



**Global Optimization for Integrated Solvent and Process
Design**

DISSERTATION

zur Erlangung des akademischen Grades

Doktoringenieur

(Dr.-Ing.)

von M.Sc. Tobias Kefler

geboren am 01.02.1990 in Magdeburg

genehmigt durch die Fakultät für Elektrotechnik und Informationstechnik
der Otto-von-Guericke-Universität Magdeburg.

Gutachter:

Prof. Dr.-Ing. Achim Kienle

Prof. Dr.-Ing. Mirko Skiborowski

Promotionskolloquium am 07.12.2021

FOREWORD

*Knowledge, like air, is vital to life.
Like air, no one should be denied it.*

— Alan Moore

This thesis is the result of the research I conducted from 2015 to 2020 at the Chair for Automation/Modeling at the Otto von Guericke University and the Process Synthesis and Dynamics Group at the Max Planck Institute for Dynamics of Complex Technical Systems in Magdeburg. The research was funded by the German Research Foundation through SFB/TRR63, which is hereby gratefully acknowledged.

ABSTRACT

This thesis aims to build a framework for integrated process and solvent design to obtain good performing, environmentally benign, and safe solvent alternatives to industrially used standard solvents. Hydroformylation of alkenes is considered as an application example in this thesis.

Hydroformylation is the standard process for the conversion of alkenes to aldehydes. Aldehydes are a crucial raw material in the chemical industry. Many large-scale plants are necessary to produce more than five megatons of aldehyde yearly. Traditionally, alkenes from petrochemical sources are used as raw materials. Nowadays, a transition to ecologically benign processes started, and alternatives to established industry standards emerged. One of them is the hydroformylation of long-chain alkenes from renewable resources.

A rhodium-based catalyst is used in a homogeneously catalyzed reaction in this process realization to achieve a high space-time yield. A big challenge is the recovery of the expensive catalyst for economic feasibility. Thermomorphic solvent systems (TMS) are a solution for the catalyst recovery problem, because they allow for a high recovery rate for relatively low costs.

The standard TMS consists of dimethylformamide (DMF) and dodecane. Although DMF is a widely used solvent in the chemical industry, it exhibits enormous potential environmental and health risks.

The main objective of this thesis is to develop a methodology for systematically identifying promising economically and ecologically benign alternatives for DMF.

Efficient methods for the optimization of each of the process parts are developed to meet the goal of finding alternative solvent candidates.

Distillation columns are employed in the process for product purification, and extraction solvent recovery. Due to the energy demand of

distillation units inefficiency is expensive. The global optimization of non-ideal distillation columns becomes tractable by employing surrogate models. The optimization study yields the insight that the column's thermodynamics can be assumed as ideal, lowering the complexity of the overall optimization problem. The computational complexity is further reduced by employing additional surrogate models and model reformulation techniques on the remaining process parts: reactor and phase separators.

Optimization of the whole process is achieved by interconnecting each individual process part. Dimethyl succinate and tetrahydropyranone, two alternative solvent candidates, and DMF are investigated concerning economic process performance. One of the solvent candidates performs on the same level as DMF while being safe and environmentally benign.

Finally, a hierarchical approach to integrated design consisting of candidate solvent generation and process optimization is used. The combinatorial problem of obtaining feasible new candidates becomes tractable by limiting the search-space to a neighborhood of already known green solvent candidates. One of the challenges is to reduce the search-space to molecules exhibiting the necessary thermodynamic behavior, i. e. a miscibility gap for the catalyst separation. For this, an efficient search-space reduction based on quantum-chemical calculations is developed.

In total, five well-performing green solvents are identified as viable alternatives to DMF for the hydroformylation process. Four of them by using the integrated design approach. They offered a performance close to or even better than that of DMF.

KURZZUSAMMENFASSUNG

Das Ziel dieser Arbeit ist das Schaffen eines Rahmens für das integrierte Prozess- und Lösungsmitteldesign zur Identifikation performanter, ökologisch unbedenklicher und sicherer Lösungsmittelalternativen zu industriell eingesetzten Standardlösungsmitteln.

Der Standardprozess für die Umwandlung von Alkenen zu Aldehyden ist die Hydroformylierung, welche in dieser Arbeit als Anwendungsbeispiel betrachtet wird. Aldehyde sind wichtige Rohmaterialien in der chemischen Industrie, für die jährliche Produktion von über fünf Megatonnen dieser Stoffe werden traditionell Alkene aus petrochemischen Quellen als Rohstoff eingesetzt. Inzwischen hat ein Übergang zu ökologisch verträglichen Verfahren begonnen und es sind Alternativen zu den etablierten Industriestandards entstanden. Eine dieser Alternativen ist die Hydroformylierung von langkettigen Alkenen aus nachwachsenden Rohstoffen.

Bei dieser Prozessrealisierung wird ein rhodiumbasierter Katalysator-komplex in einer homogen katalysierten Reaktion eingesetzt, um eine hohe Raum-Zeit-Ausbeute zu erreichen. Eine der größten Herausforderungen für die Wirtschaftlichkeit dieses Prozesses ist die Rückgewinnung des teuren Katalysators. Thermomorphe Lösungsmittelsysteme (TMS) sind eine mögliche Lösung für das Problem der Katalysatorrückgewinnung, da sie eine gute Rückgewinnung bei relativ niedrigen Kosten ermöglichen.

Das eingesetzte Standard-TMS besteht aus Dimethylformamid (DMF) und Dodekan. Obwohl DMF ein weitverbreitetes Lösungsmittel in der chemischen Industrie ist, weist es enorme potenzielle Umwelt- und Gesundheitsrisiken auf.

Das Hauptziel dieser Arbeit ist die Entwicklung einer Methodik zur systematischen Identifizierung von vielversprechenden ökonomisch und ökologisch verträglichen Alternativen zu DMF.

Um das Ziel zu erreichen, alternative Lösungsmittelkandidaten zu finden, werden effiziente Methoden zur Optimierung der einzelnen Prozesssteile entwickelt.

Im Prozess werden Destillationskolonnen für die Produktaufreinigung und die Rückgewinnung des Extraktionslösungsmittels eingesetzt. Aufgrund des Energiebedarfs von Destillationsanlagen ist ein ineffizienter Betrieb teuer und wird durch Optimierung verhindert. Die globale Optimierung von nicht-idealen Destillationskolonnen wird durch die Verwendung von Ersatzmodellen handhabbar. Die Optimierungsstudie liefert die Erkenntnis, dass die Thermodynamik der Destillationskolonnen als ideal angenommen werden kann, was die Komplexität des gesamten Optimierungsproblems senkt. Die Komplexität wird weiter reduziert, indem zusätzliche Ersatzmodelle und Modellreformulierungstechniken für die verbleibenden Prozesssteile, den Reaktor und die Phasentrenner, angewendet werden.

Die Optimierung des Gesamtprozesses wird durch die Verknüpfung der einzelnen Prozesssteile erreicht. Dimethylsuccinat und Tetrahydropyranon, zwei alternative Lösungsmittelkandidaten, sowie DMF werden hinsichtlich der wirtschaftlichen Prozessführung untersucht. Die Wirtschaftlichkeit des Prozesses mit einem der Lösungsmittelkandidaten liegt auf dem gleichen Niveau wie mit DMF, der Kandidat ist gleichzeitig sicher und umweltfreundlich.

Schließlich wird ein hierarchischer Ansatz zum integrierten Prozess- und Lösungsmitteldesign, bestehend aus Lösungsmittelkandidatengenerierung und Prozessoptimierung, verwendet. Das kombinatorisch anspruchsvolle Problem, geeignete neue Kandidaten zu generieren, wird durch die Beschränkung des Suchraums auf eine Nachbarschaft bereits bekannter grüner Lösungsmittelkandidaten handhabbar. Eine der Herausforderungen besteht darin, den Suchraum auf Moleküle zu reduzieren, die das notwendige thermodynamische Verhalten, d.h. eine Mischungslücke für die Katalysatorabscheidung, aufweisen. Hierfür wird eine effiziente Suchraumreduktion auf Basis quantenchemischer Berechnungen entwickelt.

Insgesamt werden fünf ökonomisch effiziente grüne Lösungsmittel als brauchbare Alternativen zu DMF für den Hydroformylierungsprozess identifiziert. Vier davon durch Anwendung des integrierten Designansatzes. Die mit diesen Lösungsmitteln erreichte Prozesseffizienz liegt nahe an oder sogar über der mit DMF erreichten.

CONTENTS

1	INTRODUCTION	1
1.1	Motivation	1
1.2	Outline	4
I THEORETICAL BACKGROUND		
2	METHODS	9
2.1	Numerical Optimization	9
2.2	Surrogate Modeling	13
2.2.1	Kriging Interpolation	14
2.2.2	Artificial Neural Networks	19
2.3	Group-Contribution Methods	21
2.3.1	Pure Species Properties	21
2.3.2	Mixture Properties	22
2.4	Conductor Like Screening Model	25
II PROCESS OPTIMIZATION		
3	DOWNSTREAM PROCESSING DESIGN	31
3.1	Introduction	31
3.2	Global Optimization of Distillation Columns	32
3.2.1	Definition of the Optimization Problem	33
3.2.2	Algorithmic Implementation	35
3.2.3	Column Specifications	38
3.2.4	Ideal Distillation	41
3.2.5	Moderately Non-ideal Distillation	44
3.2.6	Highly Non-ideal Distillation	46
3.3	Decanter Cascade	54
3.4	Conclusion	55
4	SIMULTANEOUS REACTOR AND DOWNSTREAM PROCESSING DESIGN	59
4.1	Introduction	59

4.2	Overall Process Optimization	60
4.2.1	Candidate Solvent Selection	60
4.2.2	Model Description	61
4.2.3	Definition of the Optimization Problem	71
4.2.4	Process Optimization Results	73
4.3	Conclusion	77
III INTEGRATED DESIGN		
5	SIMULTANEOUS PROCESS AND SOLVENT DESIGN	81
5.1	Introduction	81
5.2	Computer-Aided Molecular and Mixture Design	83
5.2.1	Structural Feasibility Constraints	84
5.2.2	Complexity Constraints	88
5.2.3	Property Estimation	90
5.3	Integrated Design	92
5.3.1	Search-Space Reduction	92
5.3.2	Process Optimization	96
5.4	Results	98
5.5	Conclusion	100
IV CONCLUDING REMARKS		
6	SUMMARY AND FUTURE PERSPECTIVE	105
6.1	Summary	105
6.2	Outlook	106
V APPENDIX		
A	GLOBAL OPTIMIZATION OF REACTOR AND DECANter	111
B	THERMODYNAMIC AND PHYSICAL DATA	119
C	MOLECULE GROUP PARAMETERS	123
D	ALTERNATIVE SEARCH-SPACE REDUCTION	125
E	HIERARCHICAL ITERATIVE GLOBAL OPTIMIZATION	129
	BIBLIOGRAPHY	135
	ABOUT THE AUTHOR	153

LIST OF FIGURES

Chapter 1

Figure 1.1	Thermomorphic solvent system	2
------------	--	---

Chapter 2

Figure 2.1	Optimization approaches	10
Figure 2.2	Kriging example	18
Figure 2.3	Visualization of an artificial neural network	19
Figure 2.4	BPP cavity and σ -profile	25

Chapter 3

Figure 3.1	Simplified process flowsheet	31
Figure 3.2	Distillation column scheme	39
Figure 3.3	Binary phase equilibria DMF, C ₁₀ an, C ₁₂ an	45
Figure 3.4	Two Sketches of response surfaces.	47
Figure 3.5	Continuation: predictor-corrector step	49
Figure 3.6	Residue curve map.	50
Figure 3.7	Continuation results	52
Figure 3.8	Ternary phase equilibrium	55

Chapter 4

Figure 4.1	Detailed process flowsheet.	62
Figure 4.2	Molar flows in mol/s	74

Chapter 5

Figure 5.1	σ -moments of solvent candidates	94
Figure 5.2	σ -moments VT-2005 database	95
Figure 5.3	Process flowsheet.	96
Figure 5.4	Spider diagram results	99

Appendix A

Figure A.1	Process flowsheet	112
------------	-----------------------------	-----

Appendix E

Figure E.1	Process flowsheet	129
Figure E.2	Surrogate	131
Figure E.3	Results	133

LIST OF TABLES

Chapter 3

Table 3.1	Parameters of Algorithm 1	38
Table 3.2	Chemical species within the mixture	38
Table 3.3	Lumped cost function parameters κ	41
Table 3.4	Computation results ideal model, 95 % recovery . .	42
Table 3.5	Computation results ideal model, 99 % recovery . .	43
Table 3.6	Computation results non-ideal model, 95 % recovery	45
Table 3.7	Computation results non-ideal model, 99 % recovery	46
Table 3.8	Computation results heavily non-ideal distillation.	54

Chapter 4

Table 4.1	Chemical species	62
Table 4.2	Degrees of freedom	72
Table 4.3	Optimization results overall process	73

Chapter 5

Table 5.1	Groups in group set i	88
Table 5.2	Chemical species	97
Table 5.3	Solvent candidates	98

Appendix A

Table A.1	Chemical species	112
Table A.2	Degree of freedom (DOF) bounds	116

Appendix B

Table B.1	Vapor pressure parameters	119
Table B.2	Relative volatilities	120
Table B.3	Cost function parameters	121

Appendix C

Table C.1	Group parameter values	123
-----------	----------------------------------	-----

NOMENCLATURE

Latin

ΔW	Exchange energy	N	Number of nodes
\dot{V}	Volumetric flow rate	n	Group occurrences
$\mathbf{B}_{S D SD}$	Presence of bond type parameter	p	Smoothness parameter
\mathbf{K}	Reaction rate parameters	P_{vap}	Partial pressures
$\mathbf{S}_{S D SD}$	Number of bonds parameter	P_{vap}^*	Vapor pressures
a, b, c	Group interaction parameters	P_{yx}	Partition coefficient
B	Optimization branch	$q(x)$	Trend
b_l	Lower bound	R	Reference matrix
b_u	Upper bound	r	Reaction rates
C	Group contribution	$s(n)$	Structural feasibility constraints
e_j	j th unit vector	t	Binodal curve parameter
F	Interpolation matrix	u	Activation potential
f	Vector-valued function	w	Weights
E_μ	Jacobian	x	Degrees of freedom
$g(x)$	Inequality constraints	y	Output
h	Weighted group interactions	$z^{(k)}$	Tangent vector
$h(x)$	Equality constraints	z_{feed}	Feed composition
$L^{(I)}, L^{(II)}$	Liquid phases	a	Antoine parameters
$M(p(\sigma))$	σ -moments	a	Slope parameter
		B	Bottom product flow rate
		$c(x_1, x_2)$	Parameterized function
		c_{cat}	Catalyst concentration

c_{hb}	Hydrogen bonding interaction	p	Property
D	Distillate flow rate	$p(\sigma)$	Charge density
F	Molar flow rate	q_s	VdW pure species surface area
J	Objective function	r	Recovery rate
L	Likelihood function	r_s	VdW Pure species volume
l_r	Number of stages, rectifying section	T	Temperature
L_s	VdW Mixture surface area	T_b	Boiling point
l_s	Number of stages, stripping section	V	Vapor flow rate
M	Molar mass	V_{reac}	Reactor volume
m_{cat}	Catalyst mass	V_s	VdW Mixture volume
P	Pressure	Z	Molar feed flow rate
		z	Coordination number

Greek

α'	Misfit energy	ψ	Activation function
δ	Search direction	ρ	Density
δ_0	Hansen solubility parameter	σ	Charged surface segment
		σ^2	Variance
ε	relative improvement threshold	$\sigma^{(k)}$	Continuation step size
η	Local parameterization parameter	σ_{hb}	Hydrogen bonding cutoff
		τ	Threshold
Γ	Segment activity coefficient	θ_{1-8}	Cost function parameters
		$\underline{\delta}$	Minimum absolute surrogate error
ω	Continuation parameter		
$\bar{\delta}$	Maximum absolute surrogate error	ϖ	Purge factor
		ς	Split factor

α	Relative volatilities	μ	Continuation starting point
β	Number of bonds	τ	Group interactions
$\beta^{(psol)}$	Split factor vector	θ	Decay parameter
χ	Liquid mole fractions	Υ	Output matrix / Vapor mole fractions
η	Weighted group interactions	ϑ	Weighted VdW surface area
γ	Activity coefficient	ζ	Constant trend
κ	Cost function parameter	Ξ	Purge stream
λ	Distillation surrogate parameters		

Abbreviations

<i>psol</i>	Polar solvent	HSP	Hansen solubility parameter
ANN	Artificial Neural Network	LD ₅₀	Oral median lethal dose
BCF	Bioconcentration factor	LLE	Liquid-liquid equilibrium
BIP	BiPhePhos	MEA	Methyl 2-ethylacetoacetate
cat	Rhodium/BiPhePhos complex	MINLP	Mixed-integer nonlinear program
COSMO	Conductor-like screening model	PEL	Permissible exposure limit
CSTR	Continuously stirred tank reactor	RHO	Rhodium
DMF	Dimethylformamide	SFB	Sonderforschungsbereich
DMG	Dimethyl glutarate	TAC	Total annualized costs
DSUC	Dimethyl succinate	THPO	Tetrahydropyranone
ELL	Ethyl levulinate	VdW	Van-der-Waals
EMM	1-Ethyl 4-methyl methylsuccinate	VLE	Vapor-liquid equilibrium
GC	Group-contribution		

INTRODUCTION

1.1 MOTIVATION

Since the 12 principles of “Green Chemistry” were postulated in 1998 (Anastas and Warner, 1998) the concept has gained an evergrowing interest (Linthorst, 2010). Nowadays, the chemical industry has to fulfill the customer’s demand for more sustainable products. The requirements include using renewable raw materials, highly selective catalysts, and ecologically benign auxiliary materials, such as solvents.

With an annual world production of more than five megatons, aldehydes are an important raw material for amines, alcohols, and carboxyl acids (Wiberg et al., 2001). The standard process for their production is the hydroformylation of olefines. Short-chain olefines from petrochemical origin are used as raw materials in industrial state-of-the-art process realizations. Due to the importance of the process, a high impact can be obtained by making it cleaner, safer, and more energy efficient. This happens to be the definition of process intensification (Stankiewicz and Mouljin, 2000).

A promising process intensification approach is using homogeneously solvated transition metals as catalysts (Behr and Vorholt, 2017). Homogeneously catalyzed reactions have the benefit of a high space-time yield. Thereby, these catalysts offer the possibility to use renewable feedstocks, where traditional processes are bound to petrochemical sources. However, the recycling of the, in most cases, expensive catalyst is a challenging task.

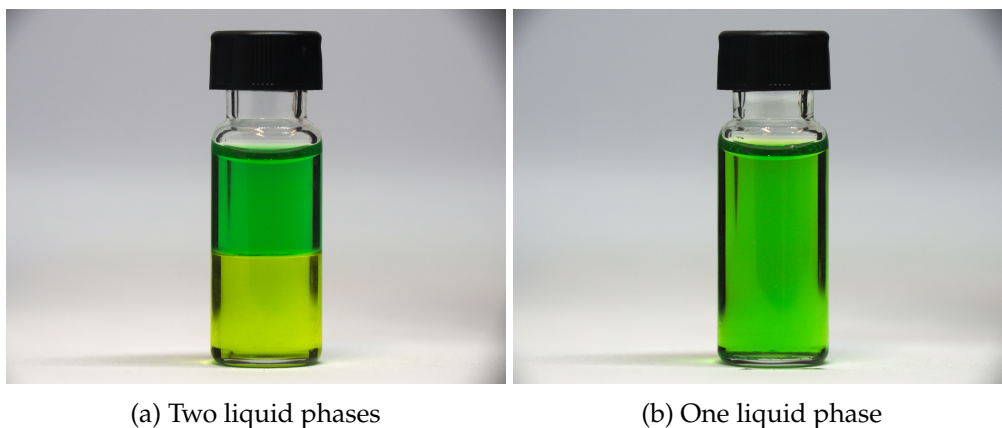


Figure 1.1: Thermomorphic solvent system (TMS) consisting of isopropanol, linseed oil, and water (Rao and Arnold, 1957). The isopropanol was colored using blue food coloring. Figure 1.1a shows the mixture at ambient temperature, two phases are present. After heating it up, the mixture exhibits one homogeneous liquid phase in Figure 1.1b.

The employment of homogeneously solved transition metals is the research topic of the collaborative research center *Sonderforschungsbereich 63; InPrompt* (Vetter, 2020). Over the last eleven years, more than 60 researchers from seven universities tried to overcome the challenges imposed by multiphase systems. The hydroformylation with varying alkene chain-lengths was exemplarily chosen as a case study (Schäfer et al., 2012; Kiedorf et al., 2014; Hentschel et al., 2015; Bianga et al., 2019). Several optimization studies were conducted (McBride and Sundmacher, 2015; Nentwich and Engell, 2016) and the process was successfully operated continuously and optimized in real-time with a thermomorphic-multiphase-system-based catalyst separation (Dreimann et al., 2017; Hernández and Engell, 2016) and with a surfactant-based catalyst separation (Illner et al., 2016; Müller et al., 2017).

In contrast to traditional hydroformylation realizations, InPrompt employs long-chain olefines as raw materials. Unsaturated oleochemicals from renewable resources can be used to derive these raw materials, and they may also be synthesized from biogas using the Fischer-Tropsch process (Kraume, 2013; Behr and Vorholt, 2012; Behr et al., 2005).

In their considerations, the InPrompt researchers use a rhodium-based catalyst for the hydroformylation process. The recovery of this catalyst using a thermomorphic solvent system (TMS) is based on a temperature difference between the reactor and a subsequent phase separator. At reaction conditions, the mixture forms a homogeneous liquid phase. In such a phase, the reaction's space-time yield increases due to good contact between catalyst and reactant. In the subsequent phase separator, the mixture's temperature is reduced to a specific value, and two phases occur. One phase holds the majority of the catalyst; the other phase holds the majority of the product. An exemplary TMS is shown in [Figure 1.1](#).

Although efficient process operation and the usage of renewable raw materials are an important step in the direction of green chemistry, there are still some challenges to overcome. One of which is the use of N,N-Dimethylformamide (DMF) as a polar solvent in the thermomorphic solvent system. This widely used solvent exhibits excellent properties for catalyst recovery and was identified as one of the best solvents for this process (McBride et al., 2016). However, DMF is toxic (Fail et al., 1998) and damaging to the liver (Kleiner, 2018). DMF fulfills the criteria of Article 57 of the European chemicals ordinance REACH and is on the list of substances of very high concern (SVHC) (European Parliament, 2006).

DMF should be replaced by a safer, ecologically benign alternative to obtain a greener process. However, the choice of a solvent with which the process operates economically efficient is non-trivial, even more so if the solvent has to fulfill additional non-economical restrictions, such as non-toxicity. Computer-aided and optimization-based methods help to overcome this problem. In the integrated process and solvent design, the process is optimized, and a new solvent is identified simultaneously. Complex mixed-integer non-linear programs (MINLPs) have to be solved for this task. Solving complex MINLPs is difficult. Hence hierarchical approaches are commonly used.

This thesis aims at developing a framework for the integrated solvent and process design applied to the hydroformylation of long-chain olefines with a TMS for catalyst separation. Solving the single parts of the hierarchical optimization problem effectively and efficiently is a prerequisite to obtaining a good performing green solvent candidate.

1.2 OUTLINE

The present thesis consists of six chapters, structured in four parts.

[Chapter 2](#) in [Part i](#) gives an overview of the theoretical background necessary for the integrated process and solvent design approach presented in the following chapters. It starts with the description of optimization methods, including rigorous global optimization, to avoid weak local minima. Then, surrogate modeling approaches required to make the optimization problem easier are introduced. The chapter closes with approaches for describing the candidate solvents molecule and mixture behavior.

[Part ii](#) introduces the employed process optimization approaches. The downstream processing design, consisting of a purification step and a phase separation step, is described in [Chapter 3](#). First, an optimization problem concerning the distillation column, necessary for the product's purification, is solved to global optimality, using an iterative Kriging approach. The procedure is extended to optimization problems concerning unstable distillation columns exhibiting state multiplicities using an implicit surrogate model formulation. The phase separation step is then added to the optimization problem, that is solved to global optimality again. The downstream process design becomes tractable in [Chapter 3](#). Consequently, a simultaneous reactor and downstream process design will be conducted in [Chapter 4](#). The reactor and a subsequent phase separator are added to the downstream processing, and all recycles are closed. As a result, the optimization problem becomes more challenging and is solved using a multi-start approach. The process optimization is done three times to make the first step towards a green alternative solvent for DMF. Two candidate solvents from McBride et al. (2018) and DMF are investigated

and compared.

The complexity is further increased in [Part iii](#). Here, the optimization methods from [Part ii](#) are used in conjunction with molecular design approaches in [Chapter 5](#) to achieve an integrated process and solvent design. A hierarchical two step approach is presented. In the first step, candidate molecules are generated in the vicinity of already known green alternatives from Linke et al. (2020). For this, some properties of their respective σ -profiles are exploited for a reduction of the search space. The second step is a rigorous process optimization for each of the generated candidate solvents.

[Part iv](#) concludes the thesis with a summary of the key results and future perspectives for the integrated solvent and process design in [Chapter 6](#).

Main results of this thesis have been published before in a series of journal papers:

- Tobias Keßler, Christian Kunde, Nick Mertens, Dennis Michaels, and Achim Kienle. "Global optimization of distillation columns using surrogate models." In: *SN Applied Sciences* 1.1 (2018), p. 11.
doi: [10.1007/s42452-018-0008-9](https://doi.org/10.1007/s42452-018-0008-9).
- Tobias Keßler, Christian Kunde, Kevin McBride, Nick Mertens, Dennis Michaels, Kai Sundmacher, and Achim Kienle. "Global optimization of distillation columns using explicit and implicit surrogate models." In: *Chemical Engineering Science* 197 (2019), pp. 235–245.
doi: [10.1016/j.ces.2018.12.002](https://doi.org/10.1016/j.ces.2018.12.002).
- Tobias Keßler, Christian Kunde, Steffen Linke, Kevin McBride, Kai Sundmacher, and Achim Kienle. "Systematic Selection of Green Solvents and Process Optimization for the Hydroformylation of Long-Chain Olefines." In: *Processes* 7.12 (2019), p. 882.
doi: [10.3390/pr7120882](https://doi.org/10.3390/pr7120882).
- Tobias Keßler, Christian Kunde, Steffen Linke, Kai Sundmacher, and Achim Kienle. "Integrated Computer-Aided Molecular and Process

Design: Green Solvents for the Hydroformylation of Long-Chain Olefines.” In: *Chemical Engineering Science* 249 (2022), 117243.
doi: [10.1016/j.ces.2021.117243](https://doi.org/10.1016/j.ces.2021.117243)

The author of this thesis (TK) was first author, all computations were done by TK, and the draft papers were written by TK with minor revisions by the other authors.

In addition, TK was involved in the following journal paper relevant for this thesis,

- Christian Kunde, Tobias Keßler, Steffen Linke, Kevin McBride, Kai Sundmacher, and Achim Kienle. “Surrogate modeling for liquid-liquid equilibria using a parametrization of the binodal curve.” In: *Processes* 7.10 (2019), p. 753.
doi: [10.3390/pr7100753](https://doi.org/10.3390/pr7100753),

and in a series of contributions to conference proceedings,

- Tobias Keßler, Nick Mertens, Christian Kunde, Corina Nentwich, Dennis Michaels, Sebastian Engell, and Achim Kienle. “Efficient global optimization of a novel hydroformylation process.” In: *Computer Aided Chemical Engineering* 40 (2017), pp. 2113–2118.
doi: [10.1016/B978-0-444-63965-3.50354-8](https://doi.org/10.1016/B978-0-444-63965-3.50354-8).
- Tobias Keßler, Christian Kunde, Steffen Linke, Kevin McBride, Kai Sundmacher, and Achim Kienle. “Computer Aided Molecular Design of Green Solvents for the Hydroformylation of Long-Chain Olefines.” In: *Computer Aided Chemical Engineering* 48 (2020), pp. 745–750.
doi: [10.1016/B978-0-12-823377-1.50125-7](https://doi.org/10.1016/B978-0-12-823377-1.50125-7).

Part I

THEORETICAL BACKGROUND

METHODS

This chapter gives an overview of the theoretical background for the approach to integrated solvent and process design presented in the following chapters.

2.1 NUMERICAL OPTIMIZATION

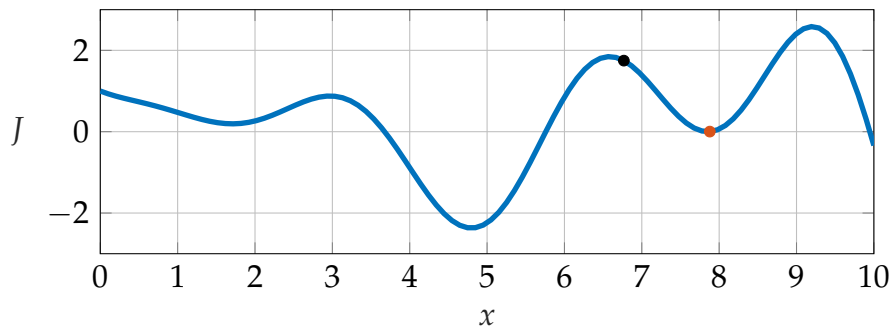
In this chapter, the optimization problems are assumed to be of the following form:

$$\begin{aligned}
 \min_{\mathbf{x}} \quad & J(\mathbf{x}), \\
 \text{s.t.} \quad & \mathbf{h}(\mathbf{x}) = 0, \\
 & \mathbf{g}(\mathbf{x}) \leq 0, \\
 & \mathbf{x} \in G, \quad G \subseteq \mathbb{R}^n, \\
 & x_i \in Z, \quad Z \subseteq \mathbb{Z}, \quad \forall i \in I.
 \end{aligned} \tag{2.1}$$

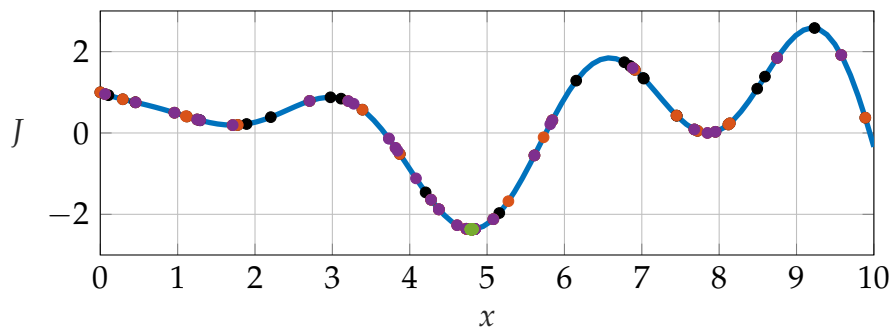
Here, J is the objective function, \mathbf{x} are the degrees of freedom, and $\mathbf{h}(\mathbf{x})$ and $\mathbf{g}(\mathbf{x})$ are the equality and inequality constraints, respectively. If a subset of \mathbf{x} , x_i , needs to fulfill integrality restrictions, i. e. it may only take discrete values, the problem is called MINLP.

If J is a non-convex function, it may exhibit multiple *local* minima. The lowest among those local minima are *global* minima \mathbf{x}^* , for which

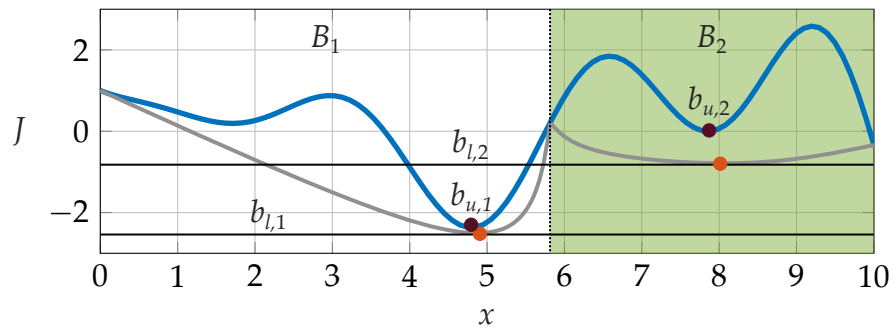
$$J(\mathbf{x}^*) \leq J(\mathbf{x}) \quad \forall \mathbf{x} \neq \mathbf{x}^*, \tag{2.2}$$



(a) Gradient-based optimization



(b) Stochastic optimization



(c) Branch-and-bound-based optimization

Figure 2.1: One-dimensional example optimization. (a) The black dot denotes the starting point x_1 , the orange dot denotes the identified local optimum x_2 . (b) Multiple generations of Matlab’s genetic algorithm (ga). Black dots denote the first generation, orange dots denote the second generation, magenta dots denote the third generation and green dots denote the 15th generation. (c) Two branches B with their respective lower bounds b_l , and upper bounds b_u . Grey lines denote convex relaxations. The branch colored in green will be pruned from the search tree in the next iteration.

holds.

Numerical optimization strategies can be roughly split into three categories:

1. calculus-based algorithms,
2. stochastic or genetic algorithms,
3. branch-and-bound-based algorithms.

An example of a non-convex objective function with multiple local minima is depicted in [Figure 2.1](#). Here, [Figure 2.1a](#) is an example of a calculus-based approach. Coming from the starting point x_1 , the next iteration is found in the direction of the steepest descent,

$$\delta = -\nabla J(x_1), \quad (2.3)$$

where δ is the search direction. The optimization delivers a sub-optimal local minimum,

$$J(x_2) > J(x^*). \quad (2.4)$$

Note that it may be possible to determine a better solution if the optimization is conducted from multiple starting points or a better initial guess. Gradient-based approaches are indirect methods. Indirect methods use first- and second-order optimality conditions, such as the well-known Karush-Kuhn-Tucker conditions (Boyd and Vandenberghe, 2013), and solve the resulting non-linear equation system to arrive at an optimal solution (Goldberg, 1993). In contrast to that, direct methods, such as the famous Nelder-Mead simplex algorithm (Nelder and Mead, 1965), solely rely on objective function evaluations to obtain an optimal solution. However, in both approaches the quality of this solution also depends on the initial guess. Numerous solvers based on direct and indirect methods were developed and are well-known in the literature (Biegler and Grossmann, 2004; Nocedal and Wright, 2006). The freely available interior point solver IPOPT (Wächter, 2002) is an example for a solver based on indirect methods.

Figure 2.1b, on the other hand, is an example of a genetic optimization. Genetic optimization algorithms try to mimic the mechanics of natural selection (Goldberg, 1993). The first generation consists of 50 randomly chosen starting points. Those points are evaluated based on their objective function value to determine their fitness, i. e. probability of survival. In the following generations, new points will emerge close to the best points of the previous generations. Furthermore, for robustness, some points will be placed randomly to account for the avoidance of poor local minima. Although this approach yields more robustness than calculus-based approaches (Goldberg, 1993), which may only find a local minimum close to the starting point, it is a stochastic approach. The resulting optimal solution may vary from one execution to another, i. e. global optimality cannot be guaranteed.

The last type is branch-and-bound-based algorithms. Typically those algorithms relax the integrality constraints of x_i and replace the non-convex objective function by convex under-estimators (Belotti et al., 2013). The idea of this relaxation is to generate an easier to solve convex sub-problem for the calculation of lower bounds b_l . The feasible solution set is subdivided into multiple parts or branches B . For each of these branches, a $b_{l,i}$ is calculated. A local optimization within the branches poses as upper bound, $b_{u,i}$. Branches with a higher $b_{l,i}$ than the lowest upper bound are pruned from the search tree, as they may not include the globally optimal solution. See Figure 2.1c for an illustration. Note that the choice of a convex under-estimator and the branching points is non-trivial and may greatly influence the performance of the algorithm (Locatelli and Schoen, 2013). The interested reader is referred to Belotti et al. (2013) for a more thorough description.

