Chemical Engineering Science 159 (2017) 106–130

Contents lists available at ScienceDirect

Chemical Engineering Science

journal homepage: www.elsevier.com/locate/ces

Designing optimal mixtures using generalized disjunctive programming: Hull relaxations

Suela Jonuzaj, Claire S. Adjiman*

Centre for Process Systems Engineering, Department of Chemical Engineering Imperial College London, London SW7 2AZ, UK

ARTICLE INFO

Keywords:
Mixture design
Generalized Disjunctive Programming
Hull Reformulation
Solubility
Liquid–liquid extraction

ABSTRACT

A general modeling framework for mixture design problems, which integrates Generalized Disjunctive Programming (GDP) into the Computer-Aided Mixture/blend Design (CAM^{b}D) framework, was recently proposed (S. Jonuzaj, P.T. Akula, P.-M. Kleniati, C.S. Adjiman, 2016. The formulation of optimal mixtures with Generalized Disjunctive Programming: A solvent design case study. AIChE Journal 62, 1616–1633). In this paper we derive Hull Relaxations (HRs) of GDP mixture design problems as an alternative to the big-M (BM) approach presented in this earlier work. We show that in restricted mixture design problems, where the number of components is fixed and their identities and compositions are optimised, BM and HR formulations are identical. For general mixture design problems, where the optimal number of mixture components is also determined, a generic approach is employed to enable the derivation and solution of the HR formulation for problems involving functions that are not defined at zero (e.g., logarithms). The design methodology is applied successfully to two solvent design case studies: the maximization of the solubility of a drug and the separation of acetic acid from water in a liquid–liquid extraction process. Promising solvent mixtures are identified in both case studies. The HR and BM approaches are found to be effective for the formulation and solution of mixture design problems, especially via the general design problem.

1. Introduction

The design of mixtures is an important and challenging problem with numerous industrial applications. Of particular interest are applications in separation processes, such as liquid–liquid extraction (Brignole and Gani, 1983; Karunanithi et al., 2005; Cignitti et al., 2015) and crystallization (Karunanithi et al., 2006, 2009), that require suitable solvents or solvent mixtures to meet given specifications, and where the choice of solvent can have a significant impact on the performance of the process. In drug manufacturing, for example, unsuitable solvents can result in undesired crystal morphology, which may affect downstream processing and product performance (Gordon and Amin, 1984; Karunanithi et al., 2006). Solvent mixtures are also used in chemical reactors to enhance the reaction rate (Folić et al., 2007; Struebing et al., 2013) or (Zhou et al., 2015) and in product design as constituents of the final product formulations (Gani, 2004a, b; Gani and Ng, 2015).

Several systematic methodologies have been developed within the Computer-Aided Mixture/blend Design (CAM^{b}D) framework (Gani, 2004a; Achenie et al., 2003) for the design of solvent mixtures (Brignole and Gani, 1983; Buxton et al., 1999; Sinha et al., 2003; Karunanithi et al., 2005; Cignitti et al., 2015; Jonuzaj et al., 2016), blends of refrigerants (Duvedi and Achenie, 1997; Churi and Achenie, 1997; Vaidyaraman and Maranas, 2002), blends of polymers (Vaidyanathan and El-Halwagi, 1996; Solvason et al., 2009; Zhang et al., 2015), blended liquid products (Yunus et al., 2014) and heat transfer fluid mixtures (Papadopoulos et al., 2013). A more detailed description of the existing methodologies for mixture design can be found in Jonuzaj et al. (2016). In spite of these advances, there remains great potential to improve existing approaches to mixture design within the CAM^{b}D framework. In current practice, the number of compounds or materials required for the design of mixtures or products is often chosen first, before other design decisions are made, and this can lead to suboptimal designs. Thus, most methodologies proposed to date have been focused on the design of mixtures with a pre-specified number of components and have been applied mostly to the design of binary mixtures (Sinha et al., 2003; Karunanithi et al., 2005, 2006; Buxton et al., 1999; Papadopoulos et al., 2013; Vaidyanathan and El-Halwagi, 1996), with some exceptions such as the work of Solvason et al. (2009), Yunus et al. (2014) and Jonuzaj et al. (2016), who have presented methodologies for the design of multicomponent mixtures. CAM^{b}D methods generally rely on Mixed Integer Nonlinear

* Corresponding author.
E-mail address: c.adjiman@imperial.ac.uk (C.S. Adjiman).

http://dx.doi.org/10.1016/j.ces.2016.08.008
Received 30 January 2016; Received in revised form 9 June 2016; Accepted 5 August 2016
Available online 10 August 2016

Programming (MINLP) techniques to model the discrete decisions inherent in mixture design problems, which are related to the number of components in the mixture and their identities. The solution of the resulting mixed integer optimization problems can be very challenging due to nonconvexities in the space of the continuous variables and a large combinatorial solution space which may lead to several numerical difficulties.

By extending the applicability of CAM^{b}D methods to generalized mixture design problems, in which the number of components in the optimal mixture is not fixed *a priori*, the explicit evaluation of every choice of the number of components can be avoided, making it possible to consider larger design spaces, especially as the number of desirable components increases. This requires developing a comprehensive and systematic mathematical programming approach for the formulation and solution of such problems. In the context of a generalized CAM^{b}D problem, we have recently proposed (Jonuzaj et al., 2016) a novel methodology for determining simultaneously the optimal number of compounds in a mixture, the specific identities of the compounds, and their composition in the mixture. The desired compounds are chosen from a list of possible molecules. Within this approach, logic-based modeling was employed to formulate the CAM^{b}D problem as a Generalized Disjunctive Program (GDP) (Raman and Grossmann, 1994), in order to address the difficulties arising from the complexity of the model and facilitate problem formulation. From this initial work, the objective of our current work is to study different strategies for the solution of the GDP problem, including the Big-M (BM) (Nemhauser and Wolsey, 1999; Raman and Grossmann, 1994) approach and Hull Reformulations (HRs) (Lee and Grossmann, 2000, 2003), in order to circumvent the combinatorial explosion that accompanies large design spaces and facilitate problem solution. The design methodology and the two different relaxation approaches are applied to two case studies of increasing complexity. In the first, simple, example, which involves solid–liquid equilibrium calculations, an optimal solvent mixture that maximizes the solubility of a drug is designed. The second case study consists of a more challenging problem, where the most effective solvent mixture to separate acetic acid from water by liquid–liquid extraction is designed. In both cases, the computational performance of the different reformulation strategies is assessed. As will be seen, the resulting problems are challenging to solve for existing optimization algorithms. Here we focus on the development of a generic formulation, with application to small-scale examples. In practice, the application of the proposed approach to formulation design implies considering a large number of ingredients (e.g., there can be 10–30 ingredients in a typical paint (Nicks and Ryan, 1975) or shampoo (Trüeb, 2007), chosen from a much larger list).

The paper is organized as follows. In Section 2, a brief overview of the GDP concepts necessary for the presentation of the problem formulations and solution strategies is provided. In Section 3, several mathematical formulations of the generalized mixture design problem are presented. Then, in Sections 4 and 5, the proposed approaches are applied to the two case studies.

2. A brief introduction to Generalized Disjunctive Programming (GDP)

In this section we describe briefly the general formulation of GDP problems, which was introduced by Raman and Grossmann (1994), and we review briefly how the GDP problem, with its Boolean variables, can be converted into mixed-integer form so that it can be solved by standard MINLP algorithms (e.g., the outer-approximation algorithm Duran and Grossmann, 1986; Fletcher and Leyffer, 1994). GDP is a logic-based approach for formulating discrete/continuous optimization problems that extends the disjunctive programming proposed by Balas (1985) and involves Boolean and continuous variables that are related via disjunctions, algebraic equations and logic propositions (Beaumont, 1991; Turkay and Grossmann, 1996). It has been employed by Grossmann and co-authors in several applications in the area of process systems engineering, such as the design of process network systems (Raman and Grossmann, 1994; Vecchietti et al., 2003; Ruiz and Grossmann, 2013; Trespalacios and Grossmann, 2015), the design of distillation columns (Grossmann and Trespalacios, 2013), strip-packing (Sawaya and Grossmann, 2005) and scheduling problems (Raman and Grossmann, 1994; Sawaya and Grossmann, 2005; Méndez et al., 2006; Castro and Grossmann, 2012).

The general formulation of a GDP problem involves an objective function to be optimised, general constraints that must hold regardless of the discrete choices, conditional constraints within disjunctions that depend on the discrete decisions, represented by Boolean variables, and logic propositions that connect the disjunctive variables. The general formulation of a GDP problem is presented as (GDP) in Appendix A for completeness. In order to exploit existing MINLP algorithms, once an appropriate GDP formulation has been obtained, it can be converted into an MINLP problem using different approaches, such as big-M or Hull Reformulation, that result in relaxations of varying strength (Lee and Grossmann, 2003; Grossmann and Trespalacios, 2013). The BM formulation (Nemhauser and Wolsey, 1999) is the simplest representation of a GDP problem in a mixed-integer form (Raman and Grossmann, 1994). The concept of a Convex Hull relaxation of a convex GDP problem was introduced by Stubbs and Mehrotra (1999) and was later extended by Lee and Grossmann (2000), Lee and Grossmann (2003), Lee and Grossmann (2005) for the derivation of Hull Relaxations for convex and nonconvex conditional constraints. Generic formulations of the big-M and Hull Relaxation approaches are presented in Appendix A as models (BM) and (HR), respectively. In the (HR) model, disjunctive constraints are transformed into mixed-integer equations via the perspective function, $y_{j,k}h_{j,k}(\nu_{j,k}/y_{j,k}) \leq 0$ (Grossmann and Trespalacios, 2013). In order to avoid the numerical difficulties (division by zero) that can arise from perspective functions, the following approximation was proposed by Sawaya (2006):

$$((1-\epsilon)y_{j,k}+\epsilon)h_{j,k}\left(\frac{\nu_{j,k}}{(1-\epsilon)y_{j,k}+\epsilon}\right)-\epsilon h_{j,k}(0)(1-y_{j,k}) \leq 0 \tag{1}$$

where $y_{j,k}$ is a binary variable that has one-to-one correspondence with the Boolean variable, $Y_{j,k}$, of model (GDP); $h_{j,k}$ is a nonlinear conditional constraint that depends on the discrete decisions; $\nu_{j,k}$ is a disaggregated variable and ϵ is a small tolerance which usually varies from 10^{-8} to 10^{-2}.

Both (BM) and (HR) have a one-to-one correspondence with model (GDP) (Lee and Grossmann, 2000), so that all three formulations have the same global and local solutions. The BM approach is known to give weak lower bounds in the case of a minimization problem (Grossmann, 2002; Lee and Grossmann, 2003; Vecchietti et al., 2003). This is due in part to the fact that it relies on the Big-M parameter, $M_{j,k}$, a bound whose value cannot always be calculated exactly but is often specified based on an approximate analysis of function ranges. As a result, it is usually given large values, so that feasible points are not excluded from the solution space. The HR formulation, on the other hand, incurs a computational cost due to the introduction of a new set of disaggregated variables, $\nu_{j,k}$, and new constraints, thereby increasing the size of the problem (Lee and Grossmann, 2000). For problems that are convex in the continuous variables, it can be proved (Lee and Grossmann, 2003) that when the discrete domain of the Hull Reformulation is relaxed (i.e. $0 \leq y_{j,k} \leq 1$), it gives bounds that are as tight as or tighter than the bounds generated with the Big-M approach.

Although HR techniques may provide tighter lower bounds than the traditional BM model, they do not always lead to more efficient solution times due to the increased number of variables and constraints (Lee and Grossmann, 2005; Lee and Leyffer, 2012; Grossmann and Trespalacios, 2013). In cases where tight variable bounds are provided, or in large problems where it is desirable not to increase the number of

variables and equations, the BM model can compete with the HR relaxation in terms of computational performance (Vecchietti et al., 2003). In recent work, Castro et al. (2014) addressed the optimization of the maintenance scheduling of a gas-fired power plant using a continuous-time GDP model, where both the BM and HR techniques were applied to recast the GDP into an MILP. It emerged from the results of the case study that the BM reformulation gave better computational performance than the Hull Reformulation.

In earlier works, Grossmann and co-authors (Lee and Grossmann, 2000; Vecchietti et al., 2003; Grossmann and Trespalacios, 2013) had compared the performance of BM and HR in various example problems (e.g., job scheduling, process network applications, strip-packing), where HR was found to outperform the BM approach and to give tighter relaxations. Ultimately, the best choice between BM or HR depends on the complexity of the problem at hand and on the type of the disjunctive constraints.

In order to improve the relaxation techniques, cutting plane methods (Vecchietti et al., 2003; Lee and Grossmann, 2005; Sawaya and Grossmann, 2005) and basic steps (i.e., hierarchy of relaxations) (Sawaya and Grossmann, 2012; Ruiz and Grossmann, 2012; Grossmann and Trespalacios, 2013) can also be used in solving GDP models. In our current work, we compare the performance of the BM and HR relaxations for solution of GDP formulations of mixture design problems, but we do not apply cutting planes or a basic step approach.

3. The generalized mixture design problem

The formulation of mixture design problems in a way which integrates GDP into a CAM^{b}D framework was introduced and described in detail in our recent work (Jonuzaj et al., 2016). A brief overview of the GDP formulations is given for completeness in the next subsection, followed by a description of BM and HR models of mixture problems in the remainder of this section.

3.1. Problem definition

The mixture problem is constructed in a systematic way and is presented here in the context of the formulation of a generalized model, where any number of components may be chosen up to a user-defined maximum, based on a given list of candidate compounds, and subject to property constraints. The design variables are the number of the components that participate in the mixture, their identities and their compositions. In order to develop the optimization problems, the following index sets need to be defined. The first set, $I = \{1, \ldots, N_c\}$, defines the number of each component in the mixture, where N_c is the total number of components. The mixture to be designed consists of fixed molecules (e.g., solutes) that must necessarily be present in the mixture and of unknown components (e.g., solvents) that are to be selected from a predefined list. We define N' as the number of fixed components in the mixture and N_{max} as the maximum number of components to be designed, which are also labeled via the set $N = \{1, \ldots, N_{max}\}$. Thus the maximum number of components in the mixture is $N_c = N' + N_{max}$. The list of compounds from which the unknown components are selected is given by the set $S = \{1, \ldots, N_s\}$, with $N_s \geq N_{max}$. The chemical identity of each molecule is defined using functional groups (building blocks such as CH_3, OH) that can be used in the calculation of relevant physical properties, and the groups are represented by the set $K = \{1, \ldots, N_K\}$. For clarity, we use the term "components" to refer to the ingredients/molecules in the mixture we are designing and the term "compounds" to refer to ingredients/molecules in the set S from which we choose the components. Those components in the mixture that are not fixed (i.e., components $N' + 1$ to N_c) are referred to as the "designed components".

3.2. GDP formulation of the CAM^{b}D problem

Following the derivation of Jonuzaj et al. (2016), the GDP formulation of the general mixture problem can be written as:

$$
\begin{aligned}
&\min_{x,Y} \quad f(x) \\
&\text{s. t.} \quad g(x) \leq 0 \\
&\quad \bigvee_{s\in S} \begin{bmatrix} Y_{i,s} \\ h_{i,s}(x) \leq 0 \end{bmatrix}, \quad i = N' + 1, \ldots, N_c \\
&\quad \underline{\vee}_{s=1,\ldots,N_s} \; Y_{N'+1,s} \\
&\quad \bigvee_{n=1,\ldots,N_{max}} \begin{bmatrix} \widetilde{Y}_n \\ \widetilde{F}_n(x) \leq 0 \\ x_i \geq x_i^L, \quad i = N' + 1, \ldots, N' + n \\ x_i = 0, \quad i = N' + n + 1, \ldots, N_c \end{bmatrix} \\
&\quad \underline{\vee}_{n=1,\ldots,N_{max}} \; \widetilde{Y}_n \\
&\quad \Omega(Y) = \text{True} \\
&\quad x \in [x^L, x^U] \subset \mathbb{R}^m \\
&\quad Y_{i,s} \in \{\text{True, False}\}, \quad i = N' + 1, \ldots, N_c, \; s \in S \\
&\quad \widetilde{Y}_n \in \{\text{True, False}\}, \quad n = 1, \ldots, N_{max}
\end{aligned}
\tag{G-GDP}
$$

where f is the objective function to be optimised and $g(x) \leq 0$ represents general constraints that must hold regardless of the discrete choices. Two different sets of disjunctions are included in this formulation. The first set involves disjunctions for selecting each component from a list of pure compounds and the assignment of a component i in the mixture from a compound s in the list is determined through Boolean variable $Y_{i,s}$ (a compound s is assigned to component i in the mixture if $Y_{i,s}$ is True and it is not if $Y_{i,s}$ is False). The second set of disjunctions involves property constraints that depend on the number of mixture ingredients, n, and that are active when the corresponding Boolean variable $\widetilde{Y}_n$ is True. The vector $h_{i,s}(x) \leq 0$ represents the constraints that depend on the identity of a compound and are active when compound s is assigned to the ith component in a mixture (i.e., $Y_{i,s}$ is True). $\widetilde{F}_n(x) \leq 0$ is a vector of constraints that depend on the number of components in the mixture. The mole fraction, x_i, of a component i is set to be greater than a user-specified threshold value x_i^L if the component is present in the mixture and is zero otherwise. This is done to avoid the generation of unrealistic mixture compositions, with many components present in very small quantities.

