Towards reliable characterization and model-based evaluation of organic solvent nanofiltration

Zur Erlangung des akademischen Grades eines
Dr.-Ing.
von der Fakultät Bio- und Chemieingenieurwesen
der Technischen Universität Dortmund
genehmigte Dissertation

vorgelegt von
M.Sc. Rebecca Goebel
(geb. van den Bongard)
aus
Lippstadt

Dortmund 2021

Danksagung

Ich möchte mich bei Prof. Dr.-Ing. Andrzej Górak und Prof. Dr.-Ing. Mirko Skiborowski für die Möglichkeit bedanken, diese Arbeit unter ihrer Aufsicht anzufertigen. Prof. Dr.-Ing. Andrzej Górak stand mir stets mit Rat und Tat zur Seite und setzte sich mit großem Engagement für seine Mitarbeiterinnen und Mitarbeiter ein. Unbestreitbar gilt mein größter Dank Herrn Prof. Dr.-Ing. Mirko Skiborowski für seine wertvollen Ideen, die fruchtbaren Diskussionen, die Unterstützung und sein Engagement zu jeder Zeit. Mein besonderer Dank gilt auch Herrn Prof. Dr. rer. nat. Dieter Vogt für die Übernahme der Funktion des Erstgutachters. Weiterhin danke ich PD Dr.-Ing. Christoph Held und Prof. Dr. David W. Agar für die Teilnahme an der Prüfungskommission und die bereichernde Diskussion.

Ein großer Dank geht an all meine Kolleginnen und Kollegen am Lehrstuhl für die schöne gemeinsame Zeit, die ständig offenen Türen und Ohren und die Entdeckung des ein oder anderen verborgenen Talentes. Besonders Danken möchte ich Thomas für die äußerst angenehme Atmosphäre im Büro, genauso wie Kai und Betti neben so vielen anderen Dingen auch für den stets freien Platz auf ihrer Rückbank. Ein großer Dank geht auch an die Techniker des Lehrstuhles, besonders an Kathrin Kissing und an die Projektpartner im Projekt ESIMEM, vor allem an Yvonne Thiermeyer für die regelmäßigen Telefontermine.

Diese Arbeit wäre außerdem nicht möglich gewesen ohne die Studenten, die mich durch Bachelor-, Masterarbeit oder HiWi-Tätigkeit unterstütz haben. Danke an Robin Wilmshöfer, Moritz Peterwitz, Laura Lingemann, Lukas Pietzka, Patrick Larysz, Moritz Imhausen, Mareike Schreiber, Ilka Niederkleine, Tobias Glaser, Se-

bastian Krick, Benedikt Wintz, Patrick Lotz, Sven Wibbeler und Oliver Großmann.

Danken möchte ich auch Laura für die offenen Ohren und die gegenseitige Motivation. Ein besonderer Dank geht an meine Eltern und meine Schwester für die zuverlässige Unterstützung zu jeder Zeit und natürlich an Fabian, der mir mit viel Geduld, Verständnis Unterstützung jederzeit bedingungslos zur Seite stand.

Abstract

The interest in organic solvent nanofiltration (OSN) increased substantially in both academia and industry during the last decades, since it provides a great potential for energy savings. However, despite the advantages, there are still limitations, that lead to the fact that OSN is rarely considered as a competitive separation operation in process design. For a reliable evaluation of process design, the uncertainties in lab-scale measurements and the quantification of model parameter precision are major factors and the prediction of flux and rejection is additionally essential in order to reduce experimental effort for feasibility studies during process development. These challenges are addressed in this thesis.

The evaluation of fluxes through multiple laboratory-scale membrane samples provides an accurate approximation of flux through an industrial- scale module. The results prove to be transferable to different membrane types. Furthermore, a collaborative study at different facilities demonstrates the comparability of experimental results obtained with a standardized procedure. Moreover, the consideration of experimental uncertainties in process design and membrane selection is proven to be as relevant as for the selection of an appropriate mass transfer model. In the second part of this work, a newly developed method for automatic development of predictive models for OSN shows promising results for prediction of solvent flux and solute rejection in pure and mixed solvents. The method derives the membrane-specific model structure and discriminates automatically between potential, easily retrievable descriptors based on available data. For the prediction of solvent flux, a comparison with existing phenomenological models from literature points out that the new models are superior and cover effects that are not included in the fixed model structure of phenomenological models. Models developed for the prediction of rejection are more complex compared to those for solvent flux but are comparable accurate.

Zusammenfassung

Durch das große Potential für Energieeinsparungen hat das Interesse an der organophilen Nanofiltration (ONF) in den letzten Jahrzehnten sowohl im akademischen Bereich als auch in der Industrie stark zugenommen. Trotz der Vorteile führen noch vorhandene Einschränkungen dazu, dass die ONF nur selten als konkurrenzfähige Trennoperation während des Prozessdesigns berücksichtigt wird. Für eine verlässliche Bewertung des Prozessdesigns sind die Unsicherheiten bei Messungen im Labormaßstab und die Quantifizierung der Präzision von Modellparametern wesentliche Faktoren. Die Vorhersage von Fluss und Rückhalt ist zusätzlich essenziell, um den experimentellen Aufwand für Machbarkeitsstudien während der Prozessentwicklung zu reduzieren. Diese Herausforderungen werden im Rahmen der vorliegenden Arbeit adressiert.

Lösungsmittelflüssen, die mit einer eine Vielzahl an Membranausschnitten im Labormaßstab ermittelt werden, liefert eine genaue Näherung des Flusses durch ein industrielles Modul. Die Ergebnisse erweisen sich als übertragbar auf verschiedene Membrantypen. Darüber hinaus zeigt eine kollaborativen Studie an verschiedenen Anlagen die Vergleichbarkeit von experimentellen, mit einem standardisierten Verfahren ermittelten Ergebnissen. Außerdem erweist sich die Berücksichtigung von experimentellen Unsicherheiten für das Prozessdesign und Membranauswahl als genauso relevant, wie für die Auswahl eines geeigneten Stofftransportmodells. Im zweiten Teil dieser Arbeit wurde eine Methode zur automatischen Entwicklung von prädiktiven Modellen für die ONF entwickelt, die vielversprechenden Ergebnisse für die Vorhersage von Lösungsmittelfluss und Rückhalt gelöster Komponenten in reinen und gemischten Lösungsmitteln zeigt. Die Methode generiert die membranspezifische Modellstruktur und diskriminiert automatisch zwischen potenziellen, leicht zugänglichen Deskriptoren auf Basis der verfügbaren Daten. Für die Vorhersage des Flusses zeigt ein Vergleich mit phänomenologischen Modellen aus der Literatur deutlich, dass die neuen Modelle überlegen sind und Effekte abdecken, die in der festen Modellstruktur der phänomenologischen Modelle nicht enthalten sind. Die für die Vorhersage des Rückhalts entwickelten Modelle sind komplexer als die für den Fluss, aber vergleichbar genau.

Contents

Nomenclature

Symbols

AIC	Akaike information criterion (-)
A_M	Membrane area (m^2)
a_0, b_0	Empirical fitting parameter in model of Darvishmanesh et al. [1] (l m^{-2})
b_1, b_2, b_3	Frictional coupling coefficient in MS model (kg m^{-2} s^{-1})
C_1, C_2, C_3	Empirical fitting parameters in model of Marchetti et al. [2]
$ConI$	Connectivity index (-)
CI	Confidence interval
c	Concentration (mol m^{-3})
$\bar{c}$	Mean concentration (mol m^{-3})
D	Diffusion coefficent (m^2 s^{-1})
DF	Driving force
d	Diameter (m)
d_{eq}	Equivalent molecular diameter (m)
$\vec{d}$	Dipole moment (debye)
f_1	Membrane parameter characteristic for NF layer [3] (m s^{-1})
f_2	Membrane parameter characteristic for UF layer [3] (m s^{-1})
f_c	Correction factor in model of Marchetti et al. [2] (-)
J	Flux (m^3 m^{-2} s^{-1})
$J_t ot$	Flux (kg h^{-1} m^{-2})
K	Sortion coefficient (-)
K_C	Steric hindrance factor for convection (-)
K_D	Steric hindrance factor for diffusion (-)
K_{HP}	Hagen-Poiseuille proportionality constant (m)
K_{MS}	Distribution coefficient (m^3 mol^{-1})
K_s	Equilibrium swelling of the membrane (-)
k	Empirical fitting parameter
k_{pol}	Polarizability (-)
L_j	Solute permeability coefficent (m^3 m^{-2} s^{-1})
L_{IM}	Mechanical permeability of the imperfections (m^3 m^{-2} s^{-1} Pa^{-1})
L_m	Mechanical permeability of the matrix (m^3 m^{-2} s^{-1} Pa^{-1})
L_p	Mechanical permeability coefficient (m^3 m^{-2} s^{-1} Pa^{-1})

M	Molar mass ($\mathrm{g\,mol^{-1}}$)
$\dot{m}$	mass flow ($\mathrm{kg\,h^{-1}}$)
N	Number (-)
N_A	Avogadro number ($\mathrm{mol^{-1}}$)
n	Empirical fitting parameter
$\dot{N}$	Molar flux ($\mathrm{mol\,m^{-2}\,s^{-1}}$)
P	Permeance, pressure based ($\mathrm{L\,m^{-2}\,h^{-1}\,Pa^{-1}}$)
P^{CSD}	Permeance in classical solution diffusion model ($\mathrm{m^3\,m^{-2}\,s^{-1}}$)
P^{SSD}	Permeance in simple solution diffusion model ($\mathrm{m^3\,m^{-2}\,s^{-1}}$)
Pe	Péclet number (-)
P_m	Diffusional permability of the matrix ($\mathrm{m^3\,m^{-2}\,s^{-1}}$)
Pol	Polarity (-)
p	Pressure (Pa)
p_m	Probability of mutation (-)
R	Rejection (-)
Ra	Radius of the interaction sphere ($\mathrm{(MPa)^{0.5}}$)
RE	Relative error (-)
R_g	Universal gas constant ($\mathrm{J\,mol^{-1}\,K^{-1}}$)
$resnorm$	Norm of residuals
r	Radius (m)
T	Temperature (K)
t	Value of Student's t-distribution (-)
V_m	Molar volume ($\mathrm{m^3\,mol^{-1}}$)
$\dot{V}$	Volumentric flow ($\mathrm{m^3\,s^{-1}}$)
W	Akaike weight (-)
W_s	Weight for parameter regression (-)
w	Mass fraction ($\mathrm{g\,g^{-1}}$)
$\bar{w}$	Mean mass fraction ($\mathrm{g\,g^{-1}}$)
X	Descriptor
x	Mole fraction ($\mathrm{mol\,mol^{-1}}$)
Y	Lumbed parameter in DSPF model (-)
y	Experimentally determined or calculated value
$\bar{y}$	Mean of experimentally determined or calculated measures
α	Polarity coefficient in model of Darvishmanesh et al. [1] (-)
β	Polarity coefficient in model of Darvishmanesh et al. [1] (-)

χ	Value of χ-distribution (-)
$\Delta\delta_M$	Difference of Hildebrand solubility parameters of solvent/solute and membrane(MPa$^{0.5}$)
δ_M	Membrane thickness (m)
δ_d	Hansen solubility parameter for dispersion (MPa$^{0.5}$)
δ_h	Hansen solubility parameter for hydrogen bonds (MPa$^{0.5}$)
δ_{HBP}	Hildebrand solubility parameter (MPa$^{0.5}$)
δ_p	Hansen solubility parameter for polarity (MPa$^{0.5}$)
ε	Surface porosity (-)
$\bar{\varepsilon}$	Mean percentage error
ε_{max}	Maximum percentage error
ϵ_f	Frictional coupling coefficient (-)
ϵ	Dielectric constant (-)
$\hat{\epsilon}$	Residuals of a model
φ	Sorption of the membrane (-)
ϕ	Factor characterizing solvent dependent pore size
γ	Surface tension (mN m^{-1})
γ_{LV}	Lliquid surface tension (mN m^{-1})
η	Viscosity (Pa s)
λ_{jp}	Ratio of solute j and pore radius (-)
λ	eigen value
$\hat{\mu}$	Estimated mean
ν	Molar volume (m^3 mol^{-1})
$\bar{v}_s$	Solvent velocity (m s^{-1})
π	Osmotic pressure (Pa)
θ	Model parameter
θ_s	Contact angle (°)
ρ	Density (kg m^{-2})
σ_{IT}	Reflection coefficient in irreversible thermodynamics models (-)
σ	Standard deviation
Θ	Steric partitioning coefficient (-)
τ	Tortuosity (-)

Superscripts

CSD	Parameter of classical solution diffusion model
eq	Equilibrium
F	Feed
$F(m)$	Feed side in membrane
P	Permeate
$P(m)$	Permeate side in membrane
SSD	Parameter of simple solution diffusion model
$*$	Scaled values

Subscripts

0	Initial value
Bh	Model of Bhanushali et al. [4]
$best$	best value
$calc$	Calculated
$comb$	Combinations
$discrete$	Discrete
$discretes$	Diskrets
exp	Experimental value
G	Model of Geens et al. [5]
i	Index
j	Solute j
k	Index
low	Lower limit
M	Membrane
mix	Mixture
$models$	Models
obs	Observations
$param$	Parameter
p	Pore
s	Solvent
$samples$	Samples
$stage$	Stage
$stages$	Stages
up	Upper limit

Abbreviations

ACN	Acetonitrile
AIC	Akaike information criterion
ANN	Artificial neural networks
ALAMO	Automated learning of algebraic models for optimization
CI	Confidence interval
CSD	Classical Solution-Diffusion
DM	DuraMem®
DSPF	Donnan Steric Pore-Flow
GAMS	General algebraic modeling system
GP	Genetic programming
HM	Hessian matrix
MEK	Methyl ethyl ketone
MF	Mikrofiltration
MIBK	Methyl isobutyl ketone
MLR	Multiple linear regression
MS	Maxwell-Stefan
MWCO	Molecular weight cutoff
NF	Nanofiltration
NL	Nonlinear
OLS	Orthogonal least squares
OSN	Organic solvent nanofiltration
PCA	Principal component analysis
PC-SAFT	Perturbed-Chain Statistical Associating Fluid Theory
PDMS	Polydimethylsiloxane
PF	Pore-Flow
PM	PuraMem®
QSAR	Quantitative structure-attribute relationships
QSPR	Quantitative structure-property relationships
RO	Reverse osmosis
RSM	Response surface models
SDI	Solution-Diffusion with imperfections
SK	Spiegler-Kedem
SSD	Simple Solution-Diffusion
UF	Ultrafiltration

1

Introduction

Global climate change is one of the greatest challenges of our time. Various studies have shown that further increases in global temperature will have significant impacts on all parts of daily life [6]. Therefore, a rethinking of current climate policy and energy use is inevitable. In the climate protection agreement of Paris, the binding target has been set to limit global warming to well below two degrees compared to pre-industrial times [7]. To achieve this goal, it is necessary to shift to environmentally friendly alternatives to energy production and, above all, to reduce energy consumption. To this end, an improvement in energy efficiency by $20\,\%$ [7] was set as a goal in 2012. According to the European Environment Agency, in 2018 the 33 member states of the agency used $24\,\%$ of energy in the industrial sector [8]. In the US, in the same year $35\,\%$ were assigned to the industrial sector [9]. Especially separation processes contribute considerably to the energy demand, but at the same time offer great potential for energy savings of up to $90\,\%$ by application of membrane processes [10].

Organic Solvent Nanofiltration (OSN) poses a promising pressure-driven membrane separation technology for organic mixtures which can be used either as a stand-alone process or in combination with conventional operations like distillation. Without the need for phase transitions or high operating temperatures, OSN offers significant energy saving potential compared to thermal separation processes. This results in possible applications in a wide variety of industrial sectors, such as the production of pharmaceutical products or in the petrochemical and food industries [11]. Developments of membrane materials and applications to new chemical systems are still mostly on laboratory-scale, but various membranes are commercially available on larger scale in specific modules and several industrial applications have been reported [12, 13].