There exist several commercially available branch-and-bound-based algorithms, such as BARON (Kılınç and Sahinidis, 2018), Antigone (Misener and Floudas, 2014), and SCIP (Gleixner et al., 2017). They allow deterministic convergence to a globally optimal solution within a finite solution time (Locatelli and Schoen, 2013). However, this time depends on the complexity of the optimization problem. Although significant improvements were made in the recent years (Boukouvala et al., 2016), the

solution time is often prohibitively large, even for relatively easy problems in the process engineering point of view (Nallasivam et al., 2016).

Numerous strategies have been developed to reduce the computational effort of branch-and-bound-based global optimization algorithms. They involve reducing the number of nonlinearities and variables, e. g. by logarithmic and implicit model reformulations, bound propagations (Puranik and Sahinidis, 2017), and relaxation strategies with simultaneous convexification of the constraint functions (Liers et al., 2020). However, these strategies can only be applied in specific cases. A more general simplification approach is to employ surrogate models to replace computationally expensive model parts or the whole model.

2.2 SURROGATE MODELING

Computationally expensive rigorous process models can be approximated by input-output or response surface models, exchanging precision for computational efficiency. In many fields of engineering and beyond, various methods of varying complexity are used. They range from basic polynomial regressions (Rawlings et al., 1998) to polynomial chaos expansion (Zhang and Sahinidis, 2013), artificial neural networks (Nentwich and Engell, 2016; Nentwich and Engell, 2019), and support vector machines (Bennett and Campbell, 2000), to name only a few. For a given task, the correct choice for a surrogate modeling approach is non-trivial, as each approach has unique characteristics. To help find the best choice, tools for automatical selection of a surrogate structure and fitting of the required parameters exist (Cozad et al., 2014; Boukouvala and Floudas, 2017). For a more thorough discussion, the interested reader is referred to literature reviews on surrogate based modeling (Bhosekar and Ierapetritou, 2018; Razavi et al., 2012) and surrogate assisted optimization (Boukouvala et al., 2016; Forrester and Keane, 2009). A comprehensive review of surrogate modeling in chemical engineering can be found in (McBride and Sundmacher, 2019). Two types of surrogate models were used in this work and will be discussed in the following: Kriging interpolation and artificial neural networks.

2.2.1 Kriging Interpolation

The roots of Kriging interpolation lie in 1951, where D.G. Krige proposed the method to analyze mining data (Krige, 1951) and thereby invented the field of geostatistics. Nowadays, it is widely used through many fields of engineering and will be described in the following.

The vector-valued function $f : \mathbb{R}^m \rightarrow \mathbb{R}^d$ is to be approximated by the Kriging interpolation $\hat{f} : \mathbb{R}^m \rightarrow \mathbb{R}^d$, which is defined by

$$f(\mathbf{x}) \approx \hat{f}(\mathbf{x}) := \mathbf{q}(\mathbf{x}) + \mathbf{z}(\mathbf{x}). \quad (2.5)$$

Here, the vector-valued polynomial $\mathbf{q}(\mathbf{x})$, which is called the “trend”, is fitted to resemble the original function $f(\mathbf{x})$. The remaining differences between the polynomial $\mathbf{q}(\mathbf{x})$ and the original function $f(\mathbf{x})$ are captured by the vector-valued function $\mathbf{z}(\mathbf{x})$. This work uses an “ordinary Kriging” approach. In ordinary Kriging, $\mathbf{q}(\mathbf{x})$ is chosen to be a zero degree polynomial, i. e. a constant non-zero vector $\boldsymbol{\zeta}$. For smooth functions, this choice does not have a high impact on the accuracy of the Kriging model because most of its information is contained in $\mathbf{z}(\mathbf{x})$ (Papalambros and Wilde, 2000).

Therefore, $\mathbf{z}(\mathbf{x})$ is of high importance and will be constructed next. *Reference points* $\bar{\mathbf{x}}^k \in \mathbb{R}^m$, $k = 1, \dots, N$ are points on which the original function $f(\mathbf{x})$ is evaluated on. A weighted sum of the difference between the original function $f(\mathbf{x})$ and the trend $\boldsymbol{\zeta}$ at the reference points defines $\mathbf{z}(\mathbf{x})$,

$$\mathbf{z}(\mathbf{x}) = \sum_{k=1}^N \left(f(\bar{\mathbf{x}}^k) - \boldsymbol{\zeta} \right) w_k(\mathbf{x}), \quad (2.6)$$

where $\boldsymbol{w}(\boldsymbol{x}) = (w_1(\boldsymbol{x}), \dots, w_N(\boldsymbol{x}))^\top$ is the weight function. With Equation (2.6) the interpolation (Equation (2.5)) can be written as

$$\hat{f}(\boldsymbol{x}) = \boldsymbol{\zeta} + \begin{pmatrix} (f(\bar{\boldsymbol{x}}^1) - \boldsymbol{\zeta})^\top \\ \vdots \\ (f(\bar{\boldsymbol{x}}^N) - \boldsymbol{\zeta})^\top \end{pmatrix}^\top \boldsymbol{w}(\boldsymbol{x}). \quad (2.7)$$

The weight function is defined as:

$$\boldsymbol{w}(\boldsymbol{x}) = \boldsymbol{R} \begin{pmatrix} c(\boldsymbol{x}, \bar{\boldsymbol{x}}^1) \\ \vdots \\ c(\boldsymbol{x}, \bar{\boldsymbol{x}}^N) \end{pmatrix}, \quad (2.8)$$

where $c : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$ is a parameterized function that depends on the distance between \boldsymbol{x} and $\bar{\boldsymbol{x}}$, and $\boldsymbol{R} \in \mathbb{R}^{N \times N}$ is a reference matrix that will be derived in the following.

The choice of c is not easy, as it significantly influences the Kriging model's accuracy. Therefore parameterized functions are employed, where the necessary parameters are usually calculated by solving an optimization problem. Numerous function types can be found in the literature; one of the most commonly used ones is the Gaussian correlation function (Caballero and Grossmann, 2008), which is of the following form:

$$c(\boldsymbol{x}^1, \boldsymbol{x}^2) = \exp\left(-\sum_{i=1}^m \theta_i \|\boldsymbol{x}_i^1 - \boldsymbol{x}_i^2\|^{p_i}\right). \quad (2.9)$$

The parameters are used to control the smoothness ($\boldsymbol{p} \in \mathbb{R}^m$) of the function, and the decay ($\boldsymbol{\theta} \in \mathbb{R}^m$) for $\boldsymbol{x}^1 \neq \boldsymbol{x}^2$. The function equals 1 for $\boldsymbol{x}^1 = \boldsymbol{x}^2$.

Equation (2.7) can be further simplified by introducing the interpolation matrix $\boldsymbol{F} \in \mathbb{R}^{d \times N}$, which is defined as

$$\boldsymbol{F} = \begin{pmatrix} f(\bar{\boldsymbol{x}}^1) - \boldsymbol{\zeta}, \dots, f(\bar{\boldsymbol{x}}^N) - \boldsymbol{\zeta} \end{pmatrix} \boldsymbol{R}, \quad (2.10)$$

thereby the final form of Equation (2.5) can be obtained as

$$\hat{f}(\boldsymbol{x}) = \boldsymbol{\zeta} + \boldsymbol{F} \begin{pmatrix} c(\boldsymbol{x}, \bar{\boldsymbol{x}}^1) \\ \vdots \\ c(\boldsymbol{x}, \bar{\boldsymbol{x}}^N) \end{pmatrix}. \quad (2.11)$$

An inherent property of Kriging models is the exact resemblance of the original function at each reference point,

$$\hat{f}(\bar{\boldsymbol{x}}^k) = f(\bar{\boldsymbol{x}}^k) \quad \text{for all } k \in \{1, \dots, N\}, \quad (2.12)$$

or, in other terms,

$$\left(f(\bar{\boldsymbol{x}}^1) - \boldsymbol{\zeta}, \dots, f(\bar{\boldsymbol{x}}^N) - \boldsymbol{\zeta} \right) \left(\left(w(\bar{\boldsymbol{x}}^1), \dots, w(\bar{\boldsymbol{x}}^N) \right) - \boldsymbol{I} \right) = \mathbf{0}. \quad (2.13)$$

This condition is satisfied by

$$\boldsymbol{R} = \begin{pmatrix} c(\bar{\boldsymbol{x}}^1, \bar{\boldsymbol{x}}^1) & \dots & c(\bar{\boldsymbol{x}}^N, \bar{\boldsymbol{x}}^1) \\ \vdots & \ddots & \vdots \\ c(\bar{\boldsymbol{x}}^1, \bar{\boldsymbol{x}}^N) & \dots & c(\bar{\boldsymbol{x}}^N, \bar{\boldsymbol{x}}^N) \end{pmatrix}^{-1}. \quad (2.14)$$

Note that function c and the reference points $\bar{\boldsymbol{x}}$ have to be chosen such that Equation (2.14) can be fulfilled, i. e. the matrix has to be invertible.

The interpolation matrix \boldsymbol{F} is constant. Using Equation (2.10) and Equation (2.14) it can be determined by

$$\boldsymbol{F} \boldsymbol{R}^{-1} = \begin{pmatrix} (f(\bar{\boldsymbol{x}}^1) - \boldsymbol{\zeta})^\top \\ \vdots \\ (f(\bar{\boldsymbol{x}}^N) - \boldsymbol{\zeta})^\top \end{pmatrix}^\top. \quad (2.15)$$

The calculation of the parameter vectors $\boldsymbol{\theta}$, \boldsymbol{p} and $\boldsymbol{\zeta}$ is called “fitting” of the surrogate model. This fitting process is, like stated above, usually done

by solving an optimization problem. One commonly used approach is maximizing the logarithmic likelihood function (Quirante et al., 2015),

$$\begin{aligned} & \min\{-\log(L) \mid p_i \in [0, 1.99], \theta_i > 0\}, \\ \log(L) = & -N/2 \left(\log(\sigma^2) + \log(2\pi) \right) \\ & - 1/2 \log(\|\mathbf{R}\|) \\ & - 1/(2\sigma^2) \left(\mathbf{Y}(\bar{\mathbf{x}}) - \mathbf{1}\boldsymbol{\zeta}^\top \right)^\top \mathbf{R}^{-1} \left(\mathbf{Y}(\bar{\mathbf{x}}) - \mathbf{1}\boldsymbol{\zeta}^\top \right), \end{aligned} \quad (2.16)$$

where

$$\sigma^2 = 1/N \left(\mathbf{Y}(\bar{\mathbf{x}}) - \mathbf{1}\boldsymbol{\zeta}^\top \right)^\top \mathbf{R}^{-1} \left(\mathbf{Y}(\bar{\mathbf{x}}) - \mathbf{1}\boldsymbol{\zeta}^\top \right). \quad (2.17)$$

The so-called output matrix $\mathbf{Y} \in \mathbb{R}^{N \times d}$ is defined as

$$\mathbf{Y}(\bar{\mathbf{x}}) = \begin{pmatrix} \mathbf{f}^\top(\bar{\mathbf{x}}^1) \\ \vdots \\ \mathbf{f}^\top(\bar{\mathbf{x}}^N) \end{pmatrix}. \quad (2.18)$$

The parameter vector $\boldsymbol{\zeta}$ is set to be the mean of the original function's output at the reference points, i.e.

$$\boldsymbol{\zeta} = \frac{1}{N} \sum_{j=1}^N \mathbf{f}(\bar{\mathbf{x}}^j). \quad (2.19)$$

Kriging models have the benefit of inherent statistics (Dubourg and Sudret, 2014). The variance σ^2 is directly available and can, to some extent, be used to evaluate the confidence interval of the surrogate at any point, with

$$\hat{\sigma}^2(\mathbf{x}) = \sigma^2 \left(1 - \left(\hat{\mathbf{f}}(\mathbf{x})^\top \quad \mathbf{r}(\mathbf{x})^\top \right) \begin{pmatrix} \mathbf{0} & \mathbf{Y}^\top \\ \mathbf{Y} & \mathbf{R} \end{pmatrix}^{-1} \begin{pmatrix} \hat{\mathbf{f}}(\mathbf{x}) \\ \mathbf{r}(\mathbf{x}) \end{pmatrix} \right), \quad (2.20)$$

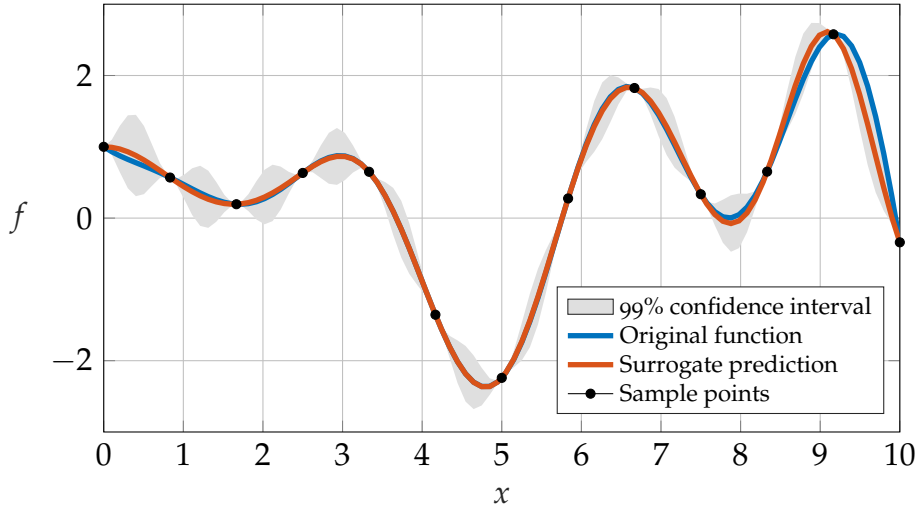


Figure 2.2: Example of a Kriging prediction.

where

$$\mathbf{r}(\mathbf{x}) = \begin{pmatrix} c(\mathbf{x}, \bar{\mathbf{x}}^1) \\ \vdots \\ c(\mathbf{x}, \bar{\mathbf{x}}^N) \end{pmatrix}. \quad (2.21)$$

However, these statistical properties of Kriging interpolation are only meaningful if the correct covariance function c is known. If it is unknown, the variance will mainly resemble the distance between one sample point and its neighbors. See [Figure 2.2](#) for an illustrative example.

Because Kriging interpolation is based on a statistical background, c is assumed to be the covariance of the process. In the more general, but mathematically equivalent (Scheuerer et al., 2013), approach of *Kernel interpolation*, this assumption is not necessary.

In the implicit surrogate formulation presented later on in this thesis, a linear correlation function is used. Hence, the approach is called *Kernel interpolation* in this specific setting. The employed kernel function is

$$c(\mathbf{x}^1, \mathbf{x}^2) = 10 \prod_{i=1}^m \left(1 - \left| \frac{x_i^1 - x_i^2}{\Delta x_i} \right| \right). \quad (2.22)$$

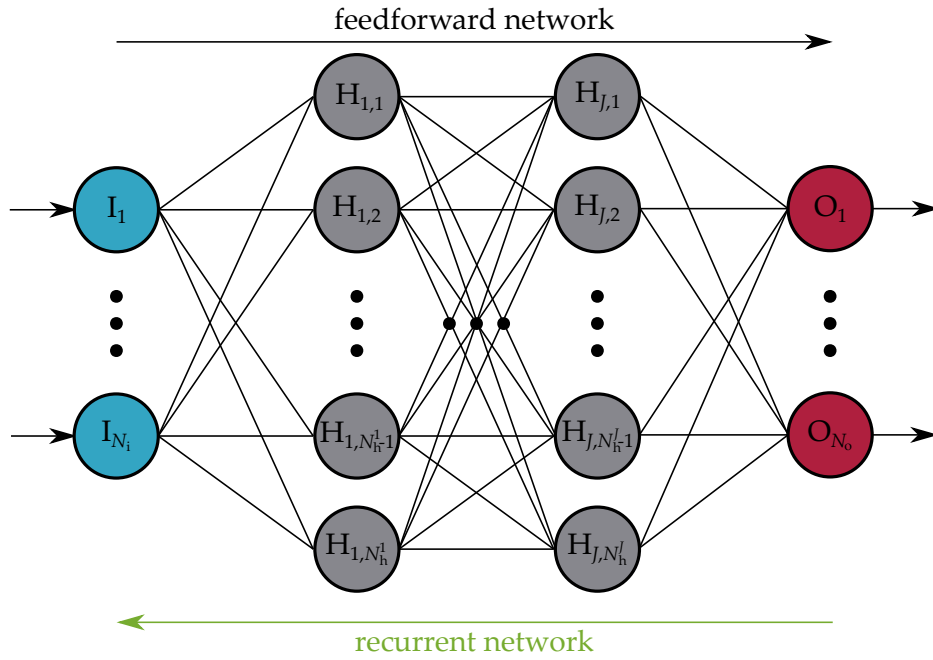


Figure 2.3: Exemplary visualization of an artificial neural network. Input nodes I are depicted in blue, hidden layer nodes H are depicted in grey and output nodes O are depicted in maroon. In a feedforward ANN the information flow is limited to the direction of the black arrow. In contrast to that, feedback or recurrent ANNs may also utilize an information flow in the other direction, as indicated by the green arrow.

2.2.2 Artificial Neural Networks

As their name already suggests, Artificial Neural Networks (ANN) are designed to mimic the function of the brain (Haykin, 1999). They are composed of several so-called neurons or nodes, which are arranged in groups or layers. A visualization of an ANN is shown in Figure 2.3. Three types of layers exist in a classic ANN: the input layer with $N_i \in \mathbb{N}$ input nodes I , $J \in \mathbb{N}_0$ hidden layers with $J \times N_h$ hidden nodes, where $N_h \in \mathbb{N}_0^J$, and the output layer with $N_o \in \mathbb{N}$ output nodes. The number of input nodes N_i and output nodes N_o depends on the inputs and outputs, respectively. The number of hidden layers J and hidden nodes N_h can be chosen freely and greatly influences the ANN's performance. Note that the number of hidden nodes may change from one hidden layer to another.

In feedforward ANNs, the nodes receive signals from nodes in preceding layers and send signals to nodes in subsequent layers. In contrast to that, feedback or recurrent networks allow for an information flow in both directions and between nodes in the same layers (Zou et al., 2009). For the sake of simplicity, the following description will only deal with feedforward networks.

Assume an arbitrary node k . The incoming signals x are summed up and weighted,

$$u_k = \sum_{i=0}^n w_{k,i} x_i, \quad (2.23)$$

where w_k are the weights of node k and u_k is the so-called activation potential (Haykin, 1999). The activation potential is used to calculate the output y_k of the node,

$$y_k = \psi(u_k), \quad (2.24)$$

where ψ is the activation function. The three common classes of activation functions are (Haykin, 1999):

1. Threshold functions: if u passes a given threshold τ , $\psi(u)$ is set to 1. Otherwise, it is set to 0,

$$\psi(u) = \begin{cases} 1, & \text{if } u \geq \tau, \\ 0, & \text{otherwise.} \end{cases} \quad (2.25)$$

2. Piecewise-linear functions: in its simplest form, $\psi(u)$ is constant except for a linear region defined by two constants τ_1 and τ_2 ,

$$\psi(u) = \begin{cases} \tau_1, & \text{if } u \geq \tau_1, \\ u - \tau_2, & \text{if } \tau_1 > u > \tau_2, \\ 0, & \text{if } u \leq \tau_2. \end{cases} \quad (2.26)$$

3. Sigmoid functions: The most commonly used activation function type (Haykin, 1999). Strictly increasing s-shaped functions, such as the logistic function,

$$\psi(u) = \frac{1}{1 + \exp(-a u)}, \quad (2.27)$$

where a is the slope parameter, or the hyperbolic tangent function,

$$\psi(u) = \frac{\exp(u) - \exp(-u)}{\exp(u) + \exp(-u)} = 1 - \frac{2}{1 + \exp(2u)}. \quad (2.28)$$

During the “learning” phase, the weights w need to be calculated such that the ANN delivers the desired output in response to a given input.

2.3 GROUP-CONTRIBUTION METHODS

Prediction of thermodynamic properties is a major challenge in computer aided molecular and process design. For this purpose, group-contribution methods are used in this thesis.

Group-contribution methods rely on the assumption that a molecule consists of parts, so-called groups. Each group is assumed to contribute a certain value to a specific property. There exist group-contribution methods for pure species properties, such as the boiling point, and for mixture properties, such as activity coefficients. Both applications will be discussed in the following.

2.3.1 Pure Species Properties

A simple way to predict pure species properties is to use group-contribution methods of first-order, where it is assumed that a group i contributes a constant value C_i to a property P ,

$$P = \sum_{i=1}^N n_i \cdot C_i, \quad (2.29)$$

where n_i denotes how many times group i occurs within the molecule. Examples of such groups are CH_3 , CH_3CO , and COOH . A large data set is required to find the correct correlations between P and C .

There already exist plenty of correlations for a large number of properties. One of the most commonly used correlations for thermophysical properties is the group-contribution method of Marrero and Gani (2001).

2.3.2 Mixture Properties

The liquid-liquid equilibrium (LLE) of a mixture is described by

$$\gamma^I x^I = \gamma^{II} x^{II}, \quad (2.30)$$

where γ is the activity coefficient and x is the mole fraction. The superscripts I and II denote the two different phases. A model to calculate the activity coefficient is necessary in order to capture the non-idealities of a mixture. A group-contribution based model to calculate the activity coefficients is UNIFAC (Fredenslund et al., 1975). The following model equations of the modified UNIFAC (Dortmund) model (Weidlich and Gmehling, 1987) are taken from Smith et al. (1996).

The activity coefficient of species s is the sum of its residual part R , which accounts for molecular interactions, and its combinatorial part C , which accounts for molecular size and shape differences,

$$\log \gamma_s = \log (\gamma_s^C) + \log (\gamma_s^R). \quad (2.31)$$

The combinatorial part of γ is described by the relative Van-der-Waals volume and surface area of the pure molecule, r_s and q_s , and the mixture, V_s and L_s ,

$$\log (\gamma_s^C) = 1 - V_s^{(3/4)} + \log (V_s^{(3/4)}) - 5q_s \cdot \left(1 - \frac{V_s}{L_s} + \log \left(\frac{V_s}{L_s} \right) \right). \quad (2.32)$$

Here, the surface area, q_s , and the volume, r_s , of the pure species are described by

$$q_s = \sum_{i=1}^N n_{i,s} \cdot Q_i, \quad (2.33)$$

$$r_s = \sum_{i=1}^N n_{i,s} \cdot R_i. \quad (2.34)$$

The surface area, L_s , and the volume, V_s , of the mixture can be calculated with

$$L_s = \frac{q_s}{\sum_{s=1}^{N^s} x_s \cdot q_s}, \quad (2.35)$$

$$V_s = \frac{r_s}{\sum_{s=1}^{N^s} x_s \cdot r_s}, \quad (2.36)$$

where N^s is the number of components in the mixture.

Because the residual part of the activity coefficient is described by the interactions between the main groups, it is necessary to know that the UNIFAC model distinguishes between main groups \bar{i} and subgroups i . As an example, the alkyl groups CH₃, CH₂, and CH are subgroups of the main group CH₂, where each sub group i has its own values of R_i and Q_i , but the group interaction parameters $a_{i,j}$, $b_{i,j}$, and $c_{i,j}$ are identical within the CH₂ main group. The residual part is defined as

$$\log(\gamma_s^R) = q_s \cdot (1 - \log(L_s)) - \sum_{\bar{i}=1}^{N^{\text{main}}} \left(\frac{h_{\bar{i},s}}{\eta_{\bar{i}}} \cdot \vartheta_{\bar{i}} - G_{\bar{i},s} \cdot \log\left(\frac{h_{\bar{i},s}}{\eta_{\bar{i}}}\right) \right), \quad (2.37)$$

where the relative Van-der-Waals surface areas of the main groups within the molecule are described by

$$G_{\bar{i},s} = \sum_{i \in \bar{i}} n_{i,s} \cdot Q_i, \quad (2.38)$$

the relative Van-der-Waals surface areas of the main groups are summed up and weighted with regard to the mixture composition,

$$\vartheta_{\bar{i}} = \sum_{s=1}^{N^s} x_s \cdot G_{\bar{i},s}, \quad (2.39)$$

and they are used to weight group interactions,

$$h_{\bar{i},s} = \sum_{\bar{j}}^{N^{\text{main}}} G_{\bar{j},s} \cdot \tau_{\bar{j},\bar{i}}, \quad (2.40)$$

which are weighted with regard to the mixture composition,

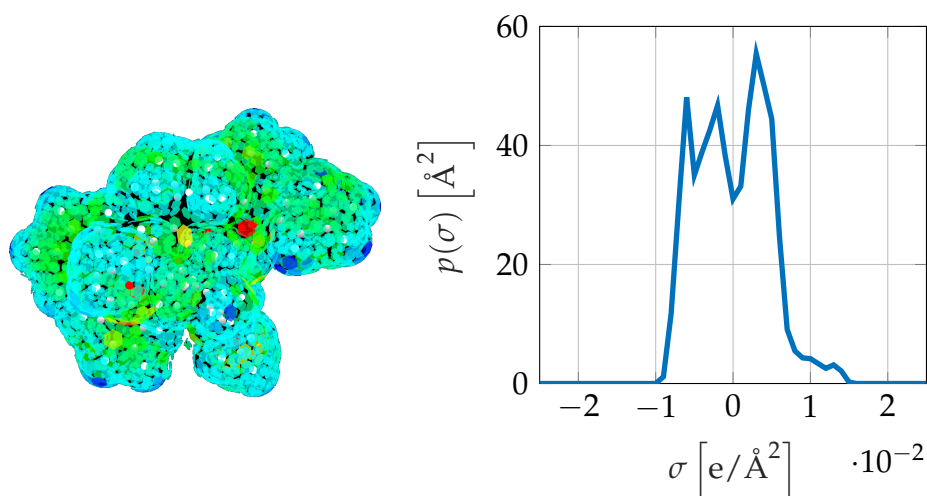
$$\eta_{\bar{i}} = \sum_{s=1}^{N^s} x_s \cdot h_{\bar{i},s}. \quad (2.41)$$

The temperature dependent group interactions are calculated with regressed group interaction parameters a , b , and c ,

$$\tau_{\bar{i},\bar{j}} = \exp \left(\frac{-a_{\bar{i},\bar{j}} - b_{\bar{i},\bar{j}} \cdot T - c_{\bar{i},\bar{j}} \cdot T^2}{T} \right). \quad (2.42)$$

Model parameters a , b , c , Q , and R for various groups can be found in the literature (Gmehling et al., 1993).

The UNIFAC method is a valuable tool for calculating phase equilibria in the absence of experimental values. However, for large molecules, the approximation quality of group-contribution methods is poor due to their assumption of additivity (Fredenslund et al., 1975). The contributions of the groups are assumed to be independent of each other. The larger a molecule, the more this assumption becomes incorrect (Struebing, 2011). Therefore, quantum chemical based methods are preferable for the description of large molecules.



(a) Cavity of charged segments of BiPhePhos as calculated using (Bell et al., 2020).

(b) σ -profile of BiPhePhos.

Figure 2.4: Example of a cavity of charged segments (a) and of a σ -profile (b). BiPhePhos is the shown molecule.

2.4 CONDUCTOR LIKE SCREENING MODEL

A quantum chemical approach for representing molecules as a cavity of charged segments in an ideal conductor is the conductor-like screening model (COSMO). The histogram of these charged surface segments σ is known as the σ -profile $p(\sigma)$. The segment charge cavity and σ -profile of BiPhePhos are shown in Figure 2.4 as an illustrative example. BiPhePhos is the ligand used in the following case studies.

The σ -profile can be seen as the fingerprint of a molecule, because it is unique. It can be used to determine the thermodynamic properties of single species or mixtures according to the COSMO-RS theory (Klamt et al., 2010).

A thermodynamic model based on σ -profiles called COSMO-SAC (segment activity coefficients) was proposed by Lin and Sandler (2002);

it will be briefly explained in the following. In this model, the activity coefficient of species i in mixture s , $\gamma_{i/s}$, is calculated by

$$\ln \gamma_{i/s} = \sum_{\sigma_m} p_i(\sigma_m) \cdot [\ln \Gamma_s(\sigma_m) - \ln \Gamma_i(\sigma_m)] + \ln \gamma_{i/s}^{\text{SG}}, \quad (2.43)$$

where $p_i(\sigma_m)$ is the charge density p of species i on segment σ_m , Γ are the activity coefficients of the segments, and $\gamma_{i/s}^{\text{SG}}$ is the Stavermann-Guggenheim combinatorial term (Stavermann, 1950; Guggenheim, 1952),

$$\ln \gamma_{i/s}^{\text{SG}} = \ln \left(\frac{\phi_i}{x_i} \right) + \frac{z}{2} q_i \cdot \ln \left(\frac{\Theta_i}{\phi_i} \right) + l_i - \frac{\phi_i}{x_i} \sum_j x_j l_j. \quad (2.44)$$

The so-called coordination number z is set to 10. The other variables are defined by

$$q_i = \frac{A_i^{(\text{q})}}{\sum_j x_j A_j^{(\text{q})}}, \quad (2.45)$$

$$r_i = \frac{V_i^{(\text{r})}}{\sum_j x_j V_j^{(\text{r})}}, \quad (2.46)$$

$$\Theta_i = \frac{x_i q_i}{\sum_j x_j q_j}, \quad (2.47)$$

$$\phi_i = \frac{x_i r_i}{\sum_j x_j r_j}, \quad (2.48)$$

$$l_i = \frac{z}{2} ((r_i - q_i)) - (r_i - 1), \quad (2.49)$$

where, $V_i^{(\text{r})}$ is the normalized cavity volume $V_i/66.69$ and $A_i^{(\text{q})}$ is the normalized cavity area $A_i/79.53$.

The segment activity coefficients can be calculated with

$$\ln \Gamma_s(\sigma_m) = -\ln \left(\sum_{\sigma_n} p_s(\sigma_n) \Gamma_s(\sigma_n) \exp \left(\frac{-\Delta W(\sigma_m, \sigma_n)}{kT} \right) \right), \quad (2.50)$$

$$\ln \Gamma_i(\sigma_m) = -\ln \left(\sum_{\sigma_n} p_i(\sigma_n) \Gamma_i(\sigma_n) \exp \left(\frac{-\Delta W(\sigma_m, \sigma_n)}{kT} \right) \right) - \quad (2.51)$$

The energy required to form a (σ_m, σ_n) is called “exchange energy”, it is denoted by $\Delta W(\sigma_m, \sigma_n)$ and defined as

$$\Delta W(\sigma_m, \sigma_n) = \frac{\alpha'}{2}(\sigma_m + \sigma_n)^2 + c_{\text{hb}} \max(0, \sigma_{\text{acc}} - \sigma_{\text{hb}}) \cdot \min(0, \sigma_{\text{don}} + \sigma_{\text{hb}}). \quad (2.52)$$

Following Lin and Sandler (2002), the parameters of Equation (2.52) are set to $\alpha' = 16467$ for the misfit energy, $c_{\text{hb}} = 85580$ (kcal/mol · Å⁴/e²) for the hydrogen bonding interaction, and $\sigma_{\text{hb}} = 0.0084$ e/Å² as hydrogen bonding cutoff value. σ_{don} and σ_{acc} denote the smaller and the larger value of σ_m and σ_n , respectively.

Note that there exist several different variations of the original COSMO-SAC approach. Silveira and Gonçalves Salau (2019) evaluated the performance of the original COSMO-SAC for small and medium sized molecules. They calculated LLEs with COSMO-SAC and UNIFAC-LLE and compared the results with experimental data. It was found that COSMO-SAC is a viable option for the calculation of LLEs. Like stated above, UNIFAC cannot be used to calculate the LLE of a mixture with large molecules due to some model assumptions. However, this limitation does not hold for COSMO-based alternatives.

Part II

PROCESS OPTIMIZATION

DOWNSTREAM PROCESSING DESIGN

This chapter was published in parts in Chemical Engineering Science, 197, Tobias Kefler, Christian Kunde, Kevin McBride, Nick Mertens, Dennis Michaels, Kai Sundmacher, and Achim Kienle. "Global optimization of distillation columns using explicit and implicit surrogate models", pp. 235–245 (2019).

3.1 INTRODUCTION

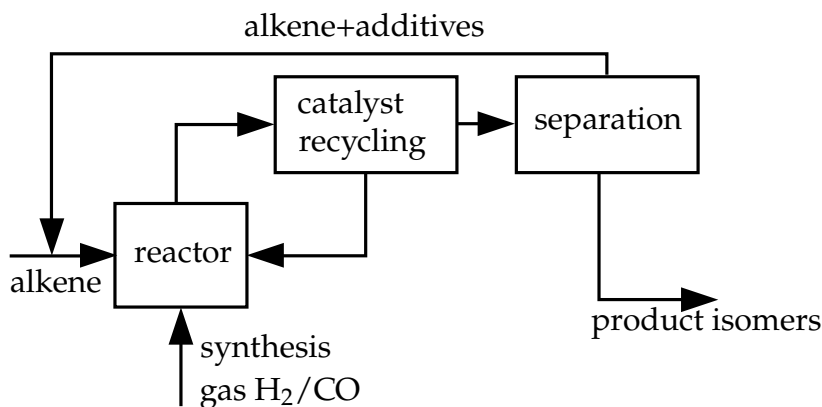


Figure 3.1: Simplified process flowsheet for the hydroformylation process.

The process considered as a case study in this work is the hydroformylation of long-chain alkenes employing a thermomorphic multiphase system to achieve a high space-time yield in the reactor and allow efficient recovery of the expensive rhodium based catalyst. A simplified process flowsheet is depicted in [Figure 3.1](#). The reactor is neglected in this chapter and will be added in [Chapter 4](#). One possible process configuration for the

catalyst recycling is an extraction cascade, consisting of several decanters for the phase separation and a distillation column for the recovery of the extraction solvent. During the separation step, multiple distillation columns are used to separate the product from unwanted side products and unused reactants.

Both process parts will be discussed in the following.

3.2 GLOBAL OPTIMIZATION OF DISTILLATION COLUMNS

Due to its high economic effect, rigorous optimization of distillation columns is of significant interest for chemical process engineers. Mixed-integer nonlinear programs (MINLP) need to be solved because of discrete and continuous decision variables in these optimization problems. As already stated in [Section 2.1](#), using local or stochastic optimization approaches bears the possibility of being trapped in a suboptimal local solution. Therefore, branch-and-bound-based deterministic global optimization methods are an appealing alternative for solving these optimization problems. Although significant progress was made in deterministic global optimization within the last years, general-purpose global optimization software is often unable to solve optimization problems interesting in the engineering point of view fastly (Nallasivam et al., 2016).

Ballerstein et al. (2015) have shown that exploiting problem-specific properties allows for substantial computational effort reductions. They derived bound-tightening strategies based on monotonous concentration profiles within ideal binary distillation columns. Mertens et al. (2018) built upon this and extended the approach to ideal multicomponent columns by introducing an innovative state transformation to obtain monotonic state variables again. In general, however, as shown in (Keßler et al., 2018), it is not possible to expand this methodology to non-ideal mixtures.

Numerous solution methods have been suggested if an effective method for determining the original model's global optimum is not applicable. Nallasivam et al. (2016) provided an algorithm based on a shortcut model to obtain minimum energy requirements for configurations of thermally coupled distillation columns. However, the approach is limited to ideal

mixtures under minimum reflux conditions. Another approach based on surrogate models was presented by Quirante et al. (2015). They use a Kriging model of a distillation column to reduce the computational effort in local and deterministic global optimization. Artificial neural networks in conjunction with support vector machines are used by Ibrahim et al. (2018) for the optimization of a petroleum refining process using a genetic optimization algorithm. A hybrid modeling approach was proposed by Eason and Biegler (2016), where Kriging models are used to replace computationally expensive process parts in a novel trust-region-based process optimization method for NLPs. As stated earlier, the choice for the correct surrogate model is non-trivial. To overcome this problem, Audet et al. (2018) proposed an error metric approach to assist in the selection of the best surrogate. Polynomial surrogate models were used by Kieslich et al. (2018) for an iterative multi-start optimization method.