The logic conditions, $\Omega(Y) = \text{True}$, included in the model are derived to avoid degeneracy of the solutions by enforcing a specific ordering of the compounds. These constraints thus restrict the feasible space by eliminating identical degenerate solutions. Logic relations are also derived to ensure that each candidate compound s is selected at most once and that at most one compound is assigned to each component i ($i = N' + 1, \ldots, N_c$) in the mixture. Additional logic propositions are required to relate the Boolean variables for the number of designed components in the mixture, $\widetilde{Y}_n$, $n = 1, \ldots, N_{max}$ to the Boolean variables for the assignment of a compound to each component, $Y_{i,s}$, $i = N' + 1, \ldots, N_c$, $s = 1, \ldots, N_s$. These logic relations can be translated into linear algebraic equations (Williams, 1985; Raman and Grossmann, 1991) by replacing the Boolean variables $Y_{i,s}$ and $\widetilde{Y}_n$ with the binary variables $y_{i,s}$ and $\widetilde{y}_n$, respectively. The logic propositions and the corresponding linear algebraic constraints are summarized in Table 1. A more detailed description of all the logic relations included in the model and how they are converted into linear inequalities is given in Jonuzaj et al. (2016).

3.3. Reformulation of problem (G-GDP) as an MINLP via Big-M

The MINLP model derived by applying the BM reformulation to the generalized problem (G-GDP) is written as follows:

Table 1
Logic propositions and algebraic constraints for the generalized mixture design problem.

Logic expressions	Linear inequalities
Eliminate degenerate solutions	
$Y_{i,s} \Rightarrow \neg\, Y_{i',s'}$	$y_{i,s} + y_{i',s'} \leq 1$
$i = N' + 1, \ldots, N_c - 1\ ;\ \ s = 2, \ldots, N_s$	$i = N' + 1, \ldots, N_c - 1\ ;\ \ s = 2, \ldots, N_s$
$i' = i + 1, \ldots, N_c\ ;\ \ s' = 1, \ldots, s$	$i' = i + 1, \ldots, N_c\ ;\ \ s' = 1, \ldots, s$
Select each compound at most once	
$Y_{i',s} \Rightarrow \neg \left(\bigvee_{\substack{i=N'+1,\ldots,N_c-1 \\ i \neq i'}} Y_{i,s} \right)$	$\sum_{i=N'+1}^{N_c} y_{i,s} \leq 1$
$i' = N' + 1, \ldots, N_c\ ;\ \ s \in S$	$s \in S$
At most one compound is assigned to components $N' + 2$ to N_c	
$Y_{i,s'} \Rightarrow \neg \left(\bigvee_{\substack{s=1,\ldots,N_s-1 \\ s \neq s'}} Y_{i,s} \right)$	$\sum_{s=1}^{N_s} y_{i,s} \leq 1$
$i = N' + 2, \ldots, N_c\ ;\ \ s' \in S$	$i = N' + 2, \ldots, N_c$
Relations between $Y_{i,s}$ and $\widetilde{Y}_n$	
$\neg \widetilde{Y}_1 \vee \left(\bigvee_{s \in S} Y_{N'+1,s} \right)$	$\widetilde{y}_1 \leq \sum_{s \in S} y_{N'+1,s}$
$(\neg \widetilde{Y}_1 \vee \neg\, Y_{N'+2,s}),\ \ s \in S$	$\widetilde{y}_1 + y_{N'+2,s} \leq 1,\ \ s \in S$
$\vdots$	$\vdots$
$\left(\neg \widetilde{Y}_1 \vee \neg\, Y_{N'+N_{max},s} \right),\ \ s \in S$	$\widetilde{y}_1 + y_{N'+N_{max},s} \leq 1,\ \ s \in S$
$\neg \widetilde{Y}_2 \vee \left(\bigvee_{s \in S} Y_{N'+1,s} \right)$	$\widetilde{y}_2 \leq \sum_{s \in S} y_{N'+1,s}$
$\neg \widetilde{Y}_2 \vee \left(\bigvee_{s \in S} Y_{N'+2,s} \right)$	$\widetilde{y}_2 \leq \sum_{s \in S} y_{N'+2,s}$
$(\neg \widetilde{Y}_2 \vee \neg\, Y_{N'+3,s}),\ \ s \in S$	$\widetilde{y}_2 + y_{N'+3,s} \leq 1,\ \ s \in S$
$\vdots$	$\vdots$
$(\neg \widetilde{Y}_2 \vee \neg\, Y_{N'+Nmax,s}),\ \ s \in S$	$\widetilde{y}_2 + y_{N'+N_{max},s} \leq 1,\ \ s \in S$
$\vdots$	$\vdots$
$\neg \widetilde{Y}_{N_{max}} \vee \left(\bigvee_{s \in S} Y_{N'+1,s} \right)$	$\widetilde{y}_{N_{max}} \leq \sum_{s \in S} y_{N'+1,s}$
$\neg \widetilde{Y}_{N_{max}} \vee \left(\bigvee_{s \in S} Y_{N'+2,s} \right)$	$\widetilde{y}_{N_{max}} \leq \sum_{s \in S} y_{N'+2,s}$
$\vdots$	$\vdots$
$\widetilde{Y}_{N_{max}} \vee \left(\bigvee_{s \in S} Y_{N'+N_{max},s} \right)$	$\widetilde{y}_{N_{max}} \leq \sum_{s \in S} y_{N'+N_{max},s}$

$$
\begin{aligned}
\min_{x,y} \quad & f(x) \\
\text{s. t.} \quad & g(x) \le 0 \\
& h_{i,s}(x) \le M_{h_{i,s}}(1 - y_{i,s}), \quad s = 1, \ldots, N_s, \; i = N' + 1, \ldots, N_c \\
& \widetilde{F}_n(x) \le M_{\widetilde{F}_n}(1 - \widetilde{y}_n), \; n = 1, \ldots, N_{max} \\
& x_i \ge x_i^L \, \widetilde{y}_n, \quad n = 1, \ldots, N_{max}, \; i = N' + 2, \ldots, N_c \\
& x_{N'+1} \ge x_{N'+1}^L \\
& \sum_{s=1}^{N_s} y_{N'+1,s} = 1 \\
& \sum_{n=1}^{N_{max}} \widetilde{y}_n = 1 \\
& Ay \le b \\
& x \in [x^L, x^U] \subset \mathbb{R}^m \\
& \widetilde{y}_n, \; y_{i,s} \in \{0, 1\}, \quad n = 1, \ldots, N_{max}, \quad i = N' + 1, \ldots, N_c \; ; \; s \in S
\end{aligned}
\tag{G-BM}
$$

where $Ay \le b$ is a set of linear inequalities resulting from the logic relations, $\Omega(Y)$, and presented in the right column of Table 1.

3.4. Reformulation of problem (G-GDP) as an MINLP via Hull Relaxation

In the HR model, the continuous variables are disaggregated into additional variables for each disjunctive term (e.g., x_i is disaggregated into $\nu_{i,n}$). Each nonlinear disjunctive constraint, $h_{i,s}(x) \le 0$, can be formulated using one of the approximation functions proposed by Sawaya and Grossmann (2007) in order to avoid singularities. The resulting MINLP model derived by applying the HR approach to the mixture design problem is written as follows:

$$
\begin{aligned}
\min_{x,y} \quad & f(x) \\
\text{s. t.} \quad & g(x) \le 0 \\
& x_i = \sum_{s=1}^{N_s} \nu_{i,s}, \; i = N' + 1, \ldots, N_c \\
& ((1 - \epsilon) y_{i,s} + \epsilon) \, h_{i,s}\left(\frac{\nu_{i,s}}{(1 - \epsilon) y_{i,s} + \epsilon} \right) - \epsilon h_{i,s}(0)(1 - y_{i,s}) \le 0, \\
& \qquad i = N' + 1, \ldots, N_c \; ; \; s \in S \\
& y_{i,s} x_i^L \le \nu_{i,s} \le y_{i,s} x_i^U, \; i = N' + 1, \ldots, N_c \; ; \; s \in S \\
& ((1 - \epsilon)\widetilde{y}_n + \epsilon)\widetilde{F}_n\left(\frac{\nu_n}{(1 - \epsilon)\widetilde{y}_n + \epsilon} \right) - \epsilon \, \widetilde{F}_n(0)(1 - \widetilde{y}_n) \le 0, \\
& \qquad n = 1, \ldots, N_{max} \\
& \nu_{i,n} \ge \nu_{i,n}^L \, \widetilde{y}_n, \quad n = 1, \ldots, N_{max}; \; i = N' + 2, \ldots, N_c \\
& \nu_{N'+1,n} \ge \nu_{N'+1,n}^L, \quad n = 1, \ldots, N_{max} \\
& \sum_{n=1}^{N_{max}} \nu_{i,n} = x_i, \; i = N' + 1, \ldots, N_c \\
& \widetilde{y}_n x_i^L \le \nu_{i,n} \le \widetilde{y}_n x_i^U, \; i = N' + 1, \ldots, N_c \; ; \; n = 1, \ldots, N_{max} \\
& \sum_{s=1}^{N_s} y_{N'+1,s} = 1 \\
& \sum_{n=1}^{N_{max}} \widetilde{y}_n = 1 \\
& Ay \le b \\
& x \in [x^L, x^U] \subset \mathbb{R}^m \\
& \widetilde{y}_n, \; y_{i,s} \in \{0, 1\}, \quad n = 1, \ldots, N_{max}, \quad i = N' + 1, \ldots, N_c \; ; \; s \in S
\end{aligned}
\tag{G-HR}
$$

The (G-HR) formulation is often not directly applicable to mixture design problems due to the presence of nonlinear phase equilibrium relations in the model which can be singular at zero. The thermodynamic expressions for some of the properties needed in phase equilibria calculations, such as chemical potentials and activity coefficients, involve logarithmic functions that can lead to numerical difficulties when a variable is equal to zero (e.g., $x \ln(x)$ cannot easily be computed numerically when x=0). This problem may occur in model (G-HR) where the disaggregated variables become zero when the corresponding binary variable is zero. In order to avoid such singularities, we employ a modified approximation function proposed by Ruiz and Grossmann (2012) when logarithms are involved. Considering a function $\widetilde{F}'_{i,n}(x_i) = \ln(x_i)$ for some $n \in \{1, \ldots, N_{max}\}$ and some $i \in \{N' + 1, \ldots, N' + n\}$, the following transformation is used:

$$x_i = x_i' + \beta \tag{2}$$

$$\widetilde{F}'_{i,n}(x_i') = \ln(x_i' + \beta) \tag{3}$$

where β is an arbitrary constant. Hence,

$$((1 - \epsilon)\widetilde{y}_n + \epsilon)\widetilde{F}'_{i,n}\left(\frac{\nu'_{i,n}}{(1 - \epsilon)\widetilde{y}_n + \epsilon} \right) - \epsilon \, \widetilde{F}'_{i,n}(x_i' = 0)(1 - \widetilde{y}_n) \le 0 \tag{4}$$

where

$$x_i' = \sum_{n=1}^{N_{max}} \nu'_{i,n} \text{ and } \widetilde{y}_n x_i'^L \le \nu'_{i,n} \le \widetilde{y}_n x_i'^U. \tag{5}$$

Based on model (G-HR), for each component $i = N' + 1, \ldots, N_c$ in the mixture, the Hull relaxation of the first set of disjunctions (for the assignment of compounds) leads to the introduction of N_s additional variables and $2N_s + 1$ additional equations in the problem. Furthermore, the Hull relaxation of the second set of disjunctions (for the number of components) leads to the introduction of N_{max} additional variables and $2N_{max} + 1$ additional constraints in the problem.

4. Case Study 1: Maximizing the solubility of Ibuprofen

In order to investigate the relative performance of the two relaxations, we first focus on the case study used by Jonuzaj et al. (2016) and present a brief summary of the salient features of the case study. The objective is to identify an optimal solvent mixture that leads to the maximum solubility of ibuprofen, x_{ibu} (i.e., the mole fraction of ibuprofen in the mixture) at 300 K and 1 atm. Ibuprofen (ibu) is a well-known anti-inflammatory drug that is used to relieve pain and symptoms of arthritis, as well as common pains such as headache, muscle ache and backache. Its manufacturing process involves crystallization of the solute from a suitable solvent or solvent mixture (Gordon and Amin, 1984; Karunanithi et al., 2006). A full description of the problem of identifying appropriate solvents or solvent mixtures that dissolve ibuprofen was given in Jonuzaj et al. (2016), and the proposed GDP formulations were converted into MINLPs by using the big-M approach. Here we extend this work by applying a Hull Relaxation and investigate the use of the two relaxation techniques in CAM^{b}D.

The model used in design includes phase equilibrium and phase stability relations that increase the complexity of the models. In this case study, however, only ibuprofen (solute) is at solid–liquid equilibrium (SLE) and this helps to avoid some difficulties as the number of phase equilibrium relations (equality of chemical potentials of each component at equilibrium in all phases) is independent of the number of components in the solvent mixture designed—in this case, one only needs to equate the chemical potentials of ibuprofen in the (pure) crystalline form and in the liquid mixture, not the chemical potentials of all components. The solubility of ibuprofen in the solvent mixture can thus be calculated as follows (Gmehling et al., 1978; Sandler, 1999):

Table 2
Indices and sets for case study 1.

Description	Index	Set	Value range
Total components in mixture	i, j	I	ibu, c_1, c_2, c_3
Fixed components in mixture		N'	ibu
Designed solvent molecules in mixture	ii	II	c_1, c_2, c_3
Pure candidate solvents	s, s'	S	1, …, 9
N°of solvent molecules in mixture	n	N	1, 2, 3
Functional groups	k, m	K	1, …, 14

$$\ln x_{\text{ibu}} + \ln \gamma_{\text{ibu}} = \frac{\Delta H_{fus}}{R}\left[\frac{1}{T_m} - \frac{1}{T}\right] \tag{6}$$

where x_{ibu} is the mole fraction of ibuprofen in the mixture, γ_{ibu} is the liquid phase activity coefficient of ibuprofen at composition x, temperature T and pressure P, ΔH_{fus} is the enthalpy of fusion of ibuprofen at temperature T_m, R is the gas constant, and T_m and T are the normal melting point of ibuprofen and the mixture temperature, respectively. The enthalpy of fusion of ibuprofen and its normal melting point are taken to be 25.5 kJ/mol and 347.15 K (Gracin and Rasmuson, 2002), respectively. The activity coefficient is evaluated using the UNIFAC (Fredenslund et al., 1975; Gmehling et al., 1978) group contribution method, where the expressions depend only on temperature and composition, but do not account for the (small) pressure dependence. The UNIFAC model proposed by Smith et al. (2001) is employed in this design problem and the relevant equations are presented in Appendix B for completeness. In order to ensure that the final mixture is in one phase, a miscibility constraint is introduced for every binary pair of solvent molecules, to ensure that they are mutually miscible in the proportions and at the temperature relevant to the mixture (Gani et al., 1991; Smith et al., 2001):

$$\left[\frac{\partial \ln \gamma_i^{i,j}}{\partial x_i^{i,j}}\right]_{T,P} + \frac{1}{x_i^{i,j}} \geq 0, \;\; i < j; \quad i = N' + 1, \ldots, N_c - 1; \quad j = i + 1, \ldots, N_c \tag{7}$$

where $\gamma_i^{i,j}$ is the activity coefficient of component i in a binary mixture of i and j, and $x_i^{i,j}$ is the mole fraction of component i in the mixture of i and j and can be calculated as follows:

$$x_i^{i,j} = \frac{x_i}{x_i + x_j}, \quad i, j \in I, \;\; i < j \tag{8}$$

$$x_i^{i,j} + x_j^{i,j} = 1, \quad i, j \in I, \;\; i < j \tag{9}$$

where x_i and x_j are the mole fractions of components i and j, respectively, in the multi-component mixture.

All the sets used in this case study are shown in Table 2. The solvent molecules and ibuprofen are built from the set of functional groups. The list of candidate solvents, the number of groups of each type k in ibuprofen ($\nu'_{\text{ibu},k}$) and the number of groups in a solvent s ($\nu'_{s,k}$) are presented in Appendix C in Tables C.1, C.2, and C.3, respectively.

The parameters used in the UNIFAC model for the prediction of the activity coefficient are taken from Poling et al. (2001).

4.1. Problem formulations

In the application of the proposed methodology, we consider two formulations: a restricted problem, where we fix the number of solvent components and design mixtures with one, two or three solvents (N=1,2, 3); the generalized problem, where the number of components in the mixture is not fixed but bounded by a maximum N_{max}, which takes a value of 3 here.

4.1.1. Restricted problem: fixed number of solvents

In a restricted problem, the decision variables are related to the identities and compositions of mixture constituents. As shown in Jonuzaj et al. (2016), the GDP formulation of the restricted problem consists of disjunctions for assigning each solvent s from the given list to components i in the mixture. In general, BM and HR formulations differ in the way conditional constraints of the form $h_{i,s}(x) \leq 0$ (cf. (GDP)) are recast. Here, conditional constraints represent the identity of the selected solvent ($n'_{i,k} = \nu'_{s,k}$), the molecular van der Waals volume ($r_i = r_s$) and the molecular surface area ($q_i = q_s$) of the mixture components. In both the BM and HR reformulations, these constraints can be expressed algebraically as products of the appropriate binary variable, $y_{i,s}$, with the variables $\nu'_{s,k}$, r_s and q_s, respectively, as given by:

$$n'_{i,k} = \sum_{s \in S} \nu'_{s,k}\, y_{i,s}, \quad i = c_1, \;\; c_2, \;\; c_3; \;\; k \in K \tag{10}$$

$$r_i = \sum_{s \in S} r_s\, y_{i,s}, \quad i = c_1, \;\; c_2, \;\; c_3 \tag{11}$$

$$q_i = \sum_{s \in S} q_s\, y_{i,s}, \quad i = c_1, \;\; c_2, \;\; c_3 \tag{12}$$

where $\nu'_{s,k}$ represents the number of groups of type k in solvent s; r_s and q_s are the van der Waals volume and surface area, respectively, for compound s in the solvent list. These relations force the variables $n'_{i,k}$, r_i and q_i to become zero when a solvent is not chosen, avoiding the use of the big-M parameter in the BM approach or of disaggregated variables in the HR. Thus, for the restricted problem, both approaches result in the same MINLP model. A more detailed description of the MINLP formulation for the design of a mixture with three solvents can be found in our previous work (Jonuzaj et al., 2016).