Besides the advantages, there are also some challenges for the implementation of a separation process based on OSN. Due to complex interactions between the components of the system, including not only the membrane but also solvents and solutes [14], it is currently necessary to experimentally investigate a membrane in the specific chemical system as there are no suitable predictive models available yet. Since a larger number of different membranes is available, a feasibility study for a possible separation process is time-consuming and costly. For the subsequent modeling of membrane separation, a sufficiently large amount of experimental data must be available to identify suitable mass transport models and to adapt the model parameters. However, there are only a few experimental databases available for different membranes, which include solvents as well as solutes with different properties, which were measured in standardized and thus transferable experiments and can be used for a first estimation of the feasibility of a separation process. Therefore, for each mixture to be separated corresponding experiments with different membranes are currently performed, which do not follow a standardized method and thus often cannot be compared with other data. Simultaneously, fluctuations in the flux were observed which are oftentimes attributed to fluctuations in the membrane material. However, a systematic investigation of both the fluctuations and the separation characteristics in various solvents and with a large number of different solutes for OSN membranes is missing.

Furthermore, the uncertainty of model parameters resulting from the adaption to experimental data measured according to common, non-standardized methods was not investigated so far. Nevertheless, these can have a significant effect on the fluxes and rejections calculated by models and should consequently be taken into account when mass transport models are applied. In addition, the parameters of current mass transfer models are adapted to experimental data for the mixture to be separated and are thus substance-specific, which prevents the use of these models for predictions for other mixtures. The integration of additional models to calculate parameters, i.e. diffusion or sorption coefficients of a substance, is a way to avoid this problem, as it was done for example by Hesse et al. [15]. However, this approach requires additional experimental data to adjust parameters of these models. Models using membrane specific parameters adapted to experimental data for the specific membrane, but different solvents and solutes can be used for predictions for further substances. The development of such predictive models is a mandatory step towards enabling the consideration of OSN in early stages of process development and the

selection of suitable membranes for the specific separation task is facilitated. For the determination of such models, however, a database is necessary, which comprises experimental data of a larger number of solvents and dissolved components, which were determined under the same, standardized experimental conditions. To make OSN more competitive in comparison to conventional separation processes, it is inevitable to address these challenges.

This thesis addresses the listed challenges by first investigating the fluctuations of the flux through a membrane using a large number of samples from one membrane batch. The measured data are statistically analyzed and evaluated, and the influence of the fluctuations on the design of processes is studied. Furthermore, experimental data for different membranes according to a standardized procedure [16] are generated and thus the available database is extended. The gathered experimental results are compared to results generated at different facilities following the same experimental procedure in order to evaluate the comparability of the results and the value of the new standardized experimental procedure. A further step is the investigation of the discrimination of different existing mass transport models for the modeling of separation processes using OSN. The parameters of common mass transfer models are determined based on experimental data for different systems and the parameter precision is investigated. Subsequently, the uncertainties for the flux and rejection determined by the model are investigated. Taking into account the knowledge gained, in the second part of the thesis a method is presented which enables an automatic, data-driven development of predictive models for OSN and thus provides an important contribution to the integration of OSN into the standard toolbox of separation methods. The method is tested and evaluated using different case studies for solvents and solvent mixtures.

2

Theoretical Background

2.1 Fundamentals of Organic Solvent Nanofiltration

Nanofiltration (NF) in general is a pressure-driven membrane process operating with transmembrane pressures ranging from 10 to 50 bar and molecules with a molecular weight between 200 and 1000 g mol^{-1} [17]. According to the pressure differences and the molecular size of the rejected component NF can be classified between reverse osmosis (RO) and ultrafiltration (UF) (cf. Figure 2.1).

In contrast to aqueous nanofiltration, organic solvent nanofiltration (OSN) deals with the separation of organic solvent mixtures [11]. In comparison to thermal separation processes, OSN only needs the pressure difference as driving force and no phase transitions. As a consequence, OSN has the potential to be energy efficient. Another benefit regarding temperature-sensitive components is a low operating temperature [11, 17].

OSN membranes are produced based on different polymers or ceramics [14]. By varying the different polymer or ceramic layers and the manufacturing process, a large number of different membranes with different properties are created. This
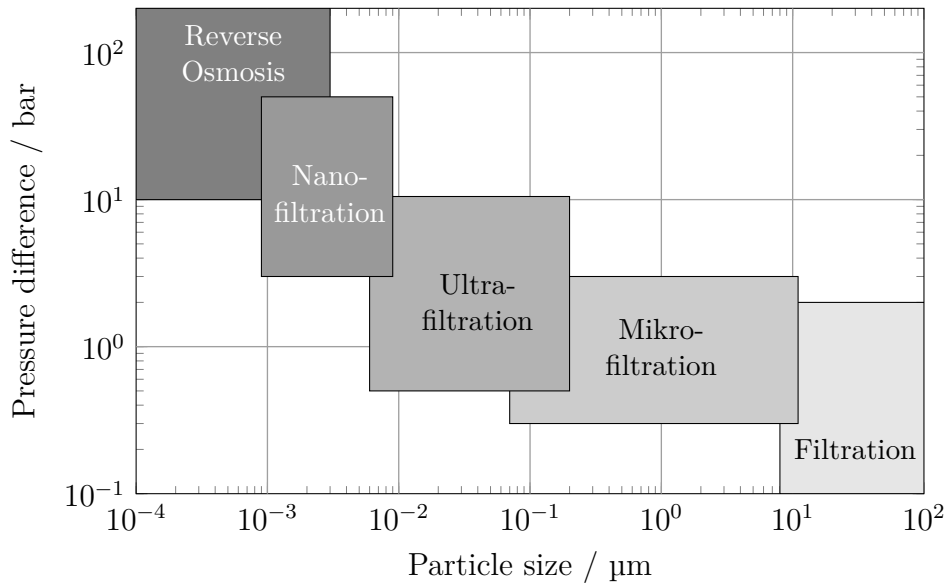
Figure 2.1: Classification of pressure-driven membrane operations (adapted from [17])

results in a variety of different application for OSN. While the number of commercially available polymeric membranes has increased significantly in recent years, the supply of ceramic membranes for OSN has not yet been as large, as the development of ceramic membranes with pore sizes below 1nm has only made significant progress in the current century. The advantages and disadvantages of the different materials and the commercially available membranes have been discussed in detail by Marchetti et al. [14]. The question of a suitable membrane material for a specific separation task must be answered individually.

2.1.1 Principle and performance factors

Figure 2.2 shows a schematic of a membrane process. The liquid feed is pressurized and fed to the module, in which it is contacted with a membrane. At least one component or a mixture of different components passes the membrane and is obtained as permeate on the low pressure side of the membrane. The rejected liquid mixture leaves the module as retentate [17].

In general, the performance of a membrane in a specific chemical system is evaluated based on different parameters. These parameters define the membrane permeate flux and the rejection of the components. The total volumetric flux J through the membrane material is defined as the ratio of the volumetric flow $\dot{V}$ and the active
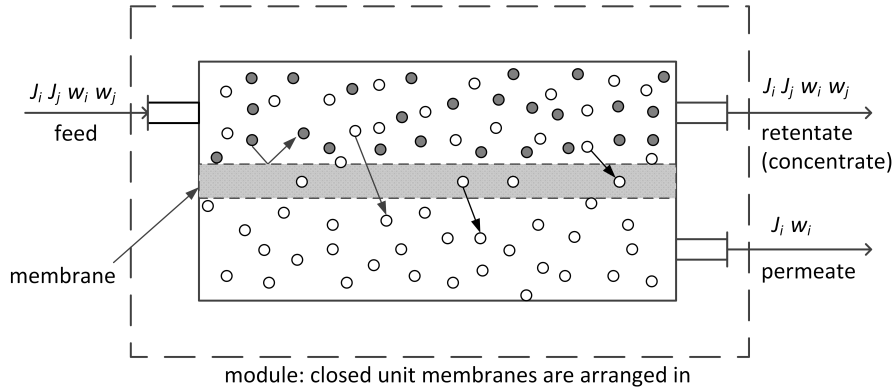
Figure 2.2: scheme of a membrane separation process (adapted from [17])

membrane area A_M:

$$J = \frac{\dot{V}}{A_M} \tag{2.1}$$

Analogously, the mass-related flux through the membrane can be calculated using mass flow and membrane area. Since the flux depends on the driving force of the process, normalizing the flux on the driving force is reasonable. However, based on the definition by Koros et al. [18] the thickness of the active membrane needs to be known, which is oftentimes not the case in OSN [11]. The permeance P_j of a component j calculated by normalizing the component specific flux J_j by the driving force DF_j (cf. Equation 2.2) is an alternative parameter describing the permeate flux [11].

$$P_j = \frac{J_j}{\Delta DF_j} \tag{2.2}$$

The rejection R_j is a measure for the separation regarding one solute j, since it is calculated based on the concentrations of one component in the permeate stream c_j^P and the feed stream c_j^F. The rejection is negative in case the component permeates preferably in comparison to the solvent and the concentration in the permeate becomes higher as in the feed mixture. On the contrary, positive values signify a decrease in the concentration from feed to permeate. The concentrations can also be replaced by molar fractions or weight fractions [17].

$$R_j = 1 - \frac{c_j^P}{c_j^F} \tag{2.3}$$

Further parameters like the permselectivity and the separation factor were suggested by Koros et al. [18] as general terminology for membrane processes in analogy to

other separation processes. Both account for the separation between two different components. Further, for the characterization of a membrane, the manufacturers usually specify the molecular weight cutoff (MWCO) of a membrane as defined by Koros et al. [18]. The MWCO corresponds to the molecular weight of a component, which is rejected by 90 %. Yet, differences in the process conditions like pressure, temperature and concentrations affect the MWCO and the separation characteristics of different membranes are hard to compare. Moreover, the properties of the solvents and solutes show a strong influence on the rejection as well as the flux in organic solvent nanofiltration. Hence, the MWCO measured in one specific chemical system cannot be transferred to other systems [14].

2.1.2 Application and Potentials

OSN is still a relatively new unit operation [12], while aqueous nanofiltration has been investigated and applied since 1980 [11]. A major challenge, which prevented the use of OSN for a significant period of time, was the lack of solvent-stable membranes. The first solvent-stable membrane was introduced by Koch Membrane Systems at the beginning of the 1990s [19]. Currently a variety of different polymeric and ceramic solvent-stable membranes are available and the number of studies focusing on non-aqueous nanofiltration increased tremendously in the past 20 years [14].

The first and largest industrial-scale application reported so far is the MAX-DEWAX^{TM} process installed by ExxonMobile, which recovers dewaxed solvents from lube oil. It was setup in 1998 and was considered superior to pervaporation during the process design due to the avoided phase transition [13]. Since the processes in refineries were intensified, dewaxing is no longer necessary and the MAX-DEWAX^{TM} process is not in operation anymore.

Recently, further application in several industries such as oleo- and petrochemicals, food and bio products, bulk chemistry and pharmaceutical industry were studied and several successful applications in industrial-scale have been reported [12]. The following exemplary studies demonstrate the promising application of OSN in various field of industry. Darvishmanesh et al. [20] developed a membrane based extraction method for edible oils using renewable solvents. They investigated four conventional polymeric membranes and demonstrated the possible replacement of non-renewable toxic solvents like n-hexane by applying OSN. Werth et al. [21] investigated the application of OSN for solvent recovery and deacidification of low-quality oils in

order to enable a new feedstock for the oleochemical industry. OSN was applied successfully for feed streams with fluctuation oil quality. Further, they found the corresponding process to reduce the energy consumption by about 70 % compared to the conventional process [22]. Hence, the application of OSN offers the possibility to use crude non-edible and waste oils as a feed stock in oleochemical industry and the competition with food industry is avoided.

Further, OSN was applied for the concentration of herbal extracts like presented by Peshev et al. [23]. They used OSN to concentrate rosemary acids and other antioxidants solved in ethanol, which were almost completely rejected by the membrane and no loss in antioxidant capacity was observed. Moreover, since the permeate contained almost no extracts, the solvent could be recycled directly to the extraction step. Another study was performed by Sereewatthanawut et al. [24] investigating the enrichment of nutritionals of rice bran by screening different solvent stable membranes. A considerable enrichment of the nutritionals was achieved in a two-stage membrane cascade.

The recovery of homogeneous catalysts is of main interest in bulk chemistry processes like hydroformylation. Dreimann et al. [25] investigated the rejection of a homogeneous transition metal by a conventional OSN membrane. The application of OSN enables an efficient separation of the catalyst from the product and energy efficient process, which can be transferred to further reactions. Peddie et al. [26] demonstrated the successful application of OSN for two homogeneous catalysts used in hydroformylations as well and found almost no catalyst in the permeate. Hence, the OSN-based process shows comparable results to the conventional distillation process, whereas the energy consumption is significantly lower. Another approach was investigated by Bertleff et al. [27] for the rejection of a catalyst in an aqueous mixture in presence of organic components.

For the production of pharmaceuticals, the purification and the non-thermal solvent recovery is an important step, which can be accomplished by OSN. The separation of genotoxic impurity during the production of APS was investigated by Székely et al. [28] by using two different commercially available membranes. They demonstrated the overall feasibility of OSN for this separation task. Further, Rundquist et al. [29] demonstrated the feasibility of OSN for the solvent recovery from a crystallization mother liquor, which can replace a distillation process. They achieve the necessary purity for the recycle of the solvent to the crystallization and calculated that OSN requires 25 % less energy than the distillation process.

The studies mentioned above have proven the successful applications of OSN in various industries. Moreover, the energy demand of industrial processes must be reduced in the upcoming years [30]. Conventional thermal separation processes like distillation or evaporation consume about 50 % of the energy needed for industrial processes [10]. The application of membrane processes can contribute significantly to the reduction of the energy consumption since membrane processes bare the potential to save up to 90 % of the energy of a distillation process [10]. Since OSN bares little potential for sharp separations as standalone process, hybrid applications are a reasonable alternative, presented by for example Micovic et al. [31], and combinations with different membrane processes [32]. However, currently intensive experimental investigations prior to the development of a new OSN process are unavoidable [33]. Thus, future studies need to focus on a systematic experimental characterization of membranes. Moreover, further methods to theoretically investigate the fundamental feasibility of an OSN process and predict the membrane performance prior to an experimental screening of membranes are required.

2.1.3 Common experimental procedure

In order to determine the feasibility of a membrane process for a specific separation task, commonly lab-scale experiments in the specific systems are performed. Hence, for the design of new processes an extensive experimental investigation of different membranes is unavoidable. A standardized experimental procedure would allow for the reconsideration of results of already performed experimental studies at different facilities for the identification of a suitable membrane for a new separation task, since the experiments are performed under the same conditions and the results are comparable. Further, the accuracy of the experimental results of the membrane performance should allow for a precisely determination of model parameters in order to verify the feasibility of the membrane performance at different process conditions. Experimental characterization of the membrane performance is commonly determined using molecules with various molecular weight, like linear and branched alkanes [34], dyes [35, 36] or polystyrene oligomers [37, 38]. However, the results of these experimental investigations are not comparable, since the molecular structure of the solutes differs as well as the underlying process conditions like pressure, temperature and concentration.

See Toh et al. [37] proposed a standard method for the characterization of OSN in order to determine reliable and comparable results for different membranes and test

different experimental set-ups. They used five different organic solvents including polar as well as non-polar solvents and styrene oligomer mixtures. Moreover, they defined specific experimental conditions for the experimental procedure in a lab-scale set-up and the analytical method. Using the oligomer styrenes See Toh et al. [37] obtained rejection curves for different membranes in the five solvents, which provides a comprehensive description of the membrane performance depending on the solvent properties. Furthermore, they demonstrated a possible application of the method to validate an experimental facility by comparison of the rejection curve to former experimental results within the same facility. Hence, leakages in the membrane cell can be determined easily. The method proposed by See Toh et al. [37] provides a first useful tool for membrane and facility comparison and systematic experimental errors can be determined. However, the method only includes styrene oligomers as solutes and thus, the influence of the solute properties on the membrane performance and the influence of different sizes of the membrane test cells were not investigated. Marchetti et al. [39] suggested to perform experiments for membrane characterization in at least one solvent per polarity class and two solvents in the group of aprotic polar solvents. Moreover, they suggested to extend the number of different solute groups in order to determine the performance of a membrane in different solvents and the influence of the solute properties. Based on the available data for different, systematically selected chemical systems, a first assessment of the separation performance of a membrane in a new, not yet investigated system is possible. If this initial assessment is conducted for different membranes, the selection of possible membrane candidates for the new separation task can be limited to the most promising ones and only these have to be examined in further experiments. Thus, although the number of necessary experiments in the first step of the general membrane characterization is larger in comparison to experiments for only one specific system, this approach leads in the following to a reduction of the number of necessary experiments for new potential application.