In most applications, the computational complexity of optimization problems can be explained by specific computationally expensive correlations. Replacing these correlations by surrogate models is a promising approach to reduce said computational complexity. The method presented in this section utilizes Kriging models for this purpose. They are iteratively solved to global optimality, where the search space is reduced in each iteration. A local optimization method is used to obtain a feasible solution for the original model.

3.2.1 Definition of the Optimization Problem

Optimization problems of distillation columns lead to MINLP problems of the following form:

$$\begin{aligned}
 & \min && J(\tilde{\mathbf{x}}), \\
 & \text{s. t.} && \mathbf{h}(\tilde{\mathbf{x}}) = 0, \\
 & && \mathbf{g}(\tilde{\mathbf{x}}) \leq 0, \\
 & && \tilde{\mathbf{x}} \in G, \quad G \subseteq \mathbb{R}^n, \\
 & && \tilde{x}_i \in Z, \quad Z \subseteq \mathbb{Z}, \quad \text{for all } i \in I,
 \end{aligned}
 \tag{OP}$$

where J is the objective function, x are the degrees of freedom, and $h(\tilde{x})$ and $g(\tilde{x})$ are the equality and inequality constraints, respectively. A subset of \tilde{x} , \tilde{x}_i , needs to fulfill integrality restrictions, i. e. it may only take discrete values. Mass balances and inequality constraints are good examples for equality and inequality constraints, reflux and molar product stream for degrees of freedom, and the number of stages for degrees of freedom with integrality restrictions.

The degrees of freedom, \tilde{x} , can either be interpreted as model input or designated model output. Therefore, \tilde{x} is split into two parts, denoted by x (input) and y (output), i.e. $\tilde{x} \hat{=} (x, y)$. Note that, depending on the problem, different choices of x and y out of \tilde{x} may be possible. This splitting is necessary for the following definition of the surrogate problem.

To approximate Problem OP by an easier to solve problem, the equality constraints $h(\tilde{x})$ are reorganized and split into two parts: a surrogate function $\hat{h}(x, y)$ to capture computationally expensive equality constraints, and the computationally less expensive equality constraints $\bar{h}(x, y) = 0$. The variables x are scaled and bounded onto the interval $[0, 1]$ to prevent extrapolation of the surrogate function.

As a result of the reorganization and scaling, the optimization problem becomes

$$\begin{aligned}
 \min \quad & J(x), \\
 \text{s. t.} \quad & \hat{h}(x, y) = 0, \\
 & \bar{h}(x, y) = 0, \\
 & g(x, y) \leq 0, & \text{AP} \\
 & x \in S, \quad S \subseteq [0, 1]^m, \\
 & x_i \in Z, \quad \text{for all } i \in I, \\
 & y \in K, \quad K \subseteq \mathbb{R}.
 \end{aligned}$$

It is assumed that all functions are continuous and that there exists a feasible solution to both Problems OP and AP.

3.2.2 Algorithmic Implementation

The optimization algorithm used in the following is implemented into an automatic framework in Matlab and calls GAMS/BARON (Kılınç and Sahinidis, 2018) as a global sub-solver. Its pseudo-code can be found in Algorithm 1. A detailed description is given in the following.

As already mentioned earlier, the number N and choice of reference points \bar{x} has a significant influence on the surrogate function's accuracy. As a standard method, space-filling approaches are used for sampling point generation. In this work, a Halton sequence (Kocis and Whiten, 1997) is used. In contrast to a random distribution of sampling points, space-filling approaches prevent clustering. As there only exist N sampling points to capture the original function's behavior, clustering may become a problem because certain essential features of the search space S , such as extreme points, are possibly neglected.

Two Halton sets, C and D , are generated for each reference point \bar{x}^k , where C captures the continuous and D the discrete variable domain. Because Halton sets are not designed to generate integer-valued fields, the entries in D are rounded to the nearest integer value. The reference set $X = C \times D$ consists of each combination of discrete and continuous points included in C and D . Additionally, to prevent the surrogate model from extrapolating, the corners of the definition domain are also included in the reference set.

Each set of integer variables defines a so-called hypersurface. Although many sampling points are necessary for this sampling technique, it leads to smooth hypersurfaces. If a hypersurface is not smooth, local minima occur, and the computational effort of global optimization increases.

In the next step, the original function $f(x)$ is evaluated at the sampling point locations \bar{x}^k . The corresponding outputs are stored in the output matrix Y . Both, \bar{x}^k and Y , are scaled onto the interval $[0, 1]$.

After this step, all necessary prerequisites to calculate a Kriging interpolation model as described in Section 2.2.1 are available. The Kriging parameters p and θ , as well as the mean of the outputs \bar{z} , i. e. the Kriging

Algorithm 1 Optimization Algorithm

```

1: Input: Optimization problem OP. Desired number of reference points
    $N$ .
2: Output: solution vector  $\mathbf{x}^{\text{opt}}$ 
3: set initial sampling region  $S = [0, 1]^m$  (corresponds to the initial
   domain for Problem AP)
4: set reference input set  $X = \{0, 1\}^m$  (all vertices of  $S$ )
5: set relative improvement  $\varepsilon = 1$ 
6: set iteration counter  $i = 0$ 
7: repeat
8:   calculate  $i = i + 1$ 
9:   use MATLAB function HALTONSET( $N, S$ ) to generate reference
   inputs  $\bar{\mathbf{x}}^1, \dots, \bar{\mathbf{x}}^N$ 
10:  for all  $\bar{\mathbf{x}} \in (X \cap S) \cup \{\bar{\mathbf{x}}^1, \dots, \bar{\mathbf{x}}^N\}$  do
11:    evaluate  $f(\bar{\mathbf{x}})$ 
12:    use MATLAB/FMINCON to solve maximum likelihood for optimal
   Kriging parameters
13:    generate surrogate function  $\hat{f}$  and resulting Problem AP
14:    if exist( $\mathbf{x}^{\text{opt}}$ ) then
15:      set optcr =  $tol_2$ 
16:      call GAMS/BARON to solve AP for  $\bar{\mathbf{x}}^{\text{opt},i}$  with initial value
    $\mathbf{x}_0 = \mathbf{x}^{\text{opt}}$ 
17:    else
18:      set optcr =  $tol_1$ 
19:      call GAMS/BARON to solve AP for  $\bar{\mathbf{x}}^{\text{opt},i}$  without initial value
    $\mathbf{x}^{\text{opt},i}$ 
20:    use MATLAB function FMINCON to solve OP for  $\mathbf{x}^{\text{opt},i}$  with initial
   value  $\mathbf{x}_0 = \bar{\mathbf{x}}^{\text{opt},i}$ 
21:    if exist( $\mathbf{x}^{\text{opt},i}$ ) then
22:      add  $\mathbf{x}^{\text{opt},i}$  to sampling set  $X = X \cup \{\mathbf{x}^{\text{opt},i}\}$ 
23:      if exist( $\mathbf{x}^{\text{opt}}$ ) then
24:        if  $J(\mathbf{x}^{\text{opt},i}) < J(\mathbf{x}^{\text{opt}})$  then
25:          calculate  $\varepsilon = |1 - J(\mathbf{x}^{\text{opt},i}) / J(\mathbf{x}^{\text{opt}})|$ 
26:          set new best known solution  $\mathbf{x}^{\text{opt}} = \mathbf{x}^{\text{opt},i}$ 
27:          define new sampling region  $S$  (contracting the domain
   for Problem AP with  $1/(2 \cdot i)$ , centered around  $\mathbf{x}^{\text{opt}}$ )
28:        else
29:          set new best known solution  $\mathbf{x}^{\text{opt}} = \mathbf{x}^{\text{opt},i}$ 
30:        else
31:          add  $\bar{\mathbf{x}}^{\text{opt},i}$  to sampling set  $X = X \cup \{\bar{\mathbf{x}}^{\text{opt},i}\}$ 
32:  until iteration  $i == iter_{\text{max}}$  or  $\varepsilon \leq reltol$ 

```

trend, and the interpolation matrix F , are calculated and written into a text file for later use by a GAMS call.

The surrogate problem **AP** is solved to global optimality using the deterministic global optimization software GAMS/BARON. The globally optimal solution of iter i is denoted by $\bar{x}^{\text{opt},i}$. Because the surrogate problem **AP** is only an approximation of the original problem **OP**, the two problems' feasible set may differ. As a result, $\bar{x}^{\text{opt},i}$ may be suboptimal or infeasible for the original model. The original model is solved using a local optimization approach with $\bar{x}^{\text{opt},i}$ as the starting point to account for this mismatch, yielding a feasible and locally optimal solution $x^{\text{opt},i}$.

The objective function value of the feasible and locally optimal solution $J(x^{\text{opt},i})$ is compared to the best-known solution x^{opt} to check whether the new solution is smaller. If so, the best-known solution is replaced by $J(x^{\text{opt},i})$, and it is added to the set of reference points X .

Regardless of the result of this comparison, the best-known solution, x^{opt} , is used as a starting point for GAMS/BARON to reduce the initial search space of the optimization problem. The search space reduction provides an initially available upper bound, thereby significantly reducing the computational effort. For the calculation of this upper bound, an additional optimization in the vicinity of x^{opt} is necessary due to minor model variations between the Matlab and GAMS implementations. This vicinity is defined as $\pm 5\%$ in the continuous variables.

The iterative optimization procedure is based on (Caballero and Grossmann, 2008). The search space S is contracted around the best-known optimal solution x^{opt} . To increase the approximation quality of the surrogate model in each iteration, the sampling region S is shrunk and centered around x^{opt} . Consequently, a new sampling set has to be generated in each iteration. Furthermore, the global optimization solver's error tolerance is reduced after the first iteration to account for the increased accuracy. The error tolerance is quantified by the optimality gap, which can be set in GAMS/BARON via the parameter "optcr".

The algorithm is stopped if a predefined number of iterations is reached or if the relative improvement of two successive feasible solutions drops below a predefined threshold.

Table 3.1: Parameters of Algorithm 1

Parameter	Description
$iter_{\max} = 5$	highest allowable iteration count
$\varepsilon = 0.02$	relative improvement threshold
$tol_1 = 0.1$	optimality gap first iteration
$tol_2 = 0.01$	optimality gap successive iterations

Table 3.2: Chemical species within the mixture

No.	Name
1	Dimethylformamide
2	Decane
3	Dodecane
4	Dodecene
5	Tridecanal

The parameters of Algorithm 1 are defined in Table 3.1.

3.2.3 Column Specifications

Distillation processes aim to separate the most volatile mixture components from the less volatile mixture components. In most cases, specific purity requirements for the product have to be fulfilled. From an economic point of view, it is crucial to reduce the total annualized cost (TAC) of the process, which usually consists of investment and operating costs. An illustration of a distillation column can be found in Figure 3.2.

The mixture considered in this work consists of five components listed in Table 3.2. Note that n- and iso- species of the components occur in the hydroformylation process, which can be treated as single components in this model due to their similar boiling point.

Here, tridecanal is the desired product of the separation. As it is the highest boiling component, it will gather at the bottom of the column.

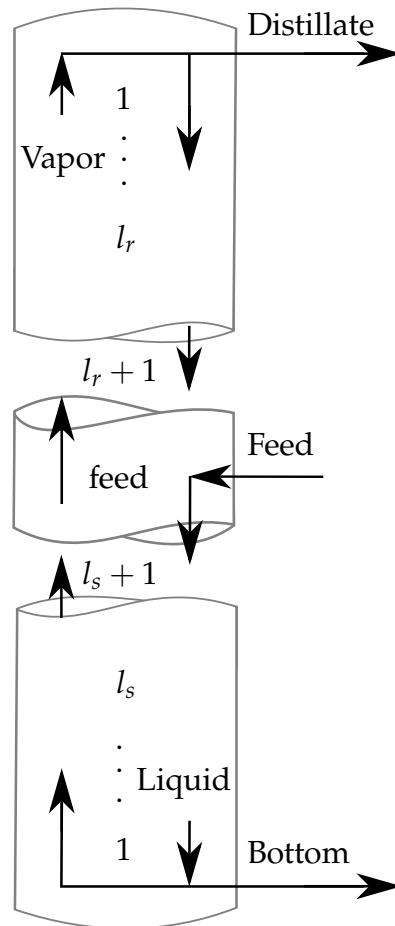


Figure 3.2: Distillation column scheme with variable number of stages in the upper and lower part of the column, i. e. in the rectifying and stripping section, respectively. The feed is located at stage $l_s + 1$ or, equivalently, $l_r + 1$. The total number of stages is $l_r + l_s + 1$.

Because tridecanal is sensitive to high temperatures, vacuum distillation is employed. The column is assumed to have a pressure of 60 mbar at the top, and a linear pressure drop of 30 mbar over the column height.

Matlab is used for the model implementation of the column. A standard tray to tray model is used, see for example the appendix of (Morari and Zafiriou, 1989). For each reference point, \bar{x}^k the steady-state solutions of the dynamic model are calculated using ode15s. The model assumptions are constant molar overflow and thermodynamic equilibrium between vapor and liquid phase. Further, the column is assumed to have a single feed stage and a total condenser.

Each reference point \bar{x}^k consists of four degrees of freedom: the vapor flow rate $V \in [18, 50]$ mol/s, the bottom product flow rate $B \in [1.233, 4.133]$ mol/s, and the rectifying and stripping section's length $l_r, l_s \in [1, 15]$, respectively. Note that the length of the stripping section also defines the feed position. The feed stream is assumed to be $F = 24.233$ mol/s, and composed of $z_{\text{feed}} = [0.0634, 0.7014, 0.1533, 0.005, 0.0769]$.

The cost function is based on the cost model of McBride and Sundmacher (2015) with slightly simplifying assumptions: in contrast to the original cost function, the column's diameter is assumed to be equal in the rectifying and stripping section; simulation data are used to calculate fixed estimates for the vaporization enthalpy Δh_{vap} , the molar mass M , and the density ρ ; the cost parameters are lumped. The resulting definition of the total annualized cost is

$$TAC = \kappa_1 \cdot V + \kappa_2 \cdot (l_{\text{col}} + \kappa_3)^{\kappa_4} \cdot V^{\kappa_5} + \kappa_6 \cdot (l_{\text{col}} - \kappa_7)^{\kappa_8} \cdot V^{\kappa_9} \text{ \$a}^{-1}. \quad (3.1)$$

The cost function takes investment costs, such as the column's length and diameter, as well as operating costs, such as condenser and reboiler duty, into account. The parameters κ are given in Table 3.3. Note that the parameters are specific to the mixture.

The purity requirements for the bottom product are set to $\chi_{5,B} \geq 0.995$. This high purity of tridecanal is chosen because of the beneficial separation characteristics. This purity requirement is identical in all test cases. They only differ in the required recovery rate r in the bottom product. The

Table 3.3: Lumped cost function parameters κ

Parameter	Value	Unit
κ_1	22824	s/mol · \$/a
κ_2	2556.2	\$/a
κ_3	7.2	1
κ_4	0.81	1
κ_5	0.525	1
κ_6	163.5543	\$/a
κ_7	0.8	1
κ_8	0.97	1
κ_9	0.725	1

scenarios are: (a) a recovery rate of 95% tridecanal, and (b) a recovery rate of 99% tridecanal.

Additional tuning parameters are the number and location of the reference points. Their choice mainly depends on the complexity of the hypersurfaces of the original function. As described in [Section 3.2.2](#), the reference points in this work are distributed using a space-filling Halton sequence. The number of reference points is heuristically set to 1175 in all case studies, consisting of 47 points in the continuous domain and 25 points in the discrete variable domain. These choices should be reevaluated for different applications, as they involve a trade-off between the surrogate model's accuracy and computational efficiency.

3.2.4 *Ideal Distillation*

For the separation of an ideal mixture, the vapor-liquid equilibrium can oftentimes be accurately described using constant relative volatilities α_i . For the specific mixture and vacuum distillation considered here, the constant relative volatilities are $\alpha = [51.35, 27, 8.43, 7.7, 1]$, where the order of the components is the same as in [Table 3.2](#). A general description of how to derive constant relative volatilities is given in [Equation \(4.8\)](#).

Table 3.4: Computation results using an ideal thermodynamic model. Test instance (a): 95 % recovery of tridecanal in the bottom product.
Left side: GAMS results and Kriging model outputs - right side: original model outputs

Iteration	B	V	l_r	l_s	TAC	time	$\chi_{5,B}$	r	$\chi_{5,B}$	r
1	1.777	25.633	2	8	748582(+17.85%)	921	1	0.9537	1.0000	0.9537
1*	1.779	22.481	2	8	665353(+4.75%)				0.9950	0.9500
2	1.779	22.481	2	6	650155(+2.36%)	14	1	0.9548	0.9948	0.9498
2*	1.779	22.486	2	6	650286(+2.38%)				0.9950	0.9500
3	1.779	22.486	2	6	650287(+2.38%)	8	1	0.9548	0.9950	0.9500
3*	1.779	22.486	2	6	650286(+2.38%)				0.9950	0.9500
Reference	1.779	22.817	1	4	635182				0.9950	0.9500

The optimization results for test instance (a) are shown in Table 3.4. The table is split into two parts, separated by a vertical bar. From left to right, the obtained optimal solution, the computation times (given in seconds), and the Kriging model's outputs are shown on the bar's left side, whereas the right side resembles the original model's outputs using the obtained optimal solution as input. The described local optimization runs are marked with an asterisk.

A reference solution is given in the last row of the table, labeled by "Reference". This solution is obtained by a rigorous global optimization of the ideal distillation column model. Mertens et al. (2018) have shown that ideal distillation column models can be solved to global optimality efficiently by employing specific model reformulations and bound tightening strategies. Mertens et al. implemented these methods in a SCIP 5.0 framework and solved the corresponding rigorous distillation column model to global optimality.

The first iteration yields a solution nearly 18 % more expensive than the reference solution. Both specifications, the purity requirement, and the recovery rate are over-fulfilled. It is noteworthy that the Kriging model's output matches the original model's output, as it indicates that the surrogate model's accuracy is high in the region of the solution. The subsequent local optimization can reduce the TAC notably; its solution lies on the desired specifications. The cost is reduced by 2.39 % in the second iteration, resulting in a third iteration. The algorithm is terminated

Table 3.5: Computation results using an ideal thermodynamic model. Test instance (b): 99 % recovery of tridecanal in the bottom product.
Left side: GAMS results and Kriging model outputs - right side: original model outputs

Iteration	B	V	l_r	l_s	TAC	time	$\chi_{5,B}$	r	$\chi_{5,B}$	r
1	1.854	26.299	4	8	782383(+14.13%)	1228	1	0.9950	1.0000	0.9950
1*	1.854	23.689	4	8	712698(+3.97%)				0.9950	0.9900
2	1.854	23.689	4	5	689538(+0.59%)	21	1	0.9950	0.9946	0.9897
2*	1.854	23.713	4	5	690175(+0.68%)				0.9950	0.9900
3	1.854	23.713	4	5	690175(+0.68%)	10	1	0.9950	0.9950	0.9900
3*	1.854	23.713	4	5	690175(+0.68%)				0.9950	0.9900
Reference	1.854	24.149	3	4	685507				0.9950	0.9900

after the third iteration because the second and third iteration's solutions are identical. The algorithm's final solution is 2.38 % more expensive than the reference solution obtained via rigorous global optimization.

The results for test case (b) are presented in Table 3.5. The table is structured identically to Table 3.4. Similar to test case (a), the first iteration is far off the reference solution (+14.13 % TAC). This discrepancy is lowered by the local optimization tremendously by roughly 10 percentage points. A further reduction of the cost is achieved in the second iteration. Note that the requirements are not met with the original model, leading to a slight cost increase in the subsequent local optimization step to achieve the fulfillment. The algorithm is terminated after the third iteration because it yields the same result as the second iteration. The algorithm's final solution is 0.68 % more expensive than the reference solution obtained via rigorous global optimization.

This case study aimed to show the suitability of the proposed Kriging-based optimization approach concerning solution quality. The algorithm's solution lies in close vicinity to the reference solution obtained via rigorous global optimization. It deviates by 2.4 % for test instance (a) and by 0.7 % for test instance (b). Although this case study is not meant to compare computation times, they are given as a total of 943 seconds for instance (a), and 1259 seconds for instance (b).

3.2.5 Moderately Non-ideal Distillation

In a next step, the vapor-liquid equilibrium (VLE) is modeled assuming an ideal gas phase and a non-ideal liquid phase. The behavior of the ideal gas phase is modeled using Dalton's law

$$P_{\text{vap}} = \mathbf{Y} \cdot P, \quad (3.2)$$

where \mathbf{Y} are the vapor mole fractions, P is total pressure, and P_{vap} are the partial pressures. The non-ideal liquid phase is described by the extended version of Raoult's law

$$P_{\text{vap}} = \chi \cdot \gamma \cdot P_{\text{vap}}^*, \quad (3.3)$$

where χ are the liquid mole fractions, P_{vap}^* are the vapor pressures, and γ are the activity coefficients, which account for non-ideal molecule interactions within the liquid phase. The vapor pressures are calculated using the extended Antoine equation

$$\log_{10}(P_{\text{vap}}^*) = \mathbf{a}_1 + \mathbf{a}_2 \cdot T^{-1} + \mathbf{a}_3 \cdot \log_{10}(T) + \mathbf{a}_4 \cdot T + \mathbf{a}_5 \cdot T^2, \quad (3.4)$$

where T is the temperature in K, P_{vap}^* are the vapor pressures in mmHg, and \mathbf{a} are the Antoine parameters, which are taken from McBride and Sundmacher (2015).

As mentioned above, the non-ideal phase behavior is captured in the activity coefficients γ . Numerous thermodynamic models have been developed for their description, for example the famous UNIQUAC and Wilson models (Poling et al., 2000). In this work, a modified UNIFAC (Dortmund) model, as described in Section 2.3.2, is employed. The necessary parameters are taken from McBride and Sundmacher (2015). The ideal and non-ideal phase equilibria for the binary mixtures DMF-Decane and Dodecane-DMF are depicted in Figure 3.3 to illustrate their differences. The ideal VLE is calculated with $\gamma = 1$, whereas the non-ideal VLE is calculated using the UNIFAC model.

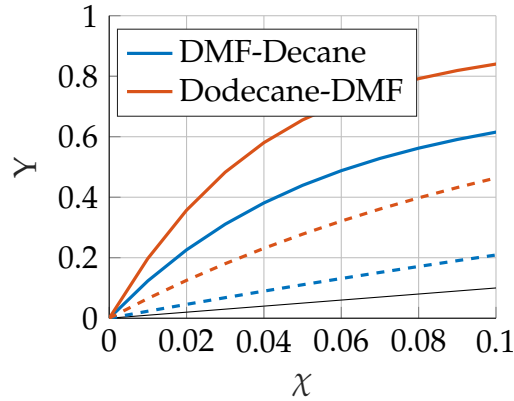


Figure 3.3: Ideal (dashed lines) and non-ideal (solid lines) binary phase equilibria for the mixtures DMF-Decane (blue) and Dodecane-DMF (orange). For the non-ideal case γ is calculated using UNIFAC, for the ideal case γ is set to 1.

Table 3.6: Computation results with VLE modeled as non-ideal. Test instance (a): 95 % recovery of tridecanal in the bottom product. Left side: GAMS results and Kriging model outputs - right side: original model outputs

Iteration	B	V	l_r	l_s	TAC	time	$\chi_{5,B}$	r	$\chi_{5,B}$	r
1	1.779	22.728	3	4	694306	1253	1	0.9500	0.9992	0.9538
1*	1.779	22.728	3	4	648854				0.9950	0.9500
2	1.779	22.728	3	4	646537	290	1	0.9500	0.9934	0.9532
2*	1.779	22.728	3	4	648854				0.9950	0.9500

The product purification column of the hydroformylation process has already been investigated in previous works, where it was assumed to be ideal (Steimel et al., 2013) and modeled using the Fenske-Underwood-Gilliland correlations (McBride and Sundmacher, 2015; McBride et al., 2017). The present case study is meant to validate these previous assumptions and demonstrate the power of the presented optimization algorithm.

As in the previous case study, Algorithm 1 is applied to solve the problem. Although global optimality cannot be guaranteed, the obtained solution is expected to lie in the vicinity of a global optimum, similar to the computational results from Section 3.2.4.

Table 3.7: Computation results with VLE modeled as non-ideal. Test instance (b): 99% recovery of tridecanal in the bottom product.
Left side: GAMS results and Kriging model outputs - right side: original model outputs

Iteration	B	V	l_r	l_s	TAC	time	$\chi_{5,B}$	r	$\chi_{5,B}$	r
1	1.852	28.705	4	4	811677	60792	1	0.9937	0.9997	0.9934
1*	1.854	23.412	4	4	674449				0.9950	0.9900
2	1.855	23.081	2	4	650082	186	1	0.9955	0.9875	0.9830
2*	1.854	24.096	2	4	676158				0.9950	0.9900

Table 3.6 summarizes the computational results for test instance (a). In the now considered non-ideal case, the problem's topology is more complex. Consequently, the computational effort to obtain a solution in the first iteration increases compared to the ideal case. The Kriging model falsely predicts that the solution lies on the specification bounds. Therefore the local optimization step reduces the cost by actually shifting the solution on said bounds. The second iteration does not yield a better solution. Hence the algorithm terminates.

Table 3.6 summarizes the computational results for test instance (b). Obtaining a solution to this problem is harder due to the stricter specifications reflected by the significantly increased computation time for the first iteration. Similar to test instance (a), the local optimization shifts the solution to lie on the specification bounds and reduces the cost. However, this time the cost reduction is more significant than in test instance (a). The second iteration result is slightly increased (0.25 %) compared to that of the first iteration. Hence the algorithm terminates.

Comparing the results of the ideal and non-ideal cases, it becomes clear that they are similar. As the optimal solution is a little bit more expensive in the ideal case, it seems like the ideal formulation is a more conservative modeling approach for this process.

3.2.6 Highly Non-ideal Distillation

The algorithm used for the previous two case studies employed a standard Kriging formulation, which can deal with smooth and continuous

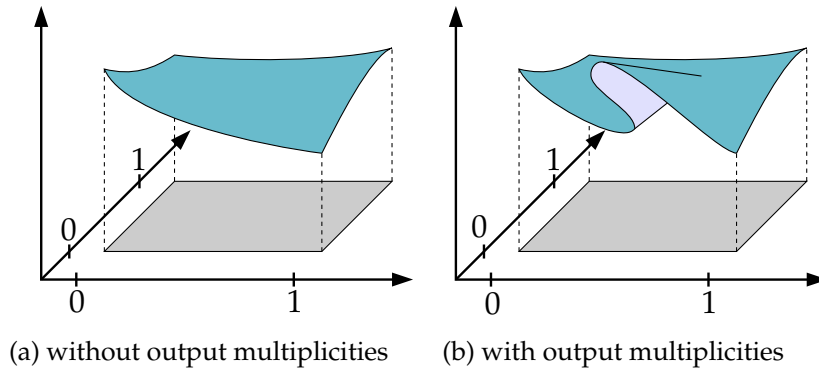


Figure 3.4: Two sketches of possible smooth and continuous response surfaces. The input set is shaded, the output is colored.

response surfaces. An example of such a response surface is shown in [Figure 3.4a](#). In the presence of output multiplicities, like shown in [Figure 3.4b](#), discontinuities, and for systems with nonlinear trends, this standard Kriging formulation behaves poorly ([Stephenson et al., 2004](#)). An implicit surrogate formulation is presented in the following to overcome the limitation regarding output multiplicities.

As already stated, standard (explicit) surrogate formulations give a specific output to a specific input $\hat{f}(x) \approx y$. The implicit surrogate formulation proposed here equals zero for feasible combinations of input and output $\hat{f}(x, y) \approx 0$. Implicit algebraic systems are already employed in bifurcation analysis ([Seydel, 2009](#)) to capture output multiplicities.

[Algorithm 1](#) needs to be slightly modified to be utilized with the implicit surrogate formulation. Owing to the different model representations, sampling, fitting, and the local optimization need to be adapted. These adaptations will be described in the following.

SAMPLING Due to output multiplicities, it is favorable to use a parameter continuation algorithm ([Rheinboldt, 1986](#); [Seydel, 2009](#)) for the sampling. Parameter continuation algorithms need valid starting points. Therefore various columns with different lengths, vapor flow, and feed stage are calculated. Then the actual parameter continuation is conducted for each of those columns.

Parameter continuation methods calculate the solution of underdetermined equation systems of the form

$$f(x, y, \omega) = 0, \quad (3.5)$$

with ω being the parameter the method is named after. ω is chosen to be the distillate product flow rate D in the present example. Solutions of Equation (3.5) are expressed as $\boldsymbol{\mu} = [x, y, \omega]$. A valid starting point $\boldsymbol{\mu}^{(k)} = [x^{(k)}, y^{(k)}, \omega^{(k)}]$ is necessary. From there, the free parameter is varied. Note that the perturbation may be positive or negative. This procedure is repeated until a predefined ω_{end} is reached. Standard continuation methods employ iterative predictor-corrector approaches, as depicted in Figure 3.5.

For the *tangent predictor* (Seydel, 2009), a tangent at solution $\boldsymbol{\mu}^{(k)}$ is calculated by solving

$$\mathbf{F}_{\boldsymbol{\mu}} \left(\mathbf{f} \left(\boldsymbol{\mu}^{(k)} \right) \right) \mathbf{z}^{(k)} = 0, \quad (3.6)$$

where $\mathbf{F}_{\boldsymbol{\mu}} \in \mathbb{R}^{m+1 \times m+2}$ is the Jacobian and $\mathbf{z}^{(k)}$ is the tangent vector. The Jacobian has to be augmented by a normalizing equation to fix the length and orientation of the tangent vector. With $\mathbf{e}_j^{\top} \mathbf{z}^{(k)} = 1$ as normalizing equation the tangent vector can be calculated with

$$\mathbf{z}^{(k)} = \begin{pmatrix} \mathbf{F}_{\boldsymbol{\mu}} \left(\mathbf{f} \left(\boldsymbol{\mu}^{(k)} \right) \right) \\ \mathbf{e}_j \end{pmatrix}^{-1} \mathbf{e}_{m+2}, \quad (3.7)$$

where \mathbf{e}_j is the j th unit vector. Now that the tangent vector is known, a prediction of the $(k+1)$ th solution can be described by

$$\hat{\boldsymbol{\mu}}^{(k+1)} = \boldsymbol{\mu}^{(k)} + \sigma^{(k)} \mathbf{z}^{(k)}, \quad (3.8)$$

where $\sigma^{(k)}$ denotes the step size of the prediction. The task of the predictor is to generate a good starting point for the corrector.

During the corrector step, the predicted solution $\hat{\boldsymbol{\mu}}^{(k+1)}$ is used to calculate the actual solution $\boldsymbol{\mu}^{(k+1)}$. This is done via parameterization.

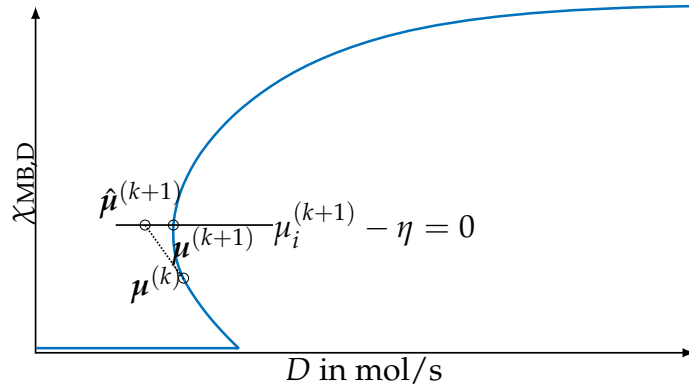


Figure 3.5: Illustration of a single parameter continuation iteration, consisting of a tangent predictor and a locally parameterized corrector step.

Various parameterization approaches exist. In this work the so-called *local parameterization* (Seydel, 2009) is utilized. The parameterized system is defined by

$$F(\boldsymbol{\mu}, \eta) = \begin{pmatrix} \mathbf{f}(\boldsymbol{\mu}) \\ \mu_i - \eta \end{pmatrix} = 0, \quad (3.9)$$

$$\hat{\mu}_i^{(k+1)} - \left(\hat{\mu}_i^{(k+1)} - \mu_i^{(k)} \right) \sigma^{(k)} = \eta,$$

where η is the additional parameter. Index i is the index of the locally most rapidly changing state. This state is most easily identified by

$$\max \left(\left| \mathbf{z}^{(k)} \right| \right) = z_i^{(k)}. \quad (3.10)$$

The parameterized system in Equation (3.9) is solved using Newton's method with $\hat{\boldsymbol{\mu}}^{(k+1)}$ as starting point.

Note that the step size of the predictor step is a tuning parameter of the parameter continuation method. As the step size choice is vital for convergence and computational efficiency, most parameter continuation methods utilize an adaptive step size control.

In contrast to dynamic simulations, the parameter continuation approach allows for the calculation of unstable solution branches. Depending on the application and specifications, the optimal solution may be located

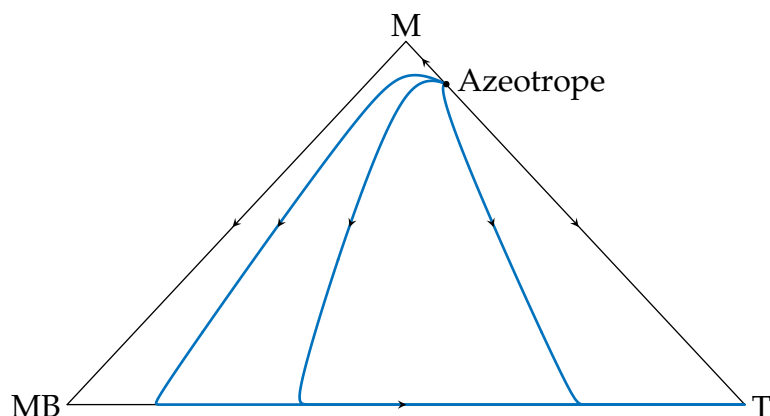


Figure 3.6: Residue curve map for the mixture methanol (M), methyl butyrate (MB), and toluol (T). The azeotrope is labeled and marked with a black dot.

in the solution space's unstable region. If the surrogate model is fitted to data from dynamic simulations, it does not capture this region.

The number of data points generated via the parameter continuation method may exceed that necessary for fitting the surrogate model because of its iterative nature. The number of sampling points is chosen to be 1145 for the present case study.

MODEL DESCRIPTION AND FITTING This case study deals with the separation of the ternary mixture of toluol, methanol, and methyl butyrate (Dorn et al., 1998). The mixture is highly non-ideal. Its residue curve map is depicted in Figure 3.6. It possesses the rare feature of an azeotrope between the heavy boiler (toluol) and the light boiler (methanol).

The original column model is implemented in Matlab, where the mixture's activity coefficients are calculated using a Wilson model. An approximation of the process' thermodynamics using constant relative volatilities is not possible because of its highly non-ideal phase behavior.

Although the azeotrope is a limiting factor concerning the product purity, it does not denote the best achievable purity. Therefore, a product specification above the azeotrope of $\sim 88\%$ methanol in the distillate is feasible. As a proof of concept, the distillate is required to have a purity of 91% methanol in this case study. Note that this product specification does not resemble the best achievable purity.

Three continuation results are shown in [Figure 3.7](#), the reader is first advised to focus on the blue line. The three plots show the liquid mole fractions of the mixture components in the distillate. The occurring output multiplicities are easy to see and manifest as hysteresis in the first two plots. The black dots denote so-called limit or turning points, i. e. points on which solution branches begin to exist (Seydel, 2009). The continuation method used to generate the plots starts with a low value of D , i. e. on the plots' left side. The purity of methanol remains at its azeotropic value of $\sim 88\%$ until the solution branch intersects with a limit point. Due to the purity requirements, this solution branch can be ignored; This renders the problem slightly more straightforward because the number of output multiplicities is reduced.