4.1.2. Generalized problem: unknown number of solvents

In the generalized case, the number of designed components in the mixture is not known in advance but rather is allowed to vary from one to three solvents. Hence, the design variables include the number of components in the mixture, their identities and compositions. The problem includes disjunctions for the assignment of each candidate solvent, disjunctions for the number of solvents selected and logic relations as described in Section 3. The disjunctions for the identity of solvent molecules in the mixture are the same as those described in the restricted problem (i.e., they include functions for the identity, the van der Waals volume and the van der Waals surface area of the components in the mixture). The disjunctions for the number of solvents selected include variables and functions that depend on the number of components in the mixture, such as the compositions of all the components, the miscibility functions and the UNIFAC model equations used to evaluating phase equilibria. The disjunctions of the

Table 3
Results obtained for case study 1–comparison of BM (Jonuzaj et al., 2016) and HR formulations.

Problem: case	x_{ibu}	c_1	x_{c_1}	c_2	x_{c_2}	c_3	x_{c_3}	CPU (s)
A1: N=1	0.31833	$CHCl_3$	0.68167					0.35
A2: N=2	0.33383	$CHCl_3$	0.52292	MeOH	0.14325			4.87
A3: N=3	0.33375	$CHCl_3$	0.52263	MeOH	0.14262	EtOH	0.00100	107.90
A4: N ≤ 3 (G-BM)	0.33383	$CHCl_3$	0.52292	MeOH	0.14325			369.06
A5: N ≤ 3 (G-HR)	0.33383	$CHCl_3$	0.52292	MeOH	0.14325			93.59

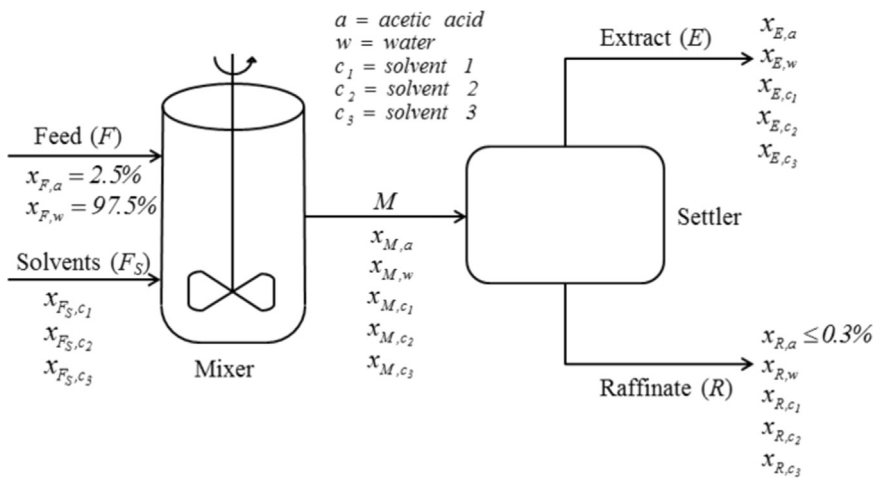

Fig. 1. Schematic of a single stage process for the extraction of acetic acid from water. Capital letters denote stream labels/flowrates and $x_{p,i}$ denotes the mole fraction of component i in stream p.

general problem are converted into mixed-integer form via BM and HR techniques. Since the SLE relations are expressed only in terms of ibuprofen, the natural logarithms that appear in the UNIFAC model (Eqs. (B.2) and (B.7)) in Appendix B are treated as general constraints and thus, do not cause any singularities when formulated via the HR approach. The MINLP formulation resulting from the BM approach is presented in Jonuzaj et al. (2016), whereas the MINLP model obtained using the HR is given as model (G1-HR) in Appendix D.

4.2. Results and discussion

All models are implemented and solved in GAMS (GAMS Development Corporation, 2014) version 24.2.3 and are run on a single core of a dual 6 core Intel Xeon X5675 machine at 3.07 GHz with 48 GB of memory. One solution approach would be to enumerate the different options on parallel computers. This is tractable for some values of N_{max} and N_s (e.g., for N_{max}=3 and N_s=9 which is the case of this example problem) but becomes challenging as the dimensionality of the problem increases. Here, we investigate the use of MINLP algorithms on a single processor for the solution of the problem. DICOPT (Kocis and Grossmann, 1989; Grossmann et al., 2002), a local MINLP solver, based on an implementation of the outer-approximation (Duran and Grossmann, 1986; Fletcher and Leyffer, 1994) was used to solve all problems, and attempts were made to solve to global optimality with ANTIGONE (Misener and Floudas, 2014) and BARON (Tawarmalani and Sahinidis, 2005). The restricted model was solved for one, two and three solvents, and the general models for at most three selected solvents; the results are summarized in Table 3. The best solution, which corresponds to the highest solubility with a value of 0.33383 in mole fraction, is achieved for a binary mixture of chloroform ($CHCl_3$) and methanol (MeOH). A mixture with three components, with ethanol (EtOH) as the third solvent, gives slightly lower solubility than the binary mixture. The mole fraction of the third component is at the user-specified lower bound of $x_{c_3}^L = 0.001$, which means that only a small amount of ethanol is added to the mixture, and thus it does not have any significant impact on the solubility of the drug.

The maximum solubility achievable in the restricted problem is more affected if tighter bounds are used for the minimum mole fraction of solvent. For lower bound values of 0.01 and 0.1 (i.e., $x_i^L = 0.01$ and $x_i^L = 0.1$), for example, the solubility (x_{ibu}) becomes 0.33312 and 0.32696, respectively in three-component solvent mixtures. These solutions are more markedly different from the performance of the binary solvent pair.

In the general problem (with the number of solvents N unknown), both the (G-BM) and (G-HR) models yield the same optimal solution and the results validate those obtained when solving the three restricted problems, confirming that the highest solubility is achieved in a binary mixture of chloroform and methanol.

Several runs were carried out from different starting points in each case and the values of the CPU time presented in Table 3 correspond to the runs where the best solutions were found. In cases where the number of mixture components is fixed (N = 1, 2, 3), the CPU time increases rapidly with the number of components, due to the increased size and complexity of the problems. The (G-BM) formulation of the general problem requires more CPU time than is needed to solve each instance of the restricted problem with a fixed number of components, from one to three. The (G-HR) model, on the other hand, is solved in less CPU time than the generalized (G-BM) problem and it also appears to be more effective than enumerating all options in the restricted problem (N = 1, 2, 3), from the perspective of computational cost. Although the CPU times are representative of most solution attempts from different starting points, the final solutions achieved cannot be guaranteed to be global and the corresponding computational times are affected by the initial guesses.

In working towards obtaining global solutions, global MINLP algorithms, ANTIGONE version 1.1 (Misener and Floudas, 2014) and BARON version 15.9.22 (Tawarmalani and Sahinidis, 2005), both accessed via GAMS version 24.6.1, were employed to solve the problems. However, only problem A1, which is the smallest in size (N=1), was solved globally with ANTIGONE and BARON in 0.02 and 5.55 CPU seconds, respectively. The results obtained with both global solvers verified the optimal solution obtained with the DICOPT MINLP solver, i.e., the optimal solvent is found to be chloroform which yields a solubility of 0.31833, as presented in Table 3. As the number of components in the mixture is increased, the problems become more challenging and thus convergence to global optimality was not reached in 36,000 CPU seconds (10 CPU hours). The global solvers did not identify better solutions than those reported with DICOPT. Although it is desirable to find a global solution, this is not yet practical with these deterministic branch-and-bound algorithms.

5. Case Study 2: Separation of acetic acid from water by liquid–liquid extraction

Acetic acid is a colorless organic compound which can be produced by methanol carbonylation, acetaldehyde oxidation, direct oxidation of ethylene, hydrocarbon oxidation or fermentation (Sano et al., 1999; Cheung et al., 2002; Gullo et al., 2014). In all cases, the purification of acetic acid from an aqueous stream is required and can be achieved via

Table 4
Indices and sets for case study 2.

Description	Index	Set	Value range
Total components in mixture	i, j	I	a, w, c_1, c_2, c_3
Fixed components in mixture		N'	a, w
Designed solvent molecules in mixture	ii	II	c_1, c_2, c_3
Pure candidate solvents	s, s'	S	1, …, 8
$N^o_=$ of solvent molecules in mixture	n	N	1, 2, 3
Functional groups	k, m	K	1, …, 11
Streams	p		F, F_S, M, E, R

several separation techniques, such as adsorption, distillation, liquid–liquid extraction, membrane separation and crystallization (Katikaneni and Cheryan, 2002; Kaur and Vohra, 2010; Cheung et al. 2012 ; Choi and Kim, 2013). Among these methods, liquid–liquid extraction is accepted as an efficient, economical and environmentally friendly method for the separation of acetic acid from water. In particular, solvent extraction is employed as a first step in the purification of mixtures with a relatively low concentration of acetic acid (e.g., mixtures of 90% mol water and 10% acetic acid), in order to reduce the energy cost associated with the vaporization of a large amount of water during distillation (Eyal and Canari, 1995; Karunanithi et al., 2005; Alkaya et al., 2009; IJmker et al., 2014). Liquid–liquid extraction is an important separation technology for a wide range of applications in the chemical process industries, in which components are separated based on their relative solubilities in two coexisting liquid phases. In a liquid–liquid extraction unit, the component(s) to be recovered (here, acetic acid) and the liquid carrier (here, water) are separated by the addition of a solvent or solvent mixture that preferentially dissolves one or more of the components therein. The degree of separation depends on the unequal distribution of the components between the two phases formed at equilibrium. The success of the extraction process depends to a large extent on the choice of a suitable solvents or solvent mixtures that meet regulatory and environmental standard. The single stage extraction of acetic acid from water is illustrated in Fig. 1, where an aqueous solution of acetic acid is first mixed with a suitable solvent or solvent mixture and is then separated into two phases in a settler (Seader et al., 2011).

Several researchers have previously investigated the design or selection of solvents for liquid–liquid extraction based on solvent properties, such as selectivity, solvent loss, solvent power and distribution coefficient (Cockrem et al., 1989; Pretel et al., 1994; Karunanithi et al., 2005; Yang and Song, 2006; Cignitti et al., 2015). In particular, Karunanithi et al. (2005) addressed the problem of identifying appropriate solvents for the extraction of acetic acid from water. Their model included physical property and process model constraints, and the amount of solvent in the feed mixture was fixed *a priori* by specifying the solvent-to-feed flowrate ratio. The objective was to minimize the quantity of acetic acid in the raffinate phase. The solvent molecules were designed from functional groups and a large list of alcohols, ketones, aldehydes, acids, esters and ethers was generated based on structural constraints. The molecules generated were then screened using pure component and mixture property constraints and only two compounds that satisfied those constraints were included in the final optimization problem. In a similar approach, Cignitti et al. (2015) applied a systematic computer-aided design methodology to the design of suitable solvents for the extraction of acetic acid from water. In their work, the solvent flowrate was fixed and the objective was to maximize the molar flowrates of acetic acid and water in the extract and raffinate, respectively. As in Karunanithi et al. (2005), a decomposition-based algorithm was used to solve the MINLP problem, where first acyclic solvent components were designed from atom groups based on structural constraints, and then the molecules were screened using pure component and mixture property constraints. Nineteen compounds passed the screening tests, and thus nineteen NLP problems including the objective function and process model constraints were solved in the final step.

In our current work, a general GDP formulation of the solvent mixture design problem for the extraction of acetic acid from water is derived based on the framework presented in Section 3. In particular, the number of components, the identities and the compositions of the mixture constituents are treated as unknowns in the model. A feed mixture with 2.5 mol% acetic acid (approximately 8 wt%) and a flowrate of 13,500 kg/h at 298 K and 1 atm is considered. The only performance constraint is that the raffinate should contain at most 0.3% mol of acetic acid. The objective is to determine the optimal solvent mixture that achieves a trade-off between a low solvent-to-feed ratio and an extract phase that contains a large fraction of acetic acid and small amount of water. This could in principle be addressed via a multi-objective formulation to explore the space of Pareto solutions (e.g., see Papadopoulos et al., 2013 and Burger et al., 2015 for examples of the use of multi-objective optimization in molecular design). Here, however, we adopt a single objective function similar, but not identical, to that proposed by Naser and Fournier (1991):

$$\min \frac{F_S}{F} + x_{E,w} - x_{E,a} \tag{13}$$

where F_S and F are the solvent and feed flowrates, respectively, while $x_{E,a}$, $x_{E,w}$ are the mole fractions of acetic acid and water in the extract phase.

The evaluation of the objective function requires the determination of two unknown liquid phases that are at equilibrium (i.e., the extract E and the raffinate R). The following equations describe the liquid–liquid equilibrium (LLE) relations between extract and raffinate phases:

$$\gamma_{E,i}(T, P, x_E)x_{E,i} = \gamma_{R,i}(T, P, x_R)x_{R,i}, \quad i \in I \tag{14}$$

where $x_{E,i}$ and $x_{R,i}$ are the mole fractions of component i in the extract and raffinate streams, respectively, and $\gamma_{p,i}$, $p \in \{E, R\}$, denotes the liquid phase activity coefficient of component i in stream p at temperature T, composition x_p of stream p and pressure P. The activity coefficients are evaluated using the modified UNIFAC (Dortmund) model (Gmehling et al., 1993, 2002); the relevant equations are presented in Appendix B. Note that the equations run over the set I of all components in the mixture, in contrast to case study 1, where the SLE equilibrium relation held only for one fixed solute, so that the number of equations in the model was independent of the number of solvents. The presence of solvent in both the raffinate and extract streams thus leads to an increased level of complexity in the problem formulation.

The following mole balances for each component in the system are also included in the formulation:

$$x_{F,a}F = x_{M,a}M \tag{15}$$

$$x_{F,w}F = x_{M,w}M \tag{16}$$

$$x_{F_S,ii}F_S = x_{M,ii}M, \quad ii \in II \tag{17}$$

$$x_{M,i}M = x_{E,i}E + x_{R,i}R, \quad i \in I \tag{18}$$

where M, E, and R are the flowrates of the stream leaving the mixer (acetic acid–water–solvents), the extract and the raffinate, respectively; $x_{p,i}$ represents the mole fraction of component i in stream p. As with the phase equilibrium model, the existence of some of the mole balance equations depend on the number of components in the solvent mixture. All the design sets used in this case study are shown in Table 4. The solvent molecules, acetic acid and water are built from the set of functional groups used in Dortmund UNIFAC.

The list of solvents in the set S is taken from Akula (2011), who screened potential solvent molecules using solvent properties previously reported in the literature (selectivity, solvent loss, solvent

power and distribution coefficient) in order to obtain an initial design space of eight candidate solvents. The organic solvents selected are all miscible or partially miscible with each other and with acetic acid, and therefore the miscibility constraints (Eq. (7)) are not considered in this case study. Here, low solvent losses to the raffinate phase (i.e., raffinate-extract immiscibility) are achieved by including this criterion in the pre-screening of the candidate compounds (Akula, 2011) used to arrive at the list of suitable solvents. We note that one could additionally consider solvent loss in the raffinate as a constraint in the design. The list of candidate solvents, the number of groups of type k in acetic acid ($\nu'_{a,k}$) and water ($\nu'_{w,k}$), and the number of groups of type k in each candidate solvent s ($\nu'_{s,k}$) are presented in Appendix C in Tables C.1, C.2., and C.3 , respectively.

5.1. Problem formulations

In this section, the formulation of the restricted problem, where the number of components is fixed, is first presented. The formulation of the general problem, with unknown number of solvents is then introduced. In the restricted problem, mixtures with one, two or three solvents are designed and in the general model, mixtures with at most three solvents are identified.

5.1.1. Restricted problem: fixed number of solvents

The formulation of the restricted problem is presented for the selection of three solvents but it can be extended to any fixed number of solvents. The model includes disjunctions for assigning solvents from the list to designed components in the mixture and the assignment of each compound is determined through Boolean variables, $Y_{i,s}$ (a compound s is assigned to component i in the mixture if $Y_{i,s}$ is True and it is not if $Y_{i,s}$ is False). The disjunctions for the assignment of solvents are shown below:

$$\bigvee_{s\in S}\begin{bmatrix} Y_{i,s} \\ n'_{i,k} = \nu'_{s,k}, \quad k \in K \\ q_i = q_s \\ r_i = r_s \end{bmatrix}, \quad i = c_1, c_2, c_3 \tag{R2-D}$$

In the above disjunctions, only one of the Boolean variables $Y_{i,s}$ can be selected, which ensure the selection of exactly three solvents. The identity of designed component i in the mixture ($n'_{i,k}$), its molecular van der Waals volume (r_i) and its molecular surface area (q_i) are represented via conditional constraints and are defined by multiplying $\nu_{s,k}$, r_s and q_s with the corresponding binary variable, as shown in Eqs. (10), (11) and (12), respectively. Thus, as discussed in the first case study (Section 4), the BM and HR approaches result in the same MINLP formulation.

In addition to the disjunctions, logic relations are derived to avoid degenerate solutions by enforcing a certain ordering of the solvents and to avoid selecting a given candidate solvent more than once, as shown in the first and second rows of Table 1. An algebraic reformulation of disjunctions (R2-D) is included in the problem formulation, resulting in the desired MINLP, whose detailed formulation is presented in Appendix D, as problem (R2-MINLP), for completeness.

