1/K	1.12	0.024	46.25

From Table 3.5 it can be seen that the rate constants, in particular the FKM one, have the same order of magnitude of the initial guess, this is reasonable because they provide for the reaction kinetics that FKM has demonstrated to be able to predict very well even with non-reliable equilibrium term present in the model. All the parameter regressed show a 95% confidence interval values that are lower than the 10% of the absolute value of parameters regressed; this a first evidence of parameters estimation quality. Since the 95% t-value is more than 25 times the reference one for all the parameters regressed, the parameters employed result to be statistically significant. It worth pointing out that the values obtained for the equilibrium constants regressed are both very close to the experimental one. The batch curves obtained employing the parameters reported in Table 3.5 are presented in Figure 3.9

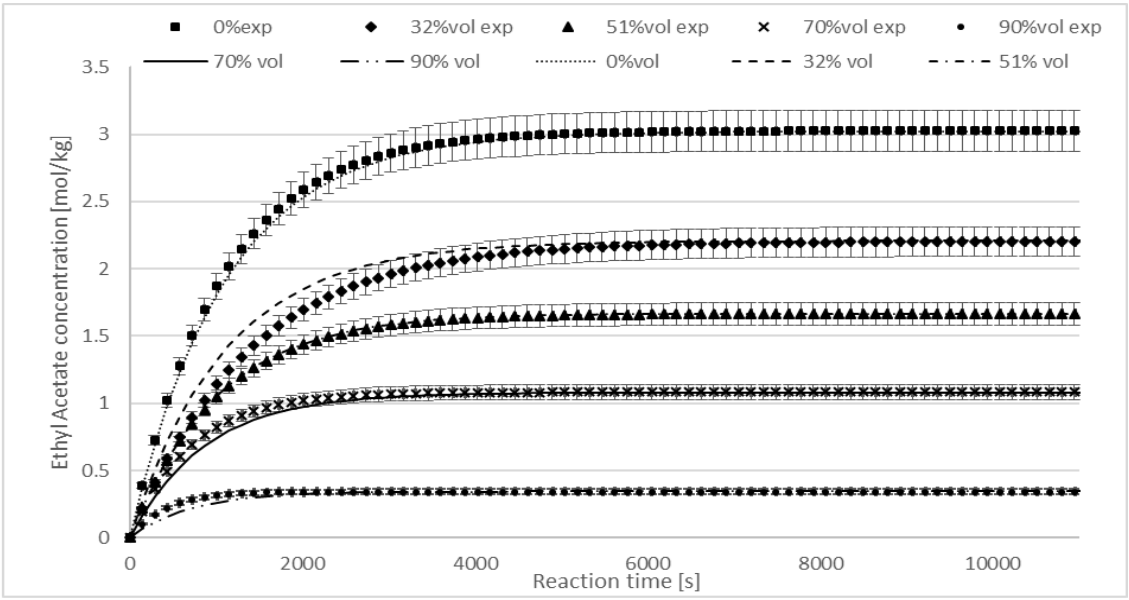


Figure 3.9 *Batch concentration - reaction time plots for ethyl acetate. Lines denote model predictions for the same reactive system characterised by variable amount of solvent. Symbols with error bars (linear error interval with coefficient0.05 and bias 10^{-4}) represent denotes experimental reading. The kinetic models employed for the calculation are: (a) FKM and (b) PLCKM. The parameter used are those in Table 3.3*

From Figure 3.9*a* it can be seen that FKM has significantly improved the kinetic predictions with respect Figure 3.7*a* and it is able to capture the equilibrium conditions. From Figure 3.9*b* the limitations of PLCKM are still evident and regardless the capability of catching the equilibrium conditions the kinetic computation remains poor. The results obtained suggest that regressing both the rate constant and the thermodynamic

equilibrium constant it is a valid option in case of non-reliable thermodynamic equilibrium constant condition if FKM is employed as kinetic model.

3.5.3 Comments and further improvements

The study carried out has underlined four aspects of the models employed: reaction equilibrium representation, system computation accuracy, reaction kinetics model description and process prediction capabilities.

The equilibrium conditions in the system analysed are characterised by a very low Gibbs free energy of reaction, for this reason it has not been possible to apply the rigorous thermodynamic expression to compute the thermodynamic equilibrium constant. From the theoretical point of view, it should be possible to apply the rigorous formulation for any reaction at any operating condition. In order to verify it, further studies are needed; they must be carried out for cases in which the errors in the Gibbs free energy value does not affect significantly the thermodynamic equilibrium value.

If a reliable equilibrium constant is used and the experimental values for all the systems considered are available, from Figure 3.5 it is evident that the batch curves obtained deploying FKM are much more accurate than PLCKM ones, especially because, while, both models allow to reach the equilibrium condition, the reaction kinetics is captured only by FKM. Therefore, it can be stated that FKM provides more accurate calculations than PLCKM when the experimental data for all the systems of interest are provided. Since the values of the rate constant for all the systems considered are very close to each other for FKM, while they depart significantly from each other for PLCKM, it has decided to compare the global rate constant to the single rate constant values obtained from different system in order to quantify the difference. Figure 3.6 shows that for FKM the rate constant remains almost unchanged over systems characterised by different solvent amount. This implies that FKM can describe satisfactorily the reaction kinetics over the entire range of solvent amount potentially present in the system. This finding proves that the fugacity driving force embeds all the different interactions that the variation of the solvent amount can cause. Thereby, FKM should be capable of predict the transesterification system studied for every amount of heptane employed without the need of any experimental data. Accordingly, Figure 3.9*a* shows the batch curves obtained deploying FKM with a constant value of rate constant over a wide range of heptane amount in the reactive system. The results obtained are in very good agreement with experimental data. Further studies should be carried out to prove that the results obtained for the different amount of the same solvent are analogous to those for different system employed in the same reactive system.

Finally, when FKM is used, even if the thermodynamic equilibrium constant is not reliable, it is still possible to obtain an accurate representation of the reaction kinetics as it is shown in Figure 3.7*a*. A second valid alternative to represent even the reaction equilibrium with satisfactory accuracy is to regress together with the rate constant the thermodynamic equilibrium constant as shown in Figure 3.9*a*

Chapter 4

Fugacity impact on complex kinetic models

In §3.1 it has been explained in a comprehensive way the complexity of the interactions involved in a multiphase catalytic system and how they differ from those involved in homogeneous liquid phase systems (e.g. transesterification). For this reason, this chapter is dedicated to the study of multiphase catalytic systems through the analysis of 4-phenil-2-butanone hydrogenation. Firstly, the industrial applications of hydrogenation systems and the objectives related to the specific case study are reported. Then an accurate description of solvent effects on multiphase catalytic systems is addressed followed by a detailed description of the case study carried out and the results achieved.

4.1 Objectives

Since this study regards a catalytic hydrogenation system, it is characterised by three different phases (liquid, solid and gas). Therefore, the interactions involved are of a different nature, thus, they cannot be fully captured only by the fugacities that are the key terms in FKM. For this reason, FKM results unable to accurately describe this system and, therefore, a third model has been considered, i.e. the Langmuir-Hinshelwood-Hougen-Watson (LHHW) model which involves additional parameters. Therefore, the main objectives for this study regards the evaluation of:

- The uncertainty decrease in the application of experimental data. This is obtained through the assessment of the increase on accuracy, in the system representation, applying LHHW model when fugacities are substituted to concentration terms.
- The impact of using fugacity terms instead of concentrations in LHHW kinetic model on the parameters estimated
- Prediction of the performance as a comparison between LHHW kinetic models employing a constant value for all the parameters involved to describe the impact of different solvents in the same reactive system

The general purpose is, therefore, to understand if changing the reaction driving force from concentration to fugacity allows to compensate for the lack in the description of thermodynamic non-idealities.

4.2 Hydrogenation case study

Hydrogenation is a chemical reaction of great importance to the petrochemical, pharmaceutical and fine chemical industries. It consist of a chemical reaction between molecular hydrogen (H_2) and another compound or element, usually in the presence of a catalyst, with the purpose of reducing or saturating a certain molecule. Some of the industrial uses of hydrogenation have been reported in the literature. In the food industry hydrogenation is employed, for instance, to completely or partially saturate the unsaturated fatty acids in vegetable oils to convert them into solid or semi-solid fats (e.g. margarine). In petrochemical processes, hydrogenation is used, for instance, to convert alkenes and aromatics into saturated alkanes (paraffins) and cycloalkanes (naphthenes), which are less toxic and less reactive. In the fine chemical and active pharmaceutical ingredient industries, the hydrogenation reaction is often an important step in producing the end product. Finally, among the chemical specialties, Xylitol, a polyol, is produced by hydrogenation of the sugar xylose. Catalytic hydrogenation, in particular, has evolved into a key process technology for the manufacture of pharmaceutical and fine chemicals, replacing chemical reduction methods that generate large quantities of waste. For this reason, hydrogenation reactions represent one of the typical and ubiquitous industrial processes, indeed, nearly 20% of all the reaction steps in a typical fine chemical synthesis are catalytic hydrogenation (Hessel, et al., 1999).

4.2.1 Solvent effect multiphase catalytic systems

Solvents are widespread used for many reactive systems such as catalytic liquid-phase reactions and crystallization process. The choice of the solvent is never arbitrary because it can be highly beneficial or detrimental to the system considered. With a focus on catalytic reactions, the solvent affects significantly both activity and selectivity. As well as influencing reactant and product behaviour, the solvent may also interact with the metal and/or support of the catalyst employed in the system. While the former interactions can be taken into account by activity or fugacity terms in the kinetic model employed the latter interactions cannot be described by such terms. The same holds true for the interactions between the substrate and the catalyst (metal plus support). What is needed, to take into account such interactions, is deriving the correct kinetic model description for specific reactions. This comprises a certain number of parameters, that regressed using specific system experimental data, embeds the system interactions. Numerous effects arise also from the solvent interactions with the substrate and ultimately from the variation of the system composition (formed mainly by the solvent) and the solubility of the hydrogen. This present a particular challenge to industry where the following problems can arise as a result of the lack in the deep knowledge of complex interactions:

- High ratio between by-products and products
- Difficult prediction of plant scale reactor performance
- Catalyst, feedstock and solvent choice constraints due to economic feasibility and environmental restrictions.