However, none of these experimental methods provide a complete standardized procedure, which includes a detailed description of all parameters for the experiment and hence, even when applying one of these methods, the operating conditions, concentrations or time slots for different tasks do differ. Moreover, the transferability of the determined experimental results to different facilities and the comparability to results determined by other persons or using different set-ups was not investigated so

far. Consequently, the available experimental data in literature might not be comparable and hence, not usable as one data set for tasks like model development.

2.2 Models for Organic Solvent Nanofiltration

In order to describe the transport of solvents and solutes through OSN membranes mainly semi-empirical models are applied. Two of the most common models are the *Solution-Diffusion Model* and the *Pore-Flow Model*. Both differ in the concepts of mass transport and are used as a basis for various refined models, which therefore correspond to one of these two model groups. A third group is based on irreversible thermodynamics assuming the membrane as a black box. The semi-empirical models are commonly fitted to experimental data of a specific solvent-solute-membrane system and are generally not able to describe the flux and rejection of a different system using the same parameter values. In contrast, a number of phenomenological models has been proposed, which correlate properties of solvent, solute and membrane specific parameters with the flux or the rejection of the components and hence, are applicable to different solvent-solute systems than the parameters were fitted to. The following sections present the different models and the underlying assumptions.

2.2.1 Pore-Flow Models

The group of Pore-Flow (PF) Models is derived from the mechanical process of filtration and considers the membrane to be porous. Pores are defined as free volume elements which do not change their shape and position in the membrane over time [40]. The concentration of the solvent as well as the solute in the membrane is assumed to be constant. Hence, the chemical potential gradient is solely generated by the pressure gradient Δp within the liquid in the pores [41] and the solvent flux can be described using the Hagen-Poiseuille equation:

$$J_s = \frac{\varepsilon r_p^2}{8\eta_s \delta_M \tau} \cdot \Delta p = \frac{K_{HP}}{\eta_s} \cdot \Delta p \qquad (2.4)$$

Since membrane specific properties like the tortuosity τ, the membrane thickness δ_M, the surface porosity ε and the pore diameter r_p are rarely known, these parameters are combined in a proportionality parameter K_{HP}. Hence, viscosity η_s is the only solvent property influencing solvent flux.

The *Donnan Steric Pore-Flow (DSPF) Model* proposed by Bowen et al. [42] is based on the combination of size exclusion and electrical effects in case of charged solutes. However, the following description is limited to uncharged solutes. For the solute transport, hindered convective and diffusive transport through the pores are considered, resulting in Equation 2.5 for the flux of an uncharged solute [43].

$$J_j = -D_{p,j} \cdot K_{D,j} \cdot \frac{dc_j}{dx} + K_{C,j} \cdot c_j \cdot \bar{v}_s \tag{2.5}$$

In this case the solvent velocity $\bar{v}_s$ is calculated based on the Hagen-Poiseuille relationships but also takes the osmotic pressure into account, which is an important factor in case of multivalent electrolytes and higher concentrations [44]. Hence, the solvent velocity is calculated as follows:

$$\bar{v}_s = \frac{\varepsilon r_p^2}{8\eta_s \delta_M \tau} \cdot (\Delta p - \Delta \pi) \tag{2.6}$$

Moreover, the steric hindrance factors for convection $K_{C,j}$(cf. Equation 2.7) and for diffusion $K_{D,j}$ (cf. Equation 2.8), which depend on the ratio $\lambda_{p,j}$ (cf. Equation 2.9) between solute radius r_j and pore radius r_p [44] are applied for calculating the solute flux.

$$K_{C,j} = (2 - \Theta_j) \cdot (1.0 + 0.054 \cdot \lambda_{p,j} - 0.988 \cdot \lambda_{p,j}^2 + 0.441 \cdot \lambda_{p,j}^3) \tag{2.7}$$

$$K_{D,j} = 1.0 - 2.30 \cdot \lambda_{p,j} + 1.154 \cdot \lambda_{p,j}^2 + 0.224 \cdot \lambda_{p,j}^3 \tag{2.8}$$

$$\lambda_{p,j} = \frac{r_j}{r_p} \tag{2.9}$$

The solute concentration at the pore inlet and outlet is not the same as in the bulk solution of the adjacent phases due to steric hindrance. This is described by a steric partitioning coefficient Θ_j, which is assumed to be identical at the feed and permeate side of the membrane (cf. Equation 2.10) [44].

$$\Theta_j = (1 - \lambda_{p,j})^2 \tag{2.10}$$

By applying the partitioning coefficient, the concentration at the pore inlet and outlet can be calculated by multiplying the coefficient with the respective bulk concentrations of the solute at each side of the membrane [44]. Since the solute flux

can also be calculated based on the solvent velocity and the concentration of the solute in the pores, a concentration gradient ca be determined based on Equation 2.5, which leads to the equation for solute rejection (cf. Equation 2.11) after integration across the membrane thickness and introducing two dimensionless parameters. The lumped parameter Y_j (cf. Equation 2.12), is assumed to be independent of solute concentration and the modified Péclet number Pe_j (cf. Equation 2.13) represents the ratio of convection and diffusion [44].

$$R_j = 1 - \frac{[(K_{C,j} - Y_j) \cdot \Theta_j]}{(1 - [1 - (K_{C,j} - Y_j) \cdot \Theta_j] \cdot exp(-Pe_j)}$$ (2.11)

$$Y_j = \frac{D_{p,j} V_{m,j} 8 \eta_p}{R_g T r_p^2}$$ (2.12)

$$Pe_j = \frac{(K_{C,j} - Y_j) \cdot r_p^2}{8 \eta_p D_{p,j}} \cdot (\Delta p - \Delta \pi)$$ (2.13)

The viscosity of the solution in the pores η_p differs from that of the bulk solution η. A mean pore viscosity is obtained by averaging across the pore radius r_p (cf. Equation 2.14) with the diameter of the solvent molecule d_s [44].

$$\frac{\eta_p}{\eta} = 1 + 18 \cdot \frac{d_s}{r_p} - 9 \cdot (\frac{d_s}{r_p})^2$$ (2.14)

Further, the diffusion coefficient in the pores $D_{p,j}$ is influenced by the changed viscosity as well [44]. It is calculated based on the bulk solution diffusion coefficient D_j and the hindrance factor for diffusion $K_{D,j}$.

$$D_{p,j} = K_{D,j} \cdot D_j$$ (2.15)

Based on the DSPF Model, solvent flux and rejection are calculated independently from each other using the Hagen-Poiseuille Equation 2.4 and Equation 2.11, respectively. The parameters obtained by regression with permeation data are the Hagen-Poiseuille constant K_{HP} and the pore radius r_p.

2.2.2 Solution-Diffusion Models

The *classical Solution-Diffusion (CSD) Model* was proposed by Lonsdale et al. [45] and reviewed by Wijmans and Baker [41]. They assumed a dense membrane and

pressure differences in the membrane material to be balanced in the same way as in liquids. Hence, the pressure is assumed constant in the membrane and equals the pressure at the feed side. On the permeate side, a sharp decrease of the pressure is assumed. Therefore, the chemical potential difference in solution-diffusion models is caused by a concentration gradient within the membrane. The sharp changes in pressure and solvent activity at the permeate side of the membrane are attributed to the equilibrium with the adjacent liquid phase. Hence, the flux of solute and solvent are calculated based on Equation 2.16.

$$J_i = \frac{D_i K_i}{\delta_M} \left[x_i^F - x_i^P \cdot exp\left(-V_{m,i}\frac{p^F - p^P}{R_g T}\right) \right] \tag{2.16}$$

The diffusion coefficient D_i and the the sorption coefficient K_i of the component i in the membrane and the membrane thickness δ either need to be known or they are handled as one parameter, namely the permeance $P_{i,CSD}$, which can be determined based on experimental data (cf. Equation 2.17).

$$J_i = P_i^{CSD} \left[x_i^F - x_i^P \cdot exp\left(-V_{m,i}\frac{\Delta p}{R_g T}\right) \right] \tag{2.17}$$

Based on the *classical Solution-Diffusion (CSD) Model* the *Simple Solution-Diffusion (SSD) Model* (cf. Equation 2.19) can be derived, which is a simplified version of the first classical one. According to Wijmans and Baker [41] in case of osmotic equilibrium, there is no solvent flux. Taking this into account, the concentration of the solvent in the permeate can be calculated according to Equation 2.18. Moreover, in case of OSN the simplification $1 - exp(x) \to x$ as $x \to 0$ is applicable. Combining this simplification with Equation 2.18 and 2.16, the simplified equation for the solvent flux (cf. Equation 2.19) is derived.

$$x_s^P = x_s^F \cdot exp\left(-V_{m,s}\frac{\Delta\pi}{R_g T}\right) \tag{2.18}$$

$$J_s = \frac{D_i K_i x_i^F V_{m,i}}{\delta_M R_g T}(\Delta p - \Delta\pi) = P_s^{SSD}(\Delta p - \Delta\pi) \tag{2.19}$$

Further, in diluted solutions the partial molar volume of the solute becomes small and the exponential term is close to 1 ($exp(x) \to 1$ as $x \to 0$), which leads to another

simplification for the solute flux equation (cf. Equation 2.20).

$$J_j = \frac{D_j K_j}{\delta_M}(x_j^F - x_j^P) = P_j^{SSD}(x_j^F - x_j^P) \tag{2.20}$$

In contrast to the CSD Model, the *Maxwell-Stefan (MS) Model* couples the fluxes of solvent and solute by friction and convection. Thus, it was considered more suitable than the CSD approach. The general Maxwell-Stefan equation was developed to describe diffusion in gases with more than two components and has been modified to be applicable to liquid and polymeric systems [46]. Three mass dependent diffusion coefficients D_{sj}, D_{sM} and D_{jM} account for the binary diffusion between solvent and solute, solvent and membrane as well as solute and membrane, respectively [46]. Further, equilibrium between the membrane surface and the adjacent phases is assumed. The frictional coupling effects are described by $\epsilon_{f,s}$ and $\epsilon_{f,j}$:

$$\epsilon_{f,s} = \frac{D_{sM}}{D_{sj}} \tag{2.21}$$

$$\epsilon_{f,j} = \frac{M_j}{M_s}\frac{D_{jM}}{D_{sj}} \tag{2.22}$$

The mass fractions of solvent and solute are calculated using the swelling of the membrane in equilibrium (cf. Equations 2.23 and 2.24) and the distribution coefficient of the solute between membrane and adjacent solution (cf. Equations 2.25 and 2.26), respectively [46]. For the mass fractions of the membrane w_M at the feed and permeate interfaces, the normalizing condition is used stating the sum of all mass fractions to unity.

$$w_s^{F(m)} = w_s^{F(m),eq} \cdot exp(-V_{m,s}\frac{\pi^F}{R_g T}) = K_s \tag{2.23}$$

$$w_s^{P(m)} = K_s \cdot exp(-V_{m,s}\frac{\Delta p}{R_g T}) \tag{2.24}$$

$$w_j^{F(m)} = K_{MS,j}c_j^{F(m)} \tag{2.25}$$

$$w_j^{P(m)} = K_{MS,j}c_j^P \cdot exp(-V_{m,j}\frac{\Delta p}{R_g T}) \tag{2.26}$$

Marchetti and Livingston [47] introduced new parameters b_1, b_2 and b_3 to reduce

the number of unknown coefficients, such as the membrane thickness δ_M:

$$b_1 = \frac{\rho_M \cdot D_{sM}}{\delta_M}; \; b_2 = \frac{\rho_M \cdot D_{jM}}{\delta_M}; \; b_3 = \frac{\rho_M \cdot D_{sj}}{\delta_M} \tag{2.27}$$

These parameters are used to replace the frictional coupling coefficients $\epsilon_{f,s}$ and $\epsilon_{f,j}$. Finally, the equations for flux of solvent J_s and solute J_j are derived:

$$J_s = \frac{1}{\rho_s} \frac{b_1(1 + \frac{b_1}{b_3}\frac{\bar{w}_j}{\bar{w}_M})(w_s^{F(m)} - w_s^{P(m)}) + b_2(\frac{b_1}{b_3}\frac{\bar{w}_s}{\bar{w}_M})(w_j^{F(m)} - w_j^{P(m)})}{\bar{w}_M(1 + \frac{b_1}{b_3}\frac{\bar{w}_j}{\bar{w}_M} + \frac{b_1}{b_3}\frac{\bar{w}_s}{\bar{w}_M})} \tag{2.28}$$

$$J_j = \frac{1}{\rho_j} \frac{b_2(1 + \frac{b_1}{b_3}\frac{\bar{w}_j}{\bar{w}_M})(w_j^{F(m)} - w_j^{P(m)}) + b_1(\frac{M_j}{M_s}\frac{b_2}{b_3}\frac{\bar{w}_s}{\bar{w}_M})(w_s^{F(m)} - w_s^{P(m)})}{\bar{w}_M(1 + \frac{b_1}{b_3}\frac{\bar{w}_j}{\bar{w}_M} + \frac{M_j}{M_s}\frac{b_2}{b_3}\frac{\bar{w}_s}{\bar{w}_M})} \tag{2.29}$$

The mean mass fractions $\bar{w}$ are determined assuming linear concentration profiles in the membrane. Thus, the arithmetical averages of the mass fractions at the feed and permeate interfaces of the membrane are calculated [46]. The mean mass fraction of the membrane $\bar{w}_M$ is obtained from the logarithmic difference of the mass fraction of the membrane at the feed and permeate interfaces [47]. In general, the *Maxwell-Stefan (MS) Model* incorporates a higher number of different sorption and diffusion parameters, which either need to be determined by regression to experimental data, in additional experiments or by using additional models. However, by introducing the parameters b_1, b_2 and b_3 (cf. Equation 2.27) as suggested by Marchetti and Livingston [47] the number of parameters, which need to be determined by regression to experimental flux data is reduced to five parameters. The three parameters b_1, b_2 and b_3 describe the coupling between solvent and solute flux by friction and convection. The equilibrium swelling of the membrane K_s and the distribution coefficient $K_{MS,j}$ can be used to determine the solvent and solute concentrations in the membrane at the feed and permeate interfaces [47].

A different approach was conducted by Hesse et al. [15] aiming to develop a predictive model based on the MS Model. The model parameters, such as the binary Maxwell-Stefan diffusion coefficients were obtained by fitting to independent solubility and sorption experiments with the respective solvent and membrane [48]. For prediction, the Perturbed-Chain Statistical Associating Fluid Theory (PC-SAFT) equation of state was used to calculate the concentration of the solvent in the membrane material [47].

Since Solution-Diffusion and Pore-Flow Models built on complementary ideal assumptions regarding the underlying transport mechanism and the structure of the membrane, these models may not be suitable to describe the mass transport through real membranes, which consist of both dense and pore-like areas. The *Solution-Diffusion with Imperfections (SDI) Model* proposed by Sherwood et al. [49] and revised by Yaroshchuk [50] combines both transport mechanisms. The structure of the membrane is considered heterogeneous and free volume elements are formed in a dense polymeric material due to swelling [50, 51], which are called imperfections. In contrast to pores, the imperfections are not fixed in place and are affected statistical fluctuations [47]. Due to the structure of the material, the transport is based on a combination of diffusion through the dense areas and convective flux through the imperfections and the following flux equations are derived [50]:

$$J_s = L_m \cdot (\Delta p - \Delta \pi) + L_{IM} \cdot \Delta p \tag{2.30}$$

$$J_j = P_m \cdot (x_j^F - x_j^P) + L_{IM} \cdot x_j^F \cdot \Delta p \tag{2.31}$$

In each of these two equations, the first term accounts for the mass transfer by diffusion through the polymeric matrix, according to the simple solution-diffusion model and the second term describes the convective transport through the free volume elements. The osmotic pressure difference is not considered since the imperfections are assumed as non-selective resulting in a constant concentration in the imperfections [47, 50].