5.1.2. Generalized problem: unknown number of solvents

In the general problem, the number of solvent components present in the mixture is allowed to vary from one to three solvents. As discussed in Section 3, the general problem includes disjunctions for the assignment of each candidate solvent and disjunctions for the number of selected solvents. The first set of disjunctions is the same as that presented in the restricted problem (disjunctions (R2-D)). The second set of disjunctions includes variables and equations that depend on the number of components in the mixture, such as the compositions, mole balances, liquid–liquid equilibrium equations and the modified UNIFAC model equations. These functions and the relevant

Table 5
Optimal objective function, optimal solvent mixtures and CPU times obtained when solving the restricted problem (for different values of N) and the general problem (with the BM and HR formulations) of case study 2. Unless otherwise stated, the lower bound on solvent mole fractions in stream F_S is set to $x^L_{F_S,i} = 0.001$, $i \in \{3, \ldots, N+2\}$.

Problem: case	Objective	Components	$x_{F,i}$	$x_{F_S,i}$	$x_{E,i}$	$x_{R,i}$	CPU (s)
B1: N=1	2.3203	Acetic acid	0.0250		0.0090	0.0030	1.00
		Water	0.9750		0.1293	0.9925	
		Hexanone		1	0.8617	0.0045	
B2: N=2	2.2299	Acetic acid	0.0250		0.0092	0.0030	17.59
		Water	0.9750		0.1978	0.9913	
		Hexanone		0.8341	0.6616	0.0040	
		Pentanol		0.1659	0.1314	0.0017	
B3: N=3	2.2305	Acetic acid	0.0250		0.0092	0.0030	470.84
		Water	0.9750		0.1977	0.9913	
		Hexanone		0.8340	0.6616	0.0040	
		Pentanol		0.1645	0.1303	0.0017	
		Heptanol		0.0015	0.0012	0.0000	
B4: N=3 ($x^L_{F_S,i} = 0.01$)	2.2342	Acetic acid	0.0250		0.0092	0.0030	175.46
		Water	0.9750		0.1971	0.9914	
		Hexanone		0.8332	0.6614	0.0040	
		Pentanol		0.1568	0.1244	0.0016	
		Heptanol		0.0100	0.0079	0.0000	
B5: $N \leq 3$ (G-BM)	2.2299	Acetic acid	0.0250		0.0092	0.0030	350.84
		Water	0.9750		0.1978	0.9913	
		Hexanone		0.8341	0.6616	0.0040	
		Pentanol		0.1659	0.1314	0.0017	
B6: $N \leq 3$ (G-HR)	2.2299	Acetic acid	0.0250		0.0092	0.0030	297.02
		Water	0.9750		0.1978	0.9913	
		Hexanone		0.8341	0.6616	0.0040	
		Pentanol		0.1659	0.1314	0.0017	

variables are placed in the appropriate disjunctions and are presented as formulation (G2-D) in Appendix D. It should be noted that although variables $n'_{i,k}$, q_i and r_i of the UNIFAC model depend on the number of solvent molecules in the mixture, the corresponding assignment equations can be placed outside the disjunctions (G2-D) and treated as general constraints. These are due to the fact that these variables are defined algebraically via the binary variables representing solvent assignment, $y_{i,s}$, (see Eqs. (10)–(12)) and that these binary variables are linked to the binary variables $\tilde{y}_n$ for the number of mixture components (see Table 1). Thus, the variables $y_{i,s}$, $s \in S$, become zero when component i is not present in the mixture.

Following the formulation steps outlined in Section 3, all logic conditions presented in Table 1 are also derived and included as algebraic equations in the MINLP models. The disjunctions (G2-D) are reformulated via BM and HR and the resulting MINLP problems are given as models (G2-BM) and (G2-HR) in Appendix D. In the Hull reformulations, disjunctive equations that include logarithmic terms, such as the natural logarithms that appear in Eqs. (B.15) and (B.21), are converted into equivalent equations using the approximation function in Eq. (4) in order to avoid numerical difficulties.

5.2. Results and discussion

All models were implemented and solved in GAMS (GAMS Development Corporation, 2014) version 24.2.3 using DICOPT (Duran and Grossmann, 1986; Kocis and Grossmann, 1989; Grossmann et al., 2002) or a global solver (ANTIGONE (Misener and Floudas, 2014) or BARON (Tawarmalani and Sahinidis, 2005)). All problems were run on the same computer as case study 1 and its specifications are given in Section 4.2.

5.2.1. Solution with DICOPT

The restricted problem was solved for one, two and three selected solvents and the results are summarized in Tables 5 and 6. The solvent that yields the minimum objective function value is a binary mixture of hexanone and pentanol. It also gives the highest recovery of acetic acid in the extract phase. When varying the number of components in the mixture, the largest difference is seen on going from a pure solvent to a binary mixture, due to a 7% decrease in solvent flowrate. The best ternary mixture found includes hexanone and pentanol, with heptanol as the third solvent; it gives a slightly worse objective function value than the binary mixture, because the solvent flowrate (F_S) increases. The mole fraction of heptanol in the mixture is very small (close to its lower bound, i.e., $x^L_{S,i}$=0.001) and this small amount of solvent does not modify significantly the results obtained with a binary mixture. The impact of the third solvent in the separation process is clearer when increasing the lower bound of the solvent mole fraction from 0.001 to 0.01, so that a greater amount of the third solvent is added to the mixture. This is seen to affect mostly the solvent mixture flowrate, as shown in Tables 5 and 6.

Hexanone, which is consistently found as the main component in the solvent mixture for separating acetic acid from water, was also identified by Karunanithi et al. (2005) as the best extractant for the same application, which indicates that a good local solution has been found. Hexanone was also found as a feasible solution in Cignitti et al. (2015) for the extraction of acetic acid from water. Based on the design criteria used by the authors, the best separation, however, with a high recovery of the solute, was achieved with butane-2,3-diyl diformate (atom groups: $2CH_3$, 2CH, 2HCOO). This molecule is not included in our solvent list (Table C4), and thus it is not identified as an optimal solution in our models. From Table 6, it can be observed that by using a mixture with two or three solvents in the process a smaller solvent flowrate is required and a larger percentage of acetic acid is recovered. However, the amount of water in the extract stream also increases. Hence, using a solvent mixture instead of a pure solvent may lead to better reduced capital and operating costs and increased acetic acid recovery but at the expense of slightly more demanding downstream separation as a large amount of water is transferred to the extract stream. One could modify the objective function to increase water recovery by putting more weight on the mole fraction of water in the extract stream.

The general problem was formulated via BM and HR and it was solved for at most three selected solvents (Table 5). Both (G-BM) and (G-HR) models yield the same optimal solution and the results are in agreement with those obtained for the restricted problem, where the best separation is achieved with a binary mixture of hexanone and pentanol.

The computational times required to obtain an optimal solution for each problem are also presented in Table 5. In the restricted problem, where the number of components is fixed, the CPU time increases with the number of components, as can be expected, due to the increasing size of the problems. For ternary mixtures, the CPU time decreases when the lower bound on the mole fraction of solvent is increased. In the general problem, where the number of components is unknown and can vary from one to three, the (G-HR) formulation requires slightly less computational time than (G-BM) model in the examples studied. Nevertheless, solving the general problem using the (G-BM) or (G-HR) model appears to be more efficient than solving the restricted problem with three solvents (N=3) in terms of computational cost. When the lower bound on the solvent mole fraction is increased ($x^L_{\tilde{S},i}$ = 0.01), the computational cost of either the (G-BM) or the (G-HR) model is higher than solving all cases (N=1, 2, 3 selected solvents) of the restricted problem. It should be noted though that the solution times observed with DICOPT are affected by the initial guesses provided in each problem and thus the application of a global optimization algorithm is required in order to conclude on the effectiveness of the solution approaches (cf. Section 5.2.2).

All the mixed solvents identified contain one or two alcohols, which may react with acetic acid, leading to the loss of acetic acid and the production of esters. We expect these reactions to be very slow in the absence of catalysts (Agreda and Zoeller, 1993; Lee et al., 2000; Cheung et al. 2012), but to minimize the risk of side reactions affecting the process, we consider the design of solvent mixtures by excluding alcohols (i.e., pentanol and heptanol) from the solvent list. The results

Table 6
Objective function value, optimal solvent, raffinate and extract flowrates, extract mole fractions and acetic acid recovery obtained when solving case study 2. The solutions of the restricted problem for different values of N and of the general problem with the BM and HR formulations are shown. All flowrates are shown in kmol/h.

Problem	Objective	F_S	E	R	$x_{E,a}$	$x_{E,w}$	%Recovery
B1	2.3203	1557.63	1805.21	460.42	0.0090	0.1293	92.20
B2	2.2299	1445.17	1820.06	333.11	0.0092	0.1978	94.35
B3	2.2305	1445.71	1820.47	333.24	0.0092	0.1977	94.35
B4	2.2342	1448.78	1822.84	333.93	0.0092	0.1971	94.34
B5	2.2299	1445.17	1820.06	333.11	0.0092	0.1978	94.35
B6	2.2299	1445.16	1820.05	333.11	0.0092	0.1978	94.35

Table 7
Optimal objective function, optimal solvent mixtures and CPU times obtained when solving the restricted problem (for different values of N) and the general problem (with the BM and HR formulations) of case study 2, with alcohols excluded from the solvent list. The lower bound on solvent mole fractions in stream F_S is set to $x^L_{F_S,i} = 0.01$, $i \in \{3, \ldots, N+2\}$.

Problem: case	Objective	Components	$x_{F,i}$	$x_{F_S,i}$	$x_{E,i}$	$x_{R,i}$	CPU (s)
C1: N=1	2.3203	Acetic acid	0.0250		0.0090	0.0030	0.01
		Water	0.9750		0.1293	0.9925	
		Hexanone		1	0.8617	0.0045	
C2: N=2	2.3265	Acetic acid	0.0250		0.0090	0.0030	0.08
		Water	0.9750		0.1303	0.9925	
		Hexanone		0.9900	0.8521	0.0045	
		Isopropyl acetate		0.0100	0.0086	0.0000	
C3: N=3	2.3355	Acetic acid	0.0250		0.0090	0.0030	9.56
		Water	0.9750		0.1309	0.9925	
		Hexanone		0.9800	0.8429	0.0044	
		Isopropyl acetate		0.0100	0.0086	0.0001	
		Butyl acetate		0.0100	0.0086	0.0000	
C4: $N \leq 3$ (G-BM)	2.3203	Acetic acid	0.0250		0.0090	0.0030	0.03
		Water	0.9750		0.1293	0.9925	
		Hexanone		1	0.8617	0.0045	
C5: $N \leq 3$ (G-HR)	2.3203	Acetic acid	0.0250		0.0090	0.0030	0.13
		Water	0.9750		0.1293	0.9925	
		Hexanone		1	0.8617	0.0045	

Table 8
Objective function value, optimal solvent, raffinate and extract flowrates, extract mole fractions and acetic acid recovery obtained when solving case study 2, with alcohols excluded from the solvent list. The solutions of the restricted problem for different values of N and of the general problem with the BM and HR formulations are shown. All flowrates are shown in kmol/h.

Problem	Objective	F_S	E	R	$x_{E,a}$	$x_{E,w}$	%Recovery
C1	2.3203	1557.63	1805.21	460.42	0.0090	0.1293	92.20
C2	2.3265	1561.29	1811.56	457.73	0.0090	0.1303	92.24
C3	2.3355	1567.19	1819.75	455.44	0.0090	0.1309	92.28
C4	2.3203	1557.63	1805.21	460.42	0.0090	0.1293	92.20
C5	2.3203	1557.63	1805.21	460.42	0.0090	0.1293	92.20

obtained with this reduced design space for the restricted and general problems are presented in Tables 7 and 8. With the removal of pentanol and heptanol, which were previously identified as the optimal solvents in the binary and ternary mixtures, the best solution found requires a pure solvent and not a solvent mixture. The optimal extractant that yields the minimum objective function value appears to be hexanone. Mixtures with two and three solvents, where isopropyl acetate and butyl acetate are identified as the second and third optimal components, respectively, lead to slightly worse results. The optimal solution obtained when solving the three restricted problems is also found with the (G-BM) and (G-HR) formulations of the general problem. The observations made on computational cost for the larger solvent list remain valid here: the CPU time required to solve the restricted problem increases when more solvents are added to the mixture (i.e., $N = 1, 2, 3$), while the (G-BM) and (G-HR) formulations require less computational time than the solution of the three restricted problems. Furthermore, the computational cost of solving every formulation is significantly smaller than that of solving the corresponding formulations when all candidate solvents are included. This lower cost may be due partly to the reduced search space, as a smaller solvent list is used, with a consequent decrease in the number of possible solvent mixtures (for instance, for N=3, there are 56 possible solvent combinations when the full list of solvents is used and only 20 combinations when alcohols are excluded). Furthermore, since all problems are solved locally, the computational cost is affected by the starting point given to the solver.

Table 9
Best known objective function value and corresponding solvent mixtures obtained with BARON when solving globally the restricted problem (for different values of N) and the general problem (with the BM and HR formulations) of case study 2, with alcohols excluded from the solvent list. In all cases except C1, convergence to global optimality is not achieved within 10 CPU hours.

Problem	Objective	Components	$x_{F,i}$	$x_{F_S,i}$	$x_{E,i}$	$x_{R,i}$
C1	2.3203	Acetic acid	0.0250		0.0090	0.0030
		Water	0.9750		0.1293	0.9925
		Hexanone		1	0.8617	0.0045
C2	2.3265	Acetic acid	0.0250		0.0090	0.0030
		Water	0.9750		0.1303	0.9925
		Hexanone		0.9900	0.8521	0.0045
		Isopropyl acetate		0.0100	0.0086	0.0000
C3	2.3355	Acetic acid	0.0250		0.0090	0.0030
		Water	0.9750		0.1309	0.9925
		Hexanone		0.9800	0.8429	0.0044
		Isopropyl acetate		0.0100	0.0086	0.0001
		Butyl acetate		0.0100	0.0086	0.0000
C4	2.3203	Acetic acid	0.0250		0.0090	0.0030
		Water	0.9750		0.1293	0.9925
		Hexanone		1	0.8617	0.0045
C5	2.3203	Acetic acid	0.0250		0.0090	0.0030
		Water	0.9750		0.1293	0.9925
		Hexanone		1	0.8617	0.0045

5.2.2. Solution with global algorithms

In order to obtain global solutions, ANTIGONE version 1.1 and BARON version 15.9.22 were also used in this case study to solve the problems. In the case where all candidate solvents were included in the solvent list (i.e., problems B1-B6), only the smallest problem, B1, was solved globally with both solvers, whereas convergence was not reached in 36,000 CPU seconds in problems B2–B6. In problem B1, both ANTIGONE and BARON found the same optimal solution as that obtained with DICOPT, i.e., the optimal extractant is hexanone and the objective value is equal to 2.3203, as presented in Table 5. Global optimality was achieved with both ANTIGONE and BARON in problem C1, when alcohols were excluded from the solvent list. The optimal solutions obtained with DICOPT were also found with BARON in problems C2-C5. However, it cannot be guaranteed that these are global solutions as convergence to global optimality was not reached within 10 CPU hours. No solutions were identified with ANTIGONE within the same time limit. The results obtained with BARON are presented in Table 9.

6. Conclusions

The design of optimal mixtures using Generalized Disjunctive Programming within the Computer Aided Mixture/blend Design framework presented in previous work (Jonuzaj et al., 2016) was briefly discussed in this paper, and the use of different MINLP reformulations of the problem was investigated. Within this systematic approach, the general CAM^{b}D problem in which the number, identities and compositions of mixture ingredients need to be determined, was formulated using GDP techniques. The number of interlinked decisions to be considered in the general mixture problem can lead to challenging mixed integer optimization problems with nonconvexities and large combinatorial spaces. In particular, when the number of components in the mixture is allowed to vary, highly nonlinear model equations whose relevance depends on this decision variable, such as phase equilibrium relations, may lead to numerical problems. In order to address these difficulties and to facilitate problem formulation and solution, the GDP framework was used to formulate the discrete choices inherent in mixture design problems, as described in Jonuzaj et al. (2016). Two different solution strategies, the Big-M approach and the Hull Relaxation, were employed in this work to transform the disjunctive constraints into mixed-integer form. The BM approach is the most common relaxation technique. For convex problems it is known that it can give poor relaxations, whereas HR leads to bounds that are at least as tight as or tighter than the BM approach, at the expense of introducing new variables and constraints. While the mixture design problems considered are nonconvex, an initial investigation of the ease of solution of the formulations was carried out.

The general mixture design methodology and the two relaxation techniques were successfully applied to two case studies for the design of optimal solvent mixtures. The solvents to be designed were selected from a given list of candidate pure compounds. In the first case study, the objective was to design the optimal solvent or solvent mixture to maximize the solubility of ibuprofen at 298 K. This problem was considered in our earlier work (Jonuzaj et al., 2016) with a BM formulation and the effectiveness of the HR approach was investigated here. The problem includes nonconvex solid–liquid equilibrium relations and stability functions. The second case study involved the design of an optimal solvent or solvent mixtures for separating acetic acid from water in a single-stage liquid–liquid extraction process. This challenging problem includes liquid–liquid equilibrium relations that must hold for all known and unknown components in the process, increasing the complexity of the models. In both case studies, the problem of finding an optimal solvent mixture with a fixed number of components (the restricted problem) was first formulated using 1, 2 or 3 solvents; then mixtures with up to three solvents were designed without specifying the number of components *a priori* (the general problem).