For this reason, the need to understand and predict the solvent effects is important and this field can be explored through methodologies that are able to probe reactions behaviour from a fundamental physical and chemical stand point but are pragmatic at the same time. Such characteristics are required in order to cover a broad spectrum of applications.

4.2.2 Solvent effect multiphase catalytic systems

This case study involves a multiphase catalytic hydrogenation system, whereby an initial reactant, 4-phenyl-2-butanone (PBN), undergoes two distinctive routes to produce the fully hydrogenated product, 4-cyclohexyl-2-butanol (CBL). Intermediate compounds are 4-phenyl-2-butanol (PBL), produced by hydrogenation of the carbonyl group and 4-cyclohexyl-2-butanone (CBN) by hydrogenation of the aromatic ring. The reaction scheme is shown in Figure 4.1. and features two reaction types, namely ketone hydrogenation (green dashed arrows) and aromatic hydrogenation (blue dashed arrows). The solvents employed in the reactive system are n-hexane, cyclohexane, ethanol and 1-propanol; all of them have been used in high purity form.

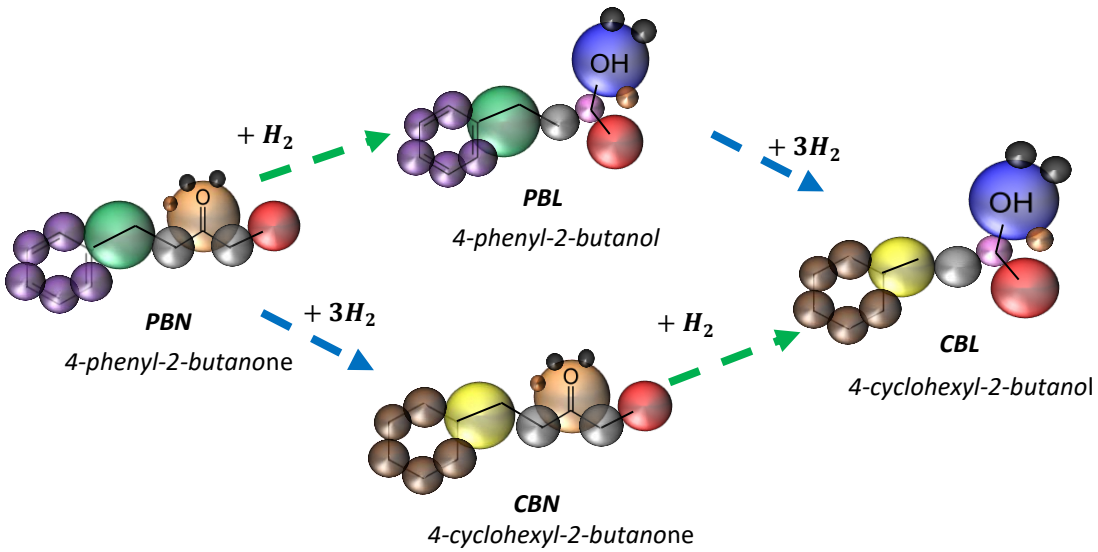


Figure 4.1 *Catalytic hydrogenation of PBN to CBL scheme: two ketone hydrogenation reactions (green dashed arrows) in series with two aromatic ring hydrogenation reactions (blue dashed line). Each compound has been represented as the contribution of its characteristic functional groups represented by colourful spheres.*

It worth pointing out that the species involved in the reaction system analysed are not present in any database available in gPROMS; what has been done is, therefore, to rely on SAFT-γ Mie advanced thermodynamic model. It allows to compute the thermodynamic properties required as density, enthalpy and fugacity coefficients for each single species and even for the mixture. In order to use SAFT-γ Mie, each molecule has been described in terms of functional group set; they can be seen as multicolour spheres in Figure 4.1.

The reactive system is carried out on a small scale (100 mL) batch reactor that would typically be used in a laboratory for the measurement and determination of reaction kinetic. Understanding the kinetics behind these reactions, specifically on how the intermediates are formed, is a critical precursor to design of a scaled-up version of this reactor and optimisation of conditions to favour a specific reaction route.

Since the real experiments were carried out by Wilkinson et al. (2015) in such a way so that mass transfer gradients and reactor adiabatic temperature rise were minimised, namely avoid mass and heat transfer limitation, the same conditions have been retained in the system representation. In order to reproduce the PBN hydrogenation in gPROMS FormulatedProducts, the following assumptions are made:

- The system is modelled as a two-phase gas-liquid multiphase stirred tank reactor (STR). Organic reactants and solvent are originally in the liquid (organic) phase and hydrogen is originally in the gas phase.
- The reactions take place in the presence of Pt/TiO_2 as catalyst. The catalyst is in a small particulate form within the liquid (slurry). This can be simplified to a pseudo-homogeneous reaction model in the liquid phase.
- Mass transfer of hydrogen between the gas and liquid phase is permitted, with the mass transfer rate constant set at a value faster than the rates of reaction to ensure it is not limiting.
- The hydrogen gas is maintained at 5 bar in the headspace of the reactor. The level and pressure maintained by a level sensor and controller.[2]
- The reactor is assumed to be isothermal at 343.15 K and heat of reaction effects are not considered. This is achieved by a temperature sensor and controller.
- Changes to liquid composition over time are assessed by a liquid composition sensor.

[2] The use of a level controller to keep the pressure constant is due to the absence of a pressure controller in gPROMS FormulatedProducts. The reason is that since the species involved in pharmaceutical applications are mainly in liquid or solid phases the effect of pressure change is neglected.

The gas-liquid hydrogenation process that occurs in a stirred tank reactor in represented by the flowsheet shown in Figure 4.2 in which temperature, level controllers and the composition analyser are reported.

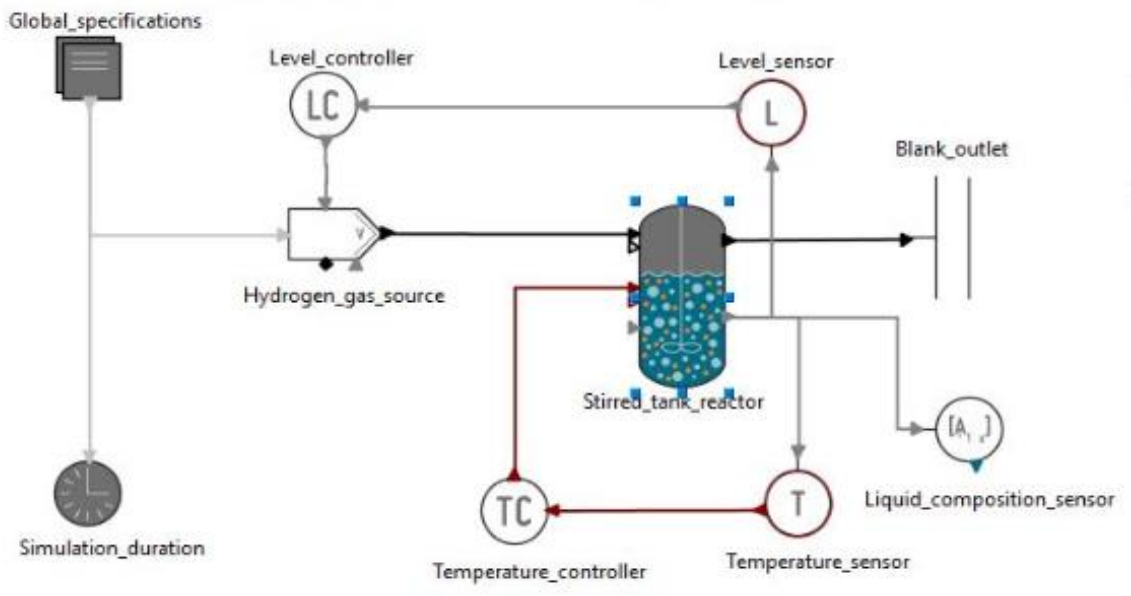


Figure 4.2 *Hydrogenation process flowsheet*

The experimental data employed in this case study are those published by Wilkinson et al. (2015); they regard system species concentration [mol/m^3] over time (from 0 to 120 minutes) in case of four different solvents employed: n-hexane, cyclohexane, 1-propanol and ethanol.

4.2.3 Model employed

The reactive system, as aforementioned, is carried out in a stirred tank reactor, gPROMS FormulatedProducts provides a unit operation called “CSTR multiphase” which contains the system of equations that describes the process. As in the transesterification case what has been changed is the kinetic model through the custom template. The assumptions made to derive “CSTR multiphase” are the same listed for “CSTR one phase liquid” plus the following:

- Partition coefficients are temperature independent.
- Mass transfer coefficients are independent from temperature and flow conditions.

For this system is not advisable employing the same kinetic model structure used to describe the transesterification system. The reason is related to the absence of any adjustable parameters that takes into account the several interactions involved in the hydrogenation system that cannot be captured by thermodynamics, thus neither by the

concentration. The rate constant, which is the only parameter in both FKM and PLCKM cannot compensate for this system representation deficiency. FKM or PLCKM can be employed in order to provide useful initial guesses for rate constant in more complicated kinetic models that deploy fugacity or concentrations as reaction driving force.