The problem possesses four degrees of freedom: the vapor flow rate $V \in [2, 13]$ mol/s, the distillate flow rate $D \in [1.1, 1.5]$ mol/s, the number of stages in the rectifying section $l_r \in [2, 25]$, and the number of stages in the stripping section $l_s \in [2, 25]$. Note that the feed position is an implicitly defined degree of freedom, which is coupled to the length of the stripping section. Degrees of freedom are denoted as x .

Parameters are the feed stream, which is assumed to be $F = 1.8$ mol/s, and the mole fractions of its composition $z_{\text{feed}} = [0.2806, 0.6566, 0.0628]$ ([toluol, methanol, methyl butyrate]).

The surrogate model takes the following variables as inputs: the degrees of freedom x , the liquid mole fraction of methyl butyrate in the distillate $\chi_{\text{MB},D}$, and the liquid mole fraction of toluol in the distillate $\chi_{\text{T},D}$. The surrogate model delivers the liquid mole fraction of methanol in the distillate and a logic variable as outputs. The logic variable equals zero for feasible solutions.

The heuristically identified kernel function

$$c(x^1, x^2) = 10 \cdot \prod_{i=1}^m 1 - |x_i^1 - x_i^2| \quad (3.11)$$

is used for fitting the surrogate model. Although this function delivers good results for the present case study, the choice should be reevaluated for different applications of the approach.

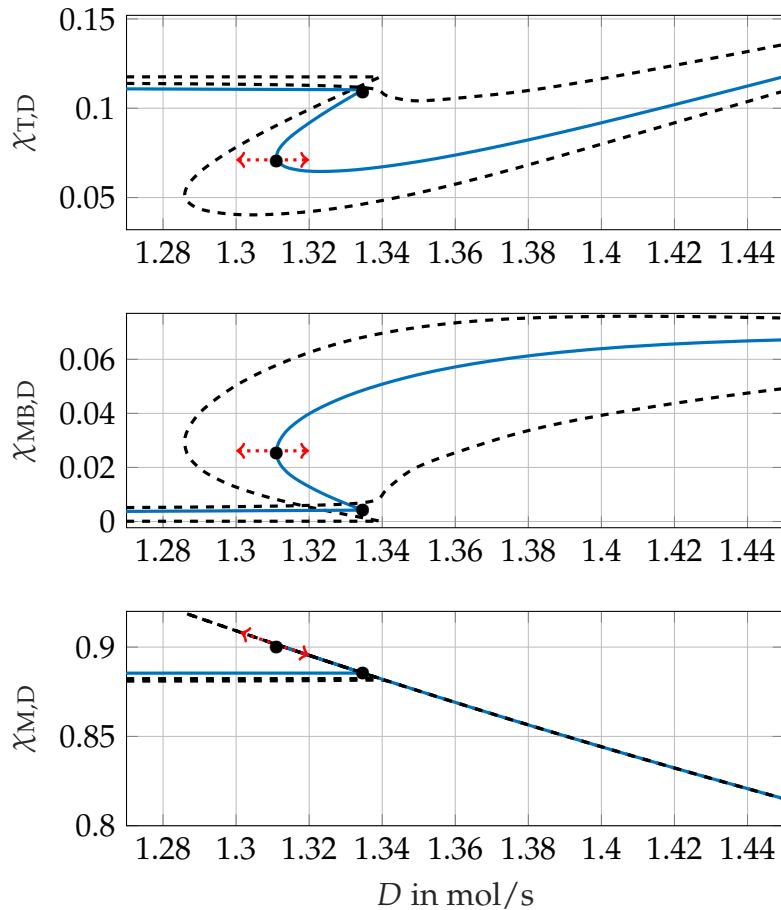


Figure 3.7: Liquid mole fractions χ of the mixture components in the distillate with respect to the distillate stream D as continuation parameter. Three continuation results are shown, one as a blue line and two as dashed black lines. The limit or turning points are denoted by black dots. The red arrows indicate the influence of two of the column's design parameters: the vapor flow rate V and the position of the feed stage. For increasing V or decreasing feed stage position, the limit point will follow the arrow pointing to the left. The arrow pointing to the right indicates the changes for decreasing V or increasing feed stage position. The dashed black lines are meant to illustrate this behavior.

LOCAL OPTIMIZATION Analogous to the case studies presented above, a local optimization step is employed. Its task is to calculate a feasible solution to the original problem and provide a feasible upper bound for the global optimization step. The model exhibits exploitable characteristics for a line search based local optimization. As depicted in [Figure 3.7](#), changes in feed location and vapor flow rate significantly influence the distillate composition. Increasing the vapor flow rate or shifting the feed stage's position shifts the marked turning point in the direction of the left arrow, i. e. the purity of methanol in the distillate is increased. Reducing the vapor flow rate or shifting the location of the feed stage upwards has the opposite effect. Therefore, shifting the feed stage location is the most cost-effective way to obtain a purer product, whereas reducing the vapor flow rate is the most cost-effective way to reduce the purity. A purity reduction may be necessary in case of an over fulfillment of the product specifications. Note that these characteristics may only be used to an extent.

For the given reasons, a tailor-made line search algorithm is used for the local optimization step in the present case study. Note that, for the same reasons, the optimal operating conditions always coincide with the turning point marked with arrows in [Figure 3.7](#). Hence the algorithm needs to identify the turning point, which is done using an underlying parameter continuation.

NUMERICAL RESULTS The numerical results are summarized in [Table 3.8](#). The vaporization rate is increased, and the feed location is shifted downwards to achieve the product specifications. Due to the increased vaporization rate, the process becomes more expensive during the local optimization step. The column becomes larger in the second iteration, which increases the investment costs, but the vaporization rate can be reduced, which results in a reduction of the overall costs. The local optimization step only needs to correct the solution slightly to achieve the required product purity. The algorithm terminates after the third iteration because the solutions of the third and second iterations are identical.

Table 3.8: Computation results for heavily non-ideal distillation.
 Left side: GAMS results and surrogate model outputs - right side:
 original model outputs

Iteration	D	V	l_r	l_s	TAC	time	$\chi_{M,D}$	$\chi_{M,D}$
1	1.29	2.25	14	6	115263	671	0.918	0.9
1*	1.3	2.7	15	5	132210	306		0.91
2	1.3	2.34	19	5	126495	1470	0.91	0.909
2*	1.3	2.387	19	5	128340	275		0.91
3	1.3	2.34	19	5	126497	4635	0.91	0.909
3*	1.3	2.385	19	5	128340	277		0.91

3.3 DECANTER CASCADE

The second part of the downstream processing is an extraction cascade consisting of a distillation column and multiple phase separators. The distillation column is necessary to recover the extraction solvent. The solvent is used in a counter-current cascade of phase separators to recover the expensive catalyst.

The catalyst recovery is crucial for an economically profitable operation of the process (McBride, 2017). Therefore, a precise phase separation model for the catalyst is necessary for a meaningful optimization of the process. As stated in Section 2.3, a group contribution based approach, e.g. UNIFAC, is not applicable here due to the size of the catalyst complex.

McBride et al. (2017) suggested using a surrogate model based on Kriging interpolation as a model for the phase separation. Data obtained from UNIFAC was used for the phase separation of the species in Table 3.2. A set of linear inequality constraints was employed to ensure that the generated data points lie within the mixture's two-phase region. The catalyst partitioning between the two phases was predicted using the commercially available software package COSMOtherm (Eckert and Klamt, 2016). Both data-sets were used for fitting the Kriging model.

This approach was further refined by Kunde et al. (2019). A parameterization-based approach was used instead of linear inequalities to model the mixture's feasible two-phase region. Two parameters were introduced to

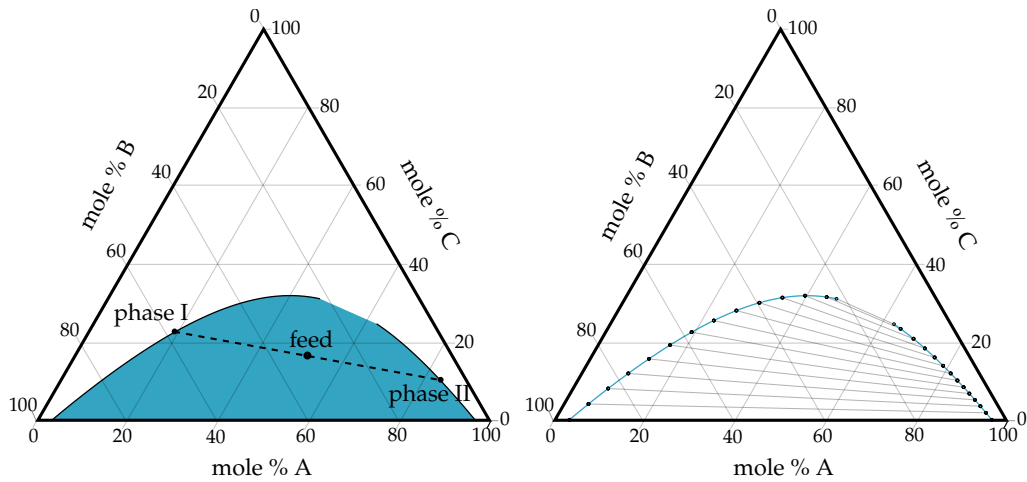


Figure 3.8: Ternary phase equilibrium of an arbitrary example mixture. On the left side, the whole area shaded in blue is necessary to describe the miscibility gap. On the right side, a parameterization of the binodal curve shaded in blue is used to describe the miscibility gap. Thereby, the dimension of the input domain is reduced by one (from plane to line).

describe the binodal curve, and a parameter continuation algorithm was employed to calculate the parameterized binodal curve. This approach reduced the dimension of the problem by one, i. e. a plane became a line. See Figure 3.8 for an illustrative example. The obtained data points were then used to train a shallow artificial neural network. With this description of the phase separators' thermodynamics, the extraction cascade could be solved to global optimality within minutes.

In the remainder of this work, the phase separators will be described as in the work by Kunde et al. (2019). A detailed description is given in Chapter 4.

3.4 CONCLUSION

This chapter presents an algorithm for the deterministic global optimization of distillation columns using surrogate models. The surrogate models are solved to global optimality using GAMS/Baron. Local optimization techniques are used to bridge the gap between the surrogate models and the original model. The iterative approach makes use of the best identified

solution to obtain a feasible upper bound for the global optimization, which results in a reduction of computational effort.

The approach is applied in two case studies: for

- (a) the product purification column of the hydroformylation process, and
- (b) the separation of methanol from toluol and methyl butyrate.

Case study (a) is additionally split into two parts. The first part assumes that the mixture's phase behavior can be described using constant relative volatilities, yielding an ideal distillation. Rigorous global optimization of ideal distillation is feasible (Mertens et al., 2018), and thus a reference solution is available. The proposed algorithm can obtain a solution that lies in a close neighborhood to this reference solution. The second part assumes that activity coefficients are necessary to describe the mixture's phase behavior. A reference solution is not available for this test case because the methods used to calculate the reference solution in the ideal case are generally not applicable to non-ideal problems. It is shown that a description of the phase behavior using constant relative volatilities is slightly more conservative and thus applicable. Both parts of case study (a) are solved using explicit surrogate model formulations.

The mixture considered in case study (b) is highly non-ideal. It possesses an azeotrope between the heavy boiler and the light boiler, resulting in steady-state multiplicities. Additionally, the product specification lies above the azeotrope. An implicit surrogate model formulation inspired by bifurcation theory is used here to overcome these problems.

It is shown that solving optimization problems regarding distillation columns, even highly non-ideal ones, to global optimality becomes tractable using the proposed method. An additional benefit of the method is that it does not rely on a good initial starting point, in contrast to local optimization strategies. Furthermore, the implicit surrogate formulation can cope with bifurcation points, whereas gradient-based local optimization approaches would fail due to the Jacobian's singularity at these points.

Concerning the process considered in this work, it became clear that the distillation column can be assumed to be ideal without conducting a

large error. Thus the computation times for optimizing the whole process can be reduced by modeling the distillation column's thermodynamic as ideal.

Furthermore, it was shown that advanced surrogate modeling based approaches exist to describe the phase behavior of phase separators.

Both findings combined make more involved overall process optimizations, as described in the following chapters, tractable.

SIMULTANEOUS REACTOR AND DOWNSTREAM PROCESSING DESIGN

This chapter was published in parts in: Processes, 7, Tobias Kefßler, Christian Kunde, Steffen Linke, Kevin McBride, Kai Sundmacher, and Achim Kienle. "Systematic Selection of Green Solvents and Process Optimization for the Hydroformylation of Long-Chain Olefines", pp. 882 (2019).

4.1 INTRODUCTION

Taking the results of [Chapter 3](#) as a foundation, the optimization problem presented in this chapter additionally includes the reactor within the process.

The difficulty of the optimization problem is increased a step further by closing all recycles. The optimization problem becomes too hard to be solved to global optimality within a reasonable amount of time, although every single part of the problem can be solved to global optimality efficiently. Therefore, a multi-start approach is used to solve the optimization problem. One major drawback of the process is the use of the polar solvent DMF, despite its toxicity. Therefore, alternative solvent candidates identified by McBride et al. (2018) are investigated regarding their process performance. The performances are compared to that of the toxic and commonly used benchmark solvent DMF.

4.2 OVERALL PROCESS OPTIMIZATION

The hydroformylation of n-decene is considered as a case study in this chapter. As in the previous chapter, a TMS is employed to achieve a high space-time yield within the reactor and allow for a good separation of the rhodium-based catalyst in a subsequent phase separator. The TMS consists of two solvents: the polar DMF and the non-polar n-dodecane (Bianga et al., 2019).

To achieve the goal of obtaining a process that can be labeled as “green chemistry”, several aspects need to be fulfilled. The optimization methods presented in the previous chapter laid the foundation since an efficiently operated process saves energy and thereby reduces the CO₂-footprint. The unsaturated oleochemicals from renewable resources used as a feedstock also help make the process ecologically benign. However, the employed polar solvent is toxic and therefore poses a high risk for the environment. In an attempt to find a good performing and non-toxic alternative, the optimization study conducted in this section also takes two additional solvent candidates into account.

4.2.1 Candidate Solvent Selection

The additional solvent candidates considered here were identified by the screening methodology proposed by McBride et al. (2018). The COSMObase database is used as the foundation of the method. It can be used in conjunction with the COSMO-RS theory (Klamt et al., 1998; Klamt et al., 2016) to predict thermodynamic properties in the liquid state.

In summary, the screening procedure consists of three parts:

1. Screening for physical properties: Some structural constraints need to be taken into account to ensure the usability of the identified solvents. Carbon double bonds are not allowed to avoid structural instabilities during the reaction. Furthermore, the boiling point is restricted due to temperature limitations within the solvent recovery unit and product purification.

2. Screening for environment, health, and safety (EHS) properties: McBride et al. (2016) identified DMF as the best performing solvent using a different screening methodology. This time the focus lies on green alternatives. Therefore, EHS properties are an essential deciding factor in the search for new solvents.
3. Evaluation of thermodynamic performance: A mixture composed of n-dodecane and a newly identified solvent candidate must exhibit the required phase behavior to form a TMS. Additionally, the solvent candidate needs to be able to extract the catalyst.

Their investigation resulted in the identification of three potential solvent alternatives. For two of them, dimethyl succinate (DSUC) and tetrahydro-4H-pyran-4-one (THPO), experiments were conducted to check their reaction feasibility and phase behavior.

As they are promising candidates, they will be compared to DMF concerning process-performance in this computational study.

4.2.2 Model Description

Figure 4.1 depicts the flowsheet of the hydroformylation process setup considered in this computational study. It consists of a continuously stirred tank reactor, an extraction cascade for the catalyst recovery, and a literal black-box for product separation and solvent/reactant recycle. Each will be described more precisely in the following. A list of chemical species present in the process is given in Table 4.1.

Reactor

The reactor model and its reaction kinetics are based on the work by Kiedorf et al. (2014) and Hentschel et al. (2015), where the hydroformylation of 1-dodecene takes place in a CSTR. The reaction occurs in a homogeneous liquid phase, and synthesis gas consisting of CO and H₂ is also present within the reactor in an additional gaseous phase.

In contrast to the original model formulations, the reactor is assumed to be operated under steady-state conditions in this study. The Perturbed-Chain Statistical Associating Fluid Theory (PC-SAFT) equation of state (Gross and Sadowski, 2001) describes the gas-liquid equilibrium. A list of all chemical species present in the process, including the new polar solvents, is given in Table 4.1.

Note that the reaction kinetics by Kiedorf et al. (2014) were meant for the hydroformylation of dodecene in a TMS with DMF as polar solvent. Nevertheless, the present study focuses on the hydroformylation of decene in a TMS with DMF and two additional polar solvents but utilizes this reaction kinetics. The error made because of the different olefin chain length is small. Additionally, McBride et al. (2016) have shown that the influence of the polar solvent on the hydroformylation is negligible.

The reactor's mass balance is given by

$$F_{\text{reac},i}^{\text{out}} = F_{\text{reac},i}^{\text{in}} + V_{\text{reac}} c_{\text{cat}} M_{\text{cat}} \sum_{j \in \text{RCT}} \nu_{i,j} r_j, \quad (4.1)$$

$$F_{\text{reac},i}^{\text{in}} = Z_i + F_{\text{product},i}^{-\text{C11al}} (1 - \omega) + F_{\text{recycle},i}, \quad (4.2)$$

where M_{cat} is the catalyst's molar mass, and c_{cat} is the concentration of the active catalyst. A total of eight reactions occur within the reactor; RCT denotes them in the summation. The eight reaction rates are represented by r , and the corresponding stoichiometric coefficients are specified by ν .

The reaction rates can be described by

$$r_j = k_j \cdot \left(\prod_i (c_i^{\nu_{j,i}}) \right) \cdot \left(1 + \sum_i (K_{i,1} \cdot c_i^{K_{i,2}}) \right)^{-1}, \quad (4.3)$$

where k are reaction rate constants derived using the Arrhenius equation, and K are reaction rate parameters from Hentschel et al. (2015).

The active catalyst's concentration c_{cat} is defined by

$$c_{\text{cat}} = \frac{c_{\text{cat,tot}}}{1 + K_{\text{cat},1} \cdot c_{\text{CO}}^{K_{\text{cat},2}}}, \quad (4.4)$$

where $c_{\text{cat,tot}}$ is the catalyst's total concentration within the reactor.

The reactor model presented here is used in [Appendix A](#) in conjunction with a phase separator model for catalyst recycling. Various techniques, such as model reformulations and bound propagations, are presented to reduce the computational effort for this process's global optimization. One example is approximating the gas-liquid equilibrium by an ANN with a piece-wise linear approximation of a hyperbolic tangent as the activation function. All techniques combined allow for global optimization of the considered model within a few minutes using GAMS/BARON. Although a rigorous global optimization is not conducted here, the modeling techniques help speed up the optimization and enable the usage of more starting points for the multi-start approach presented here.

The reactor's input $F_{\text{react}}^{\text{in}}$ consists of three parts:

1. the feed stream Z , consisting of catalyst make-up, solvent make-up, and reactants,
2. the recycle from the black-box $F_{\text{product}}^{-\text{C11al}}$, consisting of unused reactants, solvents and side-products, and
3. the recycle from the extraction cascade F_{recycle} , consisting of the recycled catalyst, unused reactants, solvents, some of the product and side-products.

Note that the recycle from the black-box is reduced by the purge ω .

Extraction Cascade

The extraction cascade, fed by the reactor's outlet $F_{\text{react}}^{\text{out}}$, consists of four phase separators and one distillation column.

DISTILLATION COLUMN The distillation column's purpose is to recover the circulated extraction solvent $F^{(psol)}$, which is the distillate product of the column. As in the previous chapter, the column is assumed to operate under vacuum due to the temperature sensitivity of undecanal. The pressure at the top is assumed to be 60 mbar, with a linear pressure drop of 50 % along the column. Following the results of [Chapter 3](#), it is

assumed to have ideal phase behavior. The distillation column can be approximated by a set of surrogate functions due to the unproblematic separation of the extraction solvent from the remaining components, as will be described in the following.

The constant relative volatilities α have to be calculated as a preliminary step. The Halton sequence approach from the previous chapter is used to generate 810 distillation columns with different design variables D , V , l_r , and l_s . They are assumed to have constant molar overflow, constant molar flows, and a total condenser. The ideal VLE is described using Dalton's law for the gas phase,

$$P_{\text{vap}} = \chi \cdot P, \quad (4.5)$$

where χ are the vapor mole fractions, P is the total pressure, and P_{vap} are the partial pressures; Raoult's law is used for the description of the liquid phase,

$$P_{\text{vap}} = \mathbf{Y} \cdot \mathbf{P}_{\text{vap}}^*, \quad (4.6)$$

where \mathbf{Y} are the liquid mole fractions and $\mathbf{P}_{\text{vap}}^*$ are the vapor pressures. The vapor pressures are calculated using the extended Antoine equation

$$\log_{10} \left(P_{\text{vap}}^* \right) = a_0 + a_1 \cdot T^{-1} + a_2 \cdot \log_{10}(T) + a_3 \cdot T + a_4 \cdot T^2, \quad (4.7)$$

with P_{vap}^* in mmHg and T in K. The parameters a_i are listed in [Table B.1](#). The constant relative volatilities can then be derived with

$$\alpha_i = \sqrt{\frac{\bar{P}_{\text{vap},B,i}^*}{\bar{P}_{\text{vap},B,\text{HK}}^*} \cdot \frac{\bar{P}_{\text{vap},D,i}^*}{\bar{P}_{\text{vap},D,\text{HK}}^*}}. \quad (4.8)$$

The vapor pressures $\bar{P}_{\text{vap},B}^*$ and $\bar{P}_{\text{vap},D}^*$ are the mean vapor pressures at the bottom and the top of the column, respectively. They are calculated using the respective mean temperatures, \bar{T}_B and \bar{T}_D , of the 810 simulation runs. The index "HK" denotes the high key component, i. e. the component with

the highest boiling point, which is undecanal in the present case study. Numerical values for α are given in [Table B.2](#).

The obtained constant relative volatilities are used in a Fenske-Underwood-Gilliland (FUG) shortcut model (Henley and Seader, 1981) of the distillation column. The shortcut model is solved to global optimality 1000 times for each solvent using GAMS/BARON (Kılınç and Sahinidis, 2018). Each of the optimization runs is conducted with a different feed stream. The purity requirements for each of the solvents are:

- DMF: 99% purity of DMF in the distillate,
- DSUC: 99% purity of DSUC + C₁₀en in the distillate, and
- THPO: 99% purity of THPO in the distillate.

Note that the purity requirement for DSUC is not as strict as for the other two solvents. This exception is necessary due to the constant relative volatilities, as shown in [Table B.2](#).

Following [Chapter 3](#), a cost function with lumped parameters is used to calculate the distillation column's TAC in \$/a,

$$J_{\text{col}} = \kappa_1 \cdot V_{\text{col}} + \kappa_2 \cdot (l_{\text{col}} + \kappa_3)^{\kappa_4} \cdot (V_{\text{col}} \cdot \kappa_5)^{\kappa_6} + \kappa_7 \cdot (l_{\text{col}} + \kappa_8)^{\kappa_9} \cdot (V_{\text{col}} \cdot \kappa_{10})^{\kappa_{11}}, \quad (4.9)$$

where $l_{\text{col}} = l_r + l_s + 1$ denotes the length of the column and V_{col} denotes the vapor flow rate. The necessary parameters κ are specific to the mixture and given in [Table B.3](#). Note that there are two additional parameters compared to [Equation \(3.1\)](#). Those parameters are 1 and were therefore omitted in [Equation \(3.1\)](#). They are given here for the sake of completeness. Additionally, the parameters for the considered mixtures differ due to differences in mean molar masses, mean densities, and mean vaporization enthalpies.

A polynomial surrogate function for the TAC is fitted for each solvent using the global optimizations data. The inputs of this surrogate functions

are the recovery rate, $rec = F^{(psol)} / F_{col,psol}^{in}$, and the extraction solvent stream $F^{(psol)}$. The functions are of the following form:

$$\begin{aligned} TAC_{col} = & \lambda_{psol,1} + \lambda_{psol,2} \cdot F^{(psol)} + \lambda_{psol,3} \cdot rec \\ & + \lambda_{psol,4} \cdot \left(F^{(psol)}\right)^2 + \lambda_{psol,5} \cdot rec^2. \end{aligned} \quad (4.10)$$

The calculated parameters λ are presented in [Table B.3](#).

PHASE SEPARATORS McBride et al. (2017) found that a cascade consisting of four phase separators in a counter-current setup is optimal concerning catalyst recovery. Therefore four phase separators are used in the present process setup. The temperature within the phase separators is fixed to 298.15 K. This temperature is low enough for the TMS to separate into two liquid phases, one of which holds most of the product and the other holds the majority of the catalyst.

The phase separator's mass balance is given by

$$F_{dec,n}^{in} = \begin{cases} L_2^{(II)} + F_{col}^{out,D}, & \text{for } n = 1, \\ L_{n-1}^{(I)} + L_{n+1}^{(II)}, & \text{for } 1 < n < n_{max}, \\ F_{reac}^{out}, & \text{for } n = n_{max}, \end{cases} \quad (4.11)$$

where $L^{(II)}$ denotes the upper, less-polar, and product-rich outlet stream of the phase separators, and $L^{(I)}$ is the lower, more-polar, and catalyst-rich outlet stream. Their respective compositions are y and x . The phase separators within the cascade are numbered according to the extraction solvent's flow from right to left, i. e. $[n_{max}, \dots, 1]$.

The solvents' thermophysical properties are not identical; therefore, the separation effort varies from solvent to solvent. The necessary separation effort can easily be seen by comparing the constant relative volatilities α . While DMF and THPO are the lowest boiling components in their respective mixtures, DSUC is not. As a result, DMF and THPO can be recovered in high purity, whereas the recovered DSUC will always be accompanied by the lowest boiling component of the mixture, decene.

This varying separation characteristics have to be taken into account in the surrogate formulation mentioned above. Therefore the distillation column's distillate flow $F_{col,i}^{out,D}$ is calculated using

$$F_{col,i}^{out,D} = \beta_i^{(psol)} F_{col,i}^{in}, \quad (4.12)$$

where F_{col}^{in} is the column's feed flow and $\beta^{(psol)}$ is a vector of split factors. The entries of $\beta^{(psol)}$ for each investigated solvent are:

$$\beta^{(psol)} = (\zeta_{psol}, C_{12an}, C_{10en}, C_{11al}, cat), \quad (4.13)$$

$$\beta^{(DMF)} = (\zeta_{DMF}, 0, 0, 0, 0), \quad (4.14)$$

$$\beta^{(DSUC)} = (\zeta_{DSUC}, 0, 1, 0, 0), \quad (4.15)$$

$$\beta^{(THPO)} = (\zeta_{THPO}, 0, 0, 0, 0). \quad (4.16)$$

The split factor $\zeta_{psol} \in [0, 1]$ is chosen based on the circulated extraction solvent stream $F^{(psol)}$ that needs to be recovered. As a high split factor leads to less solvent in the bottom of the column, it also leads to higher temperatures. High temperatures are concerning due to the product's temperature sensitivity. Dreimann et al. (2016) have shown that the unwanted side product aldol is formed if the distillation column's bottom temperature exceeds 403.15 K. The bottom temperature is constrained to temperatures equal to or below the reaction temperature of 388.15 K. As a result, the polar solvent stream entering the column needs to be larger than $F^{(psol)}$.

Another critical factor is the catalyst distribution between the two phases. The overall catalyst mass balance of the extraction cascade is defined as

$$F_{\text{reac,cat}}^{out} = L_{n_{\text{max}}}^{(I),\text{cat}} + L_{n_{\text{max}-1}}^{(I),\text{cat}} + L_1^{(II),\text{cat}}, \quad (4.17)$$

where $L^{(I),\text{cat}}$ denotes the catalyst in the more-polar catalyst-rich phase, and $L^{(II),\text{cat}}$ denotes the catalyst in the less-polar product-rich phase. The catalyst distribution itself is calculated using the partition coefficient P_{yx} :

$$\log_{10} P_{yx,n} = \log_{10} \left(\frac{L_n^{(II),\text{cat}} L_n^{(I)}}{L_n^{(II)} L_n^{(I),\text{cat}}} \right). \quad (4.18)$$

Due to the ligand's complexity and size, the partitioning has to be calculated using COSMOtherm (Klamt et al., 1998). Group-contribution based approaches, such as UNIFAC, are not suitable to describe the partitioning of the catalyst (Chen and Song, 2004). However, the distribution of the other species is calculated using a modified UNIFAC (Dortmund) (Weidlich and Gmehling, 1987) implementation to describe the activity coefficients γ . The liquid-liquid equilibrium

$$x_i \gamma_i^{(I)} - y_i \gamma_i^{(II)} = 0, \quad (4.19)$$

is solved with an accuracy of more than $1e-9$ to obtain the phase compositions.

As this procedure is computationally rather expensive, a surrogate model \hat{f} is used to describe the phase compositions during the optimization. As described in Section 3.3, the surrogate is based on the work of Kunde et al. (2019). Two parameters per phase separator, $t_1 \in [0, 1]$ and $t_2 \in [0, 1]$, are introduced to parameterize the binodal curve

$$[x_{i,n}, y_{i,n}, \log_{10} P_{y,x,n}] = \hat{f}(t_{1,n}, t_{2,n}). \quad (4.20)$$

A parameter continuation algorithm (Seydel, 2009) is used to calculate the parameterized binodal curve. The obtained data is used as a training and test set for fitting the surrogate using MATLAB's "train" command. The surrogate itself consists of a shallow artificial neural network with one hidden layer and a second-degree polynomial.

McBride et al. (2017) identified the catalyst leaching as one of the most critical cost factors of the hydroformylation process. Hence the error tolerance concerning the surrogate models' ability to resemble the catalyst distribution correctly is small. The maximum errors of the surrogate models with regard to the logarithmic partition coefficient

$\log_{10} P_{yx} = f_{psol}^{P_{yx}}(t_1, t_2)$ are given below to illustrate the surrogate models' performance.

$$\begin{aligned}\bar{\delta}_{psol} &= \max\left(f_{psol}^{P_{yx}}(t_1, t_2) - \hat{f}_{psol}^{P_{yx}}(t_1, t_2)\right), \\ \underline{\delta}_{psol} &= \min\left(f_{psol}^{P_{yx}}(t_1, t_2) - \hat{f}_{psol}^{P_{yx}}(t_1, t_2)\right), \\ \bar{\delta}_{DMF} &= 0.001087, \quad \underline{\delta}_{DMF} = -0.000684, \\ \bar{\delta}_{DSUC} &= 0.000524, \quad \underline{\delta}_{DSUC} = -0.001201, \\ \bar{\delta}_{THPO} &= 0.002302, \quad \underline{\delta}_{THPO} = -0.001656.\end{aligned}\tag{4.21}$$

Black Box

The black box represents additional downstream processing steps for the product purification. The downstream processing can be realized using various approaches. A separation of the product using a distillation column was successfully implemented by Dreimann et al. (2016) and is used in the optimization study in Appendix E. This thesis is embedded in the *Sonderforschungsbereich InPROMPT*. In the last three years, the idea of coupling the hydroformylation process with a subsequent reductive amination process (Bianga et al., 2020b; Künnemann et al., 2020) ripened, either with two subsequent reactions or in the form of a hydroaminomethylation (Bianga et al., 2020a) in a “one-pot” reaction. One of the most striking benefits of coupling these processes is that a product purification within the hydroformylation process becomes obsolete.

This chapter assumes that the aldehyde is fully separated from the other species during the reductive amination. The other species are then recycled and fed into the hydroformylation. Further, it is assumed that the catalyst leaving the hydroformylation becomes deactivated and is therefore lost.

Lumped Species

Eleven species are involved in the hydroformylation reaction, as can be seen in Table 4.1. Although they are essential for modeling the reaction,

they are not as relevant for modeling the catalyst recovery within the extraction cascade. Therefore additional model simplifications have been conducted.

1. n- and iso-components are lumped. Their ratio is assumed to be identical for reactor outlet and input.
2. dodecane and decane are lumped into one non-polar solvent for the extraction, which is assumed to behave like dodecane. Their ratio is assumed to be identical for reactor outlet and input.
3. the reactor's gaseous phase is neglected in the extraction cascade.

With these assumptions, the reactor's outlet is modeled as

$$\mathbf{F}_{\text{react}}^{\text{out}} = \begin{pmatrix} \text{psol} \\ \text{C12an} \\ \text{C10en} \\ \text{C11al} \\ \text{cat} \end{pmatrix}, \quad (4.22)$$

where *cat* denotes the rhodium/BiPhePhos complex, and *psol* denotes the investigated polar solvent candidates.

4.2.3 Definition of the Optimization Problem

The optimization problems solved in this chapter are MINLPs due to the piece-wise linear activation function of the ANN employed in the reactor. Their general form is defined as

$$\begin{aligned} \min_{\tilde{\mathbf{x}}} \quad & J(\tilde{\mathbf{x}}), \\ \text{s. t.} \quad & \mathbf{h}(\tilde{\mathbf{x}}) = 0, \\ & \mathbf{g}(\tilde{\mathbf{x}}) \leq 0, \\ & \tilde{\mathbf{x}} \in G, \quad G \subseteq \mathbb{R}^n, \\ & \tilde{\mathbf{x}}_i \in Z, \quad Z \subseteq \mathbb{Z}, \quad \text{for all } i \in I, \end{aligned} \quad (4.23)$$

Table 4.2: Degrees of freedom

name	domain	unit	description
T_r	[368.15, 388.15]	K	reactor temperature
V_{reac}	[2.1853, 4000]	m ³	reactor volume
P_r	[10, 20]	bar	reactor pressure
Z_{C10en}	[2, 3]	mol/s	C10en feed
Z_{psol}	[0, 2]	mol/s	<i>psol</i> make-up stream
Z_{C12an}	[0, 2]	mol/s	C12an make-up stream
Z_{RHO}	[0, 0.1]	mol/s	rhodium make-up stream
Z_{BIP}	[0, 0.1]	mol/s	BiPhePhos make-up stream
$F^{(\text{psol})}$	[7, 90]	mol/s	extraction solvent stream
ζ	[0, 1]	-	split factor
ω	[0.0001, 0.01]	-	purge factor

where $g(\tilde{x})$ are inequality constraints, such as purity requirements, $h(\tilde{x})$ are equality constraints, such as mass balances, $J(\tilde{x})$ is the objective function, and \tilde{x} are variables. A subset of these variables, \tilde{x}_i , may only take discrete values. The degrees of freedom are part of the variables \tilde{x} , they are given in [Table 4.2](#).

The process is required to yield 6500 metric tons of aldehyde per year, under the assumption that it operates at 220 days per year.

The objective function describes the *TAC* in \$/a and is structured in four parts:

$$\begin{aligned}
 J = & \theta_1 \sum_n \sum_i \left(\eta_i F_{\text{dec},n,i}^{\text{in}} \right)^{\theta_2} & \text{I} \\
 & + \lambda_{\text{psol},1} + \lambda_{\text{psol},2} \cdot F^{(\text{psol})} + \lambda_{\text{psol},3} \cdot \text{rec} + \lambda_{\text{psol},4} \cdot \left(F^{(\text{psol})} \right)^2 + \lambda_{\text{psol},5} \cdot \text{rec}^2 & \text{II} \\
 & + \theta_3 L_1^{(\text{II}),\text{cat}} & \text{III} \\
 & + \theta_4 m_{\text{cat}} + \theta_5 V_{\text{reac}}^{\theta_6} + \theta_7 \left(\sum_i \theta_{8,i} Z_i - 2 \theta_{8,\text{C10en}} \right). & \text{IV}
 \end{aligned}$$

The parts describe

Table 4.3: Optimization results for DMF, DSUC, and THPO.

polar solvent <i>psol</i>	reactor			recycle	extraction cascade	overall process
	volume V_{react} (m ³)	pressure P_r (bar)	temperature T_r (K)	purge ω (-)	split factor ζ (-)	costs TAC (\$/a)
DMF	273.3775	20	388.15	0.0013	0	1934522 (100%)
DSUC	3458.1775	20	388.15	0.0018	0	6154228 (+218%)
THPO	203.8447	20	388.15	0.0015	0	1748021 (-9.64%)

- I) the phase separators' operating and investment costs (McBride and Sundmacher, 2015),
- II) the extraction cascade's distillation column's operating and investment costs,
- III) the catalyst make-up stream's costs, and
- IV) the reactor's operating and investment costs (McBride and Sundmacher, 2015).