Optimal solutions were obtained for all problem instances using a local MINLP algorithm. The results showed that the proposed formulations offer a promising approach to mixture design, as the simultaneous design of the optimal number, identity and compositions of the components that participate in a mixture, can be achieved. The two different relaxation techniques (BM and HR) were applied successfully to convert the GDP formulations of the mixture problems into MINLP models and were found to lead to computationally efficient solutions, compared to enumerating the number of solvents in the mixture. Thus, on the basis of the runs performed, the solution of the general problem appears to be more advantageous than the repeated solution of the restricted problem, even when choosing up to 3 components only. For the size and type of problems considered here, no systematic trend in the relative computational performance of the BM and HR approaches could be observed. Our work highlights that the development of improved solution techniques is also an important area of research, although it is beyond our current scope. In future investigations, efforts will be directed at tackling larger problems and at achieving convergence to the global solution in order to carry out a comprehensive comparison of the proposed formulations. In addition, the application of the proposed approach to the design of formulated products (Yunus et al., 2014; Zhang et al., 2015) will be investigated.

Data statement: Data underlying this article can be accessed on Zenodo at https://zenodo.org/record/55145, and used under the Creative Commons Attribution license.

Acknowledgements

The authors gratefully acknowledge financial support from the Engineering and Physical Sciences Research Council (EPSRC) of the UK via a Leadership Fellowship (EP/J003840/1).

Appendix A. Generalized Disjunctive Programming

A.1. GDP formulation

The general formulation of a GDP involves Boolean and continuous variables, algebraic equations, disjunctions and logic propositions as shown below:

$$
\begin{aligned}
\min_{x,Y} \quad & f(x) \\
\text{s. t.} \quad & g(x) \le 0 \\
& \bigvee_{j\in J_k} \begin{bmatrix} Y_{j,k} \\ h_{j,k}(x) \le 0 \\ D_{j,k}x \le d_{j,k} \end{bmatrix} \quad k \in K \\
& \underline{\vee}_{j\in J_k} Y_{j,k}, \;\; k \in K \\
& \Omega(Y) = \text{True} \\
& x \in [x^L, x^U] \;\subset \mathbb{R}^m \\
& Y_{j,k} \in \{\text{True, False}\}, \;\; j \in J_k, \;\; k \in K
\end{aligned}
\tag{GDP}
$$

where the objective, f, is a function of continuous variables represented by the m-dimensional vector x. The set of equations $g(x) \le 0$ represents general constraints that must hold regardless of the discrete choices, while $h_{j,k}(x) \le 0$ and $D_{j,k}x \le d_{j,k}$ are vectors of nonlinear and linear conditional constraints that depend on the discrete decisions, as represented by the Boolean variables Y. The conditional constraints are included inside disjunctions and are linked by the OR ($\vee$) operator. The disjunctive terms are assigned Boolean variables, $Y_{j,k}$, and each term is active when the corresponding Boolean variable is true. Only one of the Boolean variables can be true in each disjunction $\left(\underline{\vee}_{j\in J_k} Y_{j,k}\right)$, where the notation $\underline{\vee}$ denotes the EXCLUSIVE OR operator. $\Omega(Y)$ represents a set of logic relations for the Boolean variables (i.e., these relations involve only the Boolean variables) which are expressed via propositional logic (Raman and Grossmann, 1994; Ruiz et al., 2012).

A.2. Reformulation of GDP as an MINLP via Big-M approach

The general reformulation of a GDP as an MINLP via big-M is given by:

$$
\begin{aligned}
\min_{x,y} \quad & f(x) \\
\text{s. t.} \quad & g(x) \le 0 \\
& h_{j,k}(x) \le M_{j,k}(1 - y_{j,k}), \;\; j \in J_k, \;\; k \in K \\
& D_{j,k}x - d_{j,k} \le M_{j,k}(1 - y_{j,k}), \;\; j \in J_k, \;\; k \in K \\
& Ay \le b \\
& \sum_{j=1}^{J_k} y_{j,k} = 1, \;\; k \in K \\
& x \in [x^L, x^U] \;\subset \mathbb{R}^m \\
& y_{j,k} \in \{0, 1\}, \;\; j \in J_k \; k \in K
\end{aligned}
\tag{BM}
$$

where y is a matrix of binary variables that has one-to-one correspondence with the Boolean variable matrix, Y, while the parameter matrix M is such that each element $M_{j,k}$ is a "sufficiently large" upper bound such that when $y_{j,k} = 0$ the inequality constraint is always satisfied. The logic propositions in GDP, $\Omega(Y) = \text{True}$, have been converted into linear inequalities in (BM), $Ay \le b$, using Boolean algebra rules (Williams, 1985; Raman and Grossmann, 1991).

A.3. Reformulation of GDP as an MINLP via Hull Relaxation

The general formulation for linear and/or nonlinear disjunctions is as follows:

$$
\begin{aligned}
\min_{x,y} \quad & f(x) \\
\text{s. t.} \quad & g(x) \le 0 \\
& x = \sum_{j\in J_k} \nu_{j,k}, \;\; k \in K \\
& y_{j,k} h_{j,k}\left(\frac{\nu_{j,k}}{y_{j,k}}\right) \le 0, \;\; j \in J_k, \;\; k \in K \\
& D_{j,k}\nu_{j,k} \le d_{j,k}y_{j,k}, \;\; j \in J_k, \;\; k \in K \\
& y_{j,k}x^L \le \nu_{j,k} \le y_{j,k}x^U, \;\; j \in J_k, \;\; k \in K \\
& Ay \le b \\
& \sum_{j=1}^{J_k} y_{j,k} = 1, \;\; k \in K \\
& x \in [x^L, x^U] \;\subset \mathbb{R}^m \\
& y_{j,k} \in \{0, 1\}, \;\; j \in J_k \; k \in K
\end{aligned}
\tag{HR}
$$

where v is a matrix of disaggregated variables, while the other variables, vectors and functions included in the (HR) model are the same as those described in the (GDP) and/or (BM) formulations. It has been proved that the reformulated conditional constraints, $y_{j,k}h_{j,k}(\nu_{j,k}/y_{j,k}) \leq 0$ (Lee and Grossmann, 2000) are convex in $\nu_{j,k}$, $\nu_{j,k} \in [x^L, x^U]$, if $h_{j,k}(x) \leq 0$ is convex in x, $x \in [x^L, x^U]$ (Hiriart-Urruty and Lemaréchal, 1993; Sawaya and Grossmann, 2007).

Appendix B. UNIFAC group contribution methods

These equations are proposed by Smith et al. (2001) in a form convenient for programming and they are slightly changed in order to avoid some numerical difficulties when the activity coefficients of components are calculated.

B.1. Original UNIFAC Model

Activity coefficient

$$\ln \gamma_{\text{ibu}} = \ln \gamma_{\text{ibu}}^C + \ln \gamma_{\text{ibu}}^R \tag{B.1}$$

Combinatorial part of activity coefficient

$$\ln \gamma_{\text{ibu}}^C = 1 - J_{\text{ibu}} + \ln J_{\text{ibu}} - 5q_{\text{ibu}}\left(1 - \frac{J_{\text{ibu}}}{L_{\text{ibu}}} + \ln\frac{J_{\text{ibu}}}{L_{\text{ibu}}}\right) \tag{B.2}$$

$$r_s = \sum_{k\in K} v'_{s,k}R_k \tag{B.3}$$

$$q_s = \sum_{k\in K} v'_{s,k}Q_k \tag{B.4}$$

$$J_{\text{ibu}} = \frac{r_{\text{ibu}}}{\sum_{i=1}^{N_c} r_i x_i} \tag{B.5}$$

$$L_{\text{ibu}} = \frac{q_{\text{ibu}}}{\sum_{i=1}^{N_c} q_i x_i} \tag{B.6}$$

Residual part of activity coefficient

$$\ln \gamma_{\text{ibu}}^R = q_{\text{ibu}}\left[1 - \sum_{k\in K}\left(\theta_k\frac{\beta_{\text{ibu},k}}{\omega_k} - e_{\text{ibu},k}\ln\frac{\beta_{\text{ibu},k}}{\omega_k}\right)\right] \tag{B.7}$$

$$e_{\text{ibu},k} = \frac{n'_{\text{ibu},k}Q_k}{q_{\text{ibu}}}, \quad k \in K \tag{B.8}$$

$$\beta_{\text{ibu,k}} = \sum_{m\in K}\frac{n'_{\text{ibu,m}}Q_m\psi_{m,k}}{q_{\text{ibu}}}, \quad k \in K \tag{B.9}$$

$$b_{i,k} = \sum_{m\in K} n'_{i,m}Q_m\psi_{m,k}, \quad i \in I; \ k \in K \tag{B.10}$$

$$\theta_k = \frac{\sum_{i=1}^{N_c} x_i n'_{i,k}Q_k}{\sum_{j=1}^{N_c} x_j q_j}, \quad k \in K \tag{B.11}$$

$$\omega_k = \sum_{m\in K}\theta_m\psi_{m,k}, \quad k \in K \tag{B.12}$$

$$\psi_{m,k} = \exp\left(\frac{-a_{m,k}}{T}\right), \quad m \in K; \ k \in K \tag{B.13}$$

B.2. Modified UNIFAC (Dortmund) model

Activity coefficient

$$\ln \gamma_i = \ln \gamma_i^C + \ln \gamma_i^R, \quad i \in I \tag{B.14}$$

Combinatorial part of activity coefficient

$$\ln \gamma_i^C = 1 - J_i' + \ln J_i' - 5q_i\left(1 - \frac{J_i}{L_i} + \ln\frac{J_i}{L_i}\right), \quad i \in I \tag{B.15}$$

$$r_s = \sum_{k\in K} v_{s,k}' R_k, \quad s \in S \tag{B.16}$$

$$q_s = \sum_{k\in K} v_{s,k}' Q_k, \quad s \in S \tag{B.17}$$

$$J_i = \frac{r_i}{\sum_{i=1}^{N_c} r_i x_i}, \quad i \in I \tag{B.18}$$

$$L_i = \frac{q_i}{\sum_{i=1}^{N_c} q_i x_i}, \quad i \in I \tag{B.19}$$

$$J_i' = \frac{r_i^{3/4}}{\sum_{i=1}^{N_c} r_i^{3/4} x_i}, \quad i \in I \tag{B.20}$$

Residual part of activity coefficient

$$\ln \gamma_i^R = q_i\left[1 - \sum_{k\in K}\left(\theta_k \frac{\beta_{i,k}}{\omega_k} - e_{i,k}\ln\frac{\beta_{i,k}}{\omega_k}\right)\right], \quad i \in I \tag{B.21}$$

$$e_{i,k} = \frac{n_{i,k}' Q_k}{q_i}, \quad i \in I;\ k \in K \tag{B.22}$$

$$\beta_{i,k} = \sum_{m\in K} e_{i,m}\psi_{m,k}, \quad i \in I;\ k \in K \tag{B.23}$$

$$\theta_k = \frac{\sum_{i=1}^{N_c} x_i n_{i,k}' Q_k}{\sum_{j=1}^{N_c} x_j q_j}, \quad k \in K \tag{B.24}$$

$$\omega_k = \sum_{m\in K} \theta_m \psi_{m,k}, \quad k \in K \tag{B.25}$$

$$\psi_{m,k} = \exp\left(-\frac{a_{m,k} + b_{m,k}T + c_{m,k}T^2}{T}\right), \quad m \in K;\ k \in K \tag{B.26}$$

Appendix C. Parameters of the UNIFAC model used in the case studies

C.1. Parameters for case study 1

The candidate solvents used in the first case study are included in Table C1; the number of groups of type k in ibuprofen ($v_{\mathrm{ibu,k}}'$) and in a solvent s ($v_{s,k}'$) are presented in Tables C2 and C3, respectively.

Table C1
List of candidate solvents–case study 1.

s	Compound
1	Acetone
2	Chloroform ($CHCl_3$)
3	Ethanol (EtOH)
4	Ethylacetate
5	Methanol (MeOH)
6	MIBK
7	2-Propanol
8	Toluene
9	Water

Table C2
$\nu'_{ibu,k}$, number of each group k in ibuprofen–case study 1.

$\nu'_{ibu,k}$	CH_3	CH	aCH	$aCCH_2$	aCCH	COOH
Ibuprofen	3	1	4	1	1	1

Table C3
$\nu'_{s,k}$, number of each group k in a solvent s–case study 1.

$\nu'_{s,k}$	CH_3	CH_2	CH	aCH	$aCCH_3$	CH_3OH	OH	CH_3COO	CH_3CO	$CHCl_3$	H_2O
Acetone	1								1		
Chloroform										1	
Ethanol	1	1					1				
Ethylacetate	1	1						1			
Methanol						1					
MIBK	2	1	1						1		
2-Propanol	2	1					1				
Toluene				5	1						
Water											1

C.2. Parameters for case study 2

The candidate solvents used in the second case study are presented in Table C4, whereas the number of groups of type k in acetic acid, in water and in each solvent molecule are presented in Tables C5 and C6.

Table C4
List of candidate solvents–case study 2.

s	Compound
1	Butyl acetate
2	Cyclohexyl acetate
3	Dimethyl heptene
4	Heptanol
5	Hexanone
6	Isopropyl acetate
7	Pentanol
8	Tetramethyl hexane

Table C5
$\nu'_{i,k}$, number of each group k in acetic acid ($\nu'_{a,k}$) and water ($\nu'_{w,k}$)–case study 2.

$\nu'_{i,k}$	CH_3	COOH	H_2O
Acetic acid, a	1	1	
Water, w			1

Table C6
$\nu'_{s,k}$, number of each group k in a solvent s - case study 2.

$\nu'_{s,k}$	CH_3	CH_2	CH	CH=C	CH_3CO	CH_3COO	$cyCH_2$	cyCH	OH (P)
Butyl acetate	1	3				1			
Cyclohexyl acetate						1	5	1	
Dimethyl heptene	4	2	1	1					
Heptanol	1	6							1
Hexanone	1	3			1				
Isopropyl acetate	2		1			1			
Pentanol	1	4							1
Tetramethyl hexane	6		4						

Appendix D. Problem formulations

For definition of indices and sets see Tables 2 and 4.

D.1. Case study 1: MINLP formulation via HR for the general problem ($N_{max} = 3$)

The MINLP formulation obtained using HR for the general problem of case study 1 is given below as model (G1-HR).

$$\max \quad x_{\text{ibu}}$$

$$\text{s.t.} \quad \ln x_{\text{ibu}} + \ln \gamma_{\text{ibu}} = \frac{\Delta H_{fus}}{R}\left[\frac{1}{T_m} - \frac{1}{T}\right]$$

$$\ln \gamma_{\text{ibu}} = \ln \gamma_{\text{ibu}}^{C} + \ln \gamma_{\text{ibu}}^{R}$$

$$n'_{i,k} = \sum_{s\in S} v'_{s,k} y_{i,s}, \quad i = c_1,\ c_2,\ c_3;\ k \in K$$

$$x_{\text{ibu}} \geq 0.001$$

$$\sum_{i\in I} x_i = 1$$

select only one disjunction:

$$\sum_{n=1}^{3} \tilde{y}_n = 1$$

select exactly one solvent for the first designed component:

$$\sum_{s\in S} y_{c_1,s} = 1$$

logic relations:

select at most 2 more solvents:

$$\sum_{s\in S} y_{i,s} \leq 1, \quad i = c_2, c_3$$

select each candidate solvent at most once:

$$\sum_{i\in I} y_{i,s} \leq 1, \quad s \in S;\ I = \{c_1,\ c_2,\ c_3\}$$

solvent ordering:

$$y_{c_1,s} + y_{c_2,s'} \leq 1$$

$$y_{c_1,s} + y_{c_3,s'} \leq 1, \quad \forall\, s' < s;\ s = 2, \ldots, 9;\ s' = 1, \ldots, s$$

$$y_{c_2,s} + y_{c_3,s'} \leq 1$$

using $\tilde{y}_n$ to control $y_{i,s}$:

$$\tilde{y}_1 + y_{i,s} \leq 1, \quad s \in S,\ i = c_2, c_3$$

$$\tilde{y}_2 + y_{i,s} \leq 1, \quad s \in S,\ i = c_3$$

$$\tilde{y}_2 \leq \sum_{s\in S} y_{i,s}, \quad i = c_2$$

$$\tilde{y}_3 \leq \sum_{s\in S} y_{i,s}, \quad i = c_2, c_3$$

reformulation of the linear disjunctive constraints via HR:

$$0.001\tilde{y}_n \leq \nu x_{c_1,n} \leq 0.999\tilde{y}_n, \quad n = 1, 2, 3$$

$$0.001\tilde{y}_n \leq \nu x_{c_2,n} \leq 0.999\tilde{y}_n, \quad n = 2, 3$$

$$0.001\tilde{y}_3 \leq \nu x_{c_3,3} \leq 0.999\tilde{y}_3$$

$$x_i = \sum_{n=1}^{3} \nu x_{i,n}, \quad i = c_1, c_2, c_3$$

$$\nu q_{i,n} = \tilde{y}_n \sum_{s\in S} q_s y_{i,s}, \quad i = c_1,\ c_2,\ c_3;\ n = 1, 2, 3$$

$$q_i = \sum_{n=1}^{3} \nu q_{i,n}, \quad i = c_1, c_2, c_3$$

$$\nu r_i = \tilde{y}_n \sum_{s\in S} r_s y_{i,s}, \quad i = c_1,\ c_2,\ c_3;\ n = 1, 2, 3$$

$$r_i = \sum_{n=1}^{3} \nu r_{i,n}, \quad i = c_1,\ c_2,\ c_3$$

$$\nu b_{c_1,k,n} = \tilde{y}_n \sum_{m\in K} n'_{m,c_1} Q_m \psi_{m,k}, \quad n = 2, 3;\ k \in K$$