When kinetic models are used to describe catalytic liquid multi-phase reactions, as the hydrogenation reactive system considered, it is required that such models have fundamental mechanistic basis together with estimated parameters that result to be physically meaningful in value and are statistically significant to justify their presence. Therefore, the kinetic model structure applied in this study derives from a fundamental kinetic model developed incorporating statistical analysis methods to strengthen the foundations of mechanistically sound kinetic models. It is based on 2-site model that was determined to be the most appropriate (Wilkinson et al., 2015), describing aromatic hydrogenation (postulated to be over a platinum site) and ketone hydrogenation (postulated to be at the platinum-titania interface). Therefore, from Wilkinson et al. (2015) the kinetic model structure found as the more appropriate one is that of Langmuir-Hinshelwood-Hugen-Watson (LHHW) type:

$$r_j = k'_j \cdot \frac{C_i \cdot C_{H_2}}{(1 + K_{PBN} \cdot C_{PBN} + K_{CBN} \cdot C_{CBN})^2} \tag{4.1}$$

Where k'_j is a lumped rate constant [m^3/mol s], C_i represents the concentration [mol/m^3] of the reactant that is not hydrogen in reaction j, K_{PBN} and K_{CBN} the adsorption constants [m^3/mol] related to 4-phenyl-2-butanone and 4-cyclohexyl-2-butanone respectively. The way in which the model has been developed and the presence of adsorption constants are strictly related to the need of capture the role of solvents in influencing reaction selectivity and reactant adsorption but also the interaction with the metal and support of the catalyst during reaction. Through the study the model just reported will be indicated with the acronym LHC that stands for Langmuir-Hinshelwood concentration-based model. In order to understand if the capability of capture interactions among reactive species and solvent would be beneficial to the system description, a second model has been employed in this study. It is obtained from equation 4.1 substituting the concentration terms with the ratio of fugacity over standard pressure:

$$r_j = k_j \cdot \frac{\frac{f_i}{P^\theta} \cdot \frac{f_{H_2}}{P^\theta}}{(1 + K_{PBN} \cdot f_{PBN}/P^\theta + K_{CBN} \cdot f_{CBN}/P^\theta)^2} \tag{4.2}$$

Where k_j is the rate constant [mol/m^3 s], f_i and f_{H_2} are respectively the fugacity of hydrogen and the fugacity of the reactant that is not hydrogen in reaction j in the mixture.

Through the study, the model just reported is indicated with the acronym LHF that stands for Langmuir-Hinshelwood fugacity-based model.
Both 4.1 and 4.2 should be applied to each reaction involved in the hydrogenation system; in the specific case analysed the reactions involved are four: two in series and two in parallel. Therefore, the number of parameters that characterises the system would be twelve if the two adsorption constants are considered different for each reaction.
Some assumption can be made to decrease the parameters presented in the system:

- K_{PBN} and K_{CBN} are assumed to be the same for the same type of reaction, namely ketone and aromatic hydrogenation. This assumption is justified, because the active sites involved are the same for the same type of reaction.
- K_{PBN} can be eliminated since its estimation leads to a very small t-value and a wide 95% confidence interval (Wilkinson et al., 2015)
- The rate constant is assumed to be the same for the same type of reaction, since the reactants involved in the same path have very similar structure and the same reactive groups.

The assumptions made allow to reduce the number of parameters from twelve to four. A statistical analysis has been carried out to evaluate the statistical significance of the parameters employed. The evaluation is based on 95% t-value and the 95% confidence interval. The results are reported in Table 4.1 for the four different solvents.

Table 4.1 *Statistical analysis of four parameters regressed employing LHC based on the 95% confidence interval and 95% t-value for the same reactive system with different solvent employed. The red box indicates values for which the 95% t-value is lower than the reference one and at the same time the 95% confidence interval results to be greater than the 10% absolute value of the parameters regressed. The latter condition alone is indicated by the yellow box.*

LHC: 4 Parameters	*k*_aromatic [m^3/mol s]	*k*_ketone [m^3/mol s]	Kads_aromatic (CBN) [m3/mol]	Kads_ketone (CBN) [m3/mol]
1-Propanol	3.77E-06	1.59E-05	2.23E-03	1.00E-01
95% confidence	1.40E-06	4.34E-05	8.80E-03	2.30E-01
95% t-value	2.691	0.3672	0.4363	0.2531
Cyclohexane	9.19E-06	2.44E-06	7.03E-03	6.49E-03
95% confidence	4.36E-06	1.23E-06	7.31E-03	7.38E-03
95% t-value	2.108	1.981	0.9618	0.8791
n-Hexane	1.49E-05	2.39E-06	1.98E-03	1.56E-03
95% confidence	4.46E-06	1.10E-06	2.70E-03	3.90E-03
95% t-value	3.34	2.17	0.74	0.4
Ethanol	3.11E-05	1.11E-05	7.83E-02	5.28E-02
95% confidence	1.67E-05	1.84E-05	2.71E-01	1.11E-01
95% t-value	0.2982	0.601	0.2891	0.4735
Reference t-value (95%):	1.6655			

From Table 4.1 it can be seen that the quality of the parameters estimated is and some of them are statistically meaningless as well. Therefore, considering that the adsorption constants are the parameters with the lowest t-value it has been decided to reduce the number of the parameters regress to three. This has been obtained considering a single adsorption constant for the entire reactive system. The statistical analysis has been repeated for three parameters model and the results are reported in Table 4.2.

Table 4.2 *Statistical analysis of three parameters regressed employing LHC based on the 95% confidence interval and 95% t-value for the same reactive system with different solvent employed. The red box indicates values for which the 95% t-value is lower than the reference one and at the same time the 95% confidence interval results to be greater than the 10% absolute value of the parameters regressed. The latter condition alone is indicated by the yellow box.*

LHC: 3 Parameters	*k*_aromatic [m^3/mol s]	*k*_ketone [m^3/mol s]	Kads (CBN) [m3/mol]
1-Propanol	1.33E-05	1.14E-05	0.0415136
95% confidence	6.59E-06	6.05E-06	0.02224
95% t-value	2.018	1.879	1.867
Cyclohexane	9.19E-06	2.44E-06	7.03E-03
95% confidence	3.33E-06	9.90E-07	5.10E-03
95% t-value	2.75	2.534	1.672
n-Hexane	9.82E-06	1.64E-06	1.73E-03
95% confidence	4.05E-06	7.49E-07	2.39E-03
95% t-value	3.69	3.35	0.82
Ethanol	3.11E-06	9.01E-06	0.046959
95% confidence	1.82E-06	4.80E-06	0.04319
95% t-value	1.714	1.875	1.687
Reference t-value (95%):	1.6655		

From Table 4.2 it can be seen that the quality of the parameters regressed is increased with respect the previous analysis and that only for one solvent the adsorption constant results to be statistically meaningless. Therefore, the reactive system model is characterised by equation 4.1 in which the subscript *j* represents the type of hydrogenation reaction namely, the aromatic and ketone ones and K_{PBN} is set to zero. Since the model expressed by eq. 4.2 is used as an improvement of LHC all the changes applied on that model are consequently applied also on LHF.

4.2.4 Results and discussions

The experimental data employed to regress the parameters are the concentrations [mol/L] of all the species present in the system at different time that ranges from time zero to two hours. Because of the nature of the experimental data, the variance model set is the linear

one comprising a fixed variance of 0.1 mol/L (representing ±10% of the measured value) and a constant relative term of 10^{-3} mol/L. The result reported are obtained using the same models to compute hydrogenation system calculation but are organised in order to face three main topics: improvement of calculation accuracy, impact on parameter regression, prediction capabilities due to the substitution of concentrations with fugacities in the same kinetic model.

4.2.4.1 Fugacity impact on the system representation accuracy

The model that employ eq.4.1 as kinetic model to describe the hydrogenation system results to be sufficiently accurate in representing the system when the parameters have been regressed from experimental data (Wilkinson et al., 2015). It worth pointing out that the parameters have been regressed for each solvent employed. This is possible when experimental data are available for the systems of interest. It has been carried out the system computation employed both models, LHC and LHF. In order to compare the results, parity plots for each system illustrating the CBL (final product) concentration calculated for each CBL concentration measurements over time are reported in Figure 4.3.

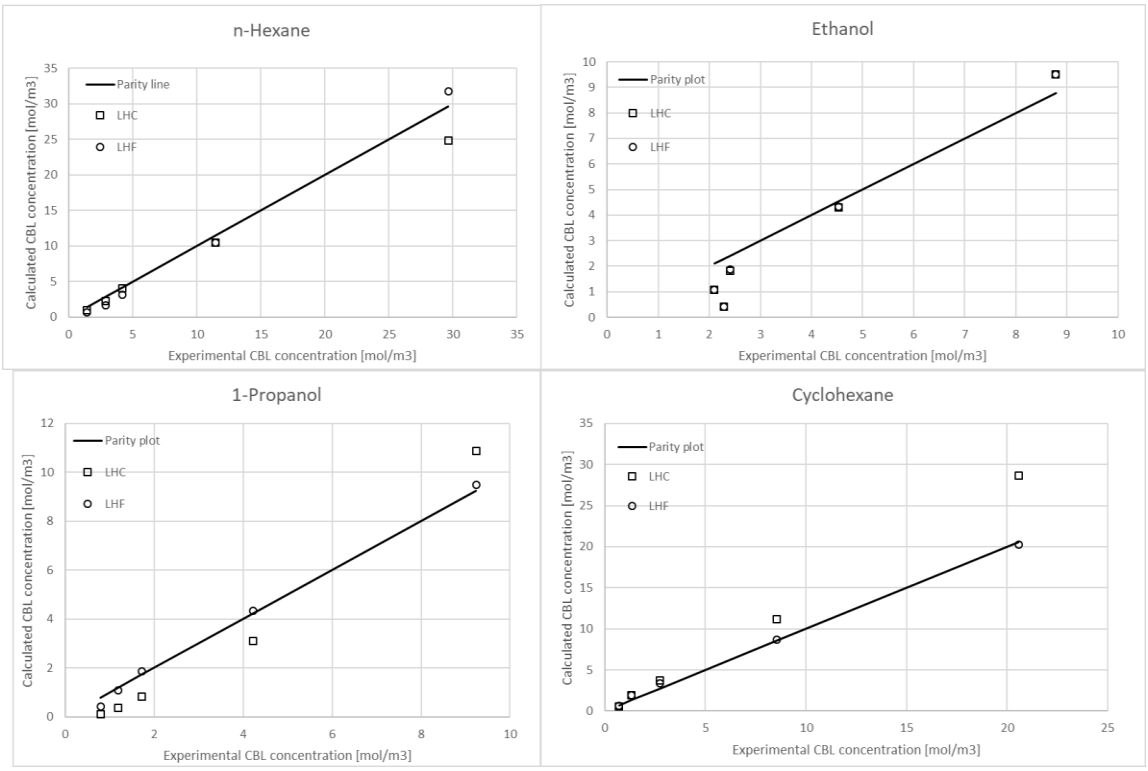


Figure 4.3 *Parity plots that illustrate the CBL concentration calculated applying LHC and LHF for each CBL concentration measurements. Each parity plot refers to the same reactive system with different solvents employed as pure: n-hexane, ethanol, 1-propanol, cyclohexane.*

From Figure 4.3 it can be seen that when alcohols are used as solvents the accuracy of the calculation is very close between the two models. In case of alkane and cycloalkane that as solvents behave in a similar way (Wilkinson et al., 2015), the LHF model is capable of representing the system more accurately than LHC. For this reason, in case of experimental availability is suggested to deploying LHF when possible instead of LHC. The reason is related to the capability of describe the interactions among reactive species and solvent through the fugacity terms.