The parameters λ , η , and θ are given in Table B.3.

The models are implemented using the GAMS 26.1.0 framework. The optimization problems are solved using the multi-start heuristic of BARON 18.11.12., CONOPT 4.09 is used as NLP sub-solver, and Cplex 12.8.0 is the employed LP/MIP sub-solver. The computations are conducted on a standard desktop PC with 3.40 GHz Intel Core i7-6700 CPU and 16 GB memory.

4.2.4 Process Optimization Results

The results of the conducted process optimizations, using BARON's multi-start heuristic with 140.000 starting points, are discussed in this section. The results for each of the three solvent candidates are presented in Figure 4.2, and Table 4.3. The figure shows the molar flow rates into and out of the reactor and the extraction cascade, and some intermediate flows within the extraction cascade. The table gives an overview of the scalar-valued degrees of freedom and includes the TAC .

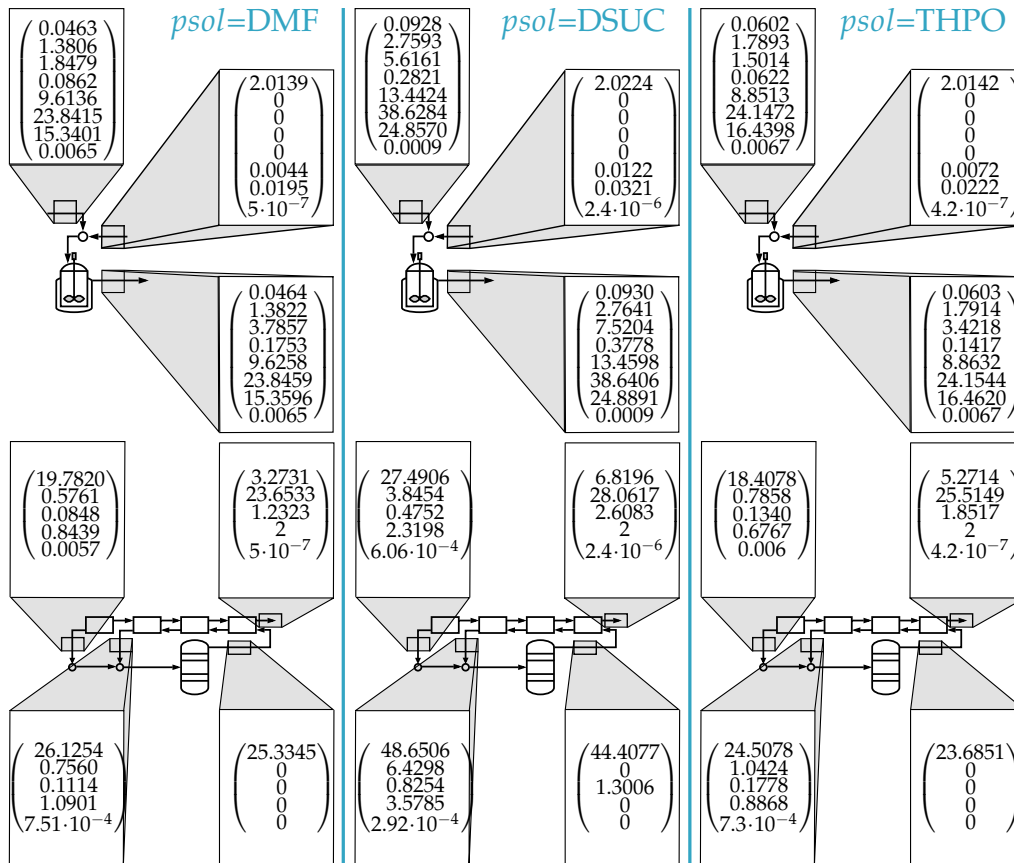


Figure 4.2: Optimization results for DMF, DSUC, and THPO. The upper half of the diagram depicts the reactor, and its molar flows in mol/s. The vectors' single entries denote the following species: (nC10en, iC10en, nC11al, iC11al, C10an, *psol*, C12an, RHO+BIP). The lower half of the diagram depicts the extraction cascade, and its molar flows in mol/s. The vectors' single entries denote the following species: (*psol*, C12an, C10en, C11al, RHO+BIP).

The results are discussed separately for each of the solvents in the following.

Dimethylformamide (DMF)

McBride et al. (2016) identified DMF as the best performing solvent using a screening methodology. Therefore it will be regarded as a benchmark in the present case study. With a catalyst loss of only 0.00008 %, DMF does an excellent job as extraction solvent, as expected. Good catalyst recovery directly influences the optimal reactor operation: A low catalyst leaching rate allows for higher catalyst concentrations within the reactor, which results in faster reaction rates. If the reaction rates are increased, the residence time can be reduced. Additional tuning parameters for increasing the reaction rates are the reactor pressure and the reactor temperature. The reduction in residence time translates into a smaller reactor volume and smaller investment costs.

Due to DMF's good catalyst extraction characteristics, the circulated extraction solvent stream within the extraction cascade is small. The feed stream of the extraction cascade's distillation column can be influenced by the split factor, ζ , which is set to zero. Consequently the distillation column bypass is deactivated and the column is fed with both streams, $F_{col,1}^{in}$ and $F_{col,2}^{in}$. This choice increases the distillation column's diameter, and therefore the investment costs, because the feed rate increases. However, the operating costs are reduced because of a lowered recovery rate. Another factor that necessitates this choice is the temperature constraint at the bottom of the column.

A fraction of the stream leaving the black box, ω , is purged. This purge is necessary to remove the side produce decane from the process. However, the operating costs are increased by the purge due to purged reactants and solvents. The make-up streams entering the reactor are used to reintroduce the solvents.

The optimal solution's *TAC* is 1934522 \$/a.

The obtained optimal solution relies on the accuracy of the surrogate employed to capture the phase behavior. The real logarithmic catalyst

partition coefficient $\log_{10} P_{yx,DMF}$ is assumed to lie within the interval $[\hat{f}(t_1, t_2) + \underline{\delta}_{DMF}, \hat{f}(t_1, t_2) + \bar{\delta}_{DMF}]$, where $\bar{\delta}_{DMF}$ and $\underline{\delta}_{DMF}$ are the maximum and minimum absolute errors given in Equation (4.21), respectively. To account for errors resulting from the phase behavior's approximation an additional worst-case optimization is conducted, with $\log_{10} P_{yx} = \hat{f}(t_1, t_2) + \bar{\delta}$. The consequence of adding the error to the surrogate is an increased catalyst loss. In this specific case, the TAC is 1939897 \$/a.

Dimethyl succinate (DSUC)

Using DSUC as polar solvent results in a catalyst loss of 0.0028 %, which is 35 times as much as in the benchmark case. This increase in catalyst leaching results in a lower catalyst concentration within the reactor, which results in an increased residence time, i.e. in an increased reactor volume; hence the investment costs are higher than in the benchmark.

The circulated extraction solvent stream is much larger than that in the benchmark case, leading to higher operating costs than the benchmark. The lower catalyst selectivity of DSUC can explain this increase in the circulated extraction solvent stream. The overall flow rates are increased consequently; hence the distillation column and the phase separators are larger. This, again, leads to an increase in the investment costs.

Due to the lower selectivity of DSUC, it transports more aldehyde through the recycle. The molar flow rate of the aldehyde entering the extraction cascade needs to be higher than the benchmark to achieve the required product flow rate. The ratios of the solvents to the other components in the reactor are fixed due to model limitations. Therefore the increase in recycled product necessitates more solvent within the reactor and, therefore, a larger reactor vessel.

An additional cost factor is the increased purge, which is necessary due to the increased side-product production.

The optimal solution's TAC is 6154228 \$/a. The TAC is increased by 218 % compared to the benchmark.

An additional worst-case optimization was conducted, analogous to the benchmark, by adding the maximum absolute error to the catalyst partitioning surrogate. The resulting *TAC* is 6177063 \$/a.

Tetrahydropyranone (THPO)

With a catalyst loss of only 0.00007% THPO does an excellent job as extraction solvent. The catalyst recovery is even slightly better than with DMF. Therefore the catalyst concentration within the reactor can also be higher, resulting in faster reaction rates and a smaller reactor vessel than in the benchmark.

Due to THPO's good catalyst selectivity, the circulated extraction solvent stream within the extraction cascade is small. Analogous to the benchmark, the the distillation column bypass is deactivated to lower the recovery rate and thereby save operating costs.

As in the previous two cases, the side-product decane has to be purged.

The optimal solution's *TAC* is 1748021 \$/a. The *TAC* is reduced by 9.64% compared to the benchmark. The small savings may be explained by the excellent catalyst recovery.

As in the previous two cases, an additional worst-case optimization was conducted by adding the maximum absolute error to the catalyst partitioning surrogate. The resulting *TAC* is 1758083 \$/a.

4.3 CONCLUSION

This chapter deals with the optimization of the hydroformylation of n-decene using a TMS for catalyst recovery. In an attempt to identify a substitute for the toxic solvent DMF, two additional solvents are tested: THPO and DSUC. The two ecologically benign solvent candidates THPO and DSUC were identified by McBride et al. (2018) using a screening approach. DMF is used as a benchmark, as its high-grade performance is known. Hence the process is optimized three times, each time with a different TMS.

As the catalyst loss is one of the most significant cost factors, it becomes evident that the polar solvent has a significant influence on the process efficiency by comparing the three solvents' performances. DSUC cannot recover the catalyst adequately, and therefore its use results in the most expensive process. THPO, on the other hand, can recover more catalyst than DMF and its use leads to the most efficient process.

An additional important factor is the solvent's extraction selectivity. While DSUC transports much of the product through the recycle, THPO and DMF are more selective. A reduced selectivity results in larger overall fluxes, more extensive facilities, and therefore higher costs.

It is usually assumed that there is a trade-off between sustainability and process performance (Sanderson, 2011), which cannot be seen in this optimization study; the green solvent candidate THPO performs on the same level as DMF.

This chapter's results underline the economic impact of the solvent's ability to recycle the catalyst selectively. This insight will play a vital role in the following chapter, where additional solvent candidates will be generated and tested using an integrated solvent and process design approach.

Part III

INTEGRATED DESIGN

SIMULTANEOUS PROCESS AND SOLVENT DESIGN

This chapter was published in parts in: Chemical Engineering Science, 249, Tobias Kessler, Christian Kunde, Steffen Linke, Kai Sundmacher, and Achim Kienle. "Integrated Computer-Aided Molecular and Process Design: Green Solvents for the Hydroformylation of Long-Chain Olefines", 117243 (2022).

5.1 INTRODUCTION

The methodologies presented in the previous chapters enable an efficient process optimization, which is the foundation for integrated solvent and process design.

Integrated solvent and process design problems are hard to solve due to their complexity. Therefore hierarchical approaches are frequently used, which decompose the problem into two sub-problems: candidate molecule identification and process optimization.

Although generating molecules is relatively easy, finding suitable molecules for a specific application is not. The most commonly used approaches for overcoming this problem can be roughly split into three categories:

1. Group-contribution (GC) (Joback and Reid, 1987) based approaches. GC methods assume that pre-defined molecular building blocks contribute a certain value to a specific property. Although this method is rather straightforward, it may easily lead to a combinatorial explosion (Joback and Stephanopoulos, 1995). Commonly used methods for overcoming this problem are hierarchical decomposition

(Karunanithi et al., 2005; Cignitti et al., 2017), the use of simplified process models (Burger et al., 2015), and search-space reduction techniques (Papadopoulos et al., 2010; Zhou et al., 2019).

2. Molecular targeting based approaches. The basic idea is that the process optimization is conducted with a physically-based state equation, where the molecule specific parameters are additional degrees of freedom. After the solver identified the "utopia" parameters, a molecule with the same or similar parameters needs to be identified. This is either be done by screening databases (Bardow et al., 2010; Stavrou et al., 2014; Wang and Lakerveld, 2018) or by using GC methods (Eden et al., 2004; Lampe et al., 2015).
3. Database screening based approaches. Commercially and open source available databases, such as COSMObase, include several thousands of compounds and their respective physical and thermodynamic properties. This information can be used to identify suitable molecule candidates by imposing specific pre-defined criteria. Subsequent process optimization for each remaining candidate allows for identifying the most efficient of those candidates (Papadopoulos and Linke, 2005; Limleamthong et al., 2016; Ten et al., 2017). This approach led to the identification of new solvents for CO₂ adsorption (Fleitmann et al., 2018), CO₂ capture and utilization (Scheffczyk et al., 2017), and extraction-distillation processes (Scheffczyk et al., 2016; Scheffczyk et al., 2018).

The thorough literature reviews by Austin et al. (2016), Papadopoulos et al. (2018), and Gertig et al. (2020) best comprehend the advances made over the last years.

The last chapter introduced two new ecologically benign solvent candidates for the hydroformylation identified by McBride et al. (2018). Linke et al. (2020) build upon this work and identified additional promising solvents through a screening methodology.

Although the databases employed for screening approaches include numerous compounds, they only allow for a limited view. The present

chapter aims to overcome this limitation and identify additional efficient and ecologically benign solvent candidates.

Group-contribution methods are employed for achieving this aim. The solvents identified by Linke et al. (2020) are used as reference solvents, the new candidates are searched in their neighborhood, which is defined by σ -moment domains. σ -moments originate from quantum chemical calculations based on the COSMO theory.

5.2 COMPUTER-AIDED MOLECULAR AND MIXTURE DESIGN

The first step of the presented approach is identifying suitable molecules through a computer-aided molecular and mixture design (CAMxD). The problem definition is as follows:

$$\begin{aligned}
 & \min_{n,x} J(n, p, q), \\
 & \text{s.t.} \quad s_1(n) \leq 0, \\
 & \quad \quad s_2(n) = 0, \\
 & \quad \quad p = f(n), \\
 & \quad \quad q = g(x, n, p), \\
 & \quad \quad h_1(p, q, n) \leq 0, \\
 & \quad \quad h_2(p, q, n) = 0, \\
 & \quad \quad p_k^L \leq p_k \leq p_k^U, \\
 & \quad \quad n_d^L \leq n_d \leq n_d^U, \\
 & \quad \quad q_j^L \leq q_j \leq q_j^U, \\
 & \quad \quad \sum_i x_i = 1.
 \end{aligned} \tag{5.1}$$

Here, s are structural feasibility constraints to ensure that the obtained molecules can exist in the real world. The molecules' properties are estimated as p , and the mixtures' properties as q . h captures thermodynamic and physical property constraints, such as a restriction in the boiling point. The mixtures' properties q , the molecules' properties p , and the number of groups n are restricted to certain upper bounds, denoted by U ,

and lower bounds, denoted by L . The optimal mixture's composition for each molecule candidate is calculated with the restriction that the molar fractions' sum has to equal 1.

The constraints are described in more detail in the following.

Note that the objective function J only plays a role in sorting the obtained molecules. Process optimizations are conducted for each feasible molecule regardless of its objective function value in this stage. Here, the objective function is chosen to be the boiling point T_b

5.2.1 Structural Feasibility Constraints

The chemical feasibility of the obtained molecules is ensured through structural feasibility constraints. The constraints used here are based on the works of Churi and Achenie (1996) and Sahinidis et al. (2003). In the following, variables and parameters named "B" are binary decision variables. Binary variables may only take values 0 or 1. Note that the restrictions imposed on the number of groups (constraint 1) and cyclic molecules (constraints 3 and 4) can be considered structural and complexity constraints.

1. Upper and lower bounds restrict the number of groups in a molecule,

$$N_{\min} \leq \sum_{i=1}^N n_i \leq N_{\max}, \quad (5.2)$$

where the i -th entry of $\mathbf{n} \in \mathbb{N}^N$ denotes the number of occurrences of group i in the molecule. In this chapter, the lower bound is defined as $N_{\min} = 2$, and the upper bound is defined as $N_{\max} = 7$.

2. To ensure consistent bonding types, the number of single bonds, $\sum_i n_i \cdot S_{S,i}$, and double bonds, $\sum_i n_i \cdot S_{D,i}$, must be even. So-called transition groups with single and double bond types, $n_i \cdot B_{SD,i} \geq 1$, are necessary if groups with single bonds but without double bonds, $n_i \cdot B_{S,i} \geq 1$, and groups with double bonds but without single bonds, $n_i \cdot B_{D,i} \geq 1$, are present in the molecule. This restriction is necessary

to guarantee that all groups may be connected to one molecule. It is imposed through the following constraints,

$$B_1 \leq \sum_{i=1}^N n_i \cdot B_{SD,i}, \quad (5.3)$$

$$\sum_{i=1}^N n_i \cdot B_{SD,i} \leq N_{\max} \cdot B_1 \cdot \sum_{i=1}^N B_{SD,i}, \quad (5.4)$$

$$B_2 \leq \sum_{i=1}^N n_i \cdot B_{S,i}, \quad (5.5)$$

$$\sum_{i=1}^N n_i \cdot B_{S,i} \leq N_{\max} \cdot B_2 \cdot \sum_{i=1}^N B_{S,i}, \quad (5.6)$$

$$B_3 \leq \sum_{i=1}^N n_i \cdot B_{D,i}, \quad (5.7)$$

$$\sum_{i=1}^N n_i \cdot B_{D,i} \leq N_{\max} \cdot B_3 \cdot \sum_{i=1}^N B_{D,i}, \quad (5.8)$$

$$1 \geq -B_1 - B_{MC} + B_2 + B_3. \quad (5.9)$$

The same rule is necessary for groups with cyclic and acyclic bonds (Sahinidis et al., 2003).

3. Only cyclic and monocyclic molecules are allowed,

$$1 = B_A + B_{MC}, \quad (5.10)$$

where B_A denotes acyclic molecules, and B_{MC} denotes monocyclic molecules

4. If the molecule is cyclic, it has to be an aromatic molecule, i. e. the number of cyclic bonds must be 12, and the number of groups having cyclic bonds must be 6. This restriction is captured with

$$B_{C,i} \cdot n_i = B_{C,i} \cdot n_i \cdot B_{MC}, \quad (5.11)$$

$$\sum_{i=1}^N n_i \cdot S_{C,i} = B_{MC} \cdot 12, \quad (5.12)$$

$$\sum_{i=1}^N n_i \cdot B_{C,i} = B_{MC} \cdot 6. \quad (5.13)$$

5. To ensure the formation of only one connected molecule, the number of bonds or double bonds must at least equal the number of groups minus one. Furthermore, the number of bonds or double bonds cannot exceed that of a complete graph in which all groups are connected (Sahinidis et al., 2003),

$$\beta_i = S_{S,i} + S_{D,i} + S_{C,i}, \quad (5.14)$$

$$\sum_{i=1}^N n_i \cdot \beta_i \geq 2 \left(\sum_{i=1}^N (n_i) - 1 \right), \quad (5.15)$$

$$\sum_{i=1}^N n_i \cdot \beta_i \leq \left(\sum_{i=1}^N n_i \right) \left(\sum_{i=1}^N (n_i) - 1 \right). \quad (5.16)$$

6. To obey the chemical octet rule that each atom bonds such that it has eight electrons in its valence shell, each group's valency needs to be satisfied with a covalent bond (Odele and Macchietto, 1993),

$$2B_A = \sum_{i=1}^N n_i \cdot (2 - \beta_i). \quad (5.17)$$

7. Because double bonds are modeled as two single bonds but as a bond type for themselves, two adjacent groups may not be linked by more than one bond or double bond. (Odele and Macchietto, 1993),

$$\sum_{j=1}^N n_j \geq n_i \cdot (\beta_i - 1) + 2 \cdot B_A. \quad (5.18)$$

8. In the following, groups present in the molecule are called active groups. The vector $\mathbf{u} \in \mathbb{N}^{N \times N_{\max}}$ captures whether a group is active, their respective entries are set to 1.

$$n_i = \sum_{m=1}^{N_{\max}} u_{i,m}, \quad i \in [1, \dots, N], \quad (5.19)$$

$$1 \geq \sum_{i=1}^N u_{i,m}, \quad m \in [1, \dots, N_{\max}], \quad (5.20)$$

$$\sum_{i=1}^N u_{i,m-1} \geq \sum_{i=1}^N u_{i,m}, \quad m \in [2, \dots, N_{\max}]. \quad (5.21)$$

9. The matrix $\mathbf{z} \in \mathbb{N}^{N_{\max} \times N_{\max}}$ captures which groups are interconnected with each other (Churi and Achenie, 1996),

$$\sum_{i=1}^N u_{i,m} \cdot \beta_i = \sum_{o=1}^{N_{\max}} z_{m,o}, \quad m \in [1, \dots, N_{\max}]. \quad (5.22)$$

10. Groups must be connected to other groups to satisfy their valency; connections within one group are not allowed. Furthermore, \mathbf{z} must be symmetric,

$$z_{m,o} = 0, \quad m, o \in [1, \dots, N_{\max}], m = o, \quad (5.23)$$

$$z_{m,o} = z_{o,m}, \quad m, o \in [1, \dots, N_{\max}]. \quad (5.24)$$

11. To ensure that only one molecule is formed, each entry m has to be connected to one of the entries before it (Churi and Achenie, 1996),

$$0 \leq w_m - 1 + \sum_{o=1}^{m-1} z_{m,o}, \quad m \in [2, \dots, N_{\max}], \quad (5.25)$$

$$N_{\max} = \sum_{i=1}^N n_i + \sum_{m=1}^{N_{\max}} w_m, \quad (5.26)$$

$$w_m \leq w_{m+1}, \quad m \in [1, \dots, N_{\max} - 1]. \quad (5.27)$$

Note that some of the constraints mentioned above are redundant, they are used to speed up the optimization.

Table 5.1: Main groups, aromatic groups, and functional groups available in the present computational study. s denotes a free single bond valency, and d denotes a free double bond valency. Aromatic C atoms are denoted as AC; they have cyclic bond valencies r.

Main groups			
sCH ₃	ssCH ₂	dCH ₂	sssCH
sdCH	ssssC	ssdC	ddC
Aromatic groups			
rrACH	rrACs	rrACCH ₃	rrACCH ₂ s
rrACCHss	rrACCHd		
Functional groups			
sCH ₃ CO	sCH ₂ COs	CH=Os	sCH ₃ COO
sOCH ₃	sOCH ₂ s	sOCHss	sCH ₂ COOs
sOCHd	sCH ₂ CN	sCOOH	

5.2.2 Complexity Constraints

In the present chapter, group-contribution methods are used to estimate specific molecule properties. As already mentioned, these methods work poorly for too large and complex molecules. Therefore their size and complexity need to be restricted. Table 5.1 gives an overview over the 25 available molecular groups used in this chapter. They are classified as main groups consisting of C and H atoms, aromatic groups with cyclic bonds, and functional groups. The imposed constraints are:

1. The number of main groups is constrained to two for monocyclic molecules,

$$2 \geq \left(\sum_{i=1}^N n_i \cdot B_{CH,i} \right) \cdot B_{MC}. \quad (5.28)$$

2. Only one C double bond and one monocyclic side-chain are allowed.

$$1 \geq \sum_{i=1}^N n_i \cdot B_{CC,i}, \quad (5.29)$$

$$1 \geq \sum_{i=1}^N n_i \cdot B_{MCside,i}. \quad (5.30)$$

3. Monocyclic molecules may only have one chain-ending group, acyclic molecules three,

$$1 \geq \left(\sum_{i=1}^N n_i \cdot B_{CE,i} \right) \cdot B_{MC}, \quad (5.31)$$

$$3 \geq \left(\sum_{i=1}^N n_i \cdot B_{CE,i} \right) \cdot B_A. \quad (5.32)$$

The name of chain-ending groups is based on their valency of one; each branch ends when they are connected to it.

4. Groups with a valency greater than one are non-chain-ending groups. Monocyclic molecules may have one non-aromatic non-chain-ending group, acyclic molecules three,

$$1 \geq \left(\sum_{i=1}^N n_i \cdot B_{NCE,i} \right) \cdot B_{MC}, \quad (5.33)$$

$$3 \geq \left(\sum_{i=1}^N n_i \cdot B_{NCE,i} \right) \cdot B_A. \quad (5.34)$$

5. As suggested by Zhou et al. (2019), the number of times a group may be used within a molecule is limited to N_i^{UP} . An additional complexity reduction is obtained by

$$3 \geq \sum_{i=1}^N \frac{n_i}{N_i^{UP}}. \quad (5.35)$$

The values for the parameters \mathbf{B} , \mathbf{N}^{UP} , and \mathbf{S} can be found in [Table C.1](#).

5.2.3 Property Estimation

The structural and complexity constraints presented above are necessary to obtain a molecule which is able to exist in the real world. Numerous molecules may be generated using the 25 available groups. To determine which of them is the best for the considered hydroformylation process, it is necessary to estimate their pure species and mixture properties.

Pure Species Properties

As a short reminder, group-contribution methods are a standard method to estimate pure species properties. In group-contribution methods, it is assumed that a particular molecular group i correlates to a decrease or increase of a specific property p via a constant contribution c_i ,

$$p = \sum_{i=1}^N n_i \cdot c_i, \quad (5.36)$$

where n_i denotes how many times group i occurs within the molecule.

An extensive data-set is necessary to correlate groups and properties. Marrero and Gani (2001) used a vast number of data points for their group-contribution method, which can estimate a large number of thermophysical properties. In this chapter, the boiling point, T_b , and the heat of vaporization, H_v , are estimated using their method.

Hukkerikar et al. (2012) proposed a group-contribution method to estimate numerous environment, health, and safety (EHS) properties of a molecule. Three of these properties are considered in the present chapter:

BCF The bioconcentration factor (BCF, in l/kg) is used to capture the molecules' impact on the environment. The greater its value, the more a substance will accumulate in water organisms; hence a small value is preferable. According to the REACH regulation of the European Chemicals Agency, a substance is said to be bioaccumulative if $BCF \geq 2000$, and very bioaccumulative if $BCF \geq 5000$ (Agency, 2017).

- PEL** The permissible exposure limit (PEL, in g/m^3) is used to capture the molecules' impact on health. The limit is defined by the Occupational Safety and Health Administration (OSHA) of the United States of America. The lower the PEL is, the unhealthier the substance is.
- LD₅₀** The oral median lethal dose ($\text{LD}_{50,\text{rat}}$, in g/kg) is used to capture the molecules' safety. The $\text{LD}_{50,\text{rat}}$ is the dose at which half a test population of rats dies within a specified test duration of exposure; hence a high $\text{LD}_{50,\text{rat}}$ is preferable.

Mixture Properties

Following the results of [Chapter 4](#), the catalyst distribution significantly influences the process efficiency. Therefore its prediction is of utmost importance.

As the solvent candidates in the last chapter were known, it was possible to use surrogate models fit to data from COSMOtherm to calculate the catalyst's partitioning coefficient. In the present chapter, a more general methodology is necessary.

The chemical species considered in the present case study can be split into two groups: already known, fixed species (decane, dodecane, decene, undecanal, and BiPhePhos), and a yet unknown polar solvent. COSMO-SAC, as described in [Section 2.4](#), is used to calculate the mixture's liquid-liquid equilibrium. The mixture's components' σ -profiles are necessary for this calculation. They are calculated from cosmo-files using the averaging algorithm from Lin and Sandler ([2002](#)).

The cosmo-files for most of the fixed species are available from the VT-2005 database presented by Mullins et al. ([2006](#)). BiPhePhos is the only component not included in the database. Therefore, a TURBOMOLE 7.1 (Turbomole, [1989-2007](#)) calculation was conducted by Steffen Linke with the same functional and basis set as in the VT-2005 database to obtain the necessary data for the calculation of the σ -profile.

The σ -profile of the unknown polar solvent is estimated using the group-contribution method from Liu et al. ([2019](#)).

5.3 INTEGRATED DESIGN

The results of [Chapter 5](#) underline the necessity of an integrated solvent and process design because the employed polar solvent's characteristics significantly influence the operating conditions and process efficiency.

Although the constraints presented in [Section 5.2](#) already reduce the search-space, it is still enormous. The optimization problem is decomposed into two sub-problems to tackle the computational complexity:

1. Generate suitable molecule candidates. The use of a TMS for catalyst recovery imposes strict additional constraints on the candidates, as the mixture has to decompose into two liquid phases at temperatures lower than the reaction conditions.
2. Process optimization for each candidate.

Due to the uncommon thermodynamics needed, numerous molecules can be neglected. A search-space reduction is conducted to exploit this in favor of the computational complexity of the problem.

5.3.1 Search-Space Reduction

The basic idea behind this search-space reduction is that similar molecules omit similar behavior. However, the similarity is not measured by the molecules' molecular structure but by quantum-chemical properties, the so-called σ -moments M . They were introduced by Klamt (2005) and can be calculated with

$$M_i(p(\sigma_m)) = \sum_{\sigma_m} \left(p(\sigma_m) \cdot \sigma_m^i \right), \quad i \in [0, \dots, 3], \quad (5.37)$$

each moment is correlated to a property of the molecule:

0. total surface area (Van der Waals surface),
1. total COSMO polarization charge,
2. total COSMO polarization energy,

3. “skewness” in σ -profile (asymmetry).

Klamt (2005) introduced two additional moments for the hydrogen bonding,

$$M_{\text{acc/don}}(p(\sigma_m)) = \sum_{\sigma_m} \left(p(\sigma_m) \cdot f_{\text{acc/don}}^{\text{hb}}(\sigma_m) \right), \quad (5.38)$$

$$f_{\text{acc/don}}^{\text{hb}} = \begin{cases} 0, & \pm\sigma_m < \sigma'_{\text{hb}}, \\ \pm\sigma_m - \sigma'_{\text{hb}}, & \pm\sigma_m \geq \sigma'_{\text{hb}}. \end{cases}, \quad (5.39)$$

where $\sigma'_{\text{hb}} = 1 \text{ e/nm}^2$ is the hydrogen bonding cut-off value.

The correlation of σ -moments to physical molecule properties was already exploited in several applications. To name a few, they were used by Lukowicz et al. (2015) to characterize oleo-chemicals, by Weinebeck et al. (2017) as descriptors for a QSPR-based lubricity model, and by Austin et al. (2017) for mixture design.

In this work, the first, the second, the third, and the hydrogen bonding acceptor moment are employed.

The 17 most promising solvent candidates identified by Linke et al. (2020) pose as a reference to check whether the σ -moments are able to capture the molecules' characteristics. Their σ -moments are shown in Figure 5.1. Note that, except for two outliers, the σ -moments lie within tight intervals: $M_1 \in [-0.004, -7.0477 \cdot 10^{-4}]$, $M_2 \in [0.0053, 0.0083]$, $M_3 \in [2.8543 \cdot 10^{-5}, 5.0783 \cdot 10^{-5}]$, and $M_{\text{acc}} \in [0.0431, 0.0843]$. We call these intervals σ -bands.

The VT-2005 database includes 1432 species. Their σ -moments are depicted in Figure 5.2. After reducing the search-space by applying the σ -bands, 13 of 1432 species remain. COSMO-SAC predicts the required phase behavior for 11 of these 13 species.

Although this proves that a σ -bands based search-space reduction is feasible and delivers suitable molecule candidates, it is worth noting that the two outliers in Figure 5.1 indicate that suitable molecule candidates might be neglected.

Note that the screening of the VT-2005 database solely poses as a verification for the search-space reduction itself. The molecules obtained

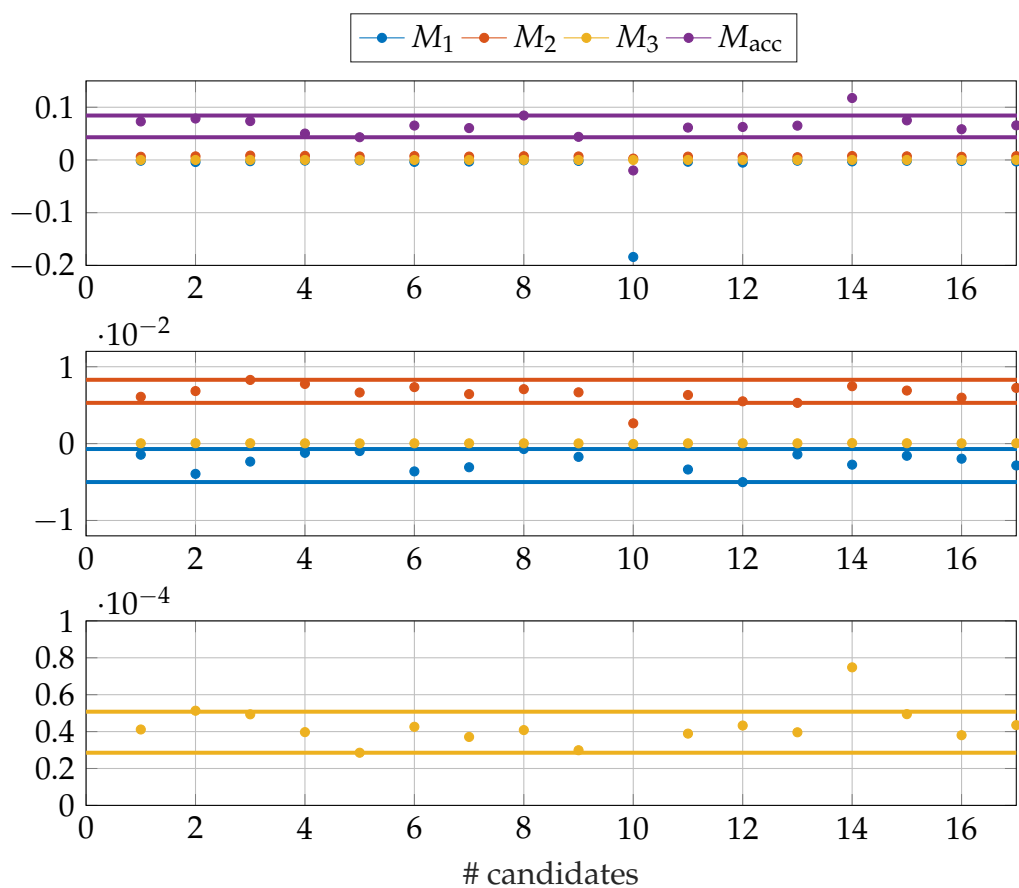


Figure 5.1: σ -moments of the 17 most promising solvent candidates identified by Linke et al. (2020) plotted on three scales. The first moment is denoted by blue dots, the second moment by orange dots, the third moment by yellow dots, and the hydrogen bond acceptor moment by purple dots. The moments lie within the so-called σ -bands, which are colored respectively.