2.2.3 Irreversible Thermodynamics Models

Kedem and Katchalsky [52] proposed a transport model based on irreversible thermodynamics. The relative movement of solvent and membrane as well as solute and membrane to each other causes a frictional force which can be quantified by a diffusion coefficient. Due to the structure of the membrane material one of the frictional forces is favored. Kedem and Katchalsky [52] concluded that solute-membrane-friction becomes more significant in dense membranes opposed to membranes with large pores. However, properties of the membrane are not represented in the irreversible thermodynamics model, it is rather seen as an irreversible process taking place in a black box.

According to Kedem and Katchalsky [52] the solvent and solute flux can be described by using three different phenomenological coefficients: the mechanical filtration coefficient L_p, the reflection coefficient σ and a coefficient L_j accounting for the permeability of the membrane regarding the solute. The solvent flux J_s and the molar solute flux $\dot{N}_j$ can be described using Equation 2.32 and 2.33, respectively. Δc_j denotes the concentration difference of the solute across the membrane and $\bar{c}_j$ the mean solute concentration, respectively.

$$J_s = L_p \cdot (\Delta p - \sigma_{IT} \cdot \Delta \pi) \tag{2.32}$$

$$\dot{N}_j = L_j \cdot \Delta c_j + (1 - \sigma_{IT}) \cdot J_s \cdot \bar{c}_j \tag{2.33}$$

Spiegler and Kedem [53] modified the model by introducing differential gradients of solute concentration as well as hydrostatic and osmotic pressure across the membrane and correspondingly differential transport coefficients introducing the *Spiegler-Kedem (SK) Model*. They found that the coefficients defined by Kedem and Katchalsky [52] depend on the solute concentration and thus are not constant over the considered pressure range. The new model for the solute rejection includes three parameters (L_p, σ and L_j), which are to be determined by regression to experimental data. The application of the equation for the solute rejection instead of an equation for solute flux might be beneficial in some cases, since the rejection is the commonly published measure. The solute rejection is determined based on Equation 2.34 and 2.35, whereas the solvent flux is calculated the same way as proposed by Kedem and Katchalsky [52] (cf. Equation 2.32).

$$R_j = \frac{(1 - F_j) \cdot \sigma_{IT}}{1 - \sigma_{IT} \cdot F_j} \tag{2.34}$$

$$F_j = exp\left(-J_s \cdot \frac{1 - \sigma_{IT}}{L_j}\right) \tag{2.35}$$

2.2.4 Phenomenological models for solvent flux

A number of phenomenological models for the prediction of pure solvent flux have been proposed, e.g. by Darvishmanesh et al. [1], Bhanushali et al. [4], Geens et al. [5], and Machado et al. [3]. These models correlate different types of descriptors, addressing the size of the permeating molecule, solvent diffusivity through the mem-

brane and the interaction between solvent and membrane. Although the same type of descriptors is considered for the individual models, the specific choice varies for each of them.

Machado et al. [3] proposed a resistance-in-series model with three resistances for an OSN membrane, located in a nanofiltration top layer, an intermediate ultrafiltration layer and a support layer. Two viscous resistances are correlated with the solvent viscosity η and size as well as the pore size of the different layers, while the surface resistance is influenced by the difference in polarities of membrane as well as solvent and is correlated to the surface tensions $\Delta\gamma$, resulting in the model for the component specific flux.

$$J = \frac{\Delta p}{\phi\left[\Delta\gamma + f_1\eta\right] + f_2\eta}.$$ (2.36)

Here, f_1 and f_2 present membrane-specific parameters that characterize the nanofiltration and ultrafiltration layer, while ϕ relates to the pore diameter of the nanofiltration layer and is considered solvent-dependent due to potential swelling of the membrane, limiting the applicability for solvents that induce different degrees of swelling.

Building on the simplified solution-diffusion model, Bhanushali et al. [4] proposed an alternative model, that assumes that the reciprocal of the viscosity η can be used as an easily measurable indicator for solvent diffusivity in the membrane matrix and that the combination with the molar volume V_m, as a size defining parameter, allows for a proportionality with the solvent permeance. In order to be applicable to hydrophobic and hydrophilic membranes, they further introduced the sorption value φ and the surface tension of the membrane γ_{sv}, to account for the interaction between membrane and solvent, resulting in the model for the component specific flux.

$$J = k\frac{V_m}{\eta\varphi^n\gamma_{sv}}\Delta p$$ (2.37)

Since the two parameters k and n are considered membrane-specific, the model is supposed to be applicable independent of the degree of swelling. However, the sorption value φ remains hard to quantify, since it depends on the combination of solvent and membrane material and therefore requires additional experimental quantification, unless it can be predicted as well.

While Geens et al. [5] recognized the model of Bhanushali et al. [4] as more appropriate for dense nanofiltration membranes, they challenged the assumptions of an inverse correlation with the sorption value and the dependence of the membrane

surface energy. As a result, they proposed a modified version of the Bhanushali et al. [4] model, which accounts for solvent-membrane interaction in a similar way as the model of Machado et al. [3], considering the difference in surface tension between the liquid solvent and the solid membrane surface $\Delta\gamma$, resulting in the following model for the local flux.

$$J = k\frac{V_m}{\eta\Delta\gamma}\Delta p. \tag{2.38}$$

Although, only a single membrane-specific parameter k needs to be regressed from experimental data, the statistical analysis of a larger data set indicated that the model is more suitable than its predecessors, particularly for hydrophobic membranes. Nevertheless, Geens et al. [5] gave the remark that for a general application, the model would need to be expanded with weighting factors to indicate the relative importance of the different contributions to solvent transport, resulting in additional fitting parameters.

Darvishmanesh et al. [1] further proposed another phenomenological model for the prediction of pure solvent flux for polymeric and inorganic membranes. Describing the transport through the membrane as a combination of diffusive and viscous flow, considering viscosity η and surface tension of the solvent γ_S as important factors for the latter. Recognizing a shortcoming of the preceding models, that similar surface tensions of membrane and solvent would result in an infinite flux, Darvishmanesh et al. [1] introduced an exponential function to compensate that. Including polarity as an important influencing factor for the diffusive flow the model was derived as:

$$J = \frac{a_0\alpha}{\eta\cdot exp(1-\beta)}(\Delta p - \Delta\pi) + \frac{b_0}{\eta\cdot exp(1-\beta)}\Delta p, \tag{2.39}$$

which for pure solvents, for which the osmotic pressure difference $\Delta\pi$ becomes zero, simplifies to

$$J = \frac{a_0\alpha + b_0}{\eta\cdot exp(1-\beta)}\Delta p. \tag{2.40}$$

In contrast to the preceding models, polar (hydrophilic) and non-polar (hydrophobic) membranes are treated differently, by a different evaluation of the non-dimensional coefficients.

$$\alpha_{polar} = \frac{\epsilon}{\epsilon_{water}} \qquad\qquad \beta_{polar} = \frac{\gamma}{\gamma_M}, \tag{2.41}$$

$$\alpha_{non-polar} = \frac{\epsilon_{n-hexane}}{\epsilon} \qquad\qquad \beta_{non-polar} = \frac{\gamma_M}{\gamma}. \tag{2.42}$$

The polarity coefficient α is determined as the ratio of dielectric constants of the solvent ϵ and either water or n-hexane as extremely polar and non-polar reference components. The ratio of the surface tension of the membrane γ_M and the solvent γ is further used for the coefficient β. The remaining two parameters a_0 and b_0 are considered as membrane-specific and need to be determined based on experimental data.

While the model of Darvishmanesh et al. [1] might, in principle, be applied to solvent mixtures, given the consideration of the osmotic pressure difference, no such application has been reported. Thus, only the phenomenological model proposed by Marchetti et al. [2], which has been developed based on permeation data of pure and mixed solvent flux for ceramic membranes, has been demonstrated for mixed solvent flux predictions. The model is based on the Hagen-Poiseuille equation, complemented by a correction factor f_c, which is supposed to lump all effects of the nanopores on the solvent flux.

$$J = \frac{K_{HP}}{\eta}\Delta p \cdot (1 + f_c),\tag{2.43}$$

$$f_c = C_1 \frac{2\gamma_{LV}cos\theta_s}{r_p} + C_2\left|\vec{d} - k_{pol}\right| + C_3\left(\frac{r_s}{r_p}\right)^2.\tag{2.44}$$

The calculation of f_c considers the effect of capillary forces, described by the liquid surface tension γ_{LV}, the contact angle θ_s and the pore radius of the membrane r_p, dipole interactions, described by the polarizability k_{pol} and the dipole moment $\vec{d}$ of the solvent, and steric forces, based on the molecular radius of the solvents r_s and the pore size of the membrane r_p. The contact angle θ_s and polarizability k_{pol} are calculated based on further equations depending on the surface tension of membrane and solvent, which have to be retrieved either from literature or dedicated experiments.

In addition to the various descriptors (properties), the comparable complex model includes four membrane-specific fitting parameters (K_{HP}, C_1, C_2, C_3), which have to be regressed from permeation data. For the prediction of mixed solvent flux a mixing rule with molar compositions x_s is applied to evaluate $f_{c,mix}$ as

$$f_{c,mix} = x_{s,1}f_{c,s,1} + (1 - x_{s,1})f_{c,s,2}.\tag{2.45}$$

2.2.5 Phenomenological models for solute rejection

In comparison to modeling solvent flux, only few phenomenological models have been proposed for solute rejection. Based on their elaborate study of solute transport in polymeric OSN membranes Geens et al. [54] analyzed the different solvent-solute-membrane interactions and evaluated the major affecting factors, providing an explanation of the rejection of six reference solutes in methanol and ethanol. Based on these findings, Geens et al. [55] further concluded that a generalized transport model needs to account for all these interactions, which are, however, difficult to measure and quantify. Concluding that solute transport in OSN membranes is mainly induced by convection they applied a simpler empirical approach, modifying different pore flow models for aqueous nanofiltration by introducing solvent depended effective pore and solute diameters, introducing an empirical correlation. Based on rejection data of different aromatic components solved in four different solvents parameter regression of the empirical correlation was performed, showing relatively good representation and prediction for the rejection of different solutes in the same solvents, with deviations to experimental data of less than 13 %. While these results are already considered satisfying, the adapted models require the evaluation of the membrane pore size for a specific solvent, as well as association parameters for the quantification of the effective solute radius, potentially limiting the prediction capability of the models. Since solute polarity was not taken into account, the models might lack accuracy in case of strong solute-membrane interactions.

In a similar way Blumenschein et al. [56] adapted the model for porous membranes developed by Bowen and Welfoot [44] to describe and predict the rejection of uncharged solutes by ceramic OSN membranes. The initial model is based on the Nernst-Planck equation and was investigated for aqueous nanofiltration by ceramic OSN membranes. This model was further adapted by Blumenschein et al. [56] based on several literature correlations for pore size dependent viscosity, pore size distribution, molar volume and solvent dependent radius of the solute. Thereby, no additional empirical parameters were introduced, while important membrane properties were determined experimentally in advance, such as the membrane pore size, determined by permporometry measurements. While the model showed excellent agreement with experimentally determined rejection curves for polystyrene standards as well as good agreement with other, more branched solutes, dissolved in tetrahydrofurane (THF), it was concluded that the model fails to describe the rejection in other solvents, such as n-heptane and ethanol. In this context, Blumenschein

et al. [56] highlighted that the experimental results did not reflect the expectations resulting from the calculated solvent-dependent solute size and that it has to be acknowledged that the rejection substantially depends on the specific solvent even for ceramic membranes.

Another model for solvent and solute transport in nanostructured membranes was recently presented Darvishmanesh and Van der Bruggen [57]. This new modeling approach is also based on the assumption of pore flow mechanism with the addition of a correction of the viscosity inside the pores as well as a correction of the size exclusion of a solute by the membrane based on the pore size and the size of the solute molecule. The latter accounts for the pore size distribution and requires the regression of one parameter based on experimental data for a specific membrane and range of solute sizes. The described model was able to predict the rejection of different dyes by ceramic membranes with maximum deviations of less than 25 %.

While the different studies showed large prospect for the prediction of solute rejection in OSN, especially for porous membranes, they also indicate potential limitations, especially in terms of generalization and applicability to different solvents. Overall, it can be summarized that accurate predictions were reported when at least some limitation to types of solvent, membrane and solute was made. While application of the developed models to polymeric membranes is possible by accounting for solvent-dependent pore size, as considered by Geens et al. [55], application of the models of Blumenschein et al. [56] and Darvishmanesh and Van der Bruggen [57] was so far only demonstrated for ceramic membranes. However, the strong solvent-solute-membrane interactions described and analyzed by Geens et al. [54] and specifically the possibility that solute transport might be affected to a considerable margin by solution and diffusion in dense membranes, raises at least concerns about the potential generalization and transfer to polymeric membranes.

3

Objectives and Approach

3.1 Gap Analysis

The previous chapter provides an overview of the current state of the art for OSN applications and available models. Although, OSN is considered frequently in academia and offers a great potential for energy savings and applications in a variety of industrial fields, industrial-scale applications are rather limited. In order to further facilitate the application of OSN in industry and enable the consideration of OSN during conceptual design of new processes, certain challenges need to be overcome. The aim of this thesis is to address some of these challenges like the reliability of lab-scale measurements and quantification of model parameter precision, which both are required for a reliable evaluation of process design studies as well as the potential prediction of flux and rejection in order to reduce experimental effort contributing to the competitiveness of OSN in early stages of process development.

Though, a large amount of experimental data for OSN membranes is already available in the literature, yet these data are usually measured in lab-scale set-ups using only a few different membrane samples and following the common experimental procedure described in Section 2.1.3. In addition, information on the reliability of the measured data or measurement errors is rarely provided, which is however inevitable considering the complex membrane structure and the possibility of fluctuations in the membrane material, as described e.g. by Hussain et al. [58], Vandezande et al. [36], and Tsibranska and Tylkowski [59]. Furthermore, different studies for both nanofiltration membranes and other membrane processes have shown significant deviations between lab-scale and module-scale measurements [60, 61]. Variations in flux are often attributed to leakage in the set-up or defects in the membrane material [37], which clearly indicates evidence for the benefit of using multiple samples

in repeated experiments and reporting measurement errors to distinguish between those measurement errors and variations in the membrane material. Moreover, the reliability of the measurements should be taken into account when comparing different studies, especially evaluating different membrane materials, adjusting model parameters and transferring the results to industrial-scale. However, so far no systematic investigation has been carried out in order to determine the variations in flux of solvents through organophilic nanofiltration membranes resulting from differences in membrane material. Furthermore, the minimum sample size required to determine an accurate estimate of the mean flux for a membrane that can be reliably used for industrial-scale transfer has not been specified yet.

In addition to reliable experimental data, reliable models are needed to develop membrane processes on an industrial-scale. The number of different models (Section 2.2) available cannot be limited at this point due to the fact that the transport of substances through an OSN membrane is influenced by different interactions and the transport mechanism has not yet been fully identified [11, 17]. In order to identify the most suitable model for a given system, the different existing models must be adapted and compared for the available experimental data, as was e.g. investigated by Marchetti and Livingston [47] for different membranes and mixtures of solute and solvent. An investigation of the physical correctness of the models was not aimed at, but the suitability of the different models to represent the experimental data was investigated. Based on the results obtained, the model complexity required to describe the OSN was determined. However, Marchetti and Livingston [47] did not examine the accuracy of the model parameters and the resulting uncertainties in the models.

One of the biggest challenges to implement OSN as a unit operation for process development is to enable the prediction of flux and rejection, which is complicated by the different underlying transport phenomena [14] and potential swelling [62]. The existing semi-empirical models are not applicable for these predictions since the model parameters are component specific and need to be adapted for the mixture to be separated. As described in Section 2.2.4 and 2.2.5 some phenomenological models have been developed so far, allowing for flux and rejection predictions, based on membrane specific parameters. These models differ both in the properties of the substances used and in their structure. Since different models are suitable for different membranes and materials [14], it is unlikely that one model is applicable for all types of OSN membranes. Therefore, the development of membrane-specific

models with an optimized model structure seems to be a promising possibility to develop predictive models. However, since manual development is complex and a comprehensive knowledge of the existing phenomena is necessary, automatic data-driven development of such models is a reasonable and promising approach, which has not yet been pursued.