4.2.4.2 Fugacity impact on the parameter regression

A further aspect for which the impact of substituting fugacity to concentration can be considered is in the effect on the parameter regressed. This comparison can be seen from two sides: the first regards the parameters regressed from each experimental data set and the second regards the parameter regressed from the entire dataset available. The first aspect can be analysed as the potential improvement in the quality of parameters regressed, namely the increase of the difference between the absolute value of the parameter and the 95% confidence interval and the increase of statistical relevance namely, the increase of the 95% t value. In order to assess these aspects Table 4.3 has been reported.

Table 4.3 *Statistical analysis of three parameters regressed employing LHF based on the 95% confidence interval and 95% t-value for the same reactive system with different solvent employed. The green boxes indicate parameters that present quality improvement with respect those in Table 4.2*

LHF: 3 Parameters	*k*_aromatic [m^3/mol s]	*k*_ketone [m^3/mol s]	Kads (CBN) [-]
1-Propanol	50.58	36.25	40071.8
95% confidence	6.12	11.36	10650
95% t-value	8.27	4.39	3.76
Cyclohexane	34.06	9.25396	1159
95% confidence	10.91	3.31	800.9
95% t-value	5.17	1.67	3.80
n-Hexane	43.99	7.77	2555.84
95% confidence	19.8	3.9	2553
95% t-value	3.42	3.06	1.63
Ethanol	16	37.59	49930.1
95% confidence	12.46	27.4	40080
95% t-value	1.68	1.67	1.60
Reference t-value (95%):	1.6655		

It worth pointing out that Table 4.3 has the same rows and columns of Table 4.2, it is therefore, easy to notice the difference in the statistical analysis results between

parameters regressed for the system employing LHC and the one deploying LHF. For simplicity sake the parameters that present quality improvement with respect those obtained applying LHC are highlighted in green boxes. These parameters are characterized therefore, by a greater difference between the 95% interval confidence and the parameters absolute value and they present a greater 95% t-value. The yellow boxes in Table 4.3 has the same meaning of that in Table 4.2. They are left highlighted in yellow because they do not result significantly improved in quality with respect those in Table 4.2. Except for the system for which ethanol is used as solvent the use of fugacity term instead of concentration provide a significant increase in the parameters regressed quality. In order to explore the prediction capability of LHF the idea of regressing the model parameters with all the experimental data available for the four different solvent is taken into account. Since LHF model has not been rigorously derived there is no theoretical base under which it should be able to predict the reactive system behavior over a wide range of solvents employed. What is of interest, therefore, is the capability of the fugacity terms to capture as much interactions as possible resulting in the minimum variation of the parameters. Since the physical meaning of the rate constant as the probability of the molecule that collides have to react still holds true, ideally, a kinetic model that is capable to describe the interactions present in the system should be able to provide accurate results using a unique value for aromatic hydrogenation rate constant and ketone hydrogenation rate constant for any kind of solvent employed in the system.

Table 4.4 *LHC and LHM model global parameters: rate constant for aromatic and ketone hydrogenation and adsorption constant regressed employing multisolvent experimental data*

	*k*_aromatic [m^3/mol s]	*k*_ketone [m^3/mol s]	*K*ads (CBN)
LHC	5.09E-06	1.33E-06	0.0001
LHF	160.44	66.88	70160.1

Figure 4.4 shows the values of the ketone and the aromatic hydrogenation rate constants of Table 4.3 for LHF model and of Table 4.2 for LHC model normalised by values of the same parameters reported in Table 4.4, obtained from the regression from the entire dataset available, which includes the four systems.

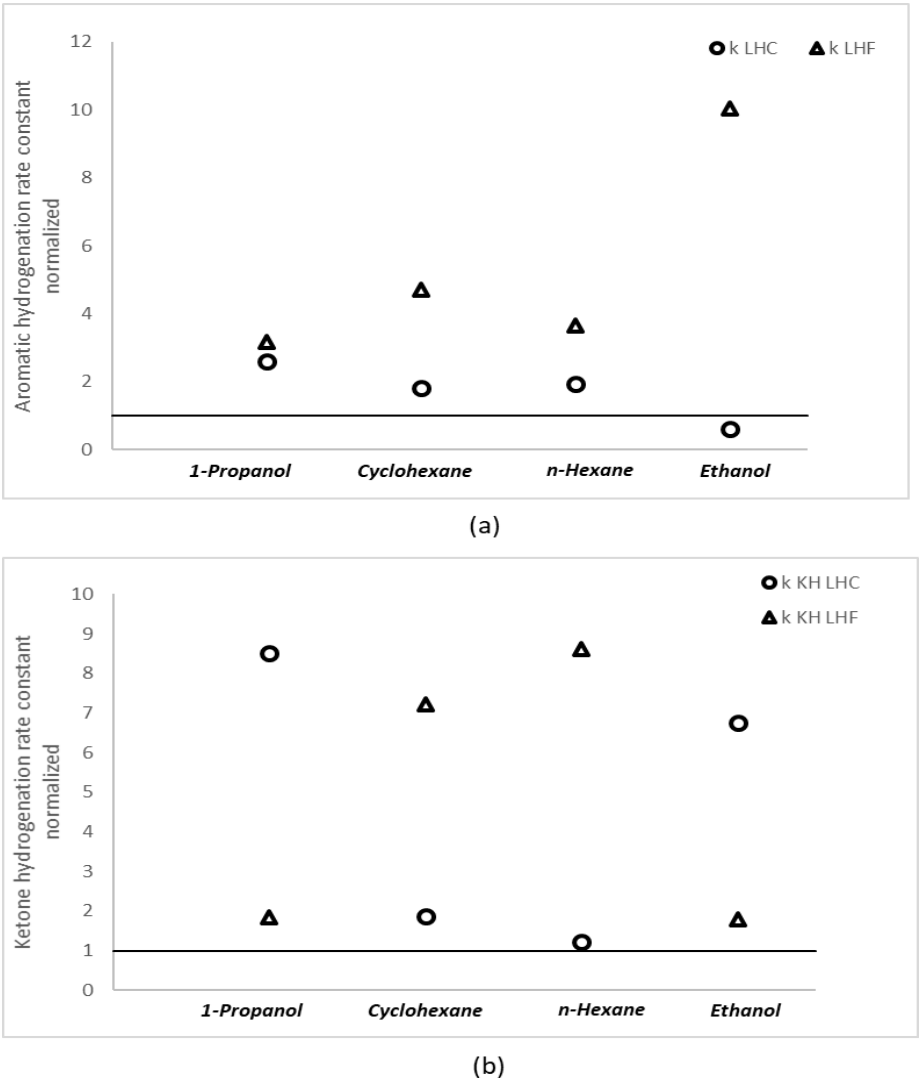


Figure 4.4 *Rate constants values of LHC and LHF obtained from single dataset regression available for each solvent normalised by the same value obtained from a unique parameter regression employing multiple dataset. The rate constant normalised refer to (a) the aromatic hydrogenation and (b) to the ketone hydrogenation.*

From both Figure 4.4*a* and Figure 4.4*b* it can be seen that the variability of the rate constant values for LHF is comparable to the one obtained for LHC. This implies that no significant improvements in the description of the system interactions can be perceived from the rate constant parameter regression point of view using fugacity terms instead of concentration. This finding proves that LHF is not capable to use the same rate constant values for aromatic and ketone hydrogenation respectively and therefore, to capture through the driving force term all the variation in the interactions that occurs changing solvent for the same reactive system. For this reason, the rate constant must compensate for this lack in the interactions description, resulting in a wide range of values assumed from the rate constants. It worth pointing out that for ketone hydrogenation reaction, as can be seen in Figure 4.4*b*, the rate constant values of both LHC and LHF models present

a wide change corresponding to the variation in the chemical family of the solvent considered, namely from alcohols to alkane/cycloalkane.
As a matter of completeness also the normalised CBN adsorption constant has been reported in Figure 3.1.