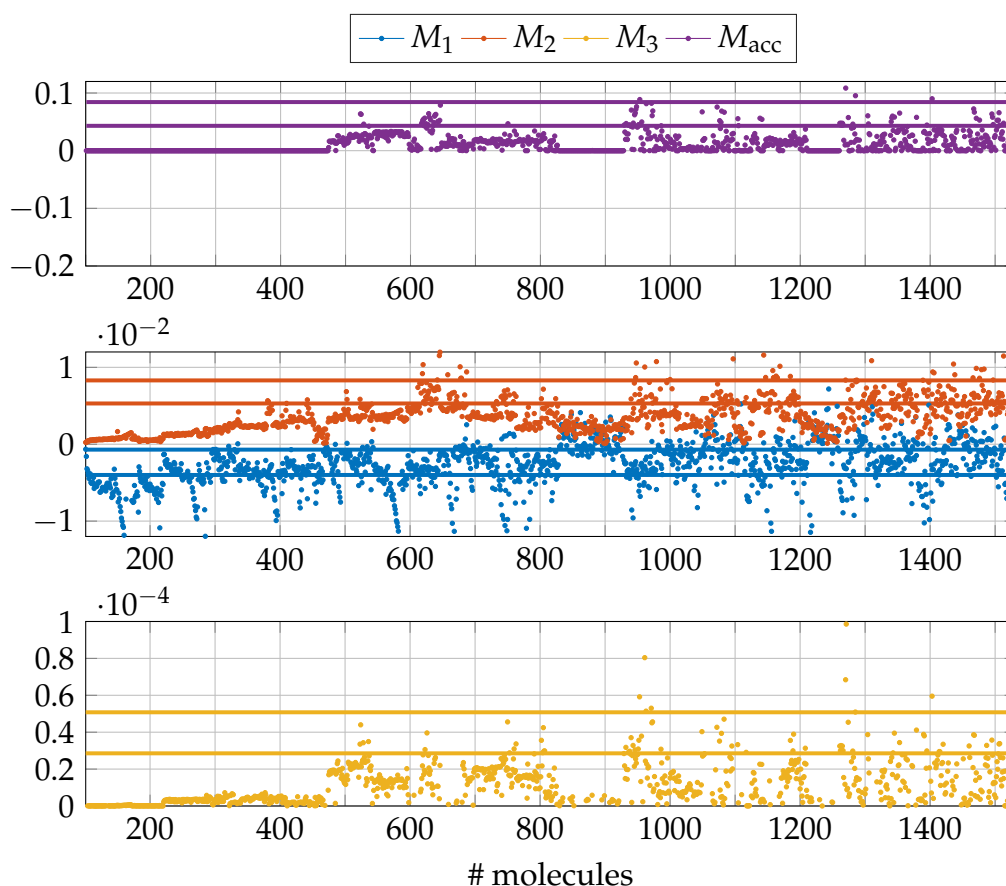


Figure 5.2: σ -moments of the 1432 species included in the VT-2005 database, plotted on three scales, and the σ -bands derived from Figure 5.1. The color-code is the same as in Figure 5.1

from the database are not considered due to additional requirements resulting from the reaction conditions.

The search-space is further reduced by limiting the predicted boiling point: $T_b \geq 400$ K.

The molecules are generated with BARON. The option “NumSol” is used to generate all feasible solutions in terms of the constraint set discussed above. To ensure that the obtained molecules differ from each other, the option “ISolTol” is set to 5.

An alternative approach for the generation of molecules based on Hansen Solubility Parameters is presented in Appendix D.

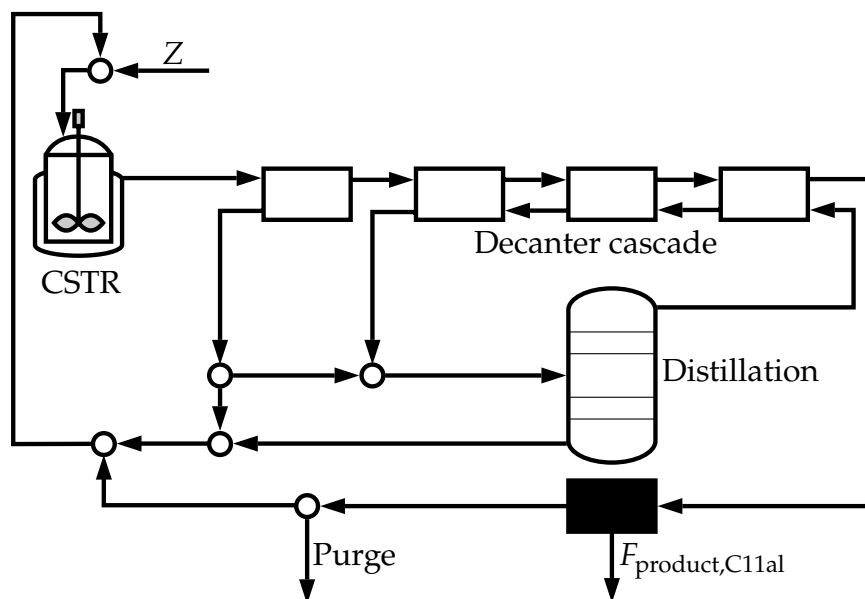


Figure 5.3: Process flowsheet.

5.3.2 Process Optimization

As in the previous chapter, the hydroformylation of *n*-decene to undecanal poses as a case study.

Figure 5.3 depicts the process flowsheet. The process parts are the same as in Chapter 4: The reactor is a continuously stirred tank reactor with external feed Z for solvent, catalyst, and reactant make-up streams; the catalyst is recovered in a counter-current extraction cascade consisting of four phase separators and a distillation column; the black-box denotes further downstream processing steps to separate the product undecanal from the other species, the catalyst is assumed to be lost in this step; the purge's purpose is to remove side-products from the recycle. Additional to the feed stream, the reactor is fed with recovered catalyst, recovered solvents, unused reactants, and side-products.

The original species involved in the process are presented in Table 5.2.

Although the process is the same as in the previous chapter, some small deviations concerning the extraction cascade and solvent costs have to be taken into account due to the increased complexity of the present optimization problem.

Table 5.2: Chemical species involved in the process. Alternatives for the polar solvent marked in red shall be identified.

Name	Purpose
n/iso-decene	reactant
n/iso-undecanal	desired/side product
dodecane	non-polar solvent
dimethylformamide	polar solvent
decane	side product
rhodium-BiPhePhos-complex	catalyst

EXTRACTION CASCADE As in the previous chapter, the distillation column within the extraction cascade is replaced with a surrogate function. However, the separation properties of the new solvents are not known a priori. Therefore it is assumed that decene and dodecane are recycled through the distillate stream. Note that this is a conservative assumption.

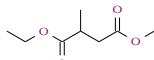
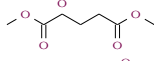
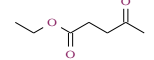
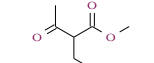
In the previous chapter, the phase behavior of the solvents was approximated by a surrogate based on UNIFAC and COSMO-RS calculations. COSMO-RS calculations are not applicable in the present setup, as they would require additional quantum-chemical calculations to generate the required cosmo-files. Therefore, the phase behavior is described by COSMO-SAC, the required σ -profiles of the new polar solvents are calculated using the above-mentioned group-contribution method from Liu et al. (2019).

SOLVENT PRICE Predicting the new solvents' price is a task burdened with many uncertainties. To the best of the author's knowledge, a cost model based solely on molecular groups does not exist. The reason is that the price greatly depends on the synthesis route and process scale. Therefore the solvents' price is fixed to the solvent price given in Section 4.2.

The process optimizations are conducted using the multi-start heuristic of the global optimization software BARON with 140,000 starting points.

The optimizations are implemented as MINLPs and solved using the GAMS 26.1.0 framework with BARON 18.11.12., Cplex 12.8.0 is used as an LP/MIP subsolver, and CONOPT 4.09 is utilized as an NLP subsolver.

Table 5.3: Solvent candidates for which the process optimization converged to a feasible solution.

Abbreviation	CAS-Number	Name	Molecular structure
EMM	204125-41-7	1-Ethyl 4-methyl 2-methylsuccinate	
DMG	1119-40-0	Dimethyl glutarate	
ELL	539-88-8	Ethyl levulinate	
MEA	51756-08-2	Methyl 2-ethylacetoacetate	

The calculations are carried out on a Linux PC with 3.40 GHz Intel Core i7-6700 CPU and 16 GB memory.

5.4 RESULTS

The imposed constraints in combination with the available group set led to the generation of six solvent candidates. Those candidates are not included in the reference screening results by Linke et al. (2020).

Process optimizations were conducted for all six candidate solvents using BARON's multi-start heuristic. For four of them, the optimization converged to a feasible solution. Those four candidates are presented in Table 5.3. The spider diagram depicted in Figure 5.4 summarizes the optimization results and the values of the considered EHS-criteria.

The remaining four solvent candidates allow for an efficient process operation, resembling the relatively low TAC. While EMM's and DMG's performance is close to that of DMF, the performances of ELL and MEA are even better. Note that this result has to be taken with some caution. As shown in the previous chapter, catalyst leaching is one of the most crucial cost factors of the process. Hence it is of utmost importance to take the catalyst partitioning into account. In the present case study, the catalyst partitioning was calculated with COSMO-SAC, where the σ -profiles were approximated using a group-contribution method. Therefore the results have to be considered in the light of the limited predictive capabilities of model-based methods. An additional uncertain cost factor is the price for

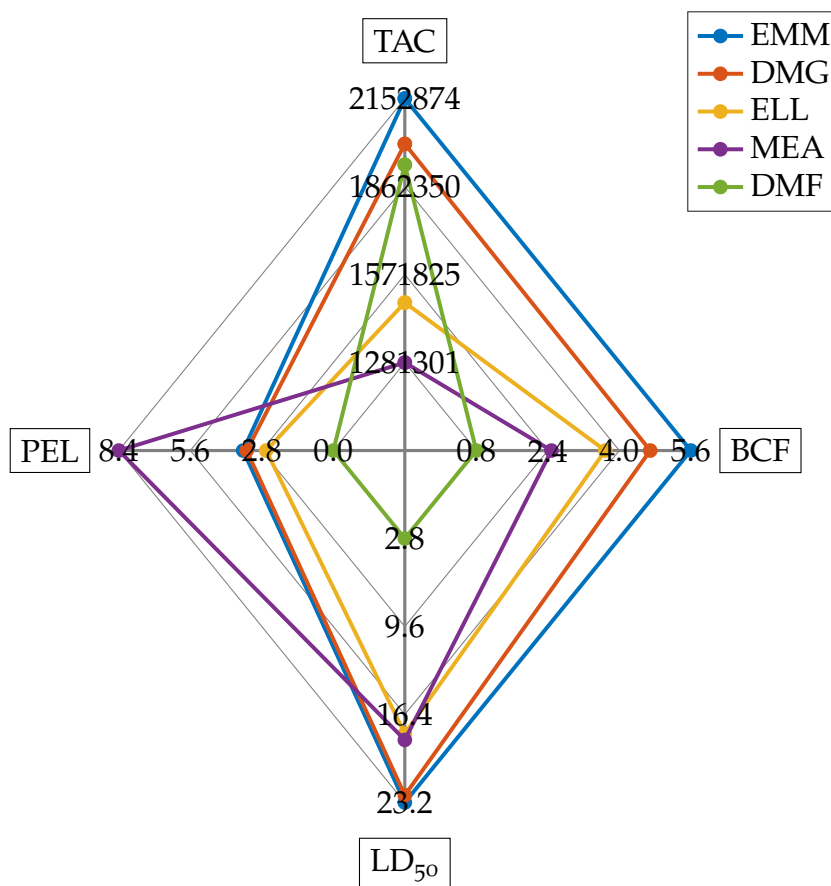


Figure 5.4: Spider diagram of the results obtained with the four candidate solvents and DMF (green) as a benchmark. The diagram includes the TAC in \$/a, LD₅₀ in g/kg, PEL in g/m³, and BCF in l/kg. An ideal candidate would have a small TAC, a high LD₅₀, a high PEL, and a small BCF.

the solvents. The assumed cost for EMM and MEA is significantly lower than the real-world price.

The considered EHS criteria indicate that all four candidates are safe and ecologically benign. One of the identified candidates, DMG, is listed as “green circle component” in the “Safer Chemical Ingredients List” of the United States Environmental Protection Agency (EPA) (EPA.gov, 2020). Chemical species in this list have to fulfill strict criteria regarding their safety and environmental impact. The green circle is awarded after an experimental validation. Additionally, it was identified as a green solvent through a model-based approach by Moity et al. (2012) and confirmed as a viable substitute for DMF in a screening approach by McBride et al. (2018).

Two of the obtained solvents, namely MEA and EMM, are not yet included in the database used for the screening approach by Linke et al. (2020). Additionally, none of the four solvent candidates is included in the VT-2005 database.

5.5 CONCLUSION

This chapter presents a new approach to integrated process and solvent design based on computational-quantum-chemistry. The approach is based on the idea that chemical species with similar σ -moments exhibit similar physical and thermodynamic properties. This similarity can be exploited using an already known set of molecule candidates to effectively and efficiently reduce the molecular search-space and obtain molecules not included in common screening databases.

The approach is tested for the hydroformylation of n-decene using a TMS for catalyst recovery, searching for a new polar solvent for the TMS. As presented in the previous chapters, this process has high demands on the polar solvent with regard to extraction selectivity. Furthermore, the mixture has to exhibit the required phase behavior for a TMS.

The relatively specific requirements on the polar solvent render many possible molecules inapplicable. The reduction of the search-space to feasible molecules is possible by defining so-called σ -bands. However, the

definition of the σ -bands requires prior knowledge of several feasible solvent candidates with the required properties regarding process efficiency and EHS-criteria. Those candidates were obtained by Linke et al. (2020) in a previous screening approach.

As a result, the approach yields four good performing green solvent candidates, one of which is already validated experimentally to be of low concern and is included in the safer chemical ingredients list of the EPA.

Note that the candidates presented in (Linke et al., 2020) are not obtained due to complexity limitations and group availability.

Part IV

CONCLUDING REMARKS

SUMMARY AND FUTURE PERSPECTIVE

6.1 SUMMARY

Homogeneously solved transition metal catalysts allow chemical processes to become more sustainable by opening the option to use renewable feedstock, where traditional process realizations are bound to petrochemical sources due to mixing problems. However, sustainability is not limited to raw materials but also involves auxiliary substances. One class of those auxiliary substances is solvents. Solvent properties directly influence a process's efficiency. Hence identifying ecologically benign and performant solvents is challenging.

This thesis presented a framework for integrated solvent and process design. This framework's two main pillars are based on efficient process optimization and a significant molecular search-space reduction.

The optimization problem was first split into several sub-problems. Solving these sub-problems to global optimality yielded the valuable process insight that the distillation columns for product purification and extraction solvent recovery can be assumed to exhibit ideal thermodynamic behavior. Different strategies were employed to achieve this. One of which is the efficient use of surrogate models for the computationally most expensive model parts.

Kriging models were employed to optimize distillation columns, thereby showing their ability to cope with discrete variables. It was possible to

obtain an optimal solution for a distillation column within its unstable region by introducing implicit surrogate model formulations.

Although each of the process parts can be optimized to global optimality, optimizing an interconnection of all process parts is computationally too expensive to be considered feasible. Therefore, a multi-start approach was employed for the optimization of the whole process. Because of the efficient model formulations, a large number of starting points could be used.

The efficient process optimization allowed to test the performance of two candidate solvents. Their efficiency was compared to the benchmark solvent DMF. This study yielded the exciting result that ecologically benign solvents do not necessarily increase the overall process costs.

The computer-aided molecular design's major problem was to limit the search-space to molecules exhibiting the required phase behavior to obtain a TMS. Two different search-space reduction techniques have been employed and tested. One was based on Hansen Solubility Parameters to arrive at molecules with a similar mixing behavior as DMF. The other was based on σ -bands. These σ -bands are intervals of specific properties of the σ -profiles from already known solvent candidates. The definition of the searched neighborhood through σ -bands proved to be the better choice. Four promising green solvent candidates were identified and tested in process optimization. Again, the green alternatives delivered a performance close to or even better than that of DMF.

6.2 OUTLOOK

Naturally, the answers presented in this thesis gave rise to additional questions, which have yet to be answered. Some of these questions are discussed in the following.

- Although global optimization of the single process parts has become feasible, the whole process, including all recycles, could not be solved to global optimality and a multi-start heuristic had to be used. One step towards a globally optimal solution could be an iterative

approach. A starting point for one of the process parts is chosen. After an optimal solution is obtained, it is fed into the subsequent process part. The procedure is repeated until a steady-state is reached. First investigations have shown that such an iterative optimization problem converges, but the convergence time greatly depends on the starting point. See [Appendix E](#) for a more thorough discussion. Further studies are necessary to elaborate on the approach.

- The green solvent candidates shown in this thesis have been tested with regards to their process performance using a number of predictive models. It lies in the nature of such models that they are fraught with some uncertainties. For a more sophisticated performance classification, experimental investigations of the solvents are necessary.
- The framework was successfully applied to the hydroformylation of long-chain olefines. A further study could build upon the results presented in this thesis and apply the methods to a different process. A worthwhile process would be the reductive amination of aldehydes, which is an option for further conversion of the hydroformylation product considered in the joint research project SFB/TRR-63.

Part V

APPENDIX

GLOBAL OPTIMIZATION OF REACTOR AND DECANter

This appendix was published in parts in: Computer Aided Chemical Engineering, 40, Tobias Kefßler, Nick Mertens, Christian Kunde, Corina Nentwich, Dennis Michaels, Sebastian Engell, and Achim Kienle. "Efficient global optimization of a novel hydroformylation process", pp. 2113-2118 (2017).

INTRODUCTION

Mathematical optimization is used in this appendix to determine optimal operating conditions for a TMS-based hydroformylation process. The objective of the optimization is to maximize profits by minimizing the leakage of catalysts and maximizing productivity. Degrees of optimization freedom include feed and recycling rates, reactor temperature, reactor volume, and reactor pressure, leading to a challenging to solve, non-linear, and non-convex optimization problem. Deterministic global optimization based on a branch-and-bound algorithm is used. In recent years, enormous progress has been made in this field, resulting in several powerful solvers, such as BARON (Kılınç and Sahinidis, 2018). Resolving this class of problems is still challenging, and computation times greater than 24 hours have been observed for the problem considered in this appendix.

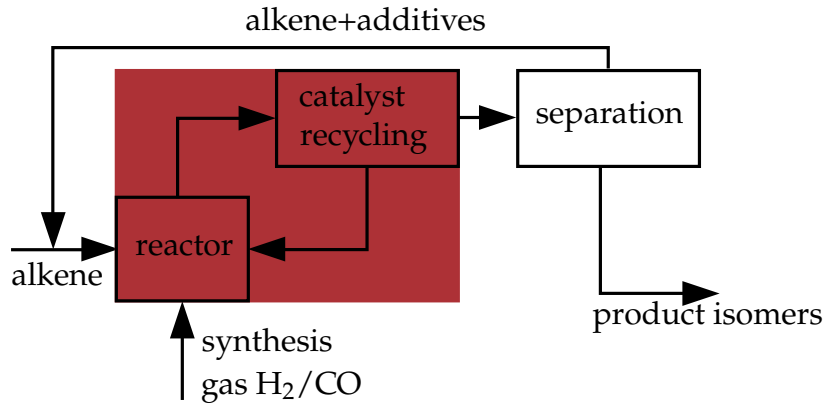


Figure A.1: Process flowsheet, reactor and phase separation highlighted in **Ma-**
roon.

Table A.1: Chemical species

No.	Name
1	n-dodecene
2	iso-dodecene
3	n-tridecanal
4	iso-tridecanal
5	dodecane
6	dimethylformamide
7	decane
8	rhodium
9	BiPhePhos
10	hydrogen
11	carbon monoxide

MODEL DESCRIPTION

This appendix focuses on is highlighted in [Figure A.1](#). The reaction process is assumed to be in a homogeneous mixture with subsequent catalyst separation and recycle.

The reactor is assumed to be an ideal continuously stirred tank reactor. There are three inputs: the alkene feed source, the synthesis gas required for the reaction, and, finally, the recycling stream of the catalyst. In order to achieve a homogeneous liquid phase in the reactor and to ease the separation of the catalyst, a TMS is utilized.

The species involved in this process are shown in [Table A.1](#). In the following, the original process model is presented. The indices in the equations refer to [Table A.1](#).

The studied optimization problem is of the following form:

$$\max_x J(\mathbf{x}, \mathbf{p}) = n_{\text{DecOut},3} - 100 \cdot n_{\text{DecOut},8} - V_{\text{R}}/500 - m_{\text{cat}}/10, \quad (\text{A.1a})$$

$$\text{s.t. } h(\mathbf{x}, \mathbf{p}) = 0, \quad (\text{A.1b})$$

$$g(\mathbf{x}, \mathbf{p}) \leq 0, \quad (\text{A.1c})$$

Therein the process variables are \mathbf{x} , $n_{\text{DecOut},3}$ is the desired product, $n_{\text{DecOut},8}$ describes the leakage of the catalyst, V_{R} is the volume of the reactor, and m_{cat} is the required catalyst mass. Furthermore the process parameters are defined by \mathbf{p} , the equality constraints, such as mass balances and reaction kinetics, by $h(\mathbf{x}, \mathbf{p})$, and the inequality constraints, such as variable intervals, by $g(\mathbf{x}, \mathbf{p})$. The process model is divided into two parts: the model of the reactor and the model of the decanter. The model of the reactor will be defined first.

The following equations describe the reactor's mass balance:

$$-\dot{n}_{\text{Out},i} + \dot{n}_{\text{In},i} + c_{\text{cat}} \cdot V_{\text{R}} \cdot m_{\text{molar,cat}} \cdot \sum_{j=1}^8 \nu_{j,i} \cdot r_j = 0, \quad 1 \leq i \leq 9, \quad (\text{A.2})$$

$$-\dot{n}_{\text{In},i} + \dot{n}_{\text{feed},i} + \dot{n}_{\text{DecReflux},i} = 0, \quad 1 \leq i \leq 9, \quad (\text{A.3})$$

$$\dot{n}_{\text{feed},k} - \dot{n}_{\text{DecOut},k} = 0, \quad 6 \leq k \leq 9, \quad (\text{A.4})$$

$$\sum_{i=1}^9 \dot{n}_{\text{feed},i} - \dot{n}_{\text{DecOut},i} = 0, \quad (\text{A.5})$$

here $\dot{n}_{\text{feed},i}$ defines the molar flow rate of component i into the system and $\dot{n}_{\text{DecOut},i}$ is the molar flow rate leaving the system, i. e. going from the decanter into the separation step for product purification (as shown in [Figure A.1](#)). $\dot{n}_{\text{Out},i}$ and $\dot{n}_{\text{In},i}$ are the molar output and input streams leaving and entering the reactor, respectively. $\nu_{j,i}$ defines the stoichiometric coefficient for reaction r_j of component i . It is worth noting that Equations (A.4) and (A.5) can be derived from Equations (A.2), (A.3), (A.12), and (A.13).

Eight reactions have to be considered, including forward and backward reactions. All eight reaction rates, r_j , are described by

$$r_j = k_j \cdot \left(\prod_i (c_i^{\nu_{j,i}}) \right) \cdot \left(1 + \sum_i (K_{i,1} \cdot c_i^{K_{i,2}}) \right)^{-1}. \quad (\text{A.6})$$

Arrhenius law is used to calculate the reaction rate constants k_j ; the necessary parameters \mathbf{K} are derived from the literature and can be found in (Hentschel et al., 2015).

The following equations are used for the calculation of the i -th species' concentration in the liquid phase, c_i ,

$$-\dot{V} + \sum_{i=1}^9 \dot{n}_{\text{Out},i} \cdot m_{\text{molar},i} / \rho_i = 0, \quad (\text{A.7})$$

$$-c_i \cdot \dot{V} + \dot{n}_{\text{Out},i} = 0, \quad (\text{A.8})$$

where ρ_i is the density of species i , $m_{\text{molar},i}$ is its molar mass, and \dot{V} is the volumetric flow rate. The active catalyst concentration, c_{cat} , is calculated with

$$c_{\text{cat}} = \frac{c_{\text{cat,tot}}}{1 + K_{\text{cat},1} \cdot c_{\text{CO}}^{K_{\text{cat},3}}}. \quad (\text{A.9})$$

Again, the parameters K_{cat} can be found in (Hentschel et al., 2015).

A description of the GLE is needed to determine the distribution of the species in both the liquid and gas phases. For the process considered, a description by Henry's law is not adequate, as the solvent's influence on the equilibrium has to be taken into account. Therefore, a PC-SAFT model with the parameters published by Vogelpohl et al. (2014) is used to describe the GLE. The complexity of this model is very high, resulting in a significant increase in computation time. An artificial neural network from Nentwich and Engell (2016) is used as a surrogate to overcome this problem.

A model of the decanter is required in addition to the reactor model, it will be described in the following:

$$-K_{\text{Dec},i} + Ar_i = 0, \quad 1 \leq i \leq 9, \quad (\text{A.10})$$

$$\exp(A_{\text{Dec},0,i} + A_{\text{Dec},1,i}/T_{\text{Dec}} + A_{\text{Dec},2,i} \cdot T_{\text{Dec}}) = Ar_i, \quad (\text{A.11})$$

$$-\dot{n}_{\text{DecOut},i} + \xi_i \cdot \dot{n}_{\text{Out},i} = 0, \quad 1 \leq i \leq 9, \quad (\text{A.12})$$

$$-\dot{n}_{\text{DecReflux},i} + (1 - \xi_i) \cdot \dot{n}_{\text{Out},i} = 0, \quad 1 \leq i \leq 9, \quad (\text{A.13})$$

Table A.2: Degree of freedom (DOF) bounds

Lower bound	DOF	Upper bound
368.15 K	T	388.15 K
263.75 K	T_{Dec}	308.15 K
1 bar	p	40 bar
2 L	V_{R}	50 L
1e−20 kg	m_{cat}	10 kg

where the equilibrium coefficients K_{Dec} describe $\xi_i = K_{\text{Dec},i} / (1 + K_{\text{Dec},i})$. The parameters $A_{\text{Dec},h,i}$, $1 \leq h \leq 3$ are derived from Kiedorf et al. (2014).

The degrees of freedom (DOF) are the reactor temperature T , reactor volume V_{R} , reactor pressure p , catalyst mass m_{cat} within the reactor, and decanter temperature T_{Dec} . The bounds of the DOF are given in Table A.2.

The model was implemented in GAMS and is optimized using GAMS/BARON. Owing to several nonlinearities and the complex GLE formulation to capture the solvent's influence on the reactions, it is suitable for simulation and local optimization but solving it to global optimality within a reasonable amount of time is not feasible. In the following, model reformulations are systematically used to reduce computational effort.

A desktop PC with 3.00 GHz Intel Xeon E5450 Processor and 30 GB memory space was used to conduct the computations. Global optimality could not be achieved within 24 hours on this equipment.

MODEL REFORMULATIONS

Four model reformulation steps are conducted in the following:

1. standard reformulations and bound propagation,
2. logarithmic reformulation and simultaneous convexification,
3. GLE approximation by an ANN, and
4. a simplification of the ANN.

STANDARD REFORMULATIONS AND BOUND PROPAGATION Reformulating explicit constraints into implicit constraints is a typical method for reducing computation times in a global optimization problem. There are two big benefits to doing this: fewer nonlinearities and certain variables can be removed. Both significantly decrease the computational effort. Equations (A.12) and (A.13) can be used as an illustrative example. ξ_i can be eliminated by an implicit reformulation,

$$\dot{n}_{\text{DecReflux},i} - K_{\text{Dec},i} \cdot \dot{n}_{\text{DecReflux},i} = 0, \quad (\text{A.14})$$

$$\dot{n}_{\text{Out},i} - \dot{n}_{\text{DecOut},i} - \dot{n}_{\text{DecReflux},i} = 0. \quad (\text{A.15})$$

An linear mass balance can be obtained via an introduction of new variables and a redefinition of existing ones,

$$-\dot{n}_{\text{Out},i} + \dot{n}_{\text{In},i} + \sum_{j=1}^8 v_{j,i} \cdot r_{\text{cat},j} = 0, \quad 1 \leq i \leq 9, \quad (\text{A.16})$$

where $m_{\text{cat}} = c_{\text{cat}} \cdot V_{\text{R}} \cdot m_{\text{molar,cat}}$ and $r_{\text{cat},j} = m_{\text{cat}} \cdot r_j$. To make the subsequent logarithmic reformulation more efficient, multiplicative terms are combined.

Real-world models are only reliable within specific intervals. The search-space can be reduced by exploiting these limitations, thereby significantly reducing the computational complexity. The approach of reducing the search-space by calculating the upper and lower bounds of dependent variables is called bound propagation. This approach is beneficial for the rate and equilibrium coefficients in the present model, as they are temperature-dependent.

The computation time is decreased from over 24 hours to 106 minutes using the above approaches.

LOGARITHMIC REFORMULATION Computationally expensive exponential terms are eliminated by using logarithmic reformulation. Nonlinear operations, such as multiplications and fractions, are also substituted by logarithmic reformulation with linear operations, such as summations.

Take, for instance, the reaction rates in Equation (A.6) that can be reformulated to

$$r_{\log,j} = \sum_i \nu_{j,i} \cdot \log(c_i) - \log \left(1 + \sum_i \left(K_{i,1} \cdot c_i^{K_{i,2}} \right) \right), \quad (\text{A.17})$$

with $r_{\log,j} = \log(r_j)$.

The computation time is decreased from 106 minutes to 34 minutes by using the logarithmic reformulation additionally to the standard approaches.

SIMULTANEOUS CONVEXIFICATION An advanced strategy for solving nonlinear optimization problems to global optimality relies on solving the feasible sets' convex relaxations. In a widely used approach for constructing those relaxations, each nonlinear term is separately replaced by a convex underestimator. Ballerstein (2013) investigated simultaneous convexification of several nonlinear terms. Certain linear combinations of nonlinear terms can be derived to obtain additional constraints. Although these constraints are redundant for the original problem, they can improve the convex relaxations. This approach is applied to the present model's reaction rates, where convex constraints can be obtained by a linear combination of the reaction rates.

Adding these constraints to the reformulated model reduces the computational time from 106 minutes to 77 minutes. Using all three approaches presented so far together reduces the computation time to 22 minutes.

ARTIFICIAL NEURAL NETWORK Additional nonlinearities are introduced into the model using a hyperbolic tangent as the activation function for the ANN. These nonlinearities are removed by a piece-wise linear approximation of the hyperbolic tangent, coming at the cost of converting the NLP into an MINLP with reduced accuracy. However, the objective function error is only 1 % if three linear subdivisions are introduced to approximate the hyperbolic tangent.

The computation time is reduced significantly from 22 minutes to 5 minutes.

APPENDIX B

THERMODYNAMIC AND PHYSICAL DATA

This appendix includes thermodynamic and physical species data, as well as cost function parameters from [Section 4.2](#).

Table B.1: Vapor pressure parameters ($10^{a_0+a_1T^{-1}+a_2\log_{10}(T)+a_3T+a_4T^2}$)[mmHg]. Vapor pressure correlations for DSUC and THPO are fitted with the method presented in (Moller et al., 2008), the other values are taken from (Yaws, 1998).

	a_0	a_1	a_2	a_3	a_4
DMF	-47.9857	$-2.385 \cdot 10^3$	28.8	$-5.8596 \cdot 10^{-2}$	$3.1386 \cdot 10^{-5}$
DSUC	117.8014	$-6.3944 \cdot 10^3$	-42.5731	$3.0869 \cdot 10^{-2}$	$-9.2995 \cdot 10^{-6}$
THPO	74.2227	$-4.2846 \cdot 10^3$	-25.9627	$1.8373 \cdot 10^{-2}$	$-5.4143 \cdot 10^{-6}$
C10en	2.2678	$-3.12 \cdot 10^3$	5.43	$-2.01 \cdot 10^{-2}$	$1.12 \cdot 10^{-5}$
C12an	-8.5899	$-3.5241 \cdot 10^3$	10.806	$-2.8161 \cdot 10^{-2}$	$1.4267 \cdot 10^{-5}$
C11al	-31.8129	$-3.14 \cdot 10^3$	20.4	$-3.73 \cdot 10^{-2}$	$1.75 \cdot 10^{-5}$

Table B.2: Relative volatilities. The indices DMF, DMS, and THPO indicate which solvent was used in the mixture. The table is to be read from top to bottom, i. e. each column holds the data for a mixture employing the respective solvent. Entries denoted by [-] are not relevant for the respective mixture. The relative volatilities were calculated following [Equation \(4.8\)](#).

	α_{DMF}	α_{DMS}	α_{THPO}
DMF	31.9942	[-]	[-]
DMS	[-]	4.0208	[-]
THPO	[-]	[-]	22.4438
C12an	3.1307	2.7766	3.0696
C10en	16.8669	11.7308	15.8739
C11al	1	1	1

Table B.3: Cost function parameters for each investigated solvent

Parameter	DMF	DSUC	THPO
κ_1	17764	15071	17783
κ_2	2463.6	2988.6	2202.3
κ_3	7.2	7.2	7.2
κ_4	0.81	0.81	0.81
κ_5	1	1	1
κ_6	0.525	0.525	0.525
κ_7	155.4357	202.9541	133.1394
κ_8	-0.8	-0.8	-0.8
κ_9	0.97	0.97	0.97
κ_{10}	1	1	1
κ_{11}	0.725	0.725	0.725
λ_1	64191	168833	51628
λ_2	20638	19656	20695
λ_3	-56984	-182134	-53368
λ_4	-9.6219	-4.2671	-12.5046
λ_5	98841	439840	71678
η_{C10en}	$0.21041 \cdot 10^{-3}$	$0.21041 \cdot 10^{-3}$	$0.21041 \cdot 10^{-3}$
η_{psol}	$0.07915 \cdot 10^{-3}$	$0.12395 \cdot 10^{-3}$	$0.08491 \cdot 10^{-3}$
η_{C12an}	$0.233642 \cdot 10^{-3}$	$0.233642 \cdot 10^{-3}$	$0.233642 \cdot 10^{-3}$
η_{C11al}	$0.195184 \cdot 10^{-3}$	$0.195184 \cdot 10^{-3}$	$0.195184 \cdot 10^{-3}$
θ_1	504155	504155	504155
θ_2	0.586667	0.586667	0.586667
θ_3	$2.9661 \cdot 10^{12}$	$2.9661 \cdot 10^{12}$	$2.9661 \cdot 10^{12}$
θ_4	85150	85150	85150
θ_5	17248.42	17248.42	17248.42
θ_6	0.62	0.62	0.62
θ_7	$2.8512 \cdot 10^7$	$2.8512 \cdot 10^7$	$2.8512 \cdot 10^7$
$\theta_{8,C10en}$	0.661	0.661	0.661
$\theta_{8,psol}$	0.0731	0.0731	0.0731
$\theta_{8,C12an}$	0.0714	0.0714	0.0714

MOLECULE GROUP PARAMETERS

This appendix gives the molecule group parameters for [Chapter 5](#).

Table C.1: Parameter values for groups in Table 5.1.

Group	S _S	S _D	S _C	B _S	B _D	B _{SD}	B _C	B _{CE}	B _{NCE}	B _{CH}	B _{MCside}	B _{CC}	N ^{UP}
sCH ₃	1	0	0	1	0	0	0	1	0	1	0	0	10
ssCH ₂	2	0	0	1	0	0	0	0	1	1	0	0	10
dCH ₂	0	1	0	0	1	0	0	1	0	1	0	0	10
sssCH	3	0	0	1	0	0	0	0	1	1	0	0	3
sdCH	1	1	0	0	0	1	0	0	1	1	0	0	3
ssssC	4	0	0	1	0	0	0	0	1	0	0	0	1
ssdC	2	1	0	0	0	1	0	0	1	0	0	0	1
rrACH	0	0	2	0	0	0	1	0	1	0	0	0	6
rrACs	1	0	2	1	0	0	1	1	0	0	1	0	1
rrACCH ₃	0	0	2	0	0	0	1	0	1	1	0	0	1
rrACCH ₂ s	1	0	2	1	0	0	1	1	0	1	1	0	1
rrACCHss	2	0	2	1	0	0	1	0	1	1	1	0	1
rrACCHd	0	1	2	0	1	0	1	1	0	1	1	0	1
sCH ₃ CO	1	0	0	1	0	0	0	1	0	0	0	0	1
sCH ₂ COs	2	0	0	1	0	0	0	0	1	0	0	0	1
CH=Os	1	0	0	1	0	0	0	1	0	0	0	0	1
sCH ₃ COO	1	0	0	1	0	0	0	1	0	0	0	0	1
sCH ₂ COOs	2	0	0	1	0	0	0	0	1	0	0	0	1
sOCH ₃	1	0	0	1	0	0	0	1	0	0	0	0	1
sOCH ₂ s	2	0	0	1	0	0	0	0	1	0	0	0	1
sOCHss	3	0	0	1	0	0	0	0	1	0	0	0	1
OCHd	0	1	0	0	1	0	0	1	0	0	0	0	1
sCH ₂ NHs	2	0	0	1	0	0	0	0	1	0	0	0	1
sCH ₂ CN	1	0	0	1	0	0	0	1	0	0	0	0	2
sCOOH	1	0	0	1	0	0	0	1	0	0	0	0	1

ALTERNATIVE SEARCH-SPACE REDUCTION

This appendix was published in parts in: Computer Aided Chemical Engineering, 48, Tobias Kessler, Christian Kunde, Steffen Linke, Kevin McBride, Kai Sundmacher, and Achim Kienle. "Efficient global optimization of a novel hydroformylation process", pp. 745-750 (2020).