3.2 Structure of the thesis

In order to overcome some of the challenges addressed in Section 3.1 investigations and developments in multiple directions have been performed. This work is summarized within this thesis in two parts. The first part focuses on the uncertainties introduced by the experimental characterization of OSN membranes. Chapter 4 presents an extensive experimental investigation on the reliability of lab-scale experiments, in order to determine the variations in flux that can be expected by following the common procedure of experimental investigations as described in Section 2.1.3. Besides the expected variations in flux, determined for a single batch of commercially available membranes, caused by fluctuations in the membrane material, the necessary sample size needed for accurate determination of the flux through a membrane is determined based on a statistical evaluation. For this, flux data measured at two different facilities following the same experimental procedure are evaluated and supplemented by experiments in industrial-scale. Additionally, the transferability to different membranes and solutes is investigated and the effect of the variations in flux on the process design and costs are studied.

In order to provide reliable experimental data for different commercially available OSN membranes, an extensive experimental study is performed following a newly developed standardized experimental procedure [16] described in Chapter 5. The gathered experimental results are examined regarding the comparability with results obtained in other facilities.

Furthermore, the study on different models for OSN by Marchetti and Livingston [47] is extended by an identifiability analysis of the model parameters, presented in Chapter 6. Following the approach of Marchetti and Livingston [47], the parameter regression is performed and the accuracy of each model is evaluated. Subsequently, the parameter precision is evaluated based on confidence intervals enabling an identifiability analysis, which is further investigated in a Monte-Carlo simulation in order to examine the distinguishability of the models.

In the second part of the thesis, a new data-driven approach for the development of predictive models for solvent flux and rejection is presented in Chapter 7, following the introduction of various data-driven methods. The novel approach is applied to different case studies including both pure and mixed solvent flux providing a comparative evaluation with respect to existing phenomenological models as presented in Section 2.2.4 and 2.2.5. Finally, the method is extended to the development of models for the rejection of solutes in pure solvents and solvent mixtures (Chapter 8) and applied to different case studies in this context as well. The thesis concludes with a final chapter that provides an outlook on potential future work (Chapter 9).

4

Reliability of lab-scale experiments for membrane characterization

While the scale-up from flat-sheet lab-scale experiments to larger modules for an industrial application requires consideration of the module geometry and hydrodynamics, it is of high importance that the lab-scale experiments provide an accurate estimate of the membrane performance at ideal hydrodynamic conditions to make a proper selection and potentially allow for a model-based evaluation of the process performance. Due to inhomogeneity of the membrane material, it is however well known that the membrane flux determined by small samples in lab-scale experiments may vary in a considerable margin. These variations are investigated with an extensive experimental screening of lab-scale samples as well as a spiral-wound module for a commercially available polymeric membrane and pure solvent flux. The experimental results confirm considerable variations for single lab-scale samples, but also indicate good comparability on the basis of a statistical evaluation of the larger set of lab-scale samples. Based on this evaluation a minimum number of samples for an accurate estimate is determined and validated for additional membrane-solvent combinations.

4.1 Introduction

While the design of OSN-based separation processes has been deemed reliable, building on multi-scale modeling and experimental data retrieved from spiral-wound modules [63], initial data on flux and separation performance for a specific chemical system and membrane is usually retrieved from lab-scale experiments with small-scale samples. The tested sample size is usually in a range of 10 to 100 cm^2, depending on the used equipment. While this data can in principle be used for parameter regression and subsequent model-based conceptual design, the reliability and the transferability of the results to industrial-size modules is a considerable concern. While a tremendous amount of such lab-scale data is reported in literature only little information on the reliability is available, especially considering the complex nature of the membrane structure and potential inhomogeneity in the membrane material. Most experimental studies report on the results of single or repeated measurements for one or in the best case a few membrane samples (two-three), as e.g. in the study of Marchetti and Livingston [47], providing either no information on uncertainty or estimates of experimental errors based on Gaussian error propagation or repeated measurements for the same membrane. E.g. See Toh et al. [37] and Sereewatthanawut et al. [64] evaluated only single samples and reported considerable differences between lab-scale and module-scale. Recent studies by Schmidt et al. [65] and Postel et al. [66] applied for instance larger sample sizes of 80 and 100 cm^2, but tested only two consecutive samples for the same conditions.

Only few studies evaluated and reported on the deviations between different nanofiltration membrane samples and between lab-scale and module-scale experiments. However, as reported by Hussain et al. [58], who investigated aqueous nanofiltration membranes, the variations in the active layer can result in considerable variations in flux for different membrane samples. The studies of Vandezande et al. [36] on a composite membrane and Tsibranska and Tylkowski [59] for the commercial DuraMem® 300 evaluated a series of test cells and reported on variations in solvent flux for the lab-scale test cells of up to 16 %, respective 20 %. Comparing lab-scale and module-scale membranes, Wang et al. [60] reported deviations in flux for ultrafiltration membrane samples of 15 %, while Schipolowski et al. [61] reported deviations for different nanofiltration and a reverse osmosis membrane of up to 23 %. While differences in flux between lab-scale samples and industrial modules are oftentimes attributed to possible leakages and defects in the membrane material [37],

the previous studies indicate the importance of multiple experiments with different samples of the same membrane, in order to achieve a higher reliability.

This reliability is not only required in order to evaluate different membranes in a screening process and get an accurate estimate of the membrane performance at larger scale, but also to compare the results of different studies. While the analysis of variance and information on confidence intervals is still considered rather as an option in most chemical engineering publications, other disciplines, like pharmacology, epidemiology, human and social sciences or geology require statistical analysis of reported data, in order to determine whether it is sufficient for the derivation of significant conclusions [67, 68]. The knowledge about the quality of published data allows the quantification of data uncertainty and hence, the quality of conclusions and models based on these data [69]. Therefore, similar requirements would improve the quality of the reported data on membrane performance as well.

In this chapter the variance of membrane flux measurements determined with small lab-scale samples is evaluated and a meaningful indication on the necessary sample size that allows for an accurate quantification of the mean membrane flux is provided. Therefore, an experimental screening of a large number of lab-scale samples of a commercially available membrane was performed in two different facilities and complemented with measurements for an industrial membrane module (Section 4.2). Based on a detailed statistical analysis the necessary sample size for accurate estimates of the mean solvent flux is evaluated (Section 4.3). Finally, Section 4.4 provides the conclusion.

4.2 Materials and Methods

In order to evaluate the necessary sample size, an experimental investigation of a large number of different lab-scale samples was performed at two different sites, namely Evonik Resource Efficiency GmbH in Marl (Evonik) and the Laboratory of Fluid Separations at TU Dortmund (FVT TUDo) and complemented by flux measurements for an industrial scale module. The experimental methods and materials are described in the following section before the statistical methods used for data analysis are presented. While further factors like spacer or potting choice for the module or solvent treatment and exchange can affect the process performance [70], the focus of this study is the quantification of variations caused by fluctuation in the membrane material itself, as indicated by Hussain et al. [58]. Therefore, only pure

solvent flux is investigated, in order to avoid any effect of hydrodynamics, especially concentration polarization.

4.2.1 Materials

Lab-scale experiments were conducted in cross-flow set-ups manufactured by Evonik MET with 2.5″ filtration cells that offer an active membrane area of 14 cm^2. In order to exclude hydrodynamic effects from the comparison between lab-scale and industrial module, only pure solvent flux was investigated. The commercially available asymmetric polyimide-based membrane DuraMem® 200 T1 (Evonik Resource Efficiency GmbH) and the polar solvent ethanol were selected as meaningful combination. Ethanol used at Evonik had a purity of 99 %, while ethanol with a purity of 100 % was used at FVT TUDo. For comparison with an industrial membrane module the ethanol flux was determined for a spiral wound 1812 module with an active membrane area 1100 cm^2 [71]. Thus, the industrial module offers about 80 times the area of the lab-scale samples.

4.2.2 Experimental procedure

For an additional analysis of comparability lab-scale experiments were performed at two different sites with slightly different experimental procedures. Thereby, it was possible to evaluate the influence of these differences that are representative for studies conducted by different research groups reported in literature. At both sites the same kind of cross-flow set-up was used and fresh membrane samples were used for each experimental run. The samples were taken from different membrane flat sheets (Din A4) of the same production batch. For all conducted experiments, the filtration cells were first equipped with new membrane samples and rinsed with ethanol at 30 bar and ambient temperature until 50 l m^{-2} permeate were collected for each cell.

At FVT TUDo a series of four filtration cells was evaluated simultaneously as shown in Figure 4.1 (right), whereby the retentate was recycled to the feed tank and the permeate was constantly withdrawn. Pressure was applied by nitrogen and the pump was only used for recirculation of feed and retentate. On the permeate side there was no valve and the permeate was collected in sample tubes but not recirculated. The flux was measured gravimetrically after a period of 4 h to ensure steady state conditions, which was validated via comparison to samples taken in previous time

intervals. The experiment was repeated seven times with four membrane cells each using fresh membrane samples. Hence, 28 samples were evaluated in total at FVT TUDo. The feed tank was refilled to maximum holdup in between the experiments. At Evonik, a series of eight filtration cells was evaluated simultaneously as shown in Figure 4.1 (left), whereby the retentate and permeate were recycled for up to 4 h, after which permeate samples were taken and the ethanol flux was determined gravimetrically as well. The experiment was repeated three times using eight membrane cells and fresh membrane samples in each experiment. Thus, the flux through 24 samples was measured at Evonik. As in the FVT TUDo set-up, the system was pressurized with nitrogen and a pump was used for recirculation of retentate. A second pump was installed in an additional permeate recycle to the feed tank. While this allows for a constant liquid level in the feed tank, the hold-up in the feed tank has no effect on the conducted experiments, as only pure solvent flux was investigated and the feed tank always had a sufficiently large holdup in the FVT TuDo set-up. Since both systems applied nitrogen for pressurization of the whole set-up, while the recirculation pump only ensured a constant feed flow rate and velocity, the pressure drop between the cells is considered negligible. This was confirmed by a trial measurement for the FVT TUDo set-up. As the temperature of the feed tank was monitored as constant and due to a high recirculation rate of more than 99 %, also a constant temperature in the system was assumed.
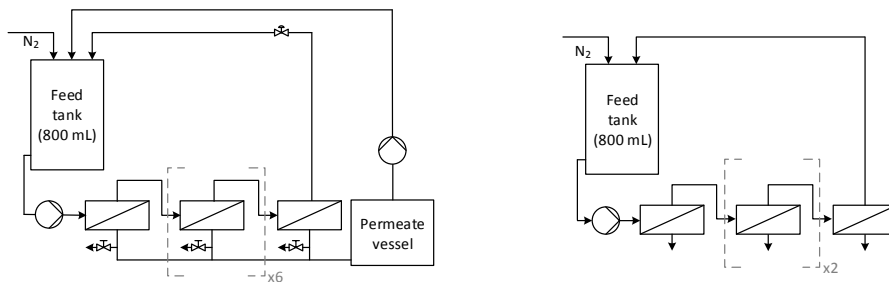
Figure 4.1: Illustration of lab-scale filtration set-ups at Evonik (left) and FVT TUDo (right).

Moreover, the large-scale experiment was not repeated since the membrane area of the model is already 80 times the area of the lab-scale samples and variations in the membrane material will balance each other. The experimental procedure, pressure and temperature were the same as for the lab-scale experiments. However, the set-up

for the experiments in large-scale differed from the lab-scale set-ups (cf. Figure 4.2). The feed tank of the plant offered a higher liquid holdup and the plant was not pressurized with nitrogen but by using a high-pressure pump. The second pump is used for recirculation of the retentate. Due to low nitrogen solubility in ethanol at high pressure (1.225 mol % at 35.671 bar [72]), the difference in the pressurization of the lab-scale and the large-scale set-ups did not influence the measurements. Moreover, the large-scale experiment was not repeated since the membrane area of the model is already 80 times the area of the lab-scale samples and variations in the membrane material will balance each other.
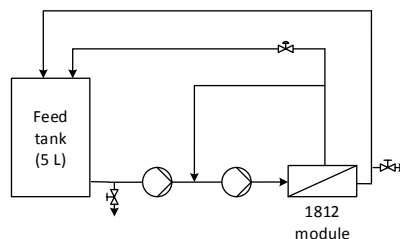
Figure 4.2: Illustration of large-scale filtration set-up for the test of the spiral wound 1812 module.

4.2.3 Data analysis

For the statistical analysis of the results, estimates of the mean $\bar{y}$ and the standard deviation σ were calculated as

$$\bar{y} = \frac{1}{N_{samples}} \sum_{i=1}^{N_{samples}} y_i, \tag{4.1}$$

$$\sigma = \sqrt{\frac{1}{N_{samples} - 1} \sum_{i=1}^{N_{samples}} (y_i - \bar{y})^2}, \tag{4.2}$$

for which $N_{samples}$ is the number of samples y_i [73]. Prior to further evaluation, the Shapiro-Wilk test was evaluated based on these estimates, in accordance with the description of Sen and Srivastava [74] in order to evaluate if the population, which the samples have been collected from, can be assumed as normally distributed. This test is applicable for the evaluation of sample sets with a size of less than 50 [74] and is characterized as strong in relation to comparable other test procedures [75].

After positive evaluation of the Shapiro-Wilk test, the 95 % confidence intervals of the mean $\bar{x}$ and the standard deviation σ were determined, in order to evaluate the accuracy of the estimates. Based on the still limited size of the sample set the Student's t-distribution was used for the calculation of the confidence interval of the mean [73]

$$\left[\bar{y} + t_{\frac{\alpha}{2},(N_{samples}-1)} \cdot \frac{\sigma}{\sqrt{N_{samples}}} \leq \bar{y} \leq \bar{y} + t_{1-\frac{\alpha}{2},(N_{samples}-1)} \cdot \frac{\sigma}{\sqrt{N_{samples}}} \right], \quad (4.3)$$

while the calculation of the confidence intervals of the standard deviation (cd. Equation 4.4), needs to be based on the χ^2-distribution [73]

$$\left[\sqrt{(N_{samples} - 1) \cdot \frac{\sigma^2}{\chi^2_{1-\frac{\alpha}{2},(N_{samples}-1)}}} \leq \sigma \leq \sqrt{(N_{samples} - 1) \cdot \frac{\sigma^2}{\chi^2_{\frac{\alpha}{2},(N_{samples}-1)}}} \right].$$
$$(4.4)$$

Additionally, the computed intervals were compared to confidence intervals determined based on the Wilcoxon test [76], which is a non-parametric test that does not assume a specific kind of distribution, in order to evaluate possible differences with respect to the assumption of a normal distribution. Furthermore, the estimated mean value and the corresponding confidence interval, computed based on the lab-scale samples, were compared with the flux measurement of the industrial module, which is considered as representative of the true mean flux. Taking into account the positive Shapiro-Wilk test and the accuracy information of the confidence interval, it was also tested, if the estimated mean and standard deviation can provide an accurate estimate of the distribution of the underlying population. Therefore, it was checked weather 95 % of the samples are located within the range of 2σ [73] around the mean value and a graphical comparison of the frequency of occurrence of specific flux measurement within a discrete set of intervals and the estimated probability density according to the normal distribution.

Finally, indicative values for the necessary sample size, for a certain accuracy of the flux estimates were determined in two ways. First, the confidence intervals according to Equation 4.3 were determined for different sample sizes from 2-15 according to the mean and standard deviation determined for the full set of lab-scale flux measurements. The width of the confidence interval thereby quantifies the range of uncertainty, with respect to the true mean value. In order to analyze the effect of single measurements, which are less probable, but are still reasonable with respect

to the analyzed distribution, another estimate of the confidence intervals for the different sample sizes was determined in a second approach, which performs a random sampling from the given full set of lab-scale flux measurements and determines estimates of mean and standard deviation, for the given sample set. The random sampling was performed 1000 times for each of the sample sizes. Based on these 1000 estimates of the mean and standard deviation and the resulting confidence intervals one overall estimate of the confidence interval was determined by calculating the 2.5 % quantile of all lower limits of the intervals and the 97.5 % quantile of the upper limits.