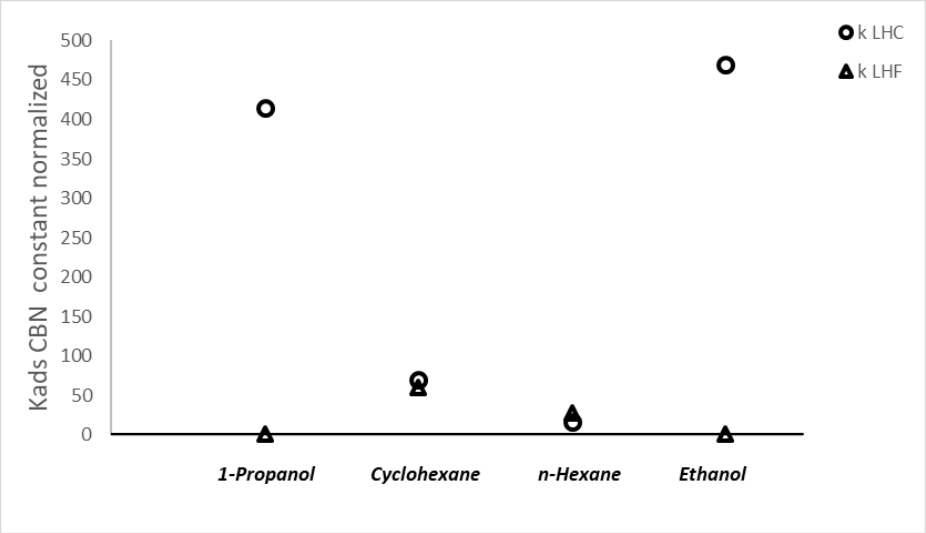
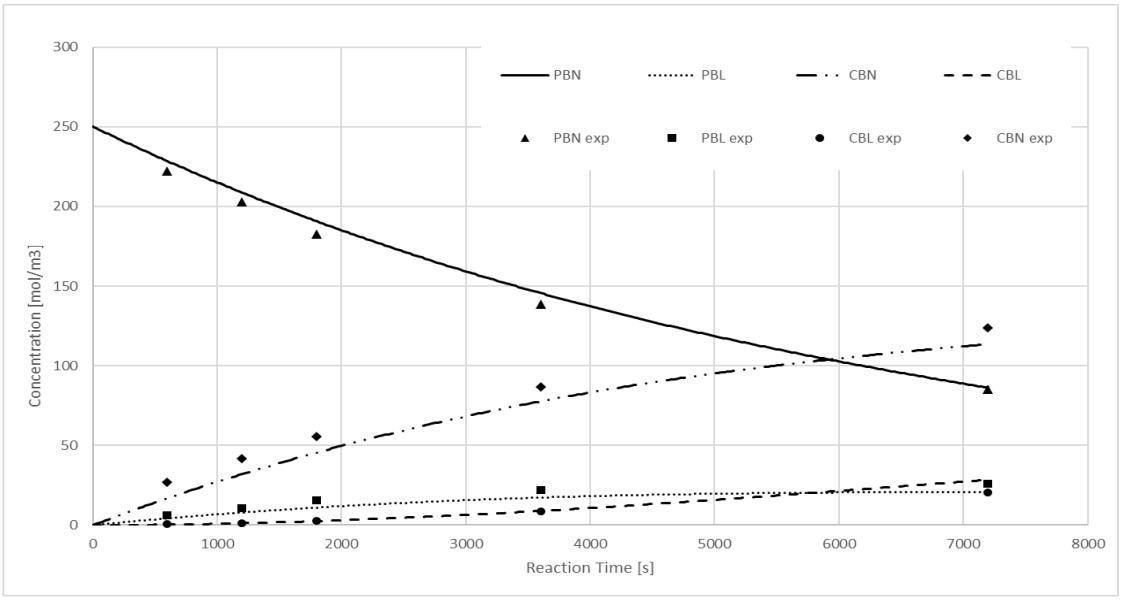


Figure 4.5 *CBN adsorption constant values of LHC and LHF obtained from single dataset regression available for each solvent normalised by the same value obtained from a unique parameter regression employing multiple dataset.*

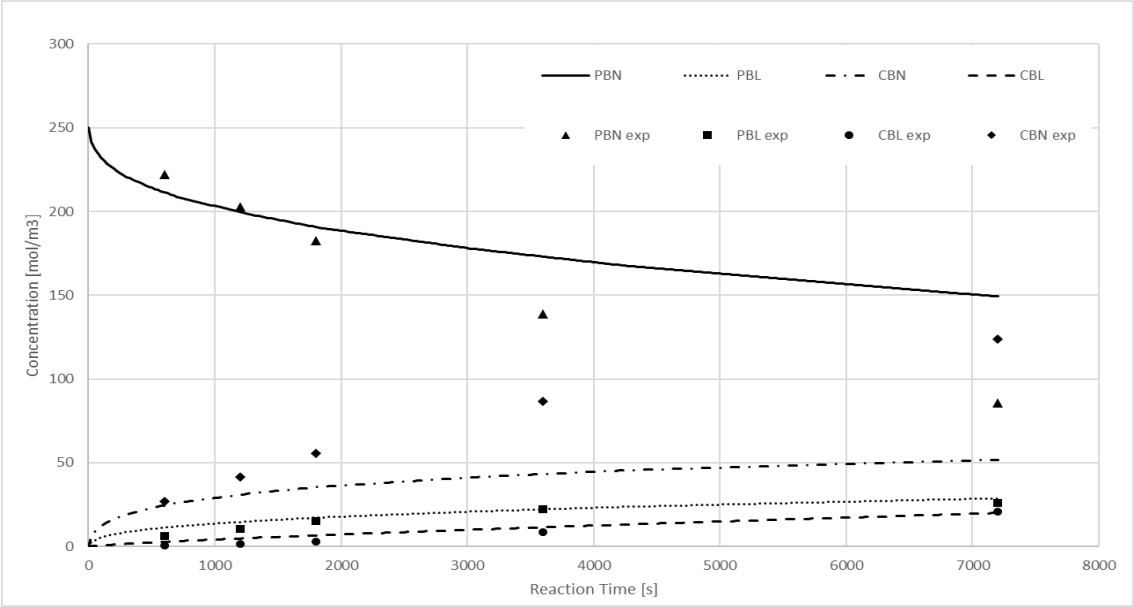
As expected, even for the CBN adsorption constant the values differ widely one another for different solvents for both models employed. The adsorption constant is the main parameter that is responsible for embedding the interactions related to the catalyst, for this reason, the fact that the parameter value respectively for alcohol solvents and alkanes solvents are relatively close while these are significantly different from each other is reasonable. Indeed, it can be noticed that, even if in two different scales, both LHC and LHF are characterised by adsorption constant values that are very close for solvents belonging to the same chemical family while they are considerably different between solvents belonging to different chemical families. A possible explanation is that in the LHC model, the different interactions impact on the system is distributed among the three parameters weighted based on the model structure. This means that a significant increase in the non-idealities and interactions such as going from alkanes solvents to alcohol solvents, is completely redistributed in the three adjustable parameters available. In case of LHF, on the other hand, the interactions that are redistributed over the parameters result to be minimized to those due to the catalyst since the others are taken into account by design from the fugacity terms.

4.2.4.3 LHF capability prediction assessment

In §4.2.4.2 it has been found that the parameters change widely for different solvents considered in the same reactive system. What is investigated here is the impact of employing parameters regressed using the entire dataset to carry out batch curves for the four systems analysed. The idea is that the parameters regressed embed a sort of average of the interactions that cannot be described by thermodynamics. The parameters regressed together with fugacity terms in LHF that take into account the variation of solvent-substrate interactions should be able to compensate for the discrepancy in system interactions description. For this reason, the accuracy of predictions carried out employing LHF should be significantly improved with respect those obtained employing LHC for which no compensation in system interactions description is present. In order to understand if employing the parameters in Table 4.4 allows for the prediction of the reactive system, batch curves for two solvents belonging to different chemical species have been reported separately. In Figure 4.6 batch curves obtained for hydrogenation system employing cyclohexane as solvent have been reported.



(a)



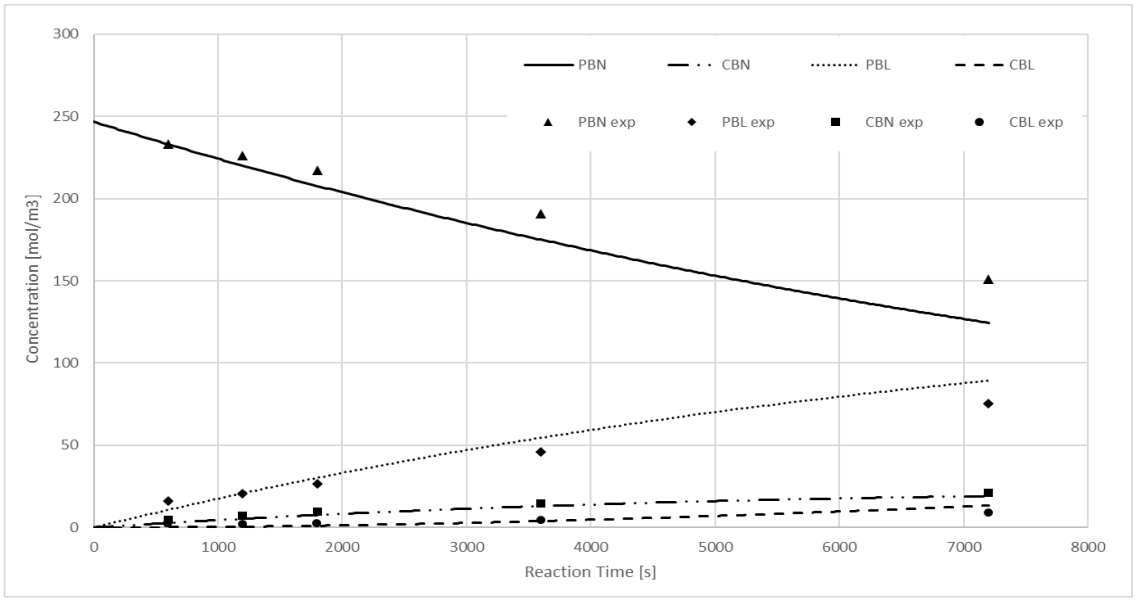
(b)