INTRODUCTION

The reduction of the search-space shown in [Section 5.3](#) is based on σ -bands. This Appendix points out how to use the Hansen Solubility Parameters to classify feasible solvent candidates as an alternative. The Hansen solubility parameter (HSP) δ_0 is calculated using a group contribution method (Stefanis and Panayiotou, 2008) for approximating the thermodynamic properties of the solvent candidate.

The desired solvent for the hydroformylation process needs to fulfill certain property constraints. The boiling point is restricted to $T_b \in [350, 600]$ K because at higher temperatures, the aldehyde is transformed into an unwanted side product. It is necessary for the new solvent to have the same characteristics as DMF regarding the catalyst and the reactant. The basic idea is that solvents with a similar HSP also have similar behavior. Therefore, the HSP is constrained to a neighborhood of the HSP value of DMF $\delta_0 \in [24, 25.6] \sqrt{\text{MPa}}$. Thereby the search-space is significantly reduced and the problem becomes feasible.

In addition, the phase compositions are constrained to $x_1^{\text{I}} \in [0.7648, 0.9671]$, and $x_2^{\text{II}} \geq 0.3$ following the optimization study in [Section 4.2](#) to ensure

that the solver is not trapped in the trivial solution of the LLE. The indices 1 and 2 refer to the candidate solvent and dodecane, respectively.

HIERARCHICAL DECOMPOSITION

To make the optimization problem tractable, it is split into two sub-problems. As the boiling point influences the process' economics, the solver searches for a molecule with a low boiling point T_b . Furthermore, as DMF should be replaced by a safe alternative, PEL and LD_{50} should be low.

$$\min_n J_1 = T_b - 20 PEL - 50 LD_{50}, \quad (\text{D.1})$$

20 molecules are generated.

The objective function [Equation \(D.1\)](#) can be interpreted as a weighted sum of objectives and can, therefore, be classified as a multi-objective optimization. However, the weights of the weighted sum are fixed, leaving a single-objective problem.

To ensure that the mixtures with the new candidates exhibit the necessary phase behavior, the objective of the second step is to maximize the catalyst distribution between the phases,

$$\max_x J_2 = x_{\text{cat}}^{\text{I}} - x_{\text{cat}}^{\text{II}}. \quad (\text{D.2})$$

The phase behavior itself is modeled using COSMO-SAC. Owing to the constraint mentioned above, this optimization problem only yields a solution for mixtures exhibiting a miscibility gap.

The optimizations are implemented as MINLPs and solved using the GAMS 26.1.0 framework with deterministic global optimization software BARON 18.11.12., Cplex 12.8.0 is used as an LP/MIP subsolver and CONOPT 4.09 is utilized as an NLP subsolver. The calculations are carried out on a Linux PC with 3.40 GHz Intel Core i7-6700 CPU and 16 GB memory. The optimization time of the second optimization step is constrained to 9000 s.

RESULTS

The first sub-problem solution yields 20 solvent candidates. Based on the COSMO-SAC estimates, 14 candidates were excluded in the second subproblem. The solver could not find feasible equilibrium compositions for the eliminated candidates within the maximum permissible computation time. The applied models predict that the boiling temperature, EHS parameters, and phase behavior are met for all remaining solvent candidates.

HIERARCHICAL ITERATIVE GLOBAL OPTIMIZATION

INTRODUCTION

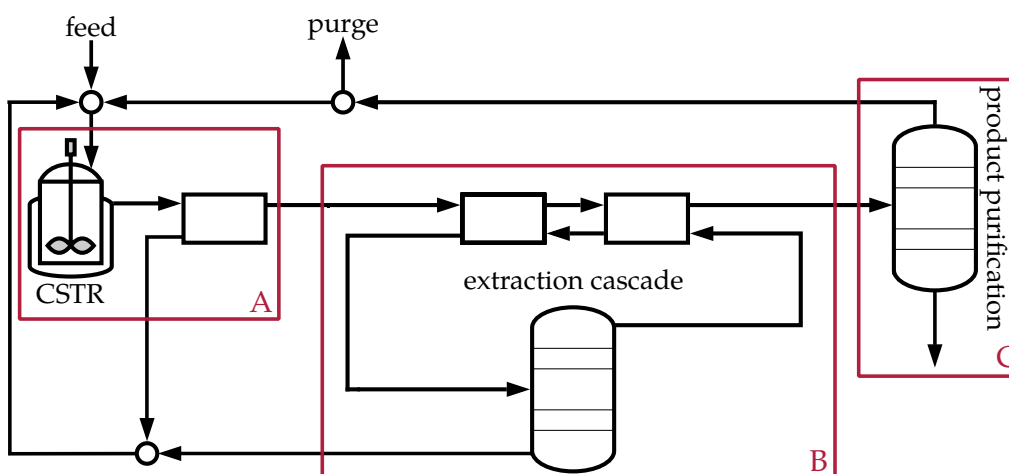


Figure E.1: Process flowsheet considered in this appendix. The process consists of three parts, numbered and highlighted in maroon: (A) reactor and phase separator, (B) extraction cascade, and (C) product purification.

The process configuration considered in this appendix is depicted in [Figure E.1](#). The process consists of a reactor with a subsequent phase separator, an extraction cascade for the recovery of the catalyst, and an additional distillation column for product separation and reactant recycle.

As described in [Part ii](#) of this thesis, each of these process parts can be efficiently optimized to global optimality within a relatively small amount of time. However, if the process parts are interconnected and the recycles are closed, the computational effort becomes infeasible for optimization studies such as the study presented in [Section 4.2](#).

Multiple approaches to overcome this problem come to mind. One approach is presented in [Section 4.2](#), where the optimization problem is solved using a multi-start approach with many starting points. Another approach is hierarchical global optimization, which will be discussed in the following.

MODEL DESCRIPTION

Each sub-model of [Figure E.1](#) will be described in the following.

(A) REACTOR AND PHASE SEPARATOR The reactor model and the species assumed to be taking part in the reaction are identical to that of the reactor model in [Appendix A](#), therefore it will not be described in more detail in this appendix.

However, the decanter model differs from the model used in [Appendix A](#). The model is taken from McBride and Sundmacher (2015). A linear regression for the estimation of the phase partition coefficients, fitted to a modified UNIFAC (Dortmund) model is employed. The decanter temperature is assumed to be constant at 298.15 K. In contrast to (McBride and Sundmacher, 2015), the calculations conducted in this appendix are carried out in mol/s rather than kmol/min.

The degrees of freedom are the feed stream F_{feed} , the reactor temperature T_r , the reactor pressure p_r , the catalyst mass m_{cat} , and the purge stream Ξ .

The cost function consists of the bare vessel costs of the reactor and the decanter, as defined in (McBride and Sundmacher, 2015), a penalty term for the purge stream, $\Xi \cdot 85150$, as well as the costs for the catalyst mass $m_{\text{cat}} \cdot 85150$ and the solvents.

(B) EXTRACTION CASCADE The cascade is assumed to consist of two decanters in a countercurrent setup, and an extraction solvent cascade.

The decanter model is the same as above.

The column is approximated through a quadratic cost function. The function is fitted to the results of 100 global optimization runs using a Fenske-Underwood-Gilliland column model. In this optimization runs, the

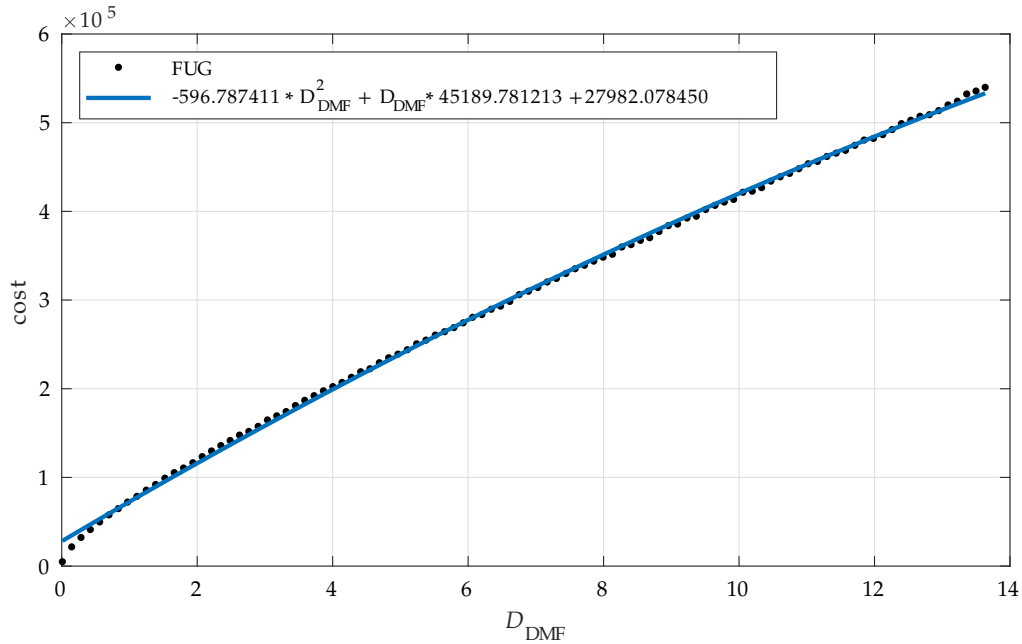


Figure E.2: Optimization results (black dots) and cost surrogate (blue line) of the extraction cascade column. The distillate flow D_{DMF} is given in mol/s.

varying input streams of the column are calculated as $F_{DMF} \in [1, 17]$ mol/s; $r \in [0.6, 0.8]$ (random number); $F_{dec} = 0.5 \cdot r$ mol/s; $F_{dod} = 0.2 \cdot r$ mol/s; $F_{tri} = 0.5 - F_{dec} - F_{dod}$ mol/s. The molar input stream of DMF is increased incrementally during the 100 optimization runs. As product specification, the distillate consists of 99 % DMF and 1 % decane. The resulting cost function is depicted in Figure E.2.

The bare vessel cost for the decanter are the same as above. Additionally, a penalty term for catalyst leaching is implemented as $F_{cat,out}^{casc} \cdot 451890.781$.

(C) PRODUCT PURIFICATION To model the product purification column, the column formulation from Mertens et al. (2018) is used. However, no bound tightening is performed.

The degrees of freedom are the distillate flow, D in mol/s, the bottom product flow rate, B in mol/s, the number of stages in the rectifying section, l_r , and the number of stages in the stripping section, l_s .

Following Hentschel et al. (2014), the plant is assumed to have an annual production of 10.000 t of n-tridecanal, with 220 production days. The

purity requirements for the bottom product stream of the column are set to 99.5 % tridecanal.

PROCESS COUPLING

Each of the three process parts has to be optimized individually. However, the individual global optima of the single process parts, and that of the whole process do not coincide. Therefore, a process coupling of some sort is necessary. As a proof of concept, this process coupling is done through additional penalty terms based on process insight. These penalty terms are:

- (A) sum of output stream and catalyst streams (both result in increase of cost for subsequent decanter cascade),
- (A) additional inequality constraint for the tridecanal output ($F_{\text{out,tri}}^{\text{reac}} \geq 1.8 \text{ mol/s}$). If not met, there would not be enough product to meet the plant specifications regarding the annual product,
- (B) additional inequality constraint for the tridecanal output ($F_{\text{out,1,tri}}^{\text{casc}} \geq 1.7772 \text{ mol/s}$), for the same reason as above,

where (A) and (B) refer to the respective process parts as depicted in [Figure E.1](#).

The inputs to the single process parts are kept constant during their respective optimizations. Therefore, as an initialization, each input of part (A) needs to be defined as a starting point.

Each process part's optimization problem is solved to global optimality using the GAMS 26.1.0 framework with the deterministic global optimization software BARON 18.11.12., Cplex 12.8.0 is used as an LP/MIP subsolver and CONOPT 4.09 is utilized as an NLP subsolver. The calculations are carried out on a Linux PC with 3.40 GHz Intel Core i7-6700 CPU and 16 GB memory.

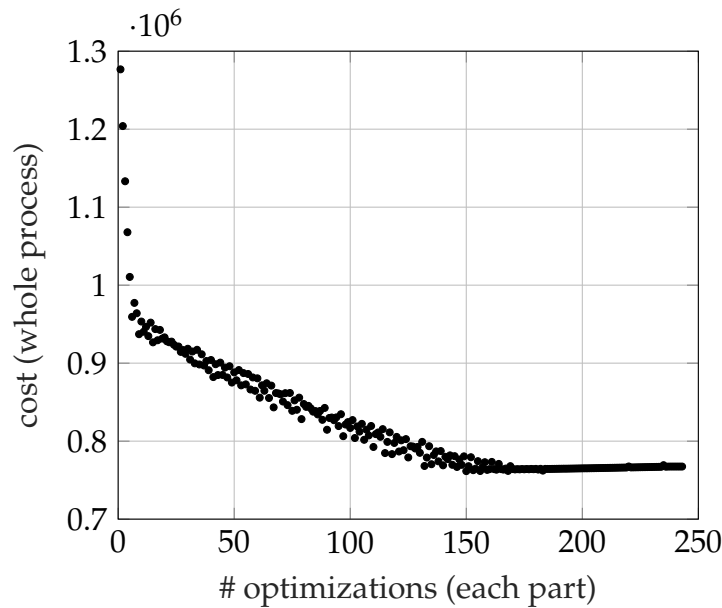


Figure E.3: Iterative optimization results. The black dots denote the whole model's cost. Each process part needs to be optimized to obtain one dot.

OPTIMIZATION RESULT

An exemplary optimization run is depicted in [Figure E.3](#). The starting point was chosen quite far away from the optimal solution: $D = 15.1679$ mol/s, $B = 3.2974$ mol/s, $D_{\text{DMF}} = 2.0719$ mol/s, $D_{\text{dec}} = 12.5772$ mol/s, $D_{\text{dod}} = 0.4945$ mol/s, $D_{\text{tri}} = 0.0243$ mol/s, $F_{\text{out},2,\text{DMF}}^{\text{casc}} = 1.4038$, $F_{\text{out},2,\text{dec}}^{\text{casc}} = 0.5345$, $F_{\text{out},2,\text{dod}}^{\text{casc}} = 0.1160$, $F_{\text{out},2,\text{tri}}^{\text{casc}} = 0.1797$, $\Xi = 0.0235$. The optimization stops if two consecutive cost values are identical. A total number of 243 optimization steps were necessary until the optimization converged. In total, these optimization steps took 63 hours.

After an initial step descent, the cost values start oscillating. These oscillations cease near iteration 170, where an almost straight line is obtained. Instead of further decreasing, the cost increases slightly from this point on.

CONCLUSION

This appendix presented preliminary results for the hierarchical iterative global optimization of a hydroformylation process. The process was split into three parts, the connection of the parts is achieved through heuristic penalty terms. Iteratively solving each process part to global optimization converges to a constant solution.

Although this approach is promising, further investigations are necessary. The heuristics may be improved or replaced by more sophisticated constraints. Furthermore, the termination condition is strict and leads to unnecessarily long computation times.

BIBLIOGRAPHY

- Agency, European Chemicals (2017). "Guidance on information requirements and chemical safety assessment." In: chap. R.11: PBT and vPvB assessment. ISBN: 978-92-9495-839-6. DOI: [10.2823/128621](https://doi.org/10.2823/128621) (cit. on p. [90](#)).
- Anastas, Paul T. and John C. Warner (1998). *Green Chemistry: Theory and Practice*. Oxford University Press. ISBN: 0-19-850234-6 (cit. on p. [1](#)).
- Audet, Charles, Michael Kokkolaras, Sébastien Le Digabel, and Bastien Talgorn (2018). "Order-based error for managing ensembles of surrogates in mesh adaptive direct search." In: *Journal of Global Optimization* 70.3, pp. 645–675. ISSN: 1573-2916. DOI: [10.1007/s10898-017-0574-1](https://doi.org/10.1007/s10898-017-0574-1) (cit. on p. [33](#)).
- Austin, Nick D., Nikolaos V. Sahinidis, and Daniel W. Trahan (2016). "Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques." In: *Chemical Engineering Research and Design* 116, pp. 2–26. DOI: [10.1016/j.cherd.2016.10.014](https://doi.org/10.1016/j.cherd.2016.10.014) (cit. on p. [82](#)).
- Austin, Nick D., Nikolaos V. Sahinidis, and Daniel W. Trahan (2017). "A COSMO-based approach to computer aided mixture design." In: *Chemical Engineering Science* 159, pp. 93–105. DOI: [10.1016/j.ces.2016.05.025](https://doi.org/10.1016/j.ces.2016.05.025) (cit. on p. [93](#)).
- Ballerstein, Martin (2013). "Convex Relaxations for Mixed-Integer Nonlinear Programs." PhD thesis (cit. on p. [118](#)).
- Ballerstein, Martin, Achim Kienle, Christian Kunde, Dennis Michaels, and Robert Weismantel (2015). "Deterministic global optimization of binary hybrid distillation/melt-crystallization processes based on relaxed MINLP formulations." In: *Optimization and Engineering* 16.2, pp. 409–440. ISSN: 1573-2924. DOI: [10.1007/s11081-014-9267-5](https://doi.org/10.1007/s11081-014-9267-5) (cit. on p. [32](#)).
- Bardow, André, Klaas Steur, and Joachim Gross (2010). "Continuous-Molecular Targeting for Integrated Solvent and Process Design." In:

- Industrial & Engineering Chemistry Research* 49.6, pp. 2834–2840. ISSN: 0888-5885. DOI: [10.1021/ie901281w](https://doi.org/10.1021/ie901281w) (cit. on p. 82).
- Behr, A. and A. J. Vorholt (2012). “Hydroformylation and Related Reactions of Renewable Resources.” In: ed. by Michael A.R. Meier, Bert M. Weckhuysen, and Pieter C. A. Bruijninx. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 103–127. ISBN: 978-3-642-28288-1. DOI: [10.1007/978-3-642-28288-1_3](https://doi.org/10.1007/978-3-642-28288-1_3) (cit. on p. 2).
- Behr, Arno, Dietmar Obst, and Alfred Westfechtel (2005). “Isomerizing hydroformylation of fatty acid esters: Formation of ω -aldehydes.” In: *European Journal of Lipid Science and Technology* 107, pp. 213–219. DOI: [10.1002/ejlt.200401123](https://doi.org/10.1002/ejlt.200401123) (cit. on p. 2).
- Behr, Arno and Andreas J Vorholt (2017). *Homogeneous Catalysis with Renewables*. Springer. DOI: [10.1007/978-3-319-54161-7](https://doi.org/10.1007/978-3-319-54161-7) (cit. on p. 1).
- Bell, Ian H., Erik Mickoleit, Chieh-Ming Hsieh, Shiang-Tai Lin, Jadran Vrabec, Cornelia Breitkopf, and Andreas Jäger (2020). “A Benchmark Open-Source Implementation of COSMO-SAC.” In: *Journal of Chemical Theory and Computation* 16.4, pp. 2635–2646. DOI: [10.1021/acs.jctc.9b01016](https://doi.org/10.1021/acs.jctc.9b01016) (cit. on p. 25).
- Belotti, Pietro, Christian Kirches, Sven Leyffer, Jeff Linderoth, James Luedtke, and Ashutosh Mahajan (2013). “Mixed-integer nonlinear optimization.” In: *Acta Numerica* 22, pp. 1–131. DOI: [10.1017/S0962492913000032](https://doi.org/10.1017/S0962492913000032) (cit. on p. 12).
- Bennett, Kristin P. and Colin Campbell (2000). “Support Vector Machines: Hype or Hallelujah?” In: *ACM SIGKDD Explorations* 2.2, pp. 1–13. DOI: [10.1145/380995.380999](https://doi.org/10.1145/380995.380999) (cit. on p. 13).
- Bhosekar, Atharv and Marianthi Ierapetritou (2018). “Advances in surrogate based modeling, feasibility analysis, and optimization: A review.” In: *Computers & Chemical Engineering* 108, pp. 250–267. ISSN: 0098-1354. DOI: [10.1016/j.compchemeng.2017.09.017](https://doi.org/10.1016/j.compchemeng.2017.09.017) (cit. on p. 13).
- Biang, Jonas, Kai U. Künnemann, Lisa Goclik, Lasse Schurm, Dieter Vogt, and Thomas Seidensticker (2020a). “Tandem Catalytic Amine Synthesis from Alkenes in Continuous Flow Enabled by Integrated Catalyst Recycling.” In: *ACS Catalysis* 10.11, pp. 6463–6472. DOI: [10.1021/acscatal.0c01465](https://doi.org/10.1021/acscatal.0c01465) (cit. on p. 70).

- Bianga, Jonas, Niklas Kopplin, Jonas Hülsmann, Dieter Vogt, and Thomas Seidensticker (Oct. 2020b). "Rhodium-Catalysed Reductive Amination for the Synthesis of Tertiary Amines." In: *Advanced Synthesis & Catalysis* 362.20, pp. 4415–4424. DOI: [10.1002/adsc.202000746](https://doi.org/10.1002/adsc.202000746) (cit. on p. 70).
- Bianga, Jonas, Kai Künnemann, Tom Gaide, Andreas J. Vorholt, Thomas Seidensticker, Jens Dreimann, and Dieter Vogt (2019). "Thermomorphic Multiphase Systems: Switchable Solvent Mixtures for the Recovery of Homogeneous Catalysts in Batch and Flow Processes." In: *Chemistry a European Journal* 25, pp. 1–24. DOI: [10.1002/chem.201902154](https://doi.org/10.1002/chem.201902154) (cit. on pp. 2, 60).
- Biegler, Lorenz T. and Ignacio E. Grossmann (2004). "Retrospective on optimization." In: *Computers & Chemical Engineering* 28, pp. 1169–1192. DOI: [10.1016/j.compchemeng.2003.11.003](https://doi.org/10.1016/j.compchemeng.2003.11.003) (cit. on p. 11).
- Boukouvala, Fani and Christodoulos A. Floudas (2017). "ARGONAUT: AlgoRithms for Global Optimization of coNstrAined grey-box compUTational problems." In: *Optimization Letters* 11.5, pp. 895–913. ISSN: 1862-4480. DOI: [10.1007/s11590-016-1028-2](https://doi.org/10.1007/s11590-016-1028-2) (cit. on p. 13).
- Boukouvala, Fani, Ruth Misener, and Christodoulos A. Floudas (2016). "Global optimization advances in Mixed-Integer Nonlinear Programming, MINLP, and Constrained Derivative-Free Optimization, CDFO." In: *European Journal of Operational Research* 252.3, pp. 701–727. DOI: [10.1016/j.ejor.2015.12.018](https://doi.org/10.1016/j.ejor.2015.12.018) (cit. on pp. 12, 13).
- Boyd, Stephen and Lieven Vandenberghe (2013). *Convex Optimization*. Cambridge University Press. DOI: [10.1017/CB09780511804441](https://doi.org/10.1017/CB09780511804441) (cit. on p. 11).
- Burger, Jakob, Vasileios Papaioannou, Smitha Gopinath, George Jackson, Amparo Galindo, and Claire S. Adjiman (2015). "A hierarchical method to integrated solvent and process design of physical CO₂ absorption using the SAFT- γ Mie approach." In: *AIChE Journal* 61, pp. 3249–3269. DOI: [10.1002/aic.14838](https://doi.org/10.1002/aic.14838) (cit. on p. 82).
- Caballero, Jose A. and I. E. Grossmann (2008). "An Algorithm for the Use of Surrogate Models in Modular Flowsheet Optimization." In: *American Institute of Chemical Engineers Journal* 54.10, pp. 2633–2650. DOI: [10.1002/aic.11579](https://doi.org/10.1002/aic.11579) (cit. on pp. 15, 37).

- Chen, Chau-Chyun and Yuhua Song (2004). "Solubility Modeling with a Nonrandom Two-Liquid Segment Activity Coefficient Model." In: *Industrial & Engineering Chemical Research* 43, pp. 8354–8362. DOI: [10.1021/ie049463u](https://doi.org/10.1021/ie049463u) (cit. on p. 69).
- Churi, Nachiket and Luke E. K. Achenie (1996). "Novel Mathematical Programming Model for Computer Aided Molecular Design." In: *Industrial & Engineering Chemistry Research* 35, pp. 3788–3794. DOI: [10.1021/ie9601920](https://doi.org/10.1021/ie9601920) (cit. on pp. 84, 87).
- Cignitti, Stefano, Seyed Soheil Mansouri, John M. Woodley, and Jens Abildskov (2017). "Systematic Optimization-Based Integrated Chemical Product–Process Design Framework." In: *Industrial and Engineering Chemistry Research* 57.2, pp. 677–688. DOI: [10.1021/acs.iecr.7b04216](https://doi.org/10.1021/acs.iecr.7b04216) (cit. on p. 82).
- Cozad, Alison, Nikolaos V. Sahinidis, and David C. Miller (2014). "Learning surrogate models for simulation-based optimization." In: *American Journal of Chemical Engineering* 60.6, pp. 2211–2227. DOI: [10.1002/aic.14418](https://doi.org/10.1002/aic.14418) (cit. on p. 13).
- Dorn, Cornelius, Thomas E. Güttinger, Gary J. Wells, Manfred Morari, Achim Kienle, Eberhard Klein, and Ernst-Dieter Gilles (1998). "Stabilization of an Unstable Distillation Column." In: *Industrial & Engineering Chemistry Research* 37.2, pp. 506–515. DOI: [10.1021/ie9703447](https://doi.org/10.1021/ie9703447) (cit. on p. 50).
- Dreimann, J. M., H. Warmeling, J. N. Weimann, K. Künnemann, A. Behr, and A. J. Vorholt (2016). "Increasing selectivity of the hydroformylation in a miniplant: Catalyst, solvent, and olefin recycle in two loops." In: *AIChE Journal* 62.12, pp. 4377–4383. ISSN: 1547-5905. DOI: [10.1002/aic.15345](https://doi.org/10.1002/aic.15345) (cit. on pp. 68, 70).
- Dreimann, Jens M., Frank Hoffmann, Mirko Skiborowski, Arno Behr, and Andreas J. Vorholt (2017). "Merging Thermomorphic Solvent Systems and Organic Solvent Nanofiltration for Hybrid Catalyst Recovery in a Hydroformylation Process." In: *Industrial & Engineering Chemical Research* 56.5. DOI: [10.1021/acs.iecr.6b04249](https://doi.org/10.1021/acs.iecr.6b04249) (cit. on p. 2).

- Dubourg, V. and B. Sudret (2014). "Meta-model-based importance sampling for reliability sensitivity analysis." In: *Structural Safety* 49, pp. 27–36. DOI: [10.1016/j.strusafe.2013.08.010](https://doi.org/10.1016/j.strusafe.2013.08.010) (cit. on p. 17).
- EPA.gov (Oct. 2020). *Safer Chemical Ingredients List*. URL: https://www.epa.gov/sites/production/files/2015-09/safer_chemical_ingredients_list.xls (cit. on p. 100).
- Eason, John P. and Lorenz T. Biegler (2016). "A trust region filter method for glass box/black box optimization." In: *AIChE Journal* 62.9, pp. 3124–3136. DOI: [10.1002/aic.15325](https://doi.org/10.1002/aic.15325) (cit. on p. 33).
- Eckert, F. and A. Klamt (2016). *COSMOtherm, version C16*. COSMOlogic GmbH & Co. KG, Leverkusen, Germany (cit. on p. 54).
- Eden, M.R., S.B. Jørgensen, R. Gani, and M.M. El-Halwagi (2004). "A novel framework for simultaneous separation process and product design." In: *Chemical Engineering and Processing* 43, pp. 595–608. DOI: [10.1016/j.cep.2003.03.002](https://doi.org/10.1016/j.cep.2003.03.002) (cit. on p. 82).
- European Parliament (2006). *Registration, Evaluation, Authorisation and Restriction of Chemicals: REACH* (cit. on p. 3).
- Fail, Patricia A., Julia D. George, Thomas B. Grizzle, and Jerrold J. Heindel (1998). "Formamide and Dimethylformamide: Reproductive Assessment by Continuous Breeding in Mice." In: *Reproductive Toxicology* 12.3. DOI: [10.1016/S0890-6238\(98\)00011-2](https://doi.org/10.1016/S0890-6238(98)00011-2) (cit. on p. 3).
- Fleitmann, Lorenz, Jan Scheffczyk, Pascal Schäfer, Christian M. Jens, Kai Leonhard, and André Bardow (2018). "Integrated Design of Solvents in Hybrid Reaction-Separation Processes Using COSMO-RS." In: *Chemical Engineering Transactions* 2018.96, pp. 559–564. DOI: [10.3303/CET1869094](https://doi.org/10.3303/CET1869094) (cit. on p. 82).
- Forrester, Alexander I.J. and Andy J. Keane (2009). "Recent advances in surrogate-based optimization." In: *Progress in Aerospace Sciences* 45.1, pp. 50–79. ISSN: 0376-0421. DOI: [10.1016/j.paerosci.2008.11.001](https://doi.org/10.1016/j.paerosci.2008.11.001) (cit. on p. 13).
- Fredenslund, Aage, Russell L. Jones, and John M. Prausnitz (1975). "Group-contribution estimation of activity coefficients in nonideal liquid mixtures." In: *American Institute of Chemical Engineers Journal* 21.6, pp. 1086–1099. DOI: [10.1002/aic.690210607](https://doi.org/10.1002/aic.690210607) (cit. on pp. 22, 24).

- Gertig, Christoph, Kai Leonhard, and Andre Bardow (2020). "Computer-aided molecular and processes design based on quantum chemistry: current status and future prospects." In: *Current Opinion in Chemical Engineering* 27, pp. 89–97. DOI: [10.1016/j.coche.2019.11.007](https://doi.org/10.1016/j.coche.2019.11.007) (cit. on p. 82).
- Gleixner, Ambros et al. (2017). *The SCIP Optimization Suite 5.0*. Tech. rep. 17-61. Takustr.7, 14195 Berlin: ZIB (cit. on p. 12).
- Gmehling, Jürgen, Jiding Li, and Martin Schiller (1993). "A modified UNIFAC model. 2. Present parameter matrix and results for different thermodynamic properties." In: *Industrial & Engineering Chemical Research* 32.1, pp. 178–193. DOI: [10.1021/ie00013a024](https://doi.org/10.1021/ie00013a024) (cit. on p. 24).
- Goldberg, David E. (1993). *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley. ISBN: 0-201-15767-5 (cit. on pp. 11, 12).
- Gross, Joachim and Gabriele Sadowski (2001). "Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules." In: *Industrial & Engineering Chemistry Research* 40, pp. 1244–1260. DOI: [10.1021/ie0003887](https://doi.org/10.1021/ie0003887) (cit. on p. 63).
- Guggenheim, Edward Armand (1952). *Mixtures: The Theory of the Equilibrium Properties of Some Simple Classes of Mixtures, Solutions and Alloys*. Clarendon Press (cit. on p. 26).
- Haykin, Simon (1999). *Neural Networks: A Comprehensive Foundation*. 2nd ed. Prentice Hall. ISBN: 978-0132733502 (cit. on pp. 19–21).
- Henley, Ernest J. and J. D. Seader (1981). *Equilibrium-Stage Separation Operations in Chemical Engineering*. John Wiley & Sons, Inc. ISBN: 978-0-471-37108-3 (cit. on p. 66).
- Hentschel, Benjamin, Gregor Kiedorf, Martin Gerlach, Christof Hamel, Andreas Seidel-Morgenstern, Hannsjörg Freund, and Kai Sundmacher (2015). "Model-Based Identification and Experimental Validation of the Optimal Reaction Route for the Hydroformylation of 1-Dodecene." In: *Industrial & Engineering Chemistry Research* 54.6, pp. 1755–1765. DOI: [10.1021/ie504388t](https://doi.org/10.1021/ie504388t) (cit. on pp. 2, 61, 63, 114, 115).
- Hentschel, Benjamin, Andreas Peschel, Hannsjörg Freund, and Kai Sundmacher (2014). "Simultaneous design of the optimal reaction and process

- concept for multiphase systems." In: *Chemical Engineering Science* 115. Special Issue: InPROMPT – Integrated Chemical Processes with Liquid Multiphase Systems, pp. 69–87. DOI: [10.1016/j.ces.2013.09.046](https://doi.org/10.1016/j.ces.2013.09.046) (cit. on p. 131).
- Hernández, R. and S. Engell (2016). "Modelling and iterative real-time optimization of a homogeneously catalyzed hydroformylation process." In: *Computer Aided Chemical Engineering* 38, pp. 1–6. DOI: [10.1016/B978-0-444-63428-3.50005-9](https://doi.org/10.1016/B978-0-444-63428-3.50005-9) (cit. on p. 2).
- Hukkerikar, Amol Shivajirao, Sawitree Kalakul, Bent Sarup, Douglas M. Young, Gürkan Sin, and Rafiqul Gani (2012). "Estimation of Environment-Related Properties of Chemicals for Design of Sustainable Processes: Development of Group-Contribution+ (GC+) Property Models and Uncertainty Analysis." In: *Journal of Chemical Information and Modeling* 52, pp. 2823–2839. DOI: [dx.d10.1021/ci300350r](https://doi.org/dx.d10.1021/ci300350r) (cit. on p. 90).
- Ibrahim, Dauda, Megan Jobson, Jie Li, and Gonzalo Guillén-Gosálbez (2018). "Optimization-based design of crude oil distillation units using surrogate column models and a support vector machine." In: *Chemical Engineering Research and Design* 134, pp. 212–225. DOI: [10.1016/j.cherd.2018.03.006](https://doi.org/10.1016/j.cherd.2018.03.006) (cit. on p. 33).
- Illner, Markus, David Müller, Erik Esche, Tobias Pogrzeba, Marcel Schmidt, Reinhard Schomäcker, Günter Wozny, and Jens-Uwe Repke (2016). "Hydroformylation in Microemulsions: Proof of Concept in a Miniplant." In: *Industrial & Engineering Chemistry Research* 55.31. DOI: [10.1021/acs.iecr.6b00547](https://doi.org/10.1021/acs.iecr.6b00547) (cit. on p. 2).
- Joback, Kevin G. and R. C. Reid (1987). "Estimation of Pure-Component Properties from Group-Contributions." In: *Chemical Engineering Communications* 56.1-6, pp. 233–243. DOI: [10.1080/00986448708960487](https://doi.org/10.1080/00986448708960487) (cit. on p. 81).
- Joback, Kevin G. and George Stephanopoulos (1995). "Searching Spaces of Discrete Solutions: The Design of Molecules Possessing Desired Physical Properties." In: *Advances in Chemical Engineering* 21, pp. 257–311. DOI: [10.1016/S0065-2377\(08\)60075-7](https://doi.org/10.1016/S0065-2377(08)60075-7) (cit. on p. 81).
- Karunanithi, Aranprakash T., Luke E. K. Achenie, and Rafiqul Gani (2005). "A New Decomposition-Based Computer-Aided Molecular/Mixture