$$\nu b_{c_2,k,3} = \tilde{y}_3 \sum_{m\in K} n'_{m,c_2} Q_m \psi_{m,k}, \quad k \in K$$

$$b_{i,k} = \sum_{n=1}^{3} \nu b_{i,k,n}, \quad i = c_1, c_2$$

$$0 \leq \nu b_{i,k,n} \leq 100\tilde{y}_n, \quad i = c_1, c_2, c_3;\ n = 1, 2, 3$$

reformulation of the nonlinear disjunctive constraints via HR:

$$A_n = (1-\epsilon)\tilde{y}_n + \epsilon, \quad n = 1, 2, 3$$
$$B_n = (1-\tilde{y}_n)\epsilon, \quad n = 1, 2, 3$$
$$\epsilon = 10^{-4}$$

miscibility for the binary pair (i, j)

$$A_n\left[\frac{\nu x_{i,n}^{i,j}}{A_n}\left(\frac{\nu x_{i,n}}{A_n} + \frac{\nu x_{j,n}}{A_n}\right) - \frac{\nu x_{i,n}}{A_n}\right] = 0, \quad i = c_1, c_2;\ j = c_2, c_3;\ i < j;\ n = 2, 3$$
$$\nu x_{i,n}^{i,j} + \nu x_{j,n}^{i,j} = \tilde{y}_n, \quad i = c_1, c_2;\ j = c_2, c_3;\ i < j;\ n = 2, 3$$
$$x_i^{i,j} = \sum_{n=1}^{3} \nu x_{i,n}^{i,j}, \quad i = c_1, c_2;\ j = c_2, c_3;\ i < j$$
$$\nu d\gamma_{i,n}^{i,j} - (\nu d\gamma_{i,n}^{i,j})^C - (\nu d\gamma_{i,n}^{i,j})^R = 0, \quad i = c_1, c_2;\ j = c_2, c_3;\ i < j;\ n = 2, 3$$
$$d\gamma_i^{i,j} = \sum_{n=1}^{3} \nu d\gamma_{i,n}^{i,j}, \quad i = c_1, c_2;\ j = c_2, c_3;\ i < j$$
$$(d\gamma_i^{i,j})^C = \sum_{n=1}^{3} (\nu d\gamma_{i,n}^{i,j})^C, \quad i = c_1, c_2;\ j = c_2, c_3;\ i < j$$
$$(d\gamma_i^{i,j})^R = \sum_{n=1}^{3} (\nu d\gamma_{i,n}^{i,j})^R, \quad i = c_1, c_2;\ j = c_2, c_3;\ i < j$$
$$A_n\left[\frac{\nu d\gamma_{i,n}^{i,j}}{A_n}\frac{\nu x_{i,n}^{i,j}}{A_n} + 1\right] - B_n \geq 0, \quad i = c_1, c_2;\ j = c_2, c_3;\ i < j;\ n = 2, 3$$
$$0 \leq \nu x_{i,n}^{i,j} \leq \tilde{y}_n, \quad i = c_1, c_2;\ j = c_2, c_3;\ i < j;\ n = 2, 3$$
$$-100\tilde{y}_n \leq \nu d\gamma_{i,n}^{i,j}, (\nu d\gamma_{i,n}^{i,j})^C, (\nu d\gamma_{i,n}^{i,j})^R \leq 100\tilde{y}_n, \quad i = c_1, c_2;\ j = c_2, c_3;\ i < j;\ n = 2, 3$$
$$0.001 \leq x_{\text{ibu}} \leq 1$$
$$0 \leq x_i \leq 1, \quad i = c_1, c_2, c_3$$
$$y_{i,s}, \tilde{y}_n \in \{0, 1\}, \quad i = c_1, c_2, c_3;\ s \in S;\ n \in N \tag{G1-HR}$$

where $d\gamma_i^{i,j}$ is the partial derivative of the natural logarithm of the activity coefficient of component i with respect to the mole fraction of i in the binary mixture, i.e. $d\gamma_i^{i,j} = \frac{\partial \ln\gamma_i^{i,j}}{\partial x_i^{i,j}}$, and it is calculated from the UNIFAC model (i. e. $d\gamma_i^{i,j} = (d\gamma_i^{i,j})^C + (d\gamma_i^{i,j})^R$, with $(d\gamma_i^{i,j})^C = \frac{\partial (\ln\gamma_i^{i,j})^C}{\partial x_i^{i,j}}$ and $(d\gamma_i^{i,j})^R = \frac{\partial (\ln\gamma_i^{i,j})^R}{\partial x_i^{i,j}}$).

D.2. Case study 2: MINLP formulation for the restricted problem (N=3)

The MINLP formulation for the restricted problem of case study 2 is given below as model (R2-MINLP).

$$\min \quad \frac{F_S}{F} + x_{E,w} - x_{E,a}$$

s. t. phase equilibrium

$$\gamma_{E,i} x_{E,i} = \gamma_{R,i} x_{R,i}, \quad i \in I$$

mole balances

$$x_{F,a} F = x_{M,a} M$$
$$x_{F,w} F = x_{M,w} M$$
$$x_{F_S,ii} F_S = x_{M,ii} M, \quad ii = c_1, c_2, c_3$$
$$x_{M,i} M = x_{E,i} E + x_{R,i} R, \quad i \in I$$
$$x_{R,a} \leq 0.003$$
$$\sum_{ii \in II} x_{S,ii} = 1$$
$$\sum_{i \in I} x_{p,i} = 1, \quad p = M, E, R$$

activity coefficients

$$\ln \gamma_{p,i} = \ln \gamma_{p,i}^C + \ln \gamma_{p,i}^R, \quad p = E, R;\ i \in I$$
$$n'_{i,k} = \sum_{s \in S} v'_{s,k}\, y_{i,s}, \quad i = c_1, c_2, c_3;\ k \in K$$
$$q_i = \sum_{s \in S} q_s\, y_{i,s}, \quad i = c_1, c_2, c_3$$
$$r_i = \sum_{s \in S} r_s\, y_{i,s}, \quad i = c_1, c_2, c_3$$

select exactly 3 solvent molecules:

$\sum_{s\in S} y_{ii,s} = 1, \;\; ii = c_1, c_2, c_3$

logic relations:

select each candidate solvent at most once:

$\sum_{ii\in II} y_{ii,s} \le 1, \quad s \in S$

solvent ordering:

$y_{c_1,s} + y_{c_2,s'} \le 1$

$y_{c_1,s} + y_{c_3,s'} \le 1, \;\; \forall\, s' < s; \;\; s = 2, \ldots, 9; \;\; s' = 1, \ldots, s$

$y_{c_2,s} + y_{c_3,s'} \le 1$

$0.001 \le x_{p,i} \le 1\,, \quad i \in I; \;\; p = F_S, M$

$10^{-6} \le x_{p,i} \le 1, \quad i \in I; \;\; p = E, R$

$y_{i,s} \in \{0, 1\}, \;\; i = c_1, c_2, c_3; \;\; s \in S$ (R2-MINLP)

D.3. Case study 2: Disjunctions for number of solvents for the general problem (N_{max}=3)

The disjunctions for number of solvents for the general problem of case study 2 is given as model (G2-D).

$$
\begin{bmatrix}
\tilde{Y}_1 \\
x_{F_S,c_1} = 1 \\
x_{p,c_2} = x_{p,c_3} = 0 \\
\text{for } p = F_S, M, E, R \\
\ln\gamma_{p,c_2} = \ln\gamma_{p,c_3} = 0 \\
\ln\gamma^C_{p,c_2} = \ln\gamma^C_{p,c_3} = 0 \\
\ln\gamma^R_{p,c_2} = \ln\gamma^R_{p,c_3} = 0 \\
\text{for } p = E, R \\
J_{p,c_2} = J_{p,c_3} = 0 \\
J'_{p,c_2} = J'_{p,c_3} = 0 \\
L_{p,c_2} = L_{p,c_3} = 0 \\
\text{for } p = E, R \\
e_{c_2,k} = e_{c_3,k} = 0 \\
\beta_{c_2,k} = \beta_{c_3,k} = 0 \\
\text{for } k \in K
\end{bmatrix}
\vee
\begin{bmatrix}
\tilde{Y}_2 \\
x_{F_S,c_1} \ge 0.001 \\
0.001 \le x_{p,c_2} \le 0.999 \\
\text{for } p = F_S, M \\
0.000001 \le x_{p,c_2} \le 0.999999 \\
\text{for } p = E, R \\
x_{p,c_3} = 0 \\
\text{for } p = F_S, M, E, R \\
x_{F_S,c_2} F_S = x_{M,c_2} M \\
x_{M,c_2} M = x_{E,c_2} E + x_{R,c_2} R \\
\ln\gamma_{p,c_2} = \ln\gamma^C_{p,c_2} + \ln\gamma^R_{p,c_2} \\
\ln\gamma^C_{p,c_2} = 1 - J'_{p,c_2} + \ln J'_{p,c_2} \\
\quad -5q_{c_2}\left(1 - \frac{J_{p,c_2}}{L_{p,c_2}} + \ln\frac{J_{p,c_2}}{L_{p,c_2}}\right) \\
\ln\gamma^R_{p,c_2} = q_{c_2}\Big[1 - \sum_{k\in K}\Big(\theta_k \frac{\beta_{p,c_2,k}}{\omega_k} \\
\quad -e_{p,c_2,k}\ln\frac{\beta_{p,c_2,k}}{\omega_k}\Big)\Big] \\
\ln\gamma_{p,c_3} = \ln\gamma^C_{p,c_3} = \ln\gamma^R_{p,c_3} = 0 \\
\text{for } p = E, R \\
J_{p,c_2} = \frac{r_{c_2}}{\sum_{i=1}^{N_c} r_i x_{p,i}} \\
J'_{p,c_2} = \frac{r_{c_2}^{3/4}}{\sum_{i=1}^{N_c} r_i^{3/4} x_{p,i}} \\
L_{p,c_2} = \frac{q_{c_2}}{\sum_{i=1}^{N_c} q_i x_{p,i}} \\
J_{p,c_3} = J'_{p,c_3} = L_{p,c_3} = 0 \\
\text{for } p = E, R \\
e_{c_2,k} = \frac{n'_{c_2,k} Q_k}{q_{c_2}} \\
\beta_{c_2,k} = \sum_{m\in K} e_{c_2,m}\psi_{m,k} \\
e_{c_3,k} = \beta_{c_3,k} = 0 \\
\text{for } k \in K
\end{bmatrix}
\vee
\begin{bmatrix}
\tilde{Y}_3 \\
x_{F_S,c_1} \ge 0.001 \\
0.001 \le x_{p,i} \le 0.999 \\
\text{for } p = F_S, M\;;\; i = c_2, c_3 \\
0.000001 \le x_{p,i} \le 0.999999 \\
\text{for } p = E, R\;;\; i = c_2, c_3 \\
x_{F_S,i} F_S = x_{M,i} M \\
x_{M,i} M = x_{E,i} E + x_{R,i} R \\
\text{for } i = c_2, c_3 \\
\ln\gamma_{p,i} = \ln\gamma^C_{p,i} + \ln\gamma^R_{p,i} \\
\ln\gamma^C_{p,i} = 1 - J'_{p,i} + \ln J'_{p,i} \\
\quad -5q_i\left(1 - \frac{J_{p,i}}{L_{p,i}} + \ln\frac{J_{p,i}}{L_{p,i}}\right) \\
\ln\gamma^R_{p,i} = q_i\Big[1 - \sum_{k\in K}\Big(\theta_k \frac{\beta_{p,i,k}}{\omega_k} \\
\quad -e_{p,i,k}\ln\frac{\beta_{p,i,k}}{\omega_k}\Big)\Big] \\
\text{for } p = E, R\;;\; i = c_2, c_3 \\
J_{p,i} = \frac{r_i}{\sum_{i=1}^{N_c} r_i x_{p,i}} \\
J'_{p,i} = \frac{r_i^{3/4}}{\sum_{i=1}^{N_c} r_i^{3/4} x_{p,i}} \\
L_{p,i} = \frac{q_i}{\sum_{i=1}^{N_c} q_i x_{p,i}} \\
\text{for } p = E, R\;;\; i = c_2, c_3 \\
e_{i,k} = \frac{n'_{i,k} Q_k}{q_i} \\
\beta_{i,k} = \sum_{m\in K} e_{i,m}\psi_{m,k} \\
\text{for } i = c_2, c_3\;;\; k \in K
\end{bmatrix}
\qquad \text{(G2-D)}
$$

D.4. Case study 2: MINLP formulation via BM for the general problem ($N_{max} = 3$)

The MINLP formulation obtained using BM for the general problem of case study 2 is given below as model (G2-BM).

$$\min \ \frac{F_S}{F} + x_{E,w} - x_{E,a}$$

s. t. phase equilibrium

$$\gamma_{E,i} x_{E,i} = \gamma_{R,i} x_{R,i}, \ \ i = a, w, c_1$$

mole balances

$$x_{F,a} F = x_{M,a} M$$
$$x_{F,w} F = x_{M,w} M$$
$$x_{F_S,c_1} F_S = x_{M,c_1} M$$
$$x_{M,i} M = x_{E,i} E + x_{R,i} R, \ \ i = a, w, c_1$$
$$x_{R,a} \leq 0.003$$
$$\sum_{ii \in II} x_{F_S,ii} = 1$$
$$\sum_{i \in I} x_{p,i} = 1, \ \ p = M, E, R$$

activity coefficients − UNIFAC equations

$$\ln \gamma_{p,i} = \ln \gamma^C_{p,i} + \ln \gamma^R_{p,i}$$
$$\ln \gamma^C_{p,i} = 1 - J'_{p,i} + \ln J'_{p,i} - 5q_i\left(1 - \frac{J_{p,i}}{L_{p,i}} + \ln\frac{J_{p,i}}{L_{p,i}}\right)$$
$$\ln \gamma^R_{p,i} = q_i\left[1 - \sum_{k \in K}\left(\theta_k \frac{\beta_{p,i,k}}{\omega_k} - e_{p,i,k} \ln\frac{\beta_{p,i,k}}{\omega_k}\right)\right]$$

for $p = E, R; \ \ i = a, w, c_1$

$$n'_{i,k} = \sum_{s \in S} v'_{s,k} y_{i,s}, \ \ i \in I; \ \ k \in K$$
$$q_i = \sum_{s \in S} q_s y_{i,s}, \ \ i \in I$$
$$r_i = \sum_{s \in S} r_s y_{i,s}, \ \ i \in I$$
$$J_{p,i} = \frac{r_i}{\sum_{i=1}^{N_C} r_i x_{p,i}}, \ \ p = E, R \ i = a, w, c_1$$
$$J'_{p,i} = \frac{r_i^{3/4}}{\sum_{i=1}^{N_C} r_i^{3/4} x_{p,i}}, \ \ p = E, R \ i = a, w, c_1$$
$$L_{p,i} = \frac{q_i}{\sum_{i=1}^{N_C} q_i x_{p,i}}, \ \ p = E, R \ i = a, w, c_1$$
$$e_{i,k} = \frac{n'_{i,k} Q_k}{q_i}, \ \ i = a, w, c_1 \ k \in K$$
$$\beta_{i,k} = \sum_{m \in K} e_{i,m} \psi_{m,k}, \ \ i = a, w, c_1 \ k \in K$$

select only one disjunction:

$$\sum_{n=1}^{3} \tilde{y}_n = 1$$

select exactly one solvent for the first designed component:

$$\sum_{s \in S} y_{c_1,s} = 1$$

logic relations:

select at most2more solvents:

$$\sum_{s \in S} y_{i,s} \leq 1, \ \ i = c_2, c_3$$

select each candidate solvent at most once:

$$\sum_{ii \in II} y_{ii,s} \leq 1, \ \ s \in S$$

solvent ordering:

$$y_{c_1,s} + y_{c_2,s'} \leq 1$$
$$y_{c_1,s} + y_{c_3,s'} \leq 1, \ \ \forall \ s' < s \ \ s = 2, \ldots, 9 \ \ s' = 1, \ldots, s$$
$$y_{c_2,s} + y_{c_3,s'} \leq 1$$

using $\tilde{y}_n$ to control$y_{i,s}$:

$$\tilde{y}_1 + y_{i,s} \leq 1, \quad s \in S, \ \ i = c_2, c_3$$
$$\tilde{y}_2 + y_{i,s} \leq 1, \quad s \in S, \ \ i = c_3$$
$$\tilde{y}_2 \leq \sum_{s \in S} y_{i,s}, \ \ i = c_2$$
$$\tilde{y}_3 \leq \sum_{s \in S} y_{i,s}, \ \ i = c_2, c_3$$

reformulation of the disjunctive constraints via BM:

mole fractions

$$x_{F_S,c_1} \geq 0.001 + 0.999\tilde{y}_1$$
$$0.001(1-\tilde{y}_1) \leq x_{p,c_2} \leq 0.999(1-\tilde{y}_1),\ \ p = F_S, M$$
$$0.001\tilde{y}_3 \leq x_{p,c_3} \leq 0.999\tilde{y}_3,\ \ p = F_S, M$$
$$0.000001(1-\tilde{y}_1) \leq x_{p,c_2} \leq 0.999999(1-\tilde{y}_1),\ \ p = E, R$$
$$0.000001\tilde{y}_3 \leq x_{p,c_3} \leq 0.999999\tilde{y}_3,\ \ p = E, R$$

phase equilibrium

$$-9999\tilde{y}_1 \leq \gamma_{E,c_2}x_{E,c_2} - \gamma_{R,c_2}x_{R,c_2} \leq 9999\tilde{y}_1$$
$$-9999(1-\tilde{y}_3) \leq \gamma_{E,c_3}x_{E,c_3} - \gamma_{R,c_3}x_{R,c_3} \leq 9999(1-\tilde{y}_3)$$