4.3 Results and Discussion

The experimental results for lab-scale membrane samples and the industrial size module are discussed in the following. Subsequent to this analysis an indicative value for the necessary sample size is determined in order to give a recommendation for flux experiments. Furthermore, the validity of the determined recommendation is evaluated for additional membranes and solvents before the effect of a consideration of the variance is discussed for some representative experimental and theoretical studies from literature.

4.3.1 Lab-Scale experiments

Overall, 52 membrane samples were evaluated in the lab-scale set-ups at FVT TUDo and Evonik. Five outliers were identified based on significant deviations of the measured flux, ranging from 150 to 470 % of the mean average flux of the remaining 47 samples. These large deviations were attributed to leakage of the cells, due to incorrect placement of the seals or damage of the membrane during installation in the cell. Hence, these outliers were not considered representative and consequently excluded. The determined flux at30 bar for all of these samples is illustrated in Figure 4.3, whereas samples evaluated at Evonik are indicated as gray bars and samples evaluated at FVT TUDo as white bars. The visual comparison confirms that there are no significant differences between the measurement results determined at each of the sides and that the measured flux for the different samples is subject to a considerable variance. This variance cannot be attributed to potential measurement errors, which based on Gaussian error propagation of balance accuracy (± 0.002 g),

the accuracy of the time measurement ($\pm 5\,\mathrm{s}$) and a potential density error ($\pm 10\,\mathrm{g\,l^{-1}}$) does not exceed $1.5\,\%$ of the flux measurement.
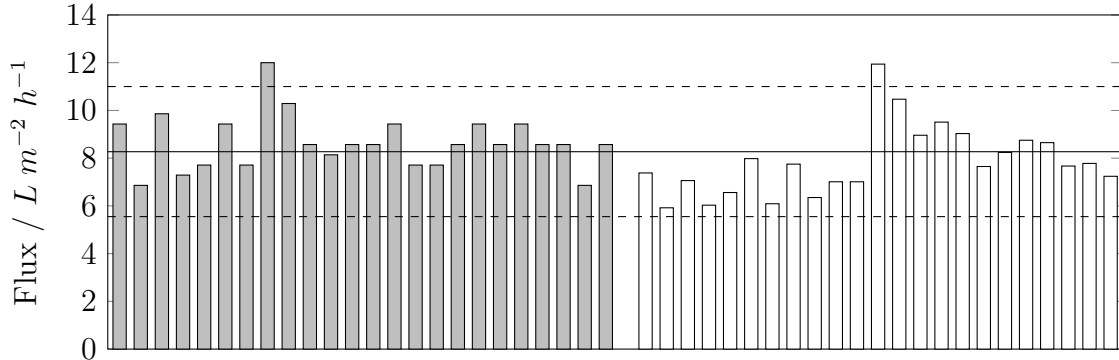
Figure 4.3: Illustration of permeate fluxes of ethanol measured for different samples of the DuraMem® 200 at 30 bar (grey bars: measurement at Evonik, white bar: measurement at FVT TUDo, solid line: mean of the measured fluxes, dashed lines: 2σ-range), with respect to the combined sample set, see Table 4.1

In order to further check whether the population of which the samples were collected may be assumed normally distributed, the Shapiro-Wilk test was performed considering a level of significance of $5\,\%$ (confidence level of $95\,\%$). Based on the number of samples within the investigated population and the level of significance a critical value for the test statistic of the Shapiro-Wilk can be derived from the corresponding tables provided by Sen and Srivastava [74]. In this case, a test statistic lower than 0.946 would result in the rejection of the hypothesis, that the underlying population is normally distributed. For the investigated population the test statistic was calculated to 0.954 which is slightly higher than the critical value. Hence, the Shapiro-Wilk test did not reject the assumption of a normally distributed population, so that the standard deviation, as well as the size of the $95\,\%$ confidence intervals were calculated for each of the experimental facilities and the combined set of samples (cf. Table 4.1).

Since the differences between the sample sets were comparably small, it was concluded that the different experimental set-ups did not have a significant effect on the measured flux data and the combined ensemble was used for further analysis. In consequence, it was tested if the estimates for the mean and standard deviation can provide an accurate estimate of the distribution of the underlying population. Therefore, it was checked weather $95\,\%$ of the samples were located within the range

Table 4.1: Estimated mean ($\hat{\mu}$), standard deviation ($\hat{\sigma}$) and respective confidence intervals (CI) determined for the flux of ethanol through the DuraMem® 200 at 30 bar for the sample sets.

	$\hat{\mu}$ $\mathrm{L\,m^{-2}\,h^{-1}}$	$CI_{\hat{\mu}}$ $\mathrm{L\,m^{-2}\,h^{-1}}$	$\hat{\sigma}$ $\mathrm{L\,m^{-2}\,h^{-1}}$	$CI_{\hat{\sigma}}$ $\mathrm{L\,m^{-2}\,h^{-1}}$	Samples
Evonik	8.66	±0.49	1.15	±0.22	24
FVT TUDo	7.87	±0.63	1.47	±0.29	23
Combined	8.27	±0.40	1.36	±0.20	47

of 2σ [73] around the estimated mean value. Figure 4.3 indicates the mean of the measured fluxes of ethanol at 30 bar and the 2σ-interval around the mean. Since only two out of 47 samples ($\approx 4\,\%$) were located slightly outside the 2σ-interval, 96 % of the samples fall within the interval. For further analysis, the samples were divided into intervals with respect to the measured flux and the relative frequency of occurrence was compared with the estimated probability according to the normal distribution with the estimated mean and standard deviation (cf. Figure 4.4).
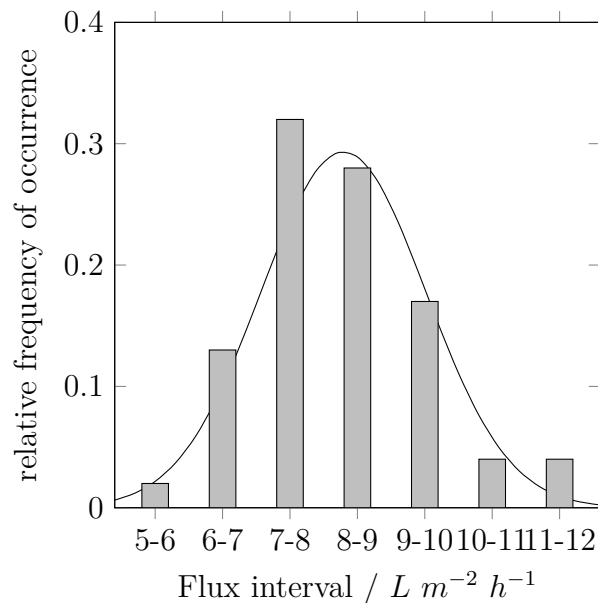
Figure 4.4: Comparison of relative frequency of occurrence of measured fluxes of ethanol at 30 bar with the probability distribution, based on the estimated mean and standard deviation.

Due to the noticeable shift to the side of lower fluxes, the relative frequency in the discredited intervals does not match perfectly to the density function of the

normal distribution. As described in Section 4.2.3 the Wilcoxon test [76] was applied additionally to evaluate possible differences of 95 % confidence intervals in case a normal distribution cannot be assumed. However, no significant deviation between the estimated confidence intervals was determined.

4.3.2 Comparison with membrane module

While the large set of samples evaluated in the lab-scale set-up provide an estimate of the mean ethanol flux and the deviation that is to be expected for such samples, it is furthermore of interest to confirm that this estimate is a good approximation of the real mean flux of the underlying population. For this purpose, a comparison was made to the industrial membrane module, which provides an active membrane area that is about 80 times the sample size of the lab-scale experiments. Thus, it was twice the size than the combined samples from the lab-scale analysis. For the spiral wound module a flux of $7.9\,\mathrm{L\,m^{-2}\,h^{-1}}$ was determined at 30 bar, which is just within the confidence interval of the estimated mean from the lab-scale experiments. While it might seem disappointing that the estimate from the combined lab-scale measurements is not an exact fit to the spiral wound module, the difference between the estimates is just $4.7\,\%$ and the fact that value for the module is located in the confidence interval of the lab-scale estimate does prove the value of the confidence interval for the analysis of the flux measurement.

Even if the measured flux from the industrial module is considered as an accurate representation of the real mean ethanol flux, the single large-scale experiment does not provide any information on the variance of the flux. Thus, the larger number of lab-scale experiments provides added information with respect to the performance quantification. For comparison purpose an additional estimate of the standard deviation was determined, considering that the flux of the spiral-wound module represents the true mean value and demanding that 95 % of all lab-scale samples are located within a 2σ-interval around this mean value. The resulting interval range of $\pm2.61\,\mathrm{L\,m^{-2}\,h^{-1}}$ was approximately 33 % of the mean ethanol flux determined for the industrial module. As illustrated in Figure 4.5 it is in very good agreement with the interval determined on the basis of the lab-scale samples.

Overall, it can be concluded that given a sufficient quantity, lab-scale experiments of appropriate quality provide an accurate estimate of the membrane specific flux, which is coherent with the performance of the industrial module. They furthermore provide insight into the possible deviations and allow for a quantification of the
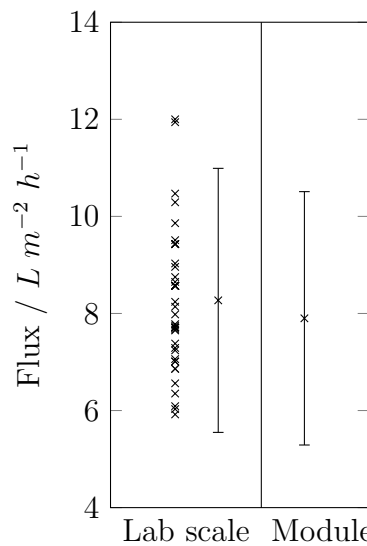
Figure 4.5: Mean of the permeate fluxes of ethanol at 30 bar determined in lab-scale experiments (left, the error bar represents the 2σ-range determined in lab-scale experiments) and for the membrane module (right, the error bars represent $\pm 33\%$ which includes 95% of all measured fluxes in lab-scale)

estimate accuracy, which is not possible through an evaluation of a single experiment with the industrial module. However, the large set of 47 samples requires multiple experimental runs and is not feasible for an implementation in the usual workflow. Therefore, the next section addresses the important question, how large a necessary sample size should be for a reliable estimate of the flux.

4.3.3 Evaluation of necessary sample size for meaningful estimates

In order to determine the necessary sample size, which allows for an acceptable accuracy of the estimates, the confidence intervals for different sample sizes of 2-15 were determined by the two different approaches described in Section 4.2.3. Table 4.2 shows the lower and upper limits ($CI_{low}|CI_{up}$) as well as the relative width (rel. width) of the derived confidence intervals. Obviously, at least three samples are required to avoid an extension of the confidence interval beyond a flux of zero. This means that any estimate that is derived from a single or two samples would need to be considered statistically insignificant. When considering the mean and standard deviation determined for the full set of lab-scale samples, a sample size of

four suffices to reduce the uncertainty with respect to the estimate of the mean to around $\pm 25\,\%$, while a sample size of five results in a further reduction to $\pm 20\,\%$. Using the sampling approach, which reflects the impact of single measurements with larger, yet reasonable, deviation to the true mean flux, the necessary size of the sampling set for a relative width of the confidence interval of $\pm 25\,\%$ increases to eight. As described in Section 4.2.1, a simultaneous evaluation of this sample size is still feasible in a lab-scale set-up like the METcell system.

Table 4.2: Estimated intervals for different sample sizes

Sample size	Based on lab-scale samples			Based on sampling approach		
	CI_{low}	CI_{up}	rel. width	CI_{low}	CI_{up}	rel. width
2	-3.95	20.49	148%	-19.96	38.70	313%
3	4.89	11.65	41%	2.26	15.73	75%
4	6.11	10.43	26%	4.55	12.83	48%
5	6.58	9.96	20%	5.48	11.73	36%
6	6.84	9.70	17%	5.89	11.18	31%
7	7.01	9.53	15%	6.15	10.78	27%
8	7.13	9.41	14%	6.39	10.63	25%
9	7.23	9.32	13%	6.60	10.40	22%
10	7.30	9.24	12%	6.68	10.27	21%
11	7.36	9.18	11%	6.74	10.08	20%
12	7.41	9.13	10%	6.87	10.03	19%
13	7.45	9.09	10%	6.90	9.90	18%
14	7.49	9.06	10%	6.96	9.78	17%
15	7.52	9.02	9%	7.02	9.70	16%

The results indicate that a good approximation of the true mean flux with high accuracy is feasible with a sample set of four, in case no large deviations occur, while otherwise an increase to a sample set of eight can compensate for such deviating experiments, which still are considered reasonable according to the resulting normal distribution. This does not include outliers that are e.g. the result of a defect of the membrane sample. Of course, this estimate is based on the small size of the membrane samples and reflects the properties of the specific membrane. Yet, the quality of this industrially produced membrane should be considered high and the size of the individual samples as representative for lab-scale studies.

4.4 Conclusion

In this chapter the material-dependent variance in solvent flux for lab-scale measurements is analyzed. For this purpose, the pure solvent flux of ethanol was evaluated for about fifty lab-scale membrane samples and an industrial spiral-wound module. The evaluation allows for the determination of a minimum sample size that should be considered for a representative evaluation of the membrane performance. Based on this investigation it could be shown that the assumption of a normal distribution with respect to the variations of the membrane material is reasonable and that the determined estimate for the average flux and its confidence interval provides a good approximation of the true mean value, represented by the industrial spiral-wound module. The ensemble of lab-scale samples provides added information on the variance of the flux, which is not attainable from the spiral-wound module. Based on the analysis it becomes apparent that a reliable flux estimate with related uncertainty of the estimate below $\pm 25\,\%$, referring to the extension of the confidence interval, requires a minimum sample size of four. Additional samples should be considered in case of higher deviations that cannot be attributed to defects and thus excludable outliers.

Since common lab-scale set-ups like the MET system or the tailored equipment of Vandezande et al. [36] easily allow for the investigation of a representative number of samples in a single experimental run, it is appropriate to utilize these set-ups. Although there is still an ongoing debate on the correct use and the potential misinterpretation of the information from statistical tests and confidence intervals [77], they do provide valuable information on the reliability of the derived estimates, which is of significant value for model-based analysis and conceptual design. The transferability of the determined uncertainty information to other solvents and membranes and the influence on process selection is investigated in the subsequent chapter.

5

Towards a comparable characterization of polymeric OSN membranes

In a cooperative study with different project partners, a standardized experimental procedure was developed and used in order to determine the comparability of flux and rejection determined at different facilities. Moreover, different categories for solutes and solvents have defined in order to create an extensive database for a variety of different components and applications. In the following chapter, the experiments contributed to this cooperative study as well as experimental results for additional solutes are presented and discussed. Moreover, these experimental results are put in the context of the cooperative study regarding the comparability. The importance of the uncertainty information derived in Chapter 4 for membrane selection and model-based conceptual process design is furthermore illustrated for case-studies based on the experimental results of the cooperative study and studies from literature.

5.1 Introduction

The feasibility of a separation process by means of organophilic nanofiltration can be evaluated using existing data. For this purpose, however, on the one hand a broad data base is required and on the other hand, these data should be comparable and reproducible. So far, it is common practice to characterize a membrane either only for a specific system or for only one or a few different systems (Section 2.1.3). Consequently, there is a lack of data that allows estimations for different systems and additionally, the experiments are not performed according to standardized methods, which makes transferability and comparability difficult.

In order to address these challenges, a standardized measurement method was developed in a cooperative project, which was used to investigate the comparability of flux and rejection. Furthermore, categories for solvents and solutes were defined and individual representatives were selected to create a broad database for OSN membranes using these components and the new measurement method. This database can be used to estimate the feasibility of separation processes or as a basis for model development.

In the following chapter, the materials used in the experiments of this thesis are introduced (Section 5.2) and the experimental procedure is described (Section 5.3). Subsequently, the results are discussed and presented in the context of the cooperative study of the project. For this, the uncertainty information determined in Chapter 4 is taken into account and extended to rejection. Moreover, the importance of taking the uncertainty information into account for process analysis is investigated. Finally, the results are summarized in Section 5.5.