Figure 4.6 *Batch concentration-reaction time plots for 4-phenyl-2-butanol hydrogenation in* ***cyclohexane*** *at 70 °C, 5bar H_2 pressure and 0.26 mol dm^{-3} [PBN]. Symbols denote experimental readings and lines denote model predictions. The kinetic model employed for the calculation is: (a) LHC and (b) LHF.*

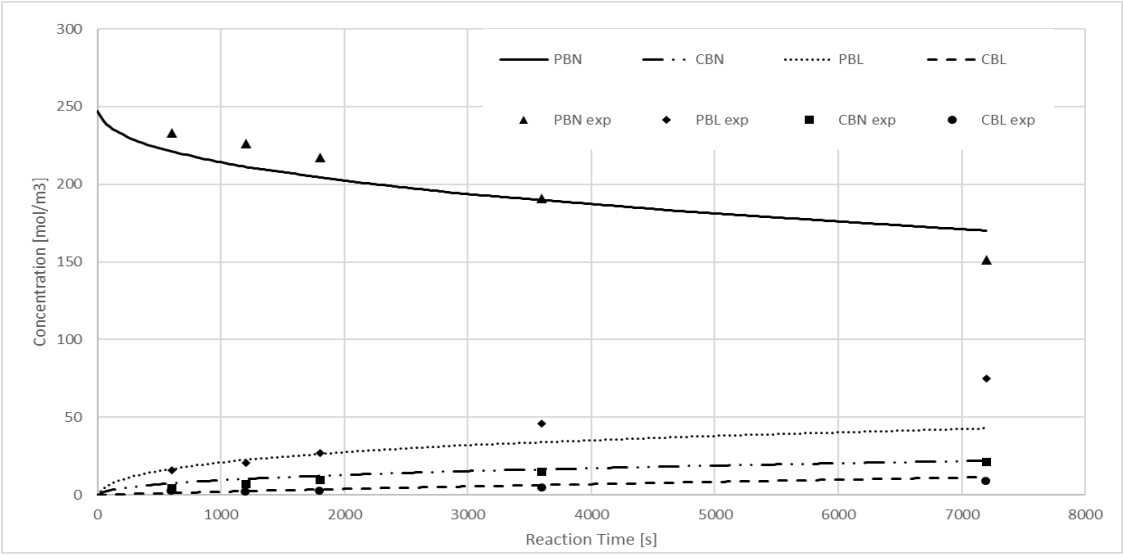
According to Wilkinson et al. (2015) if alkane/cycloalkane solvents are employed due to being apolar and aprotic solvents, they should favour the aromatic hydrogenation route giving rise to CBN and CBL selectivity. Furthermore, employing alkanes/cycloalkane as solvents allows CBN product to be easily desorbed into the liquid phase thanks to the strong solvation of CBN (hydrophobic compound) in the hydrophobic solvent. The aspects just presented can be seen in both plots in Figure 4.6, indeed, the batch curves show the selectivity towards CBN. Moreover, since CBL is obtained from PBL that

derives from the ketone hydrogenation of PBN (not favoured) and the same reaction route involving CBN, what is expected is that CBN concentration is much higher than CBL. Indeed, in both Figure 4.6*a* and Figure 4.6*b* predicted CBN is greater than predicted CBL. If the level of accuracy of the prediction is analysed, it can be clearly seen that even if both plots are qualitatively consistent with the theory batch curves in Figure 4.6*a* are significantly more accurate than those in Figure 4.6*b.*

Batch curves obtained for hydrogenation system employing ethanol as solvent have been reported as well in In Figure 4.7.



(a)



(b)

Figure 4.7 *Batch concentration - reaction time plots for 4-phenyl-2-butanol hydrogenation in* ***ethanol*** *at 70 °C, 5bar H_2 pressure and 0.26 mol dm^{-3} [PBN]. Symbols denote experimental readings and lines denote model predictions. The kinetic model employed for the calculation is: (a) LHC and (b) LHF*

In general, protic, polar solvents as alcohols favour selectivity towards ketone hydrogenation, therefore it is expected that ethanol gives rise to PBL and CBL selectivity. Furthermore, employing alcohols as solvents causes a much more difficult CBN desorption with respect the alkanes/cycloalkanes. This may be understood by the weaker solvation of CBN in the hydrophilic alcohol solvents due to its greater hydrophobicity. The aspects just presented can be seen in both plots in Figure 4.7, indeed, the batch curves show the selectivity towards PBL. Moreover, since CBL is obtained from CBN that derives from the aromatic hydrogenation of PBN (not favoured) and the same reaction route involving PBL what is expected is that PBL concentration is much higher than CBL. Indeed, in both Figure 4.7*a* and Figure 4.7*b* predicted PBL is greater than predicted CBL. If the level of accuracy of the prediction is analysed, it can be seen that even if both plots are qualitatively consistent with the theory but they are not quantitative accurate. For this system it is not clear from the batch curves if Figure 4.7*a* is characterised by a greater prediction accuracy than Figure 4.7*b*. Since four different solvents have been considered, the results which include two of them: 1-propanol and n-heptane, are reported as parity plots for the intermediates compounds, namely PBL and CBN in Figure 4.8 and Figure 4.9 respectively. These two species have been chosen because an accurate prediction of their concentration over time implies the capability of capture the solvent effect on selectivity.

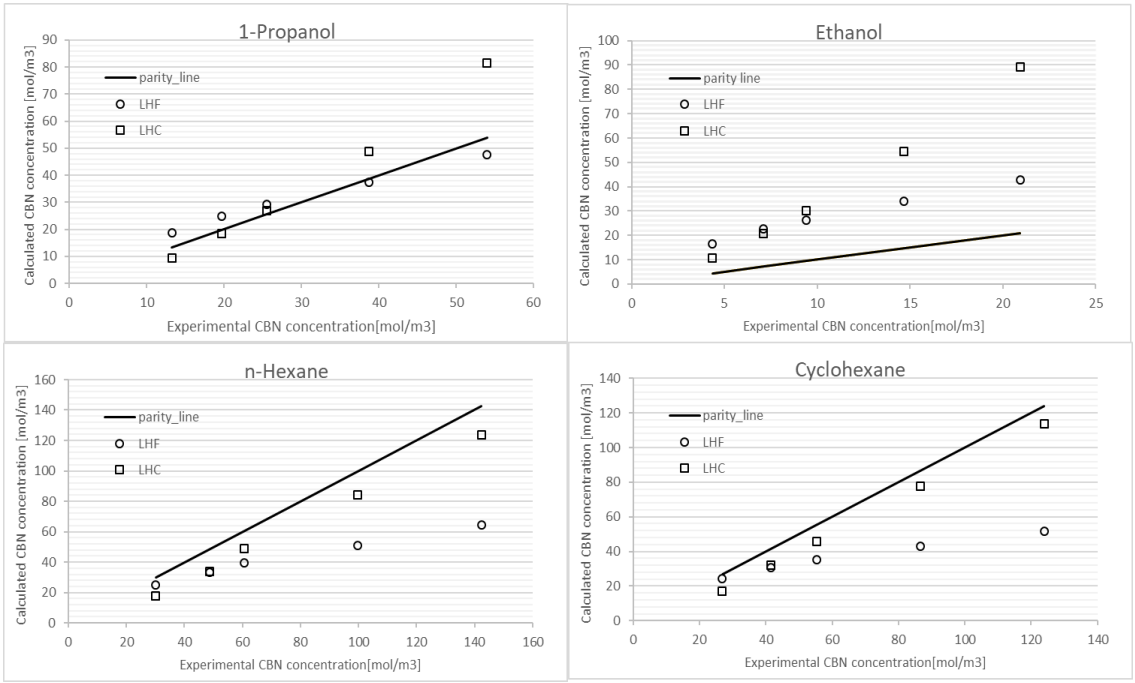


Figure 4.8 *Parity plots that illustrate the **CBN** concentration calculated applying LHC and LHF for each CBN concentration measurements. Each parity plot refers to the same reactive system with different solvents employed as pure: n-hexane, ethanol, 1-propanol, cyclohexane.*

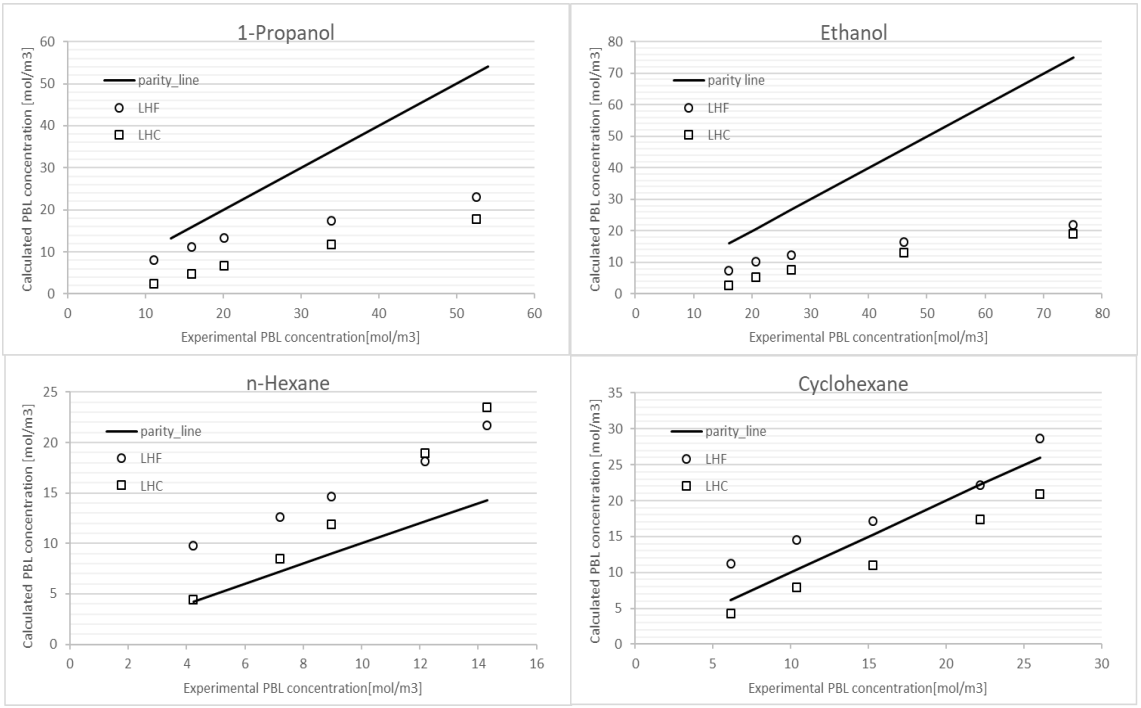


Figure 4.9 *Parity plots that illustrate the **PBL** concentration calculated applying LHC and LHF for each CBN concentration measurements. Each parity plot refers to the same reactive system with different solvents employed as pure: n-hexane, ethanol, 1-propanol, cyclohexane*