mole balances

$$-22708\tilde{y}_1 \leq x_{F_S,c_2}F_S - x_{M,c_2}M \leq 22708\tilde{y}_1$$
$$-22708(1-\tilde{y}_3) \leq x_{F_S,c_3}F_S - x_{M,c_3}M \leq 22708(1-\tilde{y}_3)$$
$$-22708\tilde{y}_1 \leq x_{M,c_2}M - x_{E,c_2}E - x_{R,c_2}R \leq 22708\tilde{y}_1$$
$$-22708(1-\tilde{y}_3) \leq x_{M,c_3}M - x_{E,c_3}E - x_{R,c_3}R \leq 22708(1-\tilde{y}_3)$$

activity coefficients − UNIFAC equations

$$-9999\tilde{y}_1 \leq \ln\gamma_{p,c_2} - \ln\gamma^C_{p,c_2} - \ln\gamma^R_{p,c_2} \leq 9999\tilde{y}_1,\ \ p = E, R$$
$$-9999(1-\tilde{y}_3) \leq \ln\gamma_{p,c_3} - \ln\gamma^C_{p,c_3} - \ln\gamma^R_{p,c_3} \leq 9999(1-\tilde{y}_3),\ \ p = E, R$$
$$-9999\tilde{y}_1 \leq \ln\gamma^C_{p,c_2} - 1 + J'_{p,c_2} - \ln J'_{p,c_2} + 5q_{c_2}\left(1 - \frac{J_{p,c_2}}{L_{p,c_2}} + \ln\frac{J_{p,c_2}}{L_{p,c_2}}\right) \leq 9999\tilde{y}_1,\ \ p = E, R$$
$$-9999(1-\tilde{y}_3) \leq \ln\gamma^C_{p,c_3} - 1 + J'_{p,c_3} - \ln J'_{p,c_3} + 5q_{c_3}\left(1 - \frac{J_{p,c_3}}{L_{p,c_3}} + \ln\frac{J_{p,c_3}}{L_{p,c_3}}\right) \leq 9999(1-\tilde{y}_3),\ \ p = E, R$$
$$-9999\tilde{y}_1 \leq \ln\gamma^R_{p,c_2} - q_{c_2}\left[1 - \sum_{k\in K}\left(\theta_k\frac{\beta_{p,c_2,k}}{\omega_k} - e_{p,c_2,k}\ln\frac{\beta_{p,c_2,k}}{\omega_k}\right)\right] \leq 9999\tilde{y}_1,\ \ p = E, R$$
$$-9999(1-\tilde{y}_3) \leq \ln\gamma^R_{p,c_3} - q_{c_3}\left[1 - \sum_{k\in K}\left(\theta_k\frac{\beta_{p,c_3,k}}{\omega_k} - e_{p,c_3,k}\ln\frac{\beta_{p,c_3,k}}{\omega_k}\right)\right] \leq 9999(1-\tilde{y}_3),\ \ p = E, R$$
$$-10\tilde{y}_1 \leq J_{p,c_2} - \frac{r_{c_2}}{\sum_{i=1}^{N_c} r_i x_{p,i}} \leq 100\tilde{y}_1,\ \ p = E, R$$
$$-10(1-\tilde{y}_3) \leq J_{p,c_3} - \frac{r_{c_3}}{\sum_{i=1}^{N_c} r_i x_{p,i}} \leq 100(1-\tilde{y}_3),\ \ p = E, R$$
$$-10\tilde{y}_1 \leq J'_{p,c_2} - \frac{r_{c_2}^{3/4}}{\sum_{i=1}^{N_c} r_i^{3/4} x_{p,i}} \leq 100\tilde{y}_1,\ \ p = E, R$$
$$-10(1-\tilde{y}_3) \leq J'_{p,c_3} - \frac{r_{c_3}^{3/4}}{\sum_{i=1}^{N_c} r_i^{3/4} x_{p,i}} \leq 100(1-\tilde{y}_3),\ \ p = E, R$$
$$-10\tilde{y}_1 \leq L_{p,c_2} - \frac{q_{c_2}}{\sum_{i=1}^{N_c} q_i x_{p,i}} \leq 100\tilde{y}_1,\ \ p = E, R$$
$$-10(1-\tilde{y}_3) \leq L_{p,c_3} - \frac{q_{c_3}}{\sum_{i=1}^{N_c} q_i x_{p,i}} \leq 100(1-\tilde{y}_3),\ \ p = E, R$$
$$-10\tilde{y}_1 \leq e_{c_2,k} - \frac{n'_{c_2,k}Q_k}{q_{c_2}} \leq 10\tilde{y}_1,\ \ k \in K$$
$$-10(1-\tilde{y}_3) \leq e_{c_3,k} - \frac{n'_{c_3,k}Q_k}{q_{c_3}} \leq 10(1-\tilde{y}_3),\ \ k \in K$$
$$-100\tilde{y}_1 \leq \beta_{c_2,k} - \textstyle\sum_{m\in K} e_{c_2,m}\psi_{m,k} \leq 100\tilde{y}_1,\ \ k \in K$$
$$-100(1-\tilde{y}_3) \leq \beta_{c_3,k} - \textstyle\sum_{m\in K} e_{c_3,m}\psi_{m,k} \leq 100(1-\tilde{y}_3),\ \ k \in K$$
$$0.001 \leq x_{p,i} \leq 1,\ \ i = a, w, c_1;\ \ p = F_S, M$$
$$10^{-6} \leq x_{p,i} \leq 1,\ \ i = a, w, c_1;\ \ p = E, R$$
$$0 \leq x_{p,i} \leq 1,\ \ i = c_2, c_3;\ \ p = F_S, M, E, R$$
$$y_{i,s}, \tilde{y}_n \in \{0, 1\},\ \ i = c_1, c_2, c_3;\ \ s \in S;\ \ n \in N \qquad \text{(G2-BM)}$$

D.5. Case study 2: MINLP formulation via HR for the general problem ($N_{max} = 3$)

The MINLP formulation obtained using HR for the general problem of case study 2 is given below as model (G2-HR).

$$\min \ \frac{F_S}{F} + x_{E,w} - x_{E,a}$$

s. t. phase equilibrium

$$\gamma_{E,i}x_{E,i} = \gamma_{R,i}x_{R,i}, \ \ i = a, w$$

mole balances

$$x_{F,a}F = x_{M,a}M$$

$$x_{F,w}F = x_{M,w}M$$

$$x_{M,i}M = x_{E,i}E + x_{R,i}R, \ \ i = a, w$$

$$x_{R,a} \leq 0.003$$

$$\sum_{ii \in II} x_{F_S,ii} = 1$$

$$\sum_{i \in I} x_{p,i} = 1, \ \ p = M, E, R$$

activity coefficients − UNIFAC equations

$$\ln\gamma_{p,i} = \ln\gamma_{p,i}^{C} + \ln\gamma_{p,i}^{R}$$

$$\ln\gamma_{p,i}^{C} = 1 - J'_{p,i} + \ln J'_{p,i} - 5q_i\left(1 - \frac{J_{p,i}}{L_{p,i}} + \ln\frac{J_{p,i}}{L_{p,i}}\right)$$

$$\ln\gamma_{p,i}^{R} = q_i\left[1 - \sum_{k \in K}\left(\theta_k\frac{\beta_{p,i,k}}{\omega_k} - e_{p,i,k}\ln\frac{\beta_{p,i,k}}{\omega_k}\right)\right]$$

for $p = E, R; \ \ i = a, w$

$$n'_{i,k} = \sum_{s \in S} v'_{s,k}y_{i,s}, \ \ i \in I; \ \ k \in K$$

$$J_{p,i} = \frac{r_i}{\sum_{i=1}^{N_c} r_i x_{p,i}}, \ \ p = E, R; \ \ i = a, w$$

$$J'_{p,i} = \frac{r_i^{3/4}}{\sum_{i=1}^{N_c} r_i^{3/4} x_{p,i}}, \ \ p = E, R; \ \ i = a, w$$

$$L_{p,i} = \frac{q_i}{\sum_{i=1}^{N_c} q_i x_{p,i}}, \ \ p = E, R; \ \ i = a, w$$

$$e_{i,k} = \frac{n'_{i,k}Q_k}{q_i}, \ \ i = a, w; \ \ k \in K$$

$$\beta_{i,k} = \sum_{m \in K} e_{i,m}\psi_{m,k}, \ \ i = a, w; \ \ k \in K$$

select only one disjunction:

$$\sum_{n=1}^{3} \tilde{y}_n = 1$$

select exactly one solvent for the first designed component:

$$\sum_{s \in S} y_{c_1,s} = 1$$

logic relations:

select at most 2 more solvents:

$$\sum_{s \in S} y_{i,s} \leq 1 \ \ i = c_2, c_3$$

select each candidate solvent at most once:

$$\sum_{ii \in II} y_{ii,s} \leq 1, \ s \in S$$

solvent ordering:

$$y_{c_1,s} + y_{c_2,s'} \leq 1$$

$$y_{c_1,s} + y_{c_3,s'} \leq 1 \quad \forall \ \ s' < s; \ \ s = 2, \ldots, 9; \ \ s' = 1, \ldots, s$$

$$y_{c_2,s} + y_{c_3,s'} \leq 1$$

using $\tilde{y}_n$ tocontrol $y_{i,s}$:

$$\tilde{y}_1 + y_{i,s} \leq 1, \ \ s \in S; \quad i = c_2, c_3$$

$$\tilde{y}_2 + y_{i,s} \leq 1 \ \ s \in S; \quad i = c_3$$

$$\tilde{y}_2 \leq \sum_{s \in S} y_{i,s}, \ \ i = c_2$$

$$\tilde{y}_3 \leq \sum_{s \in S} y_{i,s} \quad i = c_2, c_3$$

reformulation of the linear disjunctive constraints via HR:

mole fractions

$0.001\tilde{y}_n \le \nu x_{p,c_1,n} \le 0.999\tilde{y}_n,\ \ n = 1, 2, 3;\ \ p = F_S, M$

$0.000001\tilde{y}_n \le \nu x_{p,c_1,n} \le 0.999999\tilde{y}_n,\ \ n = 1, 2, 3;\ \ p = E, R$

$0.001\tilde{y}_n \le \nu x_{p,c_2,n} \le 0.999\tilde{y}_n,\ \ n = 2, 3;\ \ p = F_S, M$

$0.000001\tilde{y}_n \le \nu x_{p,c_2,n} \le 0.999999\tilde{y}_n,\ \ n = 2, 3;\ \ p = E, R$

$0.001\tilde{y}_3 \le \nu x_{p,c_3,3} \le 0.999\tilde{y}_3,\ \ p = F_S, M$

$0.000001\tilde{y}_3 \le \nu x_{p,c_3,3} \le 0.999999\tilde{y}_3,\ \ p = E, R$

$x_{p,i} = \sum_{n=1}^{3} \nu x_{p,i,n},\ \ i = c_1, c_2, c_3;\ \ p = F_S, M, E, R$

activity coefficients − UNIFAC equations

$\nu ln\gamma_{p,i,n} - \nu ln\gamma^C_{p,i,n} - \nu ln\gamma^R_{p,i,n}\ p = E, R;\ \ i = c_1, c_2, c_3;\ \ n = 1, 2, 3$

$ln\gamma_{p,i} = \sum_{n=1}^{3} \nu ln\gamma_{p,i,n},\ \ p = E, R;\ \ i = c_1, c_2, c_3$

$-99\tilde{y}_n \le \nu ln\gamma_{p,i,n} \le 99\tilde{y}_n,\ \ p = E, R;\ \ i = c_1, c_2, c_3;\ \ n = 1, 2, 3$

$\nu q_{i,n} = \tilde{y}_n \sum_{s\in S} q_s y_{i,s},\ \ i = c_1, c_2, c_3;\ \ n = 1, 2, 3$

$q_i = \sum_{n=1}^{3} \nu q_{i,n},\ \ i = c_1, c_2, c_3$

$0 \le \nu q_{i,n} \le 20\tilde{y}_n,\ \ i = c_1, c_2, c_3;\ \ n = 1, 2, 3$

$\nu r_i = \tilde{y}_n \sum_{s\in S} r_s y_{i,s},\ \ i = c_1, c_2, c_3;\ \ n = 1, 2, 3$

$r_i = \sum_{n=1}^{3} \nu r_{i,n},\ \ i = c_1, c_2, c_3$

$0 \le \nu r_{i,n} \le 20\tilde{y}_n,\ \ i = c_1, c_2, c_3;\ \ n = 1, 2, 3$

$\nu\beta_{i,k,n} - \sum_{m\in K} \nu e_{i,m,n}\psi_{m,k} = 0,\ \ i = c_1, c_2, c_3;\ \ n = i, 2, 3;\ \ k \in K$

$\beta_{i,k} = \sum_{n=1}^{3} \nu\beta_{i,k,n},\ \ i = c_1, c_2, c_3;\ \ k \in K$

$-50\tilde{y}_n \le \nu\beta_{i,k,n} \le 50\tilde{y}_n,\ \ i = c_1, c_2, c_3;\ \ n = 1, 2, 3;\ \ k \in K$

reformulation of the nonlinear disjunctive constraints via HR:

$A_n = (1 - \epsilon)\tilde{y}_n + \epsilon,\ \ n = 1, 2, 3$

$B_n = (1 - \tilde{y}_n)\epsilon,\ \ n = 1, 2, 3$

$\epsilon = 10^{-5}$

phase equilibrium

$\nu\gamma_{E,i,n}\nu x_{E,i,n} - \nu\gamma_{R,i,n}\nu x_{R,i,n} = 0,\ \ i = c_1, c_2, c_3;\ \ n = 1, 2, 3$

$-99\tilde{y}_n \le \nu\gamma_{p,i,n} \le 9999\tilde{y}_n,\ \ i = c_1, c_2, c_3;\ \ p = E, R;\ \ n = 1, 2, 3$

mole balances

$\nu x_{F_S,i,n}F_S - \nu x_{M,i,n}M = 0,\ \ i = c_1, c_2, c_3;\ \ n = 1, 2, 3$

$\nu x_{M,i,n}M - \nu x_{E,i,n}E - \nu x_{R,i,n}R\ i = c_1, c_2, c_3;\ \ n = 1, 2, 3$

activity coefficients − UNIFAC equations

$$A_n\left[\frac{\nu ln\gamma^C_{p,i,n}}{A_n} - 1 + \frac{\nu J'_{p,i,n}}{A_n} - \ln\frac{\nu J'_{p,i,n}}{A_n} + 5\frac{\nu q_{i,n}}{A_n}\left(1 - \frac{\nu J_{p,i,n}}{\nu L_{p,i,n}} + \ln\frac{\nu J_{p,i,n}}{\nu L_{p,i,n}}\right)\right] + B_n = 0$$

for $p = E, R;\ \ i = c_1, c_2, c_3;\ \ n = 1, 2, 3$

$ln\ \gamma^C_{p,i} = \sum_{n=1}^{3} \nu ln\gamma^C_{p,i,n},\ \ p = E, R;\ \ i = c_1, c_2, c_3$

$-99\tilde{y}_n \le \nu ln\gamma^C_{p,i,n} \le 99\tilde{y}_n,\ \ p = E, R;\ \ i = c_1, c_2, c_3;\ \ n = 1, 2, 3$

$$A_n\left[\frac{\nu ln\gamma^R_{p,i,n}}{A_n} - \frac{\nu q_{i,n}}{A_n}\left(1 - \sum_{k\in K}\left(\frac{\theta_k}{\omega_k}\frac{\nu\beta_{p,i,k,n}}{A_n} - \frac{\nu e_{p,i,k,n}}{A_n}\ln\frac{\nu\beta_{p,i,k,n}}{A_n\omega_k}\right)\right)\right] = 0$$

for $p = E, R;\ \ i = c_1, c_2, c_3;\ \ n = 1, 2, 3$

$ln\gamma^R_{p,i} = \sum_{n=1}^{3} \nu ln\gamma^R_{p,i,n},\ \ p = E, R;\ \ i = c_1, c_2, c_3$

$-99\tilde{y}_n \le \nu ln\gamma^R_{p,i,n} \le 99\tilde{y}_n,\ \ p = E, R\ i = c_1, c_2, c_3\ n = 1, 2, 3$

$\nu J_{p,i,n} - \frac{\nu r_{i,n}}{\sum_{i=1}^{N_c} r_i x_{p,i}} = 0,\ \ p = E, R;\ \ i = c_1, c_2, c_3;\ \ n = 1, 2, 3$

$$
\begin{aligned}
& J_{p,i} = \sum_{n=1}^{3} \nu J_{p,i,n}, \;\; p = E, R; \;\; i = c_1, c_2, c_3 \\
& 0 \le \nu J_{p,i,n} \le 20\tilde{y}_n, \;\; p = E, R; \;\; i = c_1, c_2, c_3; \;\; n = 1, 2, 3 \\
& \nu J'_{p,i,n} - \frac{\nu r_{i,n}^{3/4}}{\sum_{i=1}^{N_c} r_i^{3/4} x_{p,i}} = 0, \;\; p = E, R; \;\; i = c_1, c_2, c_3; \;\; n = 1, 2, 3 \\
& J'_{p,i} = \sum_{n=1}^{3} \nu J'_{p,i,n}, \;\; p = E, R; \;\; i = c_1, c_2, c_3 \\
& 0 \le \nu J'_{p,i,n} \le 20\tilde{y}_n, \;\; p = E, R; \;\; i = c_1, c_2, c_3; \;\; n = 1, 2, 3 \\
& \nu L_{p,i,n} - \frac{\nu q_{i,n}}{\sum_{i=1}^{N_c} q_i x_{p,i}} = 0, \;\; p = E, R; \;\; i = c_1, c_2, c_3; \;\; n = 1, 2, 3 \\
& L_{p,i} = \sum_{n=1}^{3} \nu L_{p,i,n}, \;\; p = E, R; \;\; i = c_1, c_2, c_3 \\
& 0 \le \nu L_{p,i,n} \le 20\tilde{y}_n, \;\; p = E, R; \;\; i = c_1, c_2, c_3; \;\; n = 1, 2, 3 \\
& A_n \left[\frac{\nu e_{i,k,n}}{A_n} \frac{\nu q_{i,n}}{A_n} - n'_{i,k} Q_k \right] + B_n n'_{i,k} Q_k = 0, \;\; i = c_1, c_2, c_3; \;\; k \in K; \;\; n = 1, 2, 3 \\
& e_{i,k} = \sum_{n=1}^{3} \nu e_{i,k,n}, \;\; i = c_1, c_2, c_3; \;\; k \in K \\
& 0 \le \nu e_{i,k,n} \le 5\tilde{y}_n, \;\; i = c_1, c_2, c_3; \;\; k \in K; \;\; n = 1, 2, 3 \\
& 0.001 \le x_{p,i} \le 1 \;\; i = a, w; \;\; p = F_S, M \\
& 10^{-6} \le x_{p,i} \le 1, \;\; i = a, w; \;\; p = E, R \\
& 0 \le x_{p,i} \le 1, \;\; i = c_1, c_2, c_3; \;\; p = F_S, M, E, R \\
& y_{i,s}, \tilde{y}_n \in \{0, 1\}, \;\; i = c_1, c_2, c_3; \;\; s \in S; \;\; n \in N
\end{aligned}
\tag{G2-HR}
$$

Although variables that correspond to the first designed component, c_1, can be treated as general ones, they are disaggregated when formulating via HR for consistent notation. In model (G2-HR), the approximation function of Eq. (4) is used when logarithms are involved in order to avoid numerical difficulties.