5.2 Materials

For membrane characterization the solvent flux through the membrane as well as the rejection of different solutes were determined in corresponding experiments. The different combinations of solvents and solutes used in these experiments are summarized in the following. Further, the properties of the membranes investigated in this study are specified.

5.2.1 Solvents

At first, six solvents were selected based on variations in terms of viscosity and polarity, considering the solvent dependent solute sensitivity regions reported by Thiermeyer et al. [78]. Ethanol and isopropanol were selected as common polar solvents, which are able to form hydrogen bonds. Methyl ethyl ketone (MEK) and ethyl acetate (EtAc) were selected as moderately polar solvents, which are not able to form hydrogen bonds. Finally, toluene and heptane were selected as almost non-polar solvents. The division into polar, moderately polar and non-polar solvents is in accordance with the polarity index [79] and the categorization listed in Table 5.1. Nevertheless, it should be noted that based on the Hildebrand or Hansen solubility parameters, toluene is exactly as polar as ethyl acetate.

In addition to these six solvents, acetone and n-butyl acetate (BuAc) were added in order to enlarge the data set for each chemical group of solvents. Both solvents are moderate polar according to the polarity index [79]. All solvents were acquired with warranted purities of at least 99.5 % by Merck KGaA, Carl Roth GmbH + Co. KG, ChemSolute and VWR Chemicals.

Table 5.1: Hildebrand solubility parameter [80] and Polarity [79] of the solvents used in this study.

Solvent group	Solvent	Solubility parameter MPa$^{0.5}$	Polarity %
polar	Ethanol (EtOH)	26.5	65.4%
	Isopropanol (IPA)	23.6	54.6%
moderate polar	Methyl ethyl ketone (MEK)	19.1	32.7%
	Ethyl acetate (EtAc)	18.2	23.0%
non-polar	Toluene (Tol)	18.2	9.9%
	n-Heptane (Hep)	14.9	1.2%
additional	Acetone (Ace)	19.9	35.5%
	Butyl acetate (BuAc)	17.4	24.1%

5.2.2 Solutes

The selection of solutes was based on the variation of molecular size, structure and polarity, while accounting for the necessary solubility in the selected solvents and availability. One group of solutes includes styrene oligomers with different molecular

weight. This solute group allows an insight into the rejection of different molecular sizes without functional groups changing. Moreover, it is a standard system for membrane characterization used by membrane manufacturers. For the rejection experiments the two different polymer standards PSS-ps560 and PSS-ps1.8k purchased from PSS Polymer Standards Services GmbH are used, offering a molar mass distribution between 266 and 1800 kg kmol^{-1}.
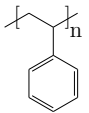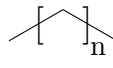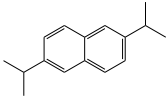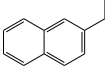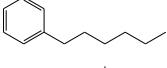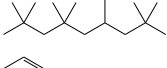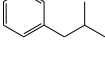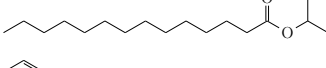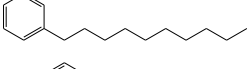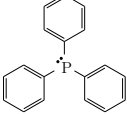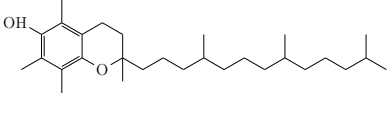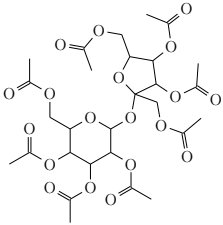
Additionally, a group of ten linear alkanes (C10-C36) in combination with sucrose octaacetate (SOA) was used. As a third group (hereinafter referred to as further solutes) aromatic compounds with different side chains (Isobutylbenzene (IBB), hexylbenzene (HB), 2-Ethylnaphthalene (EN), Phenyldecane (PD), 2,6-Diisopropylnaphthalene (DIPN) and Triphenylphosphine (TPP)) as well as vitamin E (α-tocopherol, VE), the ester Isopropylmyristate (IPM) and the branched isomer of the linear C16 alkane 2,2,4,4,6,8,8-heptamethylnonane(HMN) were selected as solutes. The alkanes represent non-polar solutes with different chain length but similar structure, while the remaining molecules are mainly moderate polar solutes with varying molecular structures. Table 5.2 provides a summary of the solutes and the corresponding structure as well as important properties like molecular weight and solubility parameter. According to the vendor specifications (Alfa Aesar, Sigma Aldrich, Arcos organics, Merck KGaA and Abcam), all solutes had a purity of at least 98 %, except for hexylbenzene for which a minimum purity of 97 % was specified. Although, the solubility of the solutes in the different solutes was taken into account, not every solute was used in each solvent.

5.2.3 Membranes

The experiments were performed using DuraMem® 150, DuraMem® 200, DuraMem® 300 and PuraMem® S600 membranes supplied by Evonik Industries. Additionally, the PDMS(2 µm) membrane manufactured by the Helmholtz-Zentrum Geestacht is used. The membranes of the DuraMem® series are made of polyimide and are suitable for applications in polar solvents. DuraMem® membranes are indicated by the vendor as stable in the solvents ethanol, isopropanol and methyl ethyl ketone, in which these membranes were used exclusively. The MWCOs of the DuraMem® 150, the DuraMem® 200 and the DuraMem® 300 membranes are listed as 150, 200 and 300 g mol^{-1}, respectively, determined using styrene oligomers dissolved in toluene [81].

The active layer of the PuraMem® S600 and PDMS(2 µm) membranes are poly-

Table 5.2: Overview of the solutes and the corresponding molar mass, structure and Hildebrand solubility parameter (Hildebrand parameter was not available for SOA)

Solute	Molar mass g mol^{-1}	Structure	Solubility parameter MPa$^{0.5}$
Styrene oligomers	266-1.8k		-
C10-C36	142.3-507.0	(n=8,10,12,14,16, 20,22,26,30,34)	15.1-15.6
DIPN	212.3		17.8
EN	156.2		19.2
HB	162.3		18.1
HMN	226.5		14.3
IBB	134.2		20.0
IPM	270.5		16.2
PD	218.4		18.0
TPP	262.3		18.7
VE	430.7		25.2
SOA	678.6		-

dimethylsiloxane (PDMS) based and these membranes are more suitable for applications in non-polar solvents. The applied solvents are specifically listed by the ven-

dor as solvents for which the PuraMem® S600 membranes can be considered stable. The technical information sheet lists a MWCO of $600\,\mathrm{g\,mol^{-1}}$ for the PuraMem® S600 based on rejection data for styrene oligomers dissolved in toluene, determined at operating conditions of 30 bar and 30 °C [82]. The PDMS(2 µm) membrane is similar to the membrane described by Ebert et al. [83] and considered stable in all solvents as well. However, since the MWCO strongly depends on the solvent used, no MWCO was listed for this membrane type so far.

5.2.4 Investigated experimental systems

The investigated experimental systems differ for the different membranes. Figure 5.1 provides an overview of the different components. Each experimental system consists of a membrane, a solvent and a solute group. Since DuraMem® membranes are not applicable in non-polar solvents and ethyl acetate, these membranes were only used in polar solvents and methyl ethyl ketone. The first experimental investigations using the DuraMem® 150 and DuraMem® 200 in ethanol with the solute group of styrene oligomers resulted in complete rejection of these solutes although a different solvent was used. A high rejection of solutes is actually very interesting from application point of view. However, in the context of this study influencing properties of solvent and solute on flux and rejection are of interest. Complete rejections of all solutes regardless of molecular size or different properties do not provide any further insight into the dependence of the rejection on different solute properties. Therefore, a broader variety in rejections is more purposeful in the context of this study. For this reason, the focus was on the membranes with a higher reported MWCO, where fluctuations in rejection were observed for different components, while the DuraMem® 150 and DuraMem® 200 membranes were not investigated further. Moreover, the styrene oligomers were not used in other solvents than ethanol and toluene in this study but investigated by project partners after a successful evaluation of the comparability of experimental results determined at different facilities [16].

The group of "further solutes" was only used in experiments with the PuraMem® S600 and PDMS(2 µm) membrane, since the DuraMem® 300 membrane was only usable in a limited number of solvents and hence, the database was small compared to the other membranes. However, the number of solvents for the experiments with the group of further solvents has been increased by adding acetone and n-butyl acetate.

Membrane
- PuraMem® S600
- PDMS(2 µm)
- DuraMem® 150[1]
- DuraMem® 200[1]
- DuraMem® 300[2]

Experimental system

Solute
- Styrene oligomers[3]
- Lineare alkanes + SOA
- Further solutes[4]

Solvent
- polar: Ethanol, Isopropanol
- moderate polar: Methyl ethyl ketone, Ethyl acetate
- nonpolar: Toluene, n-Heptane
- additional[5]: Acetone, Butyl acetate
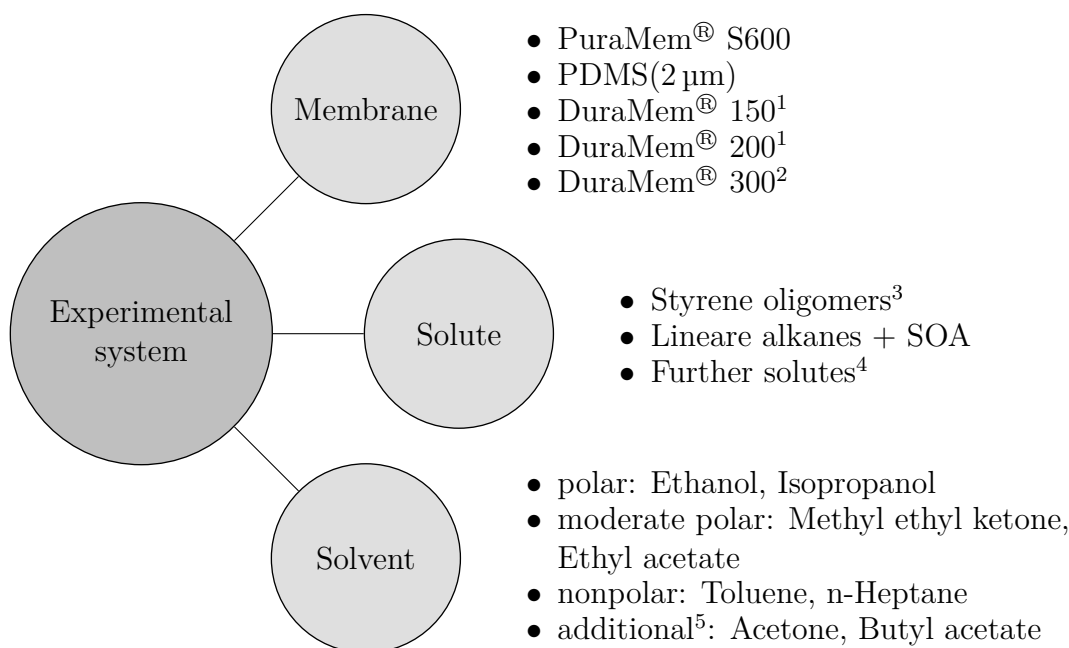
Figure 5.1: Investigated chemical systems (1: only with Ethanol + Polystyrene; 2: not with non-polar solvents and Ethyl acetate; 3: only with Ethanol and Toluene; 4: not with DuraMem®; 5: only with further solutes + PuraMem® S600 and PDMS(2 µm))

5.3 Methods

The experimental procedure used for the experiments presented in this chapter follows an standardized experimental protocol developed in a collaborative research project, in which an extensive experimental investigation of various membrane solvent and solute combinations was performed in order to analyze the comparability of flux and rejection measurements at different laboratories and experimental set-ups. The different experimental set-ups include the METcell system described in Chapter 4, but also specifically manufactured set-ups with larger membrane test cells (up to $100\,\text{cm}^2$).

The exact procedure used for determination of flux and rejection in pure solvents for different membranes is described the Section 5.3.1. The experiments using solvent mixtures are described in Appendix B. Depending on the solutes used in the experiments, either gaschromatography or high pressure liquid chromatography is used

for the determination of the concentration of the solutes in permeate and retentate (Section 5.3.2).

5.3.1 Experimental procedure following the standardized protocol

The experiments were performed in a cross-flow nanofiltration system of type LSta60-2LM, illustrated in Figure 5.2, purchased from SIMA-tec GmbH, which was already used and described by Schmidt et al. [65] and Werth et al. [21]. The cross-flow system contains a double jacket feed tank with a capacity of $7.5\,l$ and two membrane cells connected in series with an active membrane area of $80\,cm^2$ each, which both were equipped with 28 mil feed spacers to increase the turbulence on the membrane surface while the cross-flow velocity was maximized using the Hydracell G22 pump by Verder Deutschland GmbH at a frequency of $25\,Hz$. The plant is equipped with sampling ports for feed, retentate and permeate of both cells separately (cf. Figure 5.2 as well as the flow diagram of the plant reported by Schmidt et al. [65]). During the initial operation both permeate and retentate were recycled in order to maintain constant feed conditions. Moreover, the temperature in the membrane cells and the pressure in the plant were monitored using DasyLab9.0 (Measurement Computing, USA) as well as coriolis cross-flow meters (ABB, FCM2000MS-21) and gravimetric measurement.
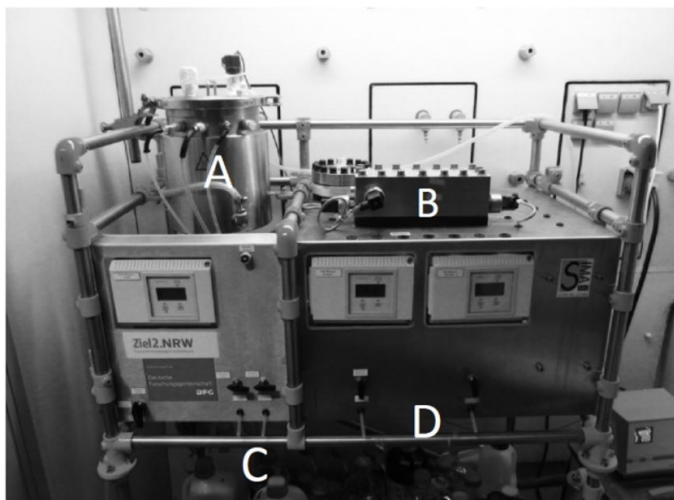
Figure 5.2: Illustration of the applied cross-flow nanofiltration system, with the double jacket feed tank (A), the membrane cells (B), sampling valves for feed and retentate (C), as well as the permeate (D).

The experiments were performed in three steps, further considered as standardized procedure [16]. First, the plant was rinsed with pure solvent to remove remaining solvents or solutes from the previous experiment. If no solvent change was performed in between experiments a single rinsing step was considered sufficient, while the plant was rinsed three times with fresh solvent in case the solvent was changed. After this initial rinsing of the plant, fresh membranes were installed and again rinsed with fresh solvent in order to remove conditioning agents from the membrane. Finally, the actual rejection experiment was performed and the feed tank was filled with $1.5\,l$ of the specific solution for the experiment. The rejection experiments were performed in series, using one group of solvents at a time. For each series $0.1\,g\,l^{-1}$ of the different solutes was solved in the solvent, thus not exceeding an initial amount of dissolved solutes of $1\,g\,l^{-1}$ in any experiment. Using the styrene oligomer mixtures, $1\,g\,l^{-1}$ of the polymer standards PSS-ps560 was solved in the corresponding solvent and in experiments using toluene as solvent, $1\,g\,l^{-1}$ of the second polymer standard PSS-ps1.8k was added, which is not soluble in ethanol.

While a constant temperature of $25\,°C$ for the feed was kept during all experiments, solvent flux and solute rejection were determined at a transmembrane pressure of 20, 30 and $40\,bar$ respectively. Both membrane cells were used with samples of the same membrane sheet in order to keep record of possible variations between the samples. Prior to the collection of retentate and permeate samples, the feed solution was circulated for at least two hours to reach steady state conditions. The flux was determined gravimetrically while the samples of retentate and both permeates used for rejection determination were analyzed applying different analytical methods depending on the solute group as described in Section 5.3.2.