It worth noting that the amount of solvent is kept constant, regardless of the type employed, and all the model parameters values are not specific of the system analysed because regressed from a dataset presenting several solvents for the same reactive system. Therefore, LHC model is allowed to consider only the degree of dilution through concentrations, indeed, these are the only terms in the model allowed to change but, they result to be the same for all the solvents. LHF model, on the other hand, can capture not only the constant solvent dilution effect, but also the difference in the system interactions moving from one solvent to the other. These considerations, together with the knowledge that system interactions moving from alkanes to alcohols as solvent significantly increase, imply that, the model that is capable to distinguish the use of different solvents based on the interaction solvent-substrate, namely LHF, should provide for more accurate results. Indeed, from both Figure 4.8 and Figure 4.9 it can be noticed that when alcohol solvents are employed the LHF model provides for greater accuracy with respect LHC in the prediction of both CBN an PBL.

4.2.5 Comments and further improvements

The analysis developed for this case study is based on a single kinetic model that has not been derived rigorously from thermodynamics and it is not thermodynamically

consistent. For these reasons, the focus has been placed in the improvement that thermodynamic quantities able to capture a certain class of interactions (eg. fugacity or activity) can bring to the empirical/semi-empirical kinetic models widely used in kinetics studies. From the results presented in § 4.2.4 the improvements related to the substitution of fugacity terms instead of concentration are manifold. They regard the increase of accuracy in the representation of the system when experimental data are available allowing for a very accurate regression of kinetic and adsorption parameters. A second improved aspect is the statistical quality of the parameter regressed that results greater for the model presented fugacities. For what regards the prediction aspects the improvements are not appreciable and in some cases the concentration-based model provides better results. It is worth pointing out that while for the transesterification case study all the group cross interaction parameter needed for the description of the system were present in the database, for the 4-phenyl-2-butanol hydrogenation some of them are approximated by combining rules. As stated in Chapter 2 the group cross interactions that involve one polar or two polar functional groups can significantly worsen the results if they are not regressed from experimental data.

Conclusions

Through the three case studies analysed it has been possible to assess the capability of thermodynamic rigorous approach to correct interpret experimental data and to predict solubility and reactive systems behaviour over a wide range of conditions when even complex molecules are involved. The general methodology followed throughout the thesis project has been to employ first principle models to describe the systems studied applying SAFT-γ Mie as thermodynamic model. The software deployed to perform the calculations is gPROMS, in particular gPROMS FormulatedProducts environment in which gSAFT material modelling has been implemented to allow employing SAFT-γ Mie as thermodynamic model. The results obtained are reported separately for each case study for clarity sake.

In order to develop the solubility prediction case study, two models have been considered and employed to carry out the calculations: melting properties approach model (MPAM) and semi empirical model (SEM). The former requires the availability of the melting temperature and the heat of fusion (at the same temperature) of the solute to predict its solubility in any solvent. SEM requires few experimental solubility data of the solute in any solvent to be able to predict its solubility in any other solvent. The solute considered is ibuprofen and its solubility has been calculated in seven different solvents (1-propanol, 2-propanol, toluene, acetone, methanol, 4-Methyl-2-Pentanone, ethanol). Both models provide very similar results; therefore, it is not possible stating which one is more accurate. Except for methanol as solvent, the predictions have the same order of magnitude as the experimental ones. The relative error between predictions, performed by the two models, and measurements is found to vary from 5% to 80%. For this reason, not all the predicted values are reliable and it has been developed a procedure to evaluate systematically the probability of the predictions obtained to be reliable, based on the type of group cross-interactions approximated by combining rules instead of regressed from experimental data. Methanol has resulted to be the only solvent not in line with the procedure developed. It has been assumed that the reason why methanol shows this behaviour is that it is characterised by appreciable difference in how it interacts with the same species in vapor-liquid phase (from which the group cross interactions are derived) and in solid-liquid phase. Therefore, as future improvement it is suggested to regress group cross interactions employing both VLE and SLE experimental data and to repeat the study to validate the procedure developed to assess the predictions reliability.

The transesterification case study is characterised by an isothermal homogeneous liquid phase reactive system described by power law concentration-based kinetic model (PLCKM) and fugacity-based kinetic model (FKM). The comparison between the two models results in FKM providing for more accurate representation of the system in case of experimental data availability and allowing for reliable predictions of the system behaviour when the experimental data available are limited. The latter aspect is demonstrated by the results carried out employing a single rate constant value for different systems characterised by the same reactive species but different amount of heptane as solvent. FKM capabilities are limited by the computation of the thermodynamic equilibrium constant that for this model depends only on reactive species formation properties. These are taken from the literature which determines that single species values may vary slightly depending on the literature source. Small errors seem to be acceptable but when the Gibbs free energy of reaction is small ($<10^4$) even slight differences lead to a significant change in the thermodynamic equilibrium constant value, thus affecting considerably the equilibrium predictions. For this reason, it has been possible to demonstrate FKM prediction capability only on reaction kinetics. Therefore, in order to develop a complete analysis (reaction equilibrium and reaction kinetic) further studies on FKM predictions capability on equilibrium conditions are required. The results of this study can be considered a general one for homogeneous liquid phase reactive systems for which the interactions involved can be fully described by fugacities, namely substrate-solvents, substrate-substrate.

The hydrogenation case study involves a multiphase system with complex interactions (not described by fugacity) due to the presence of a solid catalyst. Therefore, it has required the application of a different model able to take into account, even partially, the additional interactions. The model (Wilkinson et al., 2015) has been modified substituting fugacities (LHF) terms instead of concentrations (LHC). This upgrade has led to a greater accuracy in the description of the same reactive system with four different solvents: cyclohexane, n-hexane, ethanol, 1-propanol. However, the improvement of LHF over LHC is significant only for solvents n-hexane and cyclohexane, while in the case of alcohol solvents the improvement is limited. The impact on LHF regressed parameters over LHC ones is expressed through the increase in the quality of parameter estimates when they are regressed for each specific system through the corresponding experimental dataset. The predictions capabilities of the two models have resulted to be comparable. In fact, the regressed parameters obtained for the four different systems diverge significantly for both models. This implies a lack in the interaction capabilities description of the model driving force term. The hydrogenation study has been a first step to show how, even through just a brute substitution of fugacity instead of concentrations

terms, thermodynamics can provide for powerful instruments to allow generalizing what so far has been strictly specific and experimentally driven. In order to be able to assess if rigorous thermodynamics can provide for a general quantitative description of multiphase catalytic systems, it is required to develop new kinetic model ensuring the thermodynamic consistency at equilibrium conditions.

Thermodynamic rigorous model adoption or even the use of quantities as fugacity or activity can improve significantly the mostly empirical models used both in solubility calculation and in reactive systems description. It is important to consider that the more complex is the system to describe, the poorer the quantitative performance of the rigorous models. A further important limit encountered in these studies that prevents nowadays the general applicability of these methods is the limited availability of group cross interactions regressed from experimental data; this, as aforementioned, can worse the results significantly.

Notation

$A^{associatio}$	=	Helmholtz free energy due to the association between molecules
A^{chain}	=	Change in Helmholtz free energy associated with the formation of a molecular chain from its constituting segments
A^{ideal}	=	Ideal-gas Helmholtz free energy
$A^{monomer}$	=	Residual Helmholtz free energy due to the formation of spherical monomeric segments
A^{res}	=	Residual Helmholtz free energy
A	=	Total Helmholtz free energy
a_i	=	Symmetric activity of specie *i*
$\hat{a}_i$	=	Asymmetric activity of specie *i*
C_i	=	Molar concentration
C_p	=	Heat capacity
$\Delta C_p^{l \to s}$	=	Heat capacity variation from liquid to solid phase
$c_{p,i}^{ig}$	=	Specific heat capacity of specie *i* in the ideal gas state
$cp_{p,i}^{s}$	=	Specific heat capacity of specie *i* in the solid phase
f_i	=	Fugacity of specie *i* in the mixture
$\Delta G^{l \to s}$	=	Variation of Gibbs free energy from liquid to solid phase
ΔG_r	=	Gibbs free energy of reaction
H	=	Total enthalpy of the mixture
H^{res}	=	Residual enthalpy
$\Delta H_i^{F,ig}$	=	Ideal gas enthalpy of formation of specie *i*
$\Delta H_i^{F,s}$	=	Enthalpy of formation of specie *i* in the solid phase
$\Delta H^{l \to s}$	=	Variation of enthalpy from liquid to solid phase