References

Achenie, L.E.K., Gani, R., Venkatasubramanian, V., 2003. Computer Aided Molecular Design: Theory and Practice. Elsevier, Amsterdam,, The Netherlands.

Agreda, V.H., Zoeller, J.R., 1993. Acetic Acid and Its Derivatives. Springer-Verlag Berlin Heidelberg, New York.

Akula, P.T., 2011. Solvent mixture design problem (Ph.D. thesis). Imperial College, London.

Alkaya, E., Kaptan, S., Ozkan, L., Uludag-Demirer, S., Demirer, G.N., 2009. Recovery of acids from anaerobic acidification broth by liquid–liquid extraction. Chemosphere 77, 1137–1142.

Balas, E., 1985. Disjunctive programming and a hierarchy of relaxations for discrete optimization problems. SIAM J. Algebra. Discret. Methods 6, 466–486.

Beaumont, N., 1991. An algorithm for disjunctive programs. Eur. J. Oper. Res. 48, 362–371.

Brignole, E.A., Gani, R., 1983. Molecular design of solvents for liquid extraction based on UNIFAC. Fluid Phase Equilibria 13, 331–340.

Burger, J., Papaioannou, V., Gopinath, S., Jackson, G., Galindo, A., Adjiman, C.S., 2015. A hierarchical method to integrated solvent and process design of physical CO_2 absorption using the SAFT-γ Mie approach. AIChE J. 61, 3249–3269.

Buxton, A., Livingston, A.G., Pistikopoulos, E.N., 1999. Optimal design of solvent blends for environmental impact minimization. AIChE J. 45, 817–843.

Castro, P.M., Grossmann, I.E., 2012. Generalized disjunctive programming as a systematic modeling framework to derive scheduling formulations. Ind. Eng. Chem. Res. 51, 5781–5792.

Castro, P.M., Grossmann, I.E., Veldhuizen, P., Esplin, D., 2014. Optimal maintenance scheduling of a gas engine power plant using generalized disjunctive programming. AIChE J. 60, 2083–2097.

Cheung, H., Tanke, R.S., Torrence, G.P., 2012. Acetic Acid, in: Ullmann's Encyclopedia of Industrial Chemistry. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 209–237.

Choi, W.S., Kim, K.J., 2013. Separation of acetic acid from acetic acid–water mixture by crystallization. Sep. Sci. Technol. 48, 1056–1061.

Churi, N., Achenie, L.E.K., 1997. The optimal design of refrigerant mixtures for a two-evaporator refrigeration system. Comput. Chem. Eng. 21, S349–S354.

Cignitti, S., Zhang, L., Gani, R., 2015. Computer-aided framework for design of pure, mixed and blended products. Comput. Aided Chem. Eng. 37, 2093–2098.

Cockrem, M.C.M., Flatt, J.H., Lightfoot, E.N., 1989. Solvent selection for extraction from dilute solution. Sep. Sci. Technol. 24, 769–807.

Duran, M.A., Grossmann, I.E., 1986. An outer-approximation algorithm for a class of mixed-integer nonlinear programs. Math. Program. 36, 307–339.

Duvedi, A., Achenie, L.E.K., 1997. On the design of environmentally benign refrigerant mixtures: a mathematical programming approach. Comput. Chem. Eng. 21, 915–923.

Eyal, A.M., Canari, R., 1995. pH dependence of carboxylic and mineral acid extraction by amine-based extractants: effects of pKa, amine basicity, and diluent properties. Ind. Eng. Chem. Res. 34, 1789–1798.

Fletcher, R., Leyffer, S., 1994. Solving mixed integer nonlinear programs by outer approximation. Math. Program. 66, 327–349.

Folić, M., Adjiman, C.S., Pistikopoulos, E.N., 2007. Design of solvents for optimal reaction rate constants. AIChE J. 53, 1240–1256.

Fredenslund, A., Jones, R.L., Prausnitz, J.M., 1975. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. AIChE J. 21, 1086–1099.

GAMS Development Corporation, 2014. General Algebraic Modeling System (GAMS). URL: ⟨http://www.gams.com⟩.

Gani, R., 2004a. Chemical product design: challenges and opportunities. Comput. Chem. Eng. 28, 2441–2457.

Gani, R., 2004b. Computer-aided methods and tools for chemical product design. Chem. Eng. Res. Des. 82, 1494–1504.

Gani, R., Ng, K.M., 2015. Product design—Molecules, devices, functional products, and formulated products. Comput. Chem. Eng. 81, 70–79.

Gani, R., Nielsen, B., Fredenslund, A., 1991. A group contribution approach to computer-aided molecular design. AIChE J. 37, 1318–1332.

Gmehling, J., Anderson, T.F., Prausnitz, J.M., 1978. Solid-liquid equilibria using UNIFAC. Ind. Eng. Chem. Fundam. 17, 269–273.

Gmehling, J., Li, J., Schiller, M., 1993. A modified UNIFAC model. 2. Present parameter matrix and results for different thermodynamic properties. Ind. Eng. Chem. Res. 32, 178–193.

Gmehling, J., Wittig, R., Lohmann, J., Joh, R., 2002. A modified UNIFAC (Dortmund) model. 4. Revision and extension. Ind. Eng. Chem. Res. 41, 1678–1688.

Gordon, R.E., Amin, S.I., 1984. Crystallization of ibuprofen. US Patent. URL:⟨http://www.google.com/patents/US4476248⟩.

Gracin, S., Rasmuson, A.C., 2002. Solubility of phenylacetic acid, *p*-hydroxyphenylacetic acid, *p*-aminophenylacetic acid, *p*-hydroxybenzoic acid, and ibuprofen in pure solvents. J. Chem. Eng. Data 47, 1379–1383.

Grossmann, I.E., 2002. Review of nonlinear mixed-integer and disjunctive programming techniques. Optim. Eng. 3, 227–252.

Grossmann, I.E., Trespalacios, F., 2013. Systematic modeling of discrete-continuous optimization models through generalized disjunctive programming. AIChE J. 59, 3276–3295.

Grossmann, I.E., Viswanathan, J., Vecchietti, A., Raman, R., Kalvelagen, E., 2002. GAMS/DICOPT: A discrete continuous optimization package. URL:⟨http://www.gams.com/dd/docs/solvers/dicopt/index.html⟩.

Gullo, M., Verzelloni, E., Canonico, M., 2014. Aerobic submerged fermentation by acetic

acid bacteria for vinegar production: process and biotechnological aspects. Process Biochem. 49, 1571–1579.
Hiriart-Urruty, J.B., Lemaréchal, C., 1993. Convex Analysis and Minimization Algorithms I: Fundamentals. Marcel Dekker, Inc,, New York.
IJmker, H.M., Gramblička, M., Kersten, S.R.A., van der Ham, A.G.J., Schuur, B., 2014. Acetic acid extraction from aqueous solutions using fatty acids. Sep. Purif. Technol. 125, 256–263.
Jonuzaj, S., Akula, P.T., Kleniati, P.-M., Adjiman, C.S., 2016. The formulation of optimal mixtures with Generalized Disjunctive Programming: a solvent design case study. AIChE J. 62, 1616–1633.
Karunanithi, A.T., Achenie, L.E.K., Gani, R., 2005. A new decomposition-based computer-aided molecular/mixture design methodology for the design of optimal solvents and solvent mixtures. Ind. Eng. Chem. Res. 44, 4785–4797.
Karunanithi, A.T., Achenie, L.E.K., Gani, R., 2006. A computer-aided molecular design framework for crystallization solvent design. Chem. Eng. Sci. 61, 1247–1260.
Karunanithi, A.T., Acquah, C., Achenie, L.E.K., Sithambaram, S., Suib, S.L., 2009. Solvent design for crystallization of carboxylic acids. Comput. Chem. Eng. 33, 1014–1021.
Katikaneni, S.P.R., Cheryan, M., 2002. Purification of fermentation-derived acetic acid by liquid-liquid extraction and esterification. Ind. Eng. Chem. Res. 41, 2745–2752.
Kaur, A., Vohra, D.K., 2010. Study of bulk liquid membrane as a separation technique to recover acetic and propionic acids from dilute solutions. Indian J. Chem. Technol. 17, 133–138.
Kocis, G.R., Grossmann, I.E., 1989. Computational experience with DICOPT solving problems in process systems engineering. Comput. Chem. Eng. 13, 307–315.
Lee, J., Leyffer, S., 2012. Mixed Integer Nonlinear Programming, vol. 154, 1st ed. Springer Science & Business Media B.V; New York.
Lee, M.J., Wu, H.T., Lin, H.m., 2000. Kinetics of catalytic esterification of acetic acid and amyl alcohol over dowex. Ind. Eng. Chem. Res. 39, 4094–4099.
Lee, S., Grossmann, I.E., 2000. New algorithms for nonlinear generalized disjunctive programming. Comput. Chem. Eng. 24, 2125–2141.
Lee, S., Grossmann, I.E., 2003. Generalized convex disjunctive programming: nonlinear convex hull relaxation. Comput. Optim. Appl. 26, 83–100.
Lee, S., Grossmann, I.E., 2005. Logic-based modeling and solution of nonlinear discrete/continuous optimization problems. Ann. Oper. Res. 139, 267–288.
Méndez, C.A., Cerdá, J., Grossmann, I.E., Harjunkoski, I., Fahl, M., 2006. State-of-the-art review of optimization methods for short-term scheduling of batch processes. Comput. Chem. Eng. 30, 913–946.
Misener, R., Floudas, C.A., 2014. Antigone: algorithms for continuous/integer global optimization of nonlinear equations. J. Glob. Optim. 59, 503–526.
Naser, S.F., Fournier, R.L., 1991. A system for the design of an optimum liquid–liquid extractant molecule. Comput. Chem. Eng. 15, 397–414.
Nemhauser, G.L., Wolsey, L.A., 1999. Integer and combinatorial optimization. volume 41. Wiley & Sons, Inc., New York.
Nicks, P.F., Ryan, W.A., 1975. Paint compositions and method of their production. US Patent 3920597A. URL:⟨https://www.google.com/patents/US3920597⟩.
Papadopoulos, A.I., Stijepovic, M., Linke, P., Seferlis, P., Voutetakis, S., 2013. Toward optimum working fluid mixtures for organic Rankine cycles using molecular design and sensitivity analysis. Ind. Eng. Chem. Res. 52, 12116–12133.
Poling, B.E., Prausnitz, J.M., O'Connell, J.P., 2001. The Properties of Gases and Liquids 5th ed.. McGraw-Hill,, New York; London.
Pretel, E.J., Lopez, P.A., Bottini, S.B., Brignole, E.A., 1994. Computer-aided molecular design of solvents for separation processes. AIChE J. 40, 1349–1360.
Raman, R., Grossmann, I.E., 1991. Relation between MILP modelling and logical inference for chemical process synthesis. Comput. Chem. Eng. 15, 73–84.
Raman, R., Grossmann, I.E., 1994. Modelling and computational techniques for logic based integer programming. Comput. Chem. Eng. 18, 563–578.
Ruiz, J.P., Grossmann, I.E., 2012. A hierarchy of relaxations for nonlinear convex generalized disjunctive programming. Eur. J. Oper. Res. 218, 38–47.
Ruiz, J.P., Grossmann, I.E., 2013. Using convex nonlinear relaxations in the global optimization of nonconvex generalized disjunctive programs. Comput. Chem. Eng. 49, 70–84.
Ruiz, J.P., Jagla, J.H., Grossmann, I.E., Meeraus, A., Vecchietti, A., 2012. Generalized disjunctive programming: Solution strategies, in: Kallrath, J. (Ed.), Algebraic Modeling Systems, 1st ed. Springer - Verlag; Berlin, Heidelberg. Chapter 4, pp. 82–99.
Sandler, S.I., 1999. Chemical and Engineering Thermodynamics 3rd ed.. John Wiley & Sons,, New York; Chichester.
Sano, K., Uchida, H., Wakabayashi, S., 1999. A new process for acetic acid production by direct oxidation of ethylene. Catal. Surv. 3, 55–60.
Sawaya, N.W., 2006. Reformulations, relaxations and cutting planes for generalized disjunctive programming (Ph.D. thesis). Carnegie Mellon University.
Sawaya, N.W., Grossmann, I.E., 2005. A cutting plane method for solving linear generalized disjunctive programming problems. Comput. Chem. Eng. 29, 1891–1913.
Sawaya, N.W., Grossmann, I.E., 2007. Computational implementation of non-linear convex hull reformulation. Comput. Chem. Eng. 31, 856–866.
Sawaya, N.W., Grossmann, I.E., 2012. A hierarchy of relaxations for linear generalized disjunctive programming. Eur. J. Oper. Res. 216, 70–82.
Seader, J.D., Henley, E.J., Roper, D.K., 2011. Separation Process Principles: Chemical and Biochemical Operations 3rd ed.. Wiley & Sons, Inc.,, Hoboken, NJ.
Sinha, M., Achenie, L.E.K., Gani, R., 2003. Blanket wash solvent blend design using interval analysis. Ind. Eng. Chem. Res. 42, 516–527.
Smith, J.M., Van Ness, H.C., Abbott, M.M., 2001. Introduction to Chemical Engineering Thermodynamics 6th ed.. McGraw-Hill,, Boston.
Solvason, C.C., Chemmangattuvalappil, N.G., Eljack, F.T., Eden, M.R., 2009. Efficient visual mixture design of experiments using property clustering techniques. Ind. Eng. Chem. Res. 48, 2245–2256.
Struebing, H., Ganase, Z., Karamertzanis, P.G., Siougkrou, E., Haycock, P., Piccione, P.M., Armstrong, A., Galindo, A., Adjiman, C.S., 2013. Computer-aided molecular design of solvents for accelerated reaction kinetics. Nat. Chem. 5, 952–957.
Stubbs, R.A., Mehrotra, S., 1999. A branch-and-cut method for 0-1 mixed convex programming. Math. Program. 86, 515–532.
Tawarmalani, M., Sahinidis, N.V., 2005. A polyhedral branch-and-cut approach to global optimization. Math. Program. 103, 225–249.
Trespalacios, F., Grossmann, I.E., 2015. Improved big-M reformulation for generalized disjunctive programs. Comput. Chem. Eng. 76, 98–103.
Trüeb, R.M., 2007. Shampoos: ingredients, efficacy and adverse effects. J. der Dtsch. Dermatol. Ges. 5, 356–365.
Turkay, M., Grossmann, I.E., 1996. Logic-based MINLP algorithms for the optimal synthesis of process networks. Comput. Chem. Eng. 20, 959–978.
Vaidyanathan, R., El-Halwagi, M., 1996. Computer-aided synthesis of polymers and blends with target properties. Ind. Eng. Chem. Res. 35, 627–634.
Vaidyaraman, S., Maranas, C.D., 2002. Synthesis of mixed refrigerant cascade cycles. Chem. Eng. Commun. 189, 1057–1078.
Vecchietti, A., Lee, S., Grossmann, I.E., 2003. Modeling of discrete/continuous optimization problems: characterization and formulation of disjunctions and their relaxations. Comput. Chem. Eng. 27, 433–448.
Williams, H.P., 1985. Model Building in Mathematical Programming 2st ed.. Wiley & Sons,, California.
Yang, X., Song, H., 2006. Computer-aided molecular design of solvents for separation processes. Chem. Eng. Technol. 29, 33–43.
Yunus, N.A., Gernaey, K.V., Woodley, J.M., Gani, R., 2014. A systematic methodology for design of tailor-made blended products. Comput. Chem. Eng. 66, 201–213.
Zhang, L., Cignitti, S., Gani, R., 2015. Generic mathematical programming formulation and solution for computer-aided molecular design. Comput. Chem. Eng. 78, 79–84.
Zhou, T., McBride, K., Zhang, X., Qi, Z., Sundmacher, K., 2015. Integrated solvent and process design exemplified for a Diels-Alder reaction. AIChE J. 61, 147–158.