- Design Methodology for the Design of Optimal Solvents and Solvent Mixtures." In: *Industrial & Engineering Chemistry Research* 44.13, pp. 4785–4797. DOI: [10.1021/ie049328h](https://doi.org/10.1021/ie049328h) (cit. on p. 82).
- Keßler, Tobias, Christian Kunde, Nick Mertens, Dennis Michaels, and Achim Kienle (2018). "Global optimization of distillation columns using surrogate models." In: *SN Applied Sciences* 1.1, p. 11. ISSN: 2523-3971. DOI: [10.1007/s42452-018-0008-9](https://doi.org/10.1007/s42452-018-0008-9) (cit. on p. 32).
- Kiedorf, G., D.M. Hoang, A. Müller, A. Jörke, J. Markert, H. Arellano-Garcia, A. Seidel-Morgenstern, and C. Hamel (2014). "Kinetics of 1-dodecene hydroformylation in a thermomorphic solvent system using a rhodium-biphephos catalyst." In: *Chemical Engineering Science* 115. Special Issue: InPROMPT – Integrated Chemical Processes with Liquid Multiphase Systems, pp. 31–48. ISSN: 0009-2509. DOI: [10.1016/j.ces.2013.06.027](https://doi.org/10.1016/j.ces.2013.06.027) (cit. on pp. 2, 61, 63, 116).
- Kieslich, Chris A., Fani Boukouvala, and Christodoulos A. Floudas (2018). "Optimization of black-box problems using Smolyak grids and polynomial approximations." In: *Journal of Global Optimization*. ISSN: 1573-2916. DOI: [10.1007/s10898-018-0643-0](https://doi.org/10.1007/s10898-018-0643-0) (cit. on p. 33).
- Kılınç, Mustafa R. and Nikolaos V. Sahinidis (2018). "Exploiting integrality in the global optimization of mixed-integer nonlinear programming problems with BARON." In: *Optimization Methods and Software* 33.3, pp. 540–562. DOI: [10.1080/10556788.2017.1350178](https://doi.org/10.1080/10556788.2017.1350178) (cit. on pp. 12, 35, 66, 111).
- Klamt, Andreas (2005). *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*. Elsevier Science. ISBN: 9780444520326 (cit. on pp. 92, 93).
- Klamt, Andreas, Frank Eckert, and Wolfgang Arlt (2010). "COSMO-RS: An Alternative to Simulation for Calculating Thermodynamic Properties of Liquid Mixtures." In: *Annual Review of Chemical and Biomolecular Engineering* 1, pp. 101–122. DOI: [10.1146/annurev-chembioeng-073009-100903](https://doi.org/10.1146/annurev-chembioeng-073009-100903) (cit. on p. 25).
- Klamt, Andreas, Frank Eckert, Jens Reinisch, and Karin Wichmann (2016). "Prediction of cyclohexane-water distribution coefficients with COSMO-RS on the SAMPL5 data set." In: *Journal of Computer-Aided Molecular*

- Design* 30, pp. 959–967. DOI: [10.1007/s10822-016-9927-y](https://doi.org/10.1007/s10822-016-9927-y) (cit. on p. 60).
- Klamt, Andreas, Volker Jonas, Thorsten Bürger, and John C. W. Lohrenz (1998). “Refinement and Parametrization of COSMO-RS.” In: *The Journal of Physical Chemistry A* 102.26, pp. 5074–5085. DOI: [10.1021/jp980017s](https://doi.org/10.1021/jp980017s) (cit. on pp. 60, 69).
- Kleiner, David E. (2018). “Macswen’s Pathology of the Liver.” In: ed. by Alastair D. Burt, Linda D. Ferrell, and Stefan G. Hübscher. 7th Edition. Elsevier. Chap. 12 - Drugs and Toxins, pp. 673–779. DOI: [10.1016/B978-0-7020-6697-9.00012-1](https://doi.org/10.1016/B978-0-7020-6697-9.00012-1) (cit. on p. 3).
- Künnemann, Kai U., Jonas Bianga, Ricarda Scheel, Thomas Seidensticker, Jens M. Dreimann, and Dieter Vogt (2020). “Process Development for the Rhodium-Catalyzed Reductive Amination in a Thermomorphic Multiphase System.” In: *Organic Process Research & Development* 24.1, pp. 41–49. DOI: [10.1021/acs.oprd.9b00409](https://doi.org/10.1021/acs.oprd.9b00409) (cit. on p. 70).
- Kocis, L. and W. J. Whiten (1997). “Computational Investigations of Low-Discrepancy Sequences.” In: *ACM Transactions on Mathematical Software* 23.2, pp. 266–294. DOI: [10.1145/264029.264064](https://doi.org/10.1145/264029.264064) (cit. on p. 35).
- Kraume, Matthias (2013). “Integrierte chemische Prozesse in flüssigen Mehrphasensystemen.” In: *Chemie Ingenieur Technik* 85.10, pp. 1499–1511. DOI: [10.1002/cite.201300013](https://doi.org/10.1002/cite.201300013) (cit. on p. 2).
- Krige, D. G. (1951). “A statistical approach to some basic mine valuation problems on Witwatersrand.” In: *Journal of the Southern African Institute of Mining and Metallurgy* 52.6, pp. 119–139. DOI: [10.520/AJA0038223X_4792](https://doi.org/10.520/AJA0038223X_4792) (cit. on p. 14).
- Kunde, Christian, Tobias Keßler, Steffen Linke, Kevin McBride, Kai Sundmacher, and Achim Kienle (2019). “Surrogate modeling for liquid-liquid equilibria using a parametrization of the binodal curve.” In: *Processes* 7.10, p. 753. DOI: [10.3390/pr7100753](https://doi.org/10.3390/pr7100753) (cit. on pp. 54, 55, 69).
- Lampe, Matthias, Marina Stavrou, Johannes Schilling, Elmar Sauer, Joachim Gross, and André Bardow (2015). “Computer-aided molecular design in the continuous-molecular targeting framework using group-contribution PC-SAFT.” In: *Computers & Chemical Engineering* 81, pp. 278–

287. ISSN: 00981354. DOI: [10.1016/j.compchemeng.2015.04.008](https://doi.org/10.1016/j.compchemeng.2015.04.008) (cit. on p. 82).
- Liers, Frauke, Alexander Martin, Maximilian Merkert, Nick Mertens, and Dennis Michaels (2020). "Solving Mixed-Integer Nonlinear Optimization Problems using Simultaneous Convexification - a Case Study for Gas Networks." submitted to: *Journal of Global Optimization*. URL: <https://opus4.kobv.de/opus4-trr154/frontdoor/index/index/docId/303> (cit. on p. 13).
- Limleamthong, P., M. Gonzalez-Miquel, S. Papadokonstantakis, A. I. Papadopoulos, P. Seferlis, and G. Guillén-Gosálbez (2016). "Multi-criteria screening of chemicals considering thermodynamic and life cycle assessment metrics via data envelopment analysis: Application to CO₂ capture." In: *Green Chemistry* 18, pp. 6468–6481. DOI: [10.1039/C6GC01696K](https://doi.org/10.1039/C6GC01696K) (cit. on p. 82).
- Lin, Shiang-Tai and Stanley I. Sandler (2002). "A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model." In: *Industrial & Engineering Chemical Research* 41.5, pp. 899–913. DOI: [10.1021/ie001047w](https://doi.org/10.1021/ie001047w) (cit. on pp. 25, 27, 91).
- Linke, Steffen, Kevin McBride, and Kai Sundmacher (2020). "Systematic Green Solvent Selection for the Hydroformylation of Long-Chain Alkenes." In: *ACS Sustainable Chem. Eng.* 8.29, pp. 10795–10811. DOI: [10.1021/acssuschemeng.0c02611](https://doi.org/10.1021/acssuschemeng.0c02611) (cit. on pp. 5, 82, 83, 93, 94, 98, 100, 101).
- Linthorst, J. A. (2010). "An overview: origins and development of green chemistry." In: *Foundations of Chemistry* 12.1, pp. 55–68. DOI: [10.1007/s10698-009-9079-4](https://doi.org/10.1007/s10698-009-9079-4) (cit. on p. 1).
- Liu, Qilei, Lei Zhang, Linlin Liu, Jian Du, Qingwei Meng, and Rafiqul Gani (2019). "Computer-aided reaction solvent design based on transition state theory and COSMO-SAC." In: *Chemical Engineering Science* 202, pp. 300–317. DOI: [10.1016/j.ces.2019.03.023](https://doi.org/10.1016/j.ces.2019.03.023) (cit. on pp. 91, 97).
- Locatelli, Marco and Fabio Schoen (2013). *Global Optimization: Theory, Algorithms, and Applications*. SIAM. DOI: [10.1137/1.9781611972672](https://doi.org/10.1137/1.9781611972672) (cit. on p. 12).

- Lukowicz, Thomas, Adrien Benazzouz, Veronique Nardello-Rataj, and Jean-Marie Aubry (2015). "Rationalization and Prediction of the Equivalent Alkane Carbon Number (EACN) of Polar Hydrocarbon Oils with COSMO-RS σ -Moments." In: *Langmuir* 31.41. DOI: [10.1021/acs.langmuir.5b02545](https://doi.org/10.1021/acs.langmuir.5b02545) (cit. on p. 93).
- Marrero, Jorge and Rafiqul Gani (2001). "Group-contribution based estimation of pure component properties." In: *Fluid Phase Equilibria*, pp. 183–208. DOI: [10.1016/S0378-3812\(01\)00431-9](https://doi.org/10.1016/S0378-3812(01)00431-9) (cit. on pp. 22, 90).
- McBride, Kevin (2017). "Model-based process design and solvent selection for the efficient recovery of homogeneous catalyst in chemicals production." PhD thesis. Otto-von-Guericke University Magdeburg. DOI: [10.25673/5197](https://doi.org/10.25673/5197) (cit. on p. 54).
- McBride, Kevin, Tom Gaide, Andreas Vorholt, Arno Behr, and Kai Sundmacher (2016). "Thermomorphic solvent selection for homogeneous catalyst recovery based on COSMO-RS." In: *Chemical Engineering and Processing* 99, pp. 97–106. DOI: [10.1016/j.cep.2015.07.004](https://doi.org/10.1016/j.cep.2015.07.004) (cit. on pp. 3, 61, 63, 75).
- McBride, Kevin, Nicolas Maximilian Kaiser, and Kai Sundmacher (2017). "Integrated reaction-extraction process for the hydroformylation of long-chain alkenes with a homogeneous catalyst." In: *Computers & Chemical Engineering* 105, pp. 212–223. ISSN: 0098-1354. DOI: [10.1016/j.compchemeng.2016.11.019](https://doi.org/10.1016/j.compchemeng.2016.11.019) (cit. on pp. 45, 54, 67, 69).
- McBride, Kevin, Steffen Linke, Shuang Xu, and Kai Sundmacher (2018). "13th International Symposium on Process Systems Engineering (PSE 2018)." In: chap. Computer Aided Design of Green Thermomorphic Solvent Systems for Homogeneous Catalyst Recovery, pp. 1783–1788. ISBN: 9780444642417. DOI: [10.1016/B978-0-444-64241-7.50292-5](https://doi.org/10.1016/B978-0-444-64241-7.50292-5) (cit. on pp. 4, 59, 60, 77, 82, 100).
- McBride, Kevin and Kai Sundmacher (2015). "Data Driven Conceptual Process Design for the Hydroformylation of 1-Dodecene in a Thermomorphic Solvent System." In: *Industrial & Engineering Chemistry Research* 54.26, pp. 6761–6771. DOI: [10.1021/acs.iecr.5b00795](https://doi.org/10.1021/acs.iecr.5b00795) (cit. on pp. 2, 40, 44, 45, 73, 130).

- McBride, Kevin and Kai Sundmacher (2019). "Overview of Surrogate Modeling in Chemical Process Engineering." In: *Chemie Ingenieur Technik* 91.3, pp. 228–239. DOI: [10.1002/cite.201800091](https://doi.org/10.1002/cite.201800091) (cit. on p. 13).
- Mertens, Nick, Christian Kunde, Achim Kienle, and Dennis Michaels (2018). "Monotonic reformulation and bound tightening for global optimization of ideal multi-component distillation columns." In: *Optimization and Engineering*. ISSN: 1573-2924. DOI: [10.1007/s11081-018-9377-6](https://doi.org/10.1007/s11081-018-9377-6) (cit. on pp. 32, 42, 56, 131).
- Misener, Ruth and Christodoulos A. Floudas (2014). "ANTIGONE: Algorithms for coNTinuous / Integer Global Optimization of Nonlinear Equations." In: *Journal of Global Optimization* 59, pp. 503–526. DOI: [10.1007/s10898-014-0166-2](https://doi.org/10.1007/s10898-014-0166-2) (cit. on p. 12).
- Moity, Laurianne, Morgan Durand, Adrien Benazzouz, Christel Pierlot, Valérie Molinier, and Jean-Marie Aubry (2012). "Panorama of sustainable solvents using the COSMO-RS approach." In: *Green Chemistry* 14, pp. 1132–1145. DOI: [10.1039/C2GC16515E](https://doi.org/10.1039/C2GC16515E) (cit. on p. 100).
- Moller, Bruce, Jürgen Rarey, and Deresh Ramjugernath (2008). "Estimation of the vapour pressure of non-electrolyte organic compounds via group contributions and group interactions." In: *Journal of Molecular Liquids* 143.1, pp. 52–63. DOI: [10.1016/j.molliq.2008.04.020](https://doi.org/10.1016/j.molliq.2008.04.020) (cit. on p. 119).
- Morari, Manfred and Evangelhos Zafiriou (1989). *Robust process control*. Prentice Hall. ISBN: 0-13-782153-0 (cit. on p. 40).
- Müller, David, Markus Illner, Erik Esche, Tobias Pogrzeba, Marcel Schmidt, Reinhard Schomäcker, Lorenz T. Biegler, Günter Wozny, and Jens-Uwe Repke (2017). "Dynamic real-time optimization under uncertainty of a hydroformylation mini-plant." In: *Computers and Chemical Engineering* 106, pp. 836–848. DOI: [j.compchemeng.2017.01.041](https://doi.org/10.1016/j.compchemeng.2017.01.041) (cit. on p. 2).
- Mullins, Eric, Richard Oldland, Y. A. Liu, Shu Wang, Stanley I. Sandler, Chau-Chyun Chen, Michael Zwolak, and Kevin C. Seavey (2006). "Sigma-Profile Database for Using COSMO-Based Thermodynamic Methods." In: *Industrial & Engineering Chemical Research* 45.12, pp. 4389–4415. DOI: [10.1021/ie060370h](https://doi.org/10.1021/ie060370h) (cit. on p. 91).
- Nallasivam, Ulaganathan, Vishesh H. Shah, Anirudh A. Shenvi, Joshua Huff, M. Tawarmalani, and Rakesh Agrawal (2016). "Global optimiza-

- tion of multicomponent distillation configurations: 2. Enumeration based global minimization algorithm." In: *American Institute of Chemical Engineers Journal* 62.6, pp. 2071–2086. DOI: [10.1002/aic.15204](https://doi.org/10.1002/aic.15204) (cit. on pp. 13, 32).
- Nelder, John A. and R. Mead (1965). "A simplex method for function minimization." In: *Computer Journal* 7.4, pp. 308–313. DOI: [10.1093/comjnl/7.4.308](https://doi.org/10.1093/comjnl/7.4.308) (cit. on p. 11).
- Nentwich, C. and S. Engell (2016). "Application of surrogate models for the optimization and design of chemical processes." In: *Proceedings of the International Joint Conference on Neural Networks*. IEEE World Congress of Computational Intelligence, pp. 1291–1296. DOI: [110.1109/ijcnn.2016.7727346](https://doi.org/10.1109/ijcnn.2016.7727346) (cit. on pp. 2, 13, 115).
- Nentwich, Corina and Sebastian Engell (2019). "Surrogate modeling of phase equilibrium calculations using adaptive sampling." In: *Computers and Chemical Engineering* 126, pp. 204–217. DOI: [10.1016/j.compchemeng.2019.04.006](https://doi.org/10.1016/j.compchemeng.2019.04.006) (cit. on p. 13).
- Nocedal, Jorge and Stephen J. Wright (2006). *Numerical Optimization*. Springer. DOI: [10.1007/978-0-387-40065-5](https://doi.org/10.1007/978-0-387-40065-5) (cit. on p. 11).
- Odele, O. and Sandro Macchietto (1993). "Computer aided molecular design: a novel method for optimal solvent selection." In: *Fluid Phase Equilibria* 82, pp. 47–54. DOI: [10.1016/0378-3812\(93\)87127-M](https://doi.org/10.1016/0378-3812(93)87127-M) (cit. on p. 86).
- Papadopoulos, Athanasios I. and Patrick Linke (2005). "Multiobjective molecular design for integrated process-solvent systems synthesis." In: *AIChE Journal* 52, pp. 1057–1070. DOI: [10.1002/aic.10715](https://doi.org/10.1002/aic.10715) (cit. on p. 82).
- Papadopoulos, Athanasios I., Mirko Stijepovic, and Patrick Linke (2010). "On the systematic design and selection of optimal working fluids for Organic Rankine Cycles." In: *Applied Thermal Engineering* 30, pp. 760–769. DOI: [10.1016/j.applthermaleng.2009.12.006](https://doi.org/10.1016/j.applthermaleng.2009.12.006) (cit. on p. 82).
- Papadopoulos, Athanasios I., Ioannis Tsivintzelis, Patrick Linke, and Panos Seferlis (2018). "Reference Module in Chemistry, Molecular Sciences and Chemical Engineering." In: chap. *Computer-Aided Molecular Design: Fundamentals, Methods, and Applications*. DOI: <https://doi.org/10.1016/B978-0-12-409547-2.14342-2> (cit. on p. 82).

- Papalambros, P. Y. and D. J. Wilde (2000). *Principles of Optimal Design: Modeling and Computation*. 2nd. Cambridge: Cambridge University Press. ISBN: 978-0511626418. DOI: [10.1017/CB09780511626418](https://doi.org/10.1017/CB09780511626418) (cit. on p. 14).
- Poling, Bruce E., John M. Prausnitz, and John P. O'Connell (2000). *The Properties of Gases and Liquids*. 5th Revised. New York Citylo: Mcgraw-Hill Professional. ISBN: 978-0070116825 (cit. on p. 44).
- Puranik, Yash and Nikolaous V. Sahinidis (2017). "Domain reduction techniques for global NLP and MINLP optimization." In: *Constraints* 22, pp. 338–376. DOI: [10.1007/s10601-016-9267-5](https://doi.org/10.1007/s10601-016-9267-5) (cit. on p. 13).
- Quirante, Natalia, Juan Javaloyes, and J. A. Caballero (2015). "Rigorous Design of Distillation Columns Using Surrogate Models Based on Kriging Interpolation." In: *American Institute of Chemical Engineers Journal* 61.7, pp. 2169–2187. DOI: [10.1002/aic.14798](https://doi.org/10.1002/aic.14798) (cit. on pp. 17, 33).
- Rao, Rama Kanth and Lionel K. Arnold (1957). "Alcoholic extraction of vegetable oils. Part IV. Solubilities of vegetable oils in aqueous 2-propanol." In: *Journal of the American Oil Chemists' Society* 34, pp. 401–404. DOI: [10.1007/BF02637892](https://doi.org/10.1007/BF02637892) (cit. on p. 2).
- Rawlings, John O., Sastry G. Pantula, and David A. Dickey (1998). *Applied Regression Analysis*. Springer. Chap. Polynomial Regression, pp. 235–268. DOI: [10.1007/0-387-22753-9_8](https://doi.org/10.1007/0-387-22753-9_8) (cit. on p. 13).
- Razavi, Saman, Bryan A. Tolson, and Donald H. Burn (2012). "Review of surrogate modeling in water resources." In: *Water Resources Research* 48.7. DOI: [10.1029/2011WR011527](https://doi.org/10.1029/2011WR011527) (cit. on p. 13).
- Rheinboldt, Werner C. (1986). *Numerical Analysis of Parametrized Nonlinear Equations*. Wiley-Interscience (cit. on p. 47).
- Sahinidis, Nikolaos V., Mohit Tawarmalani, and Minrui Yu (2003). "Design of Alternative Refrigerants via Global Optimization." In: *American Institute of Chemical Engineering Journal* 49.7, pp. 1761–1775. DOI: [10.1002/aic.690490714](https://doi.org/10.1002/aic.690490714) (cit. on pp. 84–86).
- Sanderson, Katharine (2011). "Chemistry: It's not easy being green." In: *Nature* 469, pp. 18–20. DOI: [10.1038/469018a](https://doi.org/10.1038/469018a) (cit. on p. 78).
- Schäfer, Elisabeth, Yvonne Brunsch, Gabriele Sadowski, and Arno Behr (2012). "Hydroformylation of 1-Dodecene in the Thermomorphic Solvent System Dimethylformamide/Decane. Phase Behavior-Reaction

- Performance-Catalyst Recycling." In: *Industrial & Engineering Chemistry Research* 51, pp. 10296–10306. DOI: [10.1021/ie300484q](https://doi.org/10.1021/ie300484q) (cit. on p. 2).
- Scheffczyk, J., P. Schäfer, L. Fleitmann, J. Thien, C. Redepenning, K. Leonhard, W. Marquardt, and A. Bardow (2018). "COSMO-CAMPD: A framework for integrated design of molecules and processes based on COSMO-RS." In: *Molecular Systems Design & Engineering* 3.4, pp. 645–657. ISSN: 2058-9689. DOI: [10.1039/c7me00125h](https://doi.org/10.1039/c7me00125h) (cit. on p. 82).
- Scheffczyk, Jan, Christian Redepenning, Christian M. Jens, Benedikt Winter, Kai Leonhard, Wolfgang Marquardt, and André Bardow (2016). "Massive, automated solvent screening for minimum energy demand in hybrid extraction–distillation using COSMO-RS." In: *Chemical Engineering Research and Design* 115, pp. 433–442. ISSN: 02638762. DOI: [10.1016/j.cherd.2016.09.029](https://doi.org/10.1016/j.cherd.2016.09.029) (cit. on p. 82).
- Scheffczyk, Jan, Pascal Schäfer, Christian M. Jens, Kai Leonhard, and André Bardow (2017). "Integrated process and solvent design using COSMO-RS for the production of CO from CO₂ and H₂." In: *27th European Symposium on Computer Aided Process Engineering*. Vol. 40. Computer Aided Chemical Engineering. Elsevier, pp. 1765–1770. ISBN: 9780444639653. DOI: [10.1016/B978-0-444-63965-3.50296-8](https://doi.org/10.1016/B978-0-444-63965-3.50296-8) (cit. on p. 82).
- Scheuerer, M., R. Schaback, and M. Schlater (2013). "Interpolation of spatial data - A stochastic or a deterministic problem?" In: *European Journal of Applied Mathematics* 24.4, pp. 601–629. DOI: [10.1017/S0956792513000016](https://doi.org/10.1017/S0956792513000016) (cit. on p. 18).
- Seydel, Rüdiger (2009). *Practical Bifurcation and Stability Analysis*. Springer. DOI: [10.1007/978-1-4419-1740-9](https://doi.org/10.1007/978-1-4419-1740-9) (cit. on pp. 47–49, 51, 69).
- Silveira, Christian Luiz da and Nina Paula Gonçalves Salau (2019). "The UNIFAC-LLE and COSMO-SAC ternary aqueous LLE calculations." In: *Fluid Phase Equilibria* 501.112278. DOI: [10.1016/j.fluid.2019.112278](https://doi.org/10.1016/j.fluid.2019.112278) (cit. on p. 27).
- Smith, J. M., H. C. Van Ness, and M. M. Abbott (1996). *Introduction to Chemical Engineering Thermodynamics*. fifth. McGraw-Hill Chemical Engineering Series. McGraw-Hill. ISBN: 0-07-059239-X (cit. on p. 22).

- Stankiewicz, Andrzej I. and Jacob A. Mouljin (2000). "Process Intensification: Transforming Chemical Engineering." In: *Chemical Engineering Progress* 96.1, pp. 22–33. URL: https://www.aiche.org/sites/default/files/docs/news/010022_cep_stankiewicz.pdf (cit. on p. 1).
- Stavermann, A. J. (1950). "The entropy of high polymer solutions. Generalization of formulae." In: *Recueil des Travaux Chimiques des Pays-Bas* 69.2, pp. 163–174. DOI: [10.1002/recl.19500690203](https://doi.org/10.1002/recl.19500690203) (cit. on p. 26).
- Stavrou, Marina, Matthias Lampe, André Bardow, and Joachim Gross (2014). "Continuous Molecular Targeting–Computer-Aided Molecular Design (CoMT–CAMD) for Simultaneous Process and Solvent Design for CO₂ Capture." In: *Industrial & Engineering Chemistry Research* 53.46, pp. 18029–18041. ISSN: 0888-5885. DOI: [10.1021/ie502924h](https://doi.org/10.1021/ie502924h) (cit. on p. 82).
- Stefanis, Emmanuel and Costas Panayiotou (2008). "Prediction of Hansen Solubility Parameters with a New Group-Contribution Method." In: *International Journal of Thermophysics* 29.2, pp. 568–585. DOI: [10.1007/s10765-008-0415-z](https://doi.org/10.1007/s10765-008-0415-z) (cit. on p. 125).
- Steimel, Jochen, Marius Harrmann, Gerhard Schembecker, and Sebastian Engell (2013). "Model-based conceptual design and optimization tool support for the early stage development of chemical processes under uncertainty." In: *Computers & Chemical Engineering* 59, pp. 63–73. DOI: [10.1016/j.compchemeng.2013.06.017](https://doi.org/10.1016/j.compchemeng.2013.06.017) (cit. on p. 45).
- Stephenson, John, K. Gallagher, and C. C. Holmes (2004). "Beyond kriging: dealing with discontinuous spatial data fields using adaptive prior information and Bayesian partition modelling." In: *Special Publications* 239, pp. 195–209. DOI: <https://doi.org/10.1144/GSL.SP.2004.239.01.13> (cit. on p. 47).
- Struebing, Heiko (2011). "Identifying optimal solvents for reactions using quantum mechanics and computer-aided molecular design." PhD thesis. Imperial College London. URL: <http://hdl.handle.net/10044/1/9150> (cit. on p. 24).
- Ten, Joon Yoon, Mimi Haryani Hassim, Denny Kok Sum Ng, and Nishanth Gopalakrishnan Chemmangattuvalappil (2017). "A molecular design methodology by the simultaneous optimisation of performance, safety

- and health aspects." In: *Chemical Engineering Science* 159, pp. 140–153. DOI: [10.1016/j.ces.2016.03.026](https://doi.org/10.1016/j.ces.2016.03.026) (cit. on p. 82).
- Turbomole (1989-2007). "Program Package For Electronic Structure Calculations." In: URL: <https://www.turbomole.com> (cit. on p. 91).
- Vetter, Alexandra (Nov. 2020). URL: <https://www.inprompt.tu-berlin.de/> (cit. on p. 2).
- Vogelpohl, Christina, Christoph Brandenbusch, and Gabriele Sadowski (2014). "High-pressure gas solubility in multicomponent solvent systems for Hydroformylation. Part II: Syngas solubility." In: *The Journal of Supercritical Fluids* 88, pp. 74–84. DOI: [10.1016/j.supflu.2014.01.017](https://doi.org/10.1016/j.supflu.2014.01.017) (cit. on p. 115).
- Wang, Jiayuan and Richard Lakerveld (2018). "Integrated solvent and process design for continuous crystallization and solvent recycling using PC-SAFT." In: *AIChE Journal* 64.4, pp. 1205–1216. ISSN: 00011541. DOI: [10.1002/aic.15998](https://doi.org/10.1002/aic.15998) (cit. on p. 82).
- Wächter, Andreas (2002). "An interior point algorithm for large-scale nonlinear optimization with applications in process engineering." PhD thesis. Carnegie Mellon University. URL: <https://users.iems.northwestern.edu/~andreasw/pubs/thesis.pdf> (cit. on p. 11).
- Weidlich, Ulrich and Jürgen Gmehling (1987). "A modified UNIFAC model. 1. Prediction of VLE, h^E and γ^∞ ." In: *Industrial & Engineering Chemical Research* 26.7, pp. 1372–1381. DOI: [10.1021/ie00067a018](https://doi.org/10.1021/ie00067a018) (cit. on pp. 22, 69).
- Weinebeck, Alexander, Sebastian Kaminski, Hubertus Murrenhoff, and Kai Leonhard (2017). "A new QSPR-based prediction model for biofuel lubricity." In: *Tribology International* 115, pp. 274–284. DOI: [10.1016/j.triboint.2017.05.005](https://doi.org/10.1016/j.triboint.2017.05.005) (cit. on p. 93).
- Wiberg, Nils, A. F. Holleman, and Egon Wiberg (2001). *Inorganic Chemistry*. Academic Press. ISBN: 978-0123526519 (cit. on p. 1).
- Yaws, Carl (1998). *Chemical Properties Handbook: Physical, Thermodynamics, Environmental Transport, Safety & Health Related Properties for Organic & Inorganic Chemical*. McGraw-Hill Education. ISBN: 978-0070734012 (cit. on p. 119).

- Zhang, Yan and Nikolaos V. Sahinidis (2013). "Uncertainty Quantification in CO₂ Sequestration Using Surrogate Models from Polynomial Chaos Expansion." In: *Industrial & Engineering Chemistry Research* 52.9, pp. 3121–3132. DOI: [10.1021/IE300856P](https://doi.org/10.1021/IE300856P) (cit. on p. 13).
- Zhou, Teng, Zhen Song, Xiang Zhang, Rafiqul Gani, and Kai Sundmacher (2019). "Optimal Solvent Design for Extractive Distillation Processes: A Multiobjective Optimization-Based Hierarchical Framework." In: *Industrial & Engineering Chemical Research* 58.15, pp. 5777–5786. DOI: [10.1021/acs.iecr.8b04245](https://doi.org/10.1021/acs.iecr.8b04245) (cit. on pp. 82, 89).
- Zou, Jinming, Yi Han, and Sung-Sau So (2009). "Overview of Artificial Neural Networks." In: *Artificial Neural Networks: Methods and Applications*. Ed. by David J. Livingstone. Humana Press, pp. 14–22. DOI: [10.1007/978-1-60327-101-1_2](https://doi.org/10.1007/978-1-60327-101-1_2) (cit. on p. 20).

ABOUT THE AUTHOR

DECLARATION OF HONOR

I hereby declare that I produced this thesis without prohibited external assistance and that none other than the listed references and tools have been used. I did not make use of any commercial consultant concerning graduation. A third party did not receive any nonmonetary perquisites neither directly nor indirectly for activities which are connected with the contents of the presented thesis. All sources of information are clearly marked, including my own publications. In particular I have not consciously:

- Fabricated data or rejected undesired results
- Misused statistical methods with the aim of drawing other conclusions than those warranted by the available data
- Plagiarized data or publications
- Presented the results of other researchers in a distorted way

I do know that violations of copyright may lead to injunction and damage claims of the author and also to prosecution by the law enforcement authorities. I hereby agree that the thesis may need to be reviewed with an electronic data processing for plagiarism. This work has not yet been submitted as a doctoral thesis in the same or a similar form in Germany or in any other country. It has not yet been published as a whole.

Magdeburg, December 07, 2021

M.Sc. Tobias Keßler

SCIENTIFIC CONTRIBUTIONS

PUBLICATIONS IN PEER-REVIEWED JOURNALS

- Tobias Keßler, Filip Logist, and Michael Mangold. “Bi-objective optimization of dynamic systems by continuation methods.” In: *Computers & Chemical Engineering* 98 (2017), pp. 89–99.
doi: [10.1016/j.compchemeng.2016.11.025](https://doi.org/10.1016/j.compchemeng.2016.11.025).
- Tobias Keßler, Christian Kunde, Nick Mertens, Dennis Michaels, and Achim Kienle. “Global optimization of distillation columns using surrogate models.” In: *SN Applied Sciences* 1.1 (2018), p. 11.
doi: [10.1007/s42452-018-0008-9](https://doi.org/10.1007/s42452-018-0008-9).
- Tobias Keßler, Christian Kunde, Kevin McBride, Nick Mertens, Dennis Michaels, Kai Sundmacher, and Achim Kienle. “Global optimization of distillation columns using explicit and implicit surrogate models.” In: *Chemical Engineering Science* 197 (2019), pp. 235–245.
doi: [10.1016/j.ces.2018.12.002](https://doi.org/10.1016/j.ces.2018.12.002).
- Tobias Keßler, Christian Kunde, Steffen Linke, Kevin McBride, Kai Sundmacher, and Achim Kienle. “Systematic Selection of Green Solvents and Process Optimization for the Hydroformylation of Long-Chain Olefines.” In: *Processes* 7.12 (2019), p. 882.
doi: [10.3390/pr7120882](https://doi.org/10.3390/pr7120882).
- Christian Kunde, Tobias Keßler, Steffen Linke, Kevin McBride, Kai Sundmacher, and Achim Kienle. “Surrogate modeling for liquid-liquid equilibria using a parametrization of the binodal curve.” In: *Processes* 7.10 (2019), p. 753.
doi: [10.3390/pr7100753](https://doi.org/10.3390/pr7100753).
- Tobias Keßler, Christian Kunde, Steffen Linke, Kai Sundmacher, and Achim Kienle. “Integrated Computer-Aided Molecular and Process Design: Green Solvents for the Hydroformylation of Long-Chain Olefines.” In: *Chemical Engineering Science* 249 (2022), 117243.
doi: [10.1016/j.ces.2021.117243](https://doi.org/10.1016/j.ces.2021.117243).

PUBLICATIONS IN CONFERENCE PROCEEDINGS

- Tobias Keßler, Filip Logist, and Michael Mangold. "Use of predictor corrector methods for multi-objective optimization of dynamic systems." In: *Computer Aided Chemical Engineering* 38 (2016), pp. 313–318. doi: [10.1016/b978-0-444-63428-3.50057-6](https://doi.org/10.1016/b978-0-444-63428-3.50057-6).
- Tobias Keßler, Nick Mertens, Christian Kunde, Corina Nentwich, Dennis Michaels, Sebastian Engell, and Achim Kienle. "Efficient global optimization of a novel hydroformylation process." In: *Computer Aided Chemical Engineering* 40 (2017), pp. 2113–2118. doi: [10.1016/B978-0-444-63965-3.50354-8](https://doi.org/10.1016/B978-0-444-63965-3.50354-8).
- Tobias Keßler, Christian Kunde, Steffen Linke, Kevin McBride, Kai Sundmacher, and Achim Kienle. "Computer Aided Molecular Design of Green Solvents for the Hydroformylation of Long-Chain Olefines." In: *Computer Aided Chemical Engineering* 48 (2020), pp. 745–750. doi: [10.1016/B978-0-12-823377-1.50125-7](https://doi.org/10.1016/B978-0-12-823377-1.50125-7).

FURTHER CONFERENCE TALKS

- Tobias Keßler, Nick Mertens, Achim Kienle, Christian Kunde, and Dennis Michaels*. Optimization of non-ideal Multi-component Distillation Processes using Kriging Interpolation. At: International Conference on Operations Research (2017), Berlin (Germany).
- Tobias Keßler*, Christian Kunde, Nick Mertens, Dennis Michaels, and Achim Kienle. "Global optimization of distillation columns using surrogate models." At: *International Conference on Evolutionary and Deterministic Methods for Design, Optimization and Control with Applications to Industrial and Societal Problems; EUROGEN* (2017), Madrid (Spain).
- Christian Kunde*, Tobias Keßler, Steffen Linke, Kevin McBride, Kai Sundmacher, and Achim Kienle. Global optimization of catalyst

*presenter

recovery in a thermomorphic multiphase system using surrogate models. At: *12th European Congress of Chemical Engineering; ECCE* (2019), Florence (Italy).