$\Delta H_i^{F,l\theta}$	=	Enthalpy of formation of specie *i* in pure liquid state at standard pressure and temperature
$\Delta \widehat{H}_i^{F,\infty}$	=	Enthalpy of formation of specie *i* in the infinite dilution state
$K_{kl,ab}$	=	Unlike bounding volume
K_j	=	Thermodynamic equilibrium constant of reaction *j*
$K_j{}''$	=	Apparent equilibrium constant of reaction *j*
k	=	Chemical functional group
K_{PBN}	=	PBN adsorption constant
K_{CBN}	=	CBN adsorption constant
$k_{1,j}$	=	Intrinsic forward rate constant
$k_{-1,j}$	=	Intrinsic backward rate constant
$k'_{1,j}$	=	Forward rate constant
M_i	=	Molar mass of specie *i*
m_i	=	Mass of specie *i*
N	=	Total number of molecules
NP	=	Number of phases
n_i	=	Number of molecules of specie *i*
P	=	Pressure
P^{θ}	=	Standard pressure
P_m	=	Pressure at which the melting temperature is measured
R	=	Universal gas constant
r_j	=	Rate of reaction *j*
r_{kl}	=	Intersegment distance between segment *k* and *l*
S	=	Total entropy of the mixture
S^{res}	=	Residual entropy

$\Delta S_i^{F,ig}$	=	Ideal gas entropy of formation of specie *i*
$\Delta S_i^{F,s}$	=	Entropy of formation of specie *i* in the solid phase
$\Delta S^{l \to s}$	=	Variation of entropy from liquid to solid phase
$\Delta S_i^{F,l\theta}$	=	Entropy of formation of specie *i in* pure liquid state at standard pressure and temperature
S_k	=	Shape factor of the group *k*
T	=	Temperature
T^{θ}	=	Standard temperature
T_m	=	Melting temperature
V	=	Volume
w_i	=	Trial mole fraction
x_i	=	Molar fraction of specie *i*
Y_i	=	Absolute yield of solute *i*
$Y_i^{rel}\%$	=	Relative yield of solute *i*

Greek symbols

γ_i	=	Symmetric activity coefficient of specie *i*
$\hat{\gamma}_i$	=	Asymmetric activity coefficient of specie *i*
ϵ_{kl}	=	Depth of the potential well
λ_{kl}^r	=	Repulsive exponent of the segment-segment interactions
λ_{kl}^a	=	Attractive exponent of the segment-segment interactions
$\mu_i^{[k]}$	=	Chemical potential of specie *i* in phase *k*
μ_i^{ig}	=	Chemical potential of specie *i* at ideal gas state
μ_i^{res}	=	Residual chemical potential of specie *i*

μ_i^{trial}	=	Trial chemical potential of specie *i*
$\hat{\mu}_i^{w,\infty}$	=	Chemical potential of specie *i* in the infinite dilution state
ν_i	=	Stoichiometry coefficient of species *i*
ξ	=	Extent of reaction
φ_i	=	Fugacity coefficient of species *i* as pure
$\overline{\varphi_i}$	=	Fugacity coefficient of species *i* in the mixture
σ_{kl}	=	Unlike segment diameter
σ_{kk}	=	Segment diameter
$\Phi_{kl,ab}^{HB}$	=	Association interaction between two square-well association sites of type *a* in segment *k* and *b* in segment *l*
Φ	=	Maximum likelihood objective function
ω_i	=	Mass fraction of specie *i*

Acronyms

API	=	Active Pharmaceutical Ingredients
BIPs	=	Binary Interaction Parameters
CBL	=	4-cyclohexyl-2-butanol
CBN	=	4-cyclohexyl-2-butanone
CSTR	=	Continuous Stirred Tank Reactor
EoS	=	Equation of state
FKM	=	Fugacity-based Kinetic Model
GC	=	Group Contribution
GSA	=	Global Sensitivity Analysis
gFP	=	gPROMS FormulatedProducts

gFPPP	=	gPROMS FormulatedProducts Properties Package
ICH	=	International Conference on Harmonisation
LHC	=	Langmuir-Hinshelwood Concentration-based model
LHF	=	Langmuir-Hinshelwood Fugacity-based model
LHHW	=	Langmuir-Hinshelwood-Hougen-Watson model
LLE	=	Liquid-Liquid Equilibria
PBL	=	4-phenyl-2-butanol
PBN	=	4-phenyl-2-butanone
PLCKM	=	Power Law Concentration based Model
SAFT	=	Statistical Associating Fluid Theory
SLE	=	Solid-Liquid Equilibria
VLE	=	Vapour-Liquid Equilibria

Bibliography

Bermingham S.K. (2018), Mechanistic models for efficient and effective Digital Design and Digital Operation of Formulated Products and their Manufacturing Processes. Presented at *Advanced Process Modelling Forum*, London (UK), 17-18 April 2018.

Brown, C.J., McGlone, T., Yerdelen, S., Srirambhatla, V., Mabbott, F., Gurung, R., Briuglia, M.L., Ahmed, B., Polyzois, H., McGinty, J., Perciballi, F., Fysikopoulos, D., MacFhionnghaile, P., Siddique, H., Raval, V., Harrington, T.S., Vassileiou, A.D., Robertson, M., Prasad, E., Johnston, A., Johnston, B., Nordon, A., Srai, J.S., Halbert, G., Horst, J.H., Price, C.J., Rielly C.D., Sefcik, J. and Florence, A.J. (2018). Enabling precision manufacturing of active pharmaceutical ingredients: workflow for seeded cooling continuous crystallisations. *Molecular Systems Design & Engineering.*, **3**, 518-549.

Dufal, S., Lafitte, T., Haslam, A.J., Galindo, A., Clark, G.N.I., Vega, C., Jackson, G., (2015), The A in SAFT: developing the contribution of association to the Helmholtz free energy within a Wertheim TPT1 treatment of generic Mie fluids. *Molecular Physics*. **113**, 948-984.

Gracin, S. and Rasmuson, A. C. (2002). Solubility of Phenylacetic Acid, p-Hydroxyphenylacetic Acid, p-Aminophenylacetic Acid, p-Hydroxybenzoic Acid, and Ibuprofen in Pure Solvents. *Journal of Chemical & Engineering Data*., **47**, 1379-1383.

Hessel, V.; Ehrfeld, W.; Golbig, K.; Haverkamp, V.; Lowe, H.; Storz, M.; Wille, Ch.; Guber, A.; Jahnisch, K.; Baerns, M. (1999). In :*Microreaction Tecnology:Industrial Prospects* (W. Ehrfeld). Springer., Frankfurt (DE), p.526.

Joback K.G., Reid R.C. (1987). Estimation of Pure-Component Properties from Group-Contributions. *Chemical Engineering Communications*., **57**, 233–243.

Lafitte, T., Papaioannou, V., Dufal, S., Pantelides, C.C. (2017). A general framework for solid-liquid equilibria in pharmaceutical systems. Process Systems Enterprise Ltd., 26-28 Hammersmith Grove, London (U.K).

Lafitte T., (2017). *Internal report*, Process System Enterprise Ltd.

Lafitte T., Papaioannou V., Dufal S. and Pantelides C. C. (2017). gSAFT: Advanced physical property prediction for process modelling. *Computer Aided Chemical Engineering.*, **40**, 1003-1008.

Lemberg, M., Scomacker, R., Sadowski, G. (2017). Thermodynamic prediction of the solvent effect on a transesterification. *Chemical Engineering Science* **176**, 264-269.

M.L. Michelsen and J.M. Mollerup (2007). *Thermodynamic Models: Fundamentals & Computational Aspects* (2nd ed.). Tie-Line Publications., Holte (DK), p.231.

Papaioannou V., Lafitte T.,Avendaño C. ,Adjiman C. N. ,Jackson G. ,Müller E. A. and Galindo A. (2014). Group contribution methodology based on the statistical association fluid theory for heteronuclear molecules formed from Mie segments. *The Journal of Chemical Physics.*, **140**, 054107.

Rowlinson J. S. and Swinton F. L. (1982). *Liquids and Liquid Mixtures* (3rd ed.). Butterworth-Heinemann., Oxford (U.K), p.180.

Schuchardt, U., Sercheli, R. and Vargas, R.M. (1998). Transesterification of Vegetable Oils: a Review. *Journal of the Brazilian Chemical Society.*, **9**, 199-210.

Schmidt, J., Reusch, D., Elgeti, K., Schomäcker, R. (1999). Kinetik der Umesterung von Ethanol und Butylacetat – Ein Modellsystem für die Reaktivrektifikation. *Chemie Ingenieur Technik.*, **71**, 704–708.

Wang, S., Song, Z., Wang, J., Dong, Y. and Wu, M. (2010). Solubilities of Ibuprofen in Different Pure Solvents. *Journal of Chemical & Engineering Data.*, **55**, 5283–5285.

Wertheim, M.S. (1984) Fluids with highly directional attractive forces. I. Statistical thermodynamics. *Journal of Statistical Physics.*, **35**, 19-34.

Wilkinson, S.K., McManus, I., Thompson, J.M., Hardacre, C., Sadaie Bonab, N., ten Dam, J., Simmons, M.J.H., D'Agostino, C., McGregor, J., Gladden, L.F., Stitt, E.H. (2015). A kinetic analysis methodology to elucidate the roles of metal, support and solvent for the hydrogenation of 4-phenyl-2-butanone over Pt/TiO_2. *Journal of Catalysis* **330**, 362-373.

Zar, J.H. (1984) *Biostatistical Analysis.* Prentice-Hall International, Upper Saddle River (U.S.A.), pp 43–45.

Acknowledgements

I would like to acknowledge and thank my Supervisor Dr. Sam Wilkinson for his constant guidance and his careful supervision throughout the development of the master's project. You have encouraged and believed in me and this has made me try harder and overcome my limits.

I would also like to thank Niall Mitchell, Meera Mahadevan, Thomas Lafitte and Vasileios Papaioannou for the help they have provided in specific area of expertise. Without your technical support the development of the project wouldn't have been possible.

A very special thank you to all my colleagues, who enriched this experience, in particular: Stanley, Artur, Victoria, Joao, Mauro, Dimitrios, Henry, Sarah, Cristian and Lylia. I am grateful for the thoughts and experiences we shared.

Finally, I owe my deepest gratitude to my Professor (and also Supervisor) Dr. Fabrizio Bezzo for the invaluable opportunity he gave me together with his considerable support during my master's thesis writing.