5.3.2 Analytical methods for solute concentrations

For the determination of the rejection, the two permeate samples of each cell and the retentate sample were either analyzed by a gas chromatography (GC) or by high pressure liquid chromatography (HPLC) depending on the type of solutes used in the specific experiment. For the determination of the rejection using styrene oligomers HPLC of Agilent Series 1200 was used. The HPLC was equipped with a column of the type Agilent ZOBRAX Eclipse Plus C18 (4.6 mm x 250 mm, 5 µm) and combined with an UV/vis detector set to a wavelength of 264 nm. The analytical method complies with the method described by See Toh et al. [37] but was modified in order to achieve higher resolutions for the present set-up. As mobile phase a mix-

ture of analytical grade water and tetrahydrofurane (THF) are used with 0.1 vol%
triflouroacetic acid. During the analysis of each sample, the composition of the mo-
bile phase was changed according to Table 5.3 to achieve a better resolution and
after each set of two permeate samples and the corresponding retentate sample, one
sample with pure (THF) was inserted to rinse the column.

Since the peak of toluene overlays peaks of the solutes in the chromatogram, toluene
was completely evaporated of the permeate and retentate samples and the samples
were resolved in THF. For samples containing ethanol, this step was not necessary.
The different solutes were identified based on the curves proved by PSS Polymer
Standards Services GmbH. Since the area of the peaks for each solute size corre-
sponds to the concentration, the rejection of each solute is determined based on the
ratio of the areas of the peaks in the permeate and the retentate sample. Each anal-
ysis is repeated three times and the mean of the analyses is used for concentration
determination.

Table 5.3: Composition of mobile phase for the analysis of styrene oligomer samples
with HPLC

Time	Solvent composition
Minute 0 to 13	65 % THF + 35 % H_2O
Minute 13 to 23	Change to 90 % THF + 10 % H_2O (gradient)
Minute 23 to 28	90 % THF + 10 % H_2O
Minute 28 to 33	65 % THF + 35 % H_2O (rinsing)

For the determination of the rejection of alkanes, SOA and all further solutes, the
sample were analyzed by offline GC, using a SHIMADZU chromatograph equipped
with a liquid auto sampler, flame ionization detector (FID) and a capillary column
FS Supreme 5ms-HT. The column length was 30 m, the inner capillary diameter is
0.32 mm and the film thickness of the stationary phase is 0.25 μm. Each sample
was prepared for analysis in vials containing 1.1 g of the sample and 0.1 g of the
internal standard dibuthyl ether. The injection volume of 1.0 μl was chosen and
a split ration of 1:100 was applied. The injection temperature was set to 340 °C
and temperature of the flame ionization detector (FID) was 360 °C. During the
separation the temperature of the column was regulated by an individual profile
depending on the solutes in the sample.

For the group of alkanes and SOA the column temperature was set to 80 °C in the

beginning and raised to $350\,°C$ at a rate of $20\,°C\,min^{-1}$. The final temperature was held for $5.5\,min$ in order to provide enough time for each solute to pass the column. For group of "further" solutes the temperature program was adapted in order to ensure a high resolution. The initial temperature of the column was $100\,°C$ and raised to $200\,°C$ with a rate of $30\,°C\,min^{-1}$. In a second step, the rate was changed to $10\,°C\,min^{-1}$ until a column temperature of $270\,°C$ was reached. In the last step, the temperature rate of $30\,°C\,min^{-1}$ was applied again and the final column temperature of $370\,°C$ was kept constant for $2.5\,min$. Each analysis is repeated three times and the mean of the analyses is used for concentration determination. The composition of each sample was determined based on a calibration curve determined beforehand for each solute.

5.4 Results and Discussion

The results for solvent flux and rejection of the alkanes and further solutes measured with the PuraMem® S600 are presented and discussed in the following section. Experimental results for flux and rejection determined with different membranes and with solvent mixtures are reported in Appendix B.

5.4.1 Experimental results for flux and rejection using pure solvents

The measured flux of the different solvents for the membrane PuraMem® S600 are discussed below based on the values measured in the experiments with the group of further solutes at $30\,bar$. These values correspond to those measured with other solutes at $30\,bar$ and show the same trends as the fluxes at 20 and $40\,bar$. Figure 5.3 illustrates the fluxes plotted over the reciprocal of the viscosity and additionally over the Hildebrand solubility parameter of the solvents.

As already observed in different studies (e.g. [3, 84]), the flux increases significantly with decreasing viscosity of the solvent. Only for heptane the measured flux does not conform this profile. Therefore, measurements for the flux of heptane through the PuraMem® S600 membrane were repeated and measurements of the same membrane-solvent system but measured in a different experimental setup were taken into account, to confirm the results presented here. It was found that similar values for heptane were measured in these further experiments. Hence, a measure-
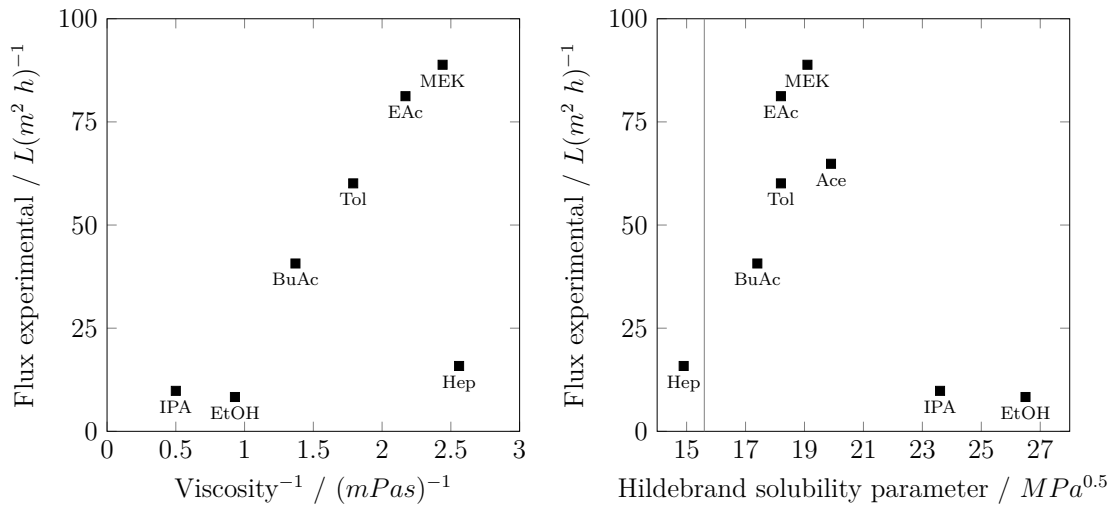
Figure 5.3: Flux of different solvents through the PuraMem® S600 membrane at 30 bar vs. reciprocal of the solvent viscosity and the Hildebrand solubility parameter. The vertical line illustrates the Hildebrand solubility parameter of PDMS.

ment error was excluded at first sight and the low flux of heptane may be the result of further influencing factors such as molecule structure and effective size of the molecule or wetting of the membrane. Compared to the other solvents presented in Figure 5.3, heptane is a more chain-like and the longest molecule. Heptane has the lowest dipole moment as well as surface tension compared to the other solvents, which may result in a less good wetting of the membrane surface and a reduction of the flux. Since Figure 5.3 only provides insights for the dependency of the flux on the reciprocal of the viscosity and the Hildebrand parameter, a strong influence of the factors mentioned above is not apparent. Additionally, the properties of the support material of the membrane may influence the flux through the membrane. Further, a combination of different factors resulting in a lower flux is also conceivable but not easy to determine by combining different factors manually. Hence, a method determining possible influencing factors based on experimental data could provide further insights. However, the low flux of heptane should be used with caution and it is reasonable to repeat the experiment for heptane again at another facility or with membrane samples of another batch to ensure the results found for heptane here.

Furthermore, a maximum in flux at a solubility parameter of the solvent of approximately $19\,\mathrm{MPa}^{0.5}$ is evident. Zeidler et al. [84], Schmidt et al. [65], and Postel et

al. [66] observed a higher flux of solvents with similar solubility parameters as the membrane. For the active PDMS layer of the membrane, the solubility parameter is $15.6\,\text{MPa}^{0.5}$ which does not correspond to the maximum in flux. On the one hand, the reason for this can be a significantly higher influence of the viscosity on the flux, whereby the maximum in the flux corresponds to a higher solubility parameter than expected, which, however, goes in line with the lowest viscosity. On the other hand, an influence of the polarity of the supporting layer of the membrane cannot be excluded. If the supporting layer is more polar than the active layer, the total polarity of the membrane may shift, whereby the maximum flux also shifts to more polar solvents. Of course, other factors can also influence the flux, which are not considered here or do not appear significant at first glance. Using data driven methods, further influencing factors can be identified and considered in models as discussed in Chapter 7 and 8.

The determined solute rejection data for the six solvents and the linear alkanes as well as the group of further solutes using the PuraMem® S600 membrane are illustrated in Figure 5.4. The data is presented in three subfigures, accounting for the different polarity ranges of the solvents, showcasing that similar general trends can be observed, as reported by Thiermeyer et al. [78]. The measurement errors are not explicitly illustrated, since both the standard deviation determined from the two membrane cells and the measurement error determined by gaussian error propagation were evaluated as smaller than an absolute deviation of $3\,\%$ from the absolute mean of the measured solute rejection, except for the rejection of alkanes in ethanol were an absolute deviation of $5\,\%$ from the absolute mean of the measured solute rejection was not exceeded. Assuming the maximum absolute deviation of $3\,\%$ for all rejection value and comparing these with the results found for the variation in flux (cf. Chapter 4), these errors are significantly smaller for high rejections but comparable high for measured rejections around zero. A detailed discussion on uncertainties in rejection will follow in Section 5.4.3.

While overall solute rejections vary between $-40\,\%$ and $100\,\%$, solvent polarity clearly affects solute rejection, whereas specifically for polar solvents no correlation between the molecular weight and the solute rejection can be observed. As apparent from the falling rejection of the alkanes with increasing chain length, the rejection in the polar solvents is dominated by solute polarity, rather than solute size. This is further supported by the strong variations in solute rejection for solutes of similar molecular weight. While a positive correlation of solute rejection

and molecular weight can be observed for the moderately polar solvents, rejections for shorter chain lengths of the alkanes are still negative and the rejection data of solutes with similar molecular weight in one solvent still vary to a considerable extend. Only for the non-polar solvents, there seems to be more distinct correlation between the molecular weight and rejection of a solute in the different solvents.
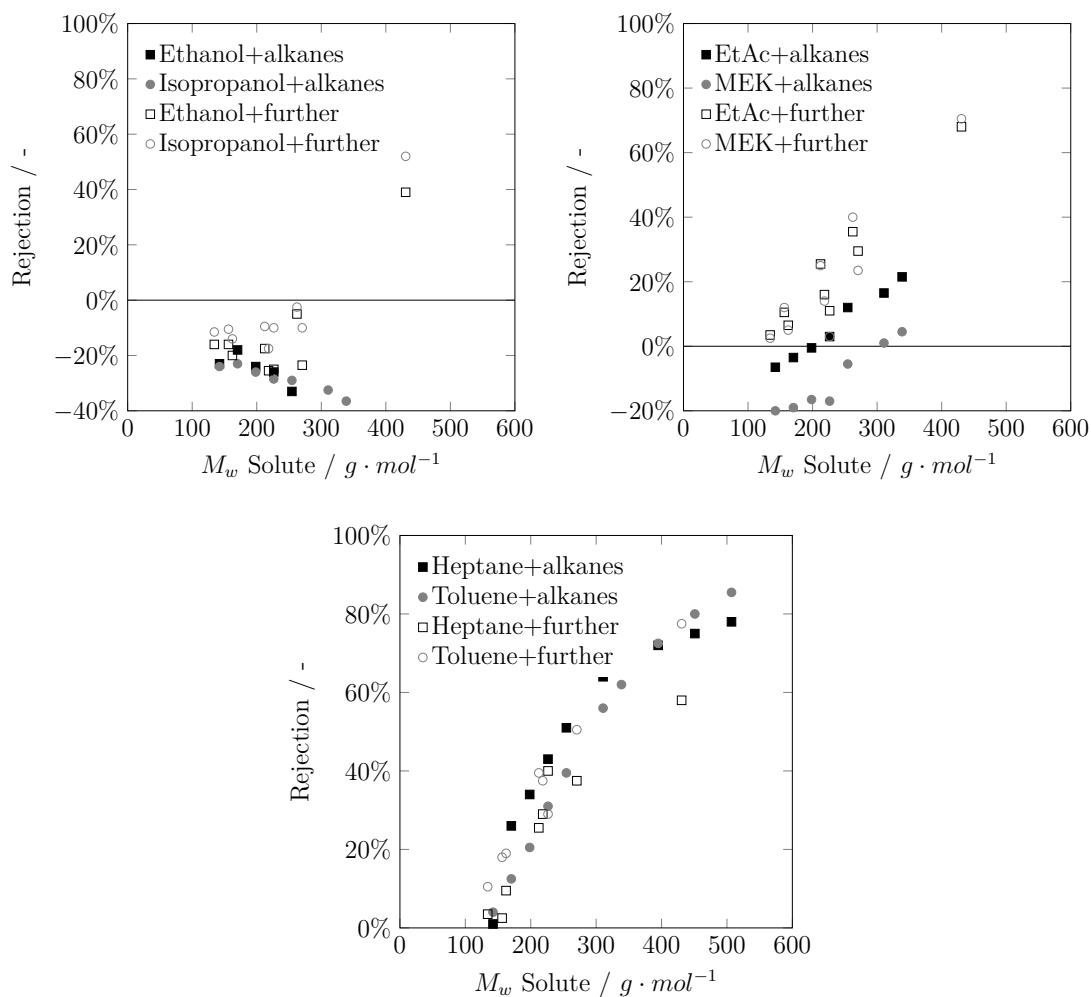
Figure 5.4: Solute rejection data of the 19 considered solutes in the 6 considered solvents for the membrane PuraMem® S600, divided into three groups according to solvent polarity (top left: polar solvents, top right: moderate polar solvents, bottom: non-polar solvents).

The experimentally determined solute rejections for the additionally considered solvents acetone and n-butyl acetate are illustrated in Figure 5.5. As observed for the moderately polar solvents ethyl acetate and MEK, the solute rejection corre-

lates with the molecular weight, while the rejection varies for solutes with the same molecular weight.
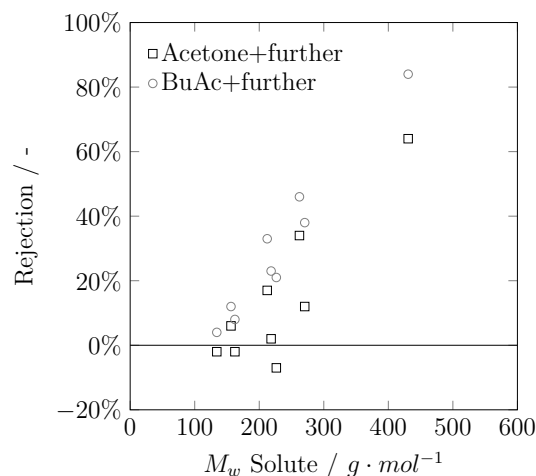
Figure 5.5: Solute rejection data of the group of further solutes in the additionally investigated solvents acetone and n-butyl acetate (BuAc) for the membrane PuraMem® S600.

5.4.2 Transferability of uncertainty information on flux for various membranes and solvents

In order to validate transferability of the uncertainty information gained in Chapter 4 to other membrane and solvent combinations, the recommendations for appropriate sample sizes were evaluated based on the experimental flux measurements determined in the cooperative study. For the following investigation the measured flux data for ethanol and toluene for four different commercially available OSN membranes (a PDMS membrane, the PuraMem® S600, DuraMem® 300 and DuraMem® 200) was considered. For each of the solvent-membrane combinations at least 15 data points were determined at different experimental facilities providing the necessary data for the evaluation. The distribution of the flux measurements is illustrated in Figure 5.6.

In order to test the validity of the suggested sample size, a random sampling of four and eight samples out of the data set for each solvent-membrane system was performed. Based on a series of 100 random sample sets and the confidence intervals determined for each combination, Table 5.4 lists the percentage of confidence intervals which were smaller than ±25 % of the estimated mean. It was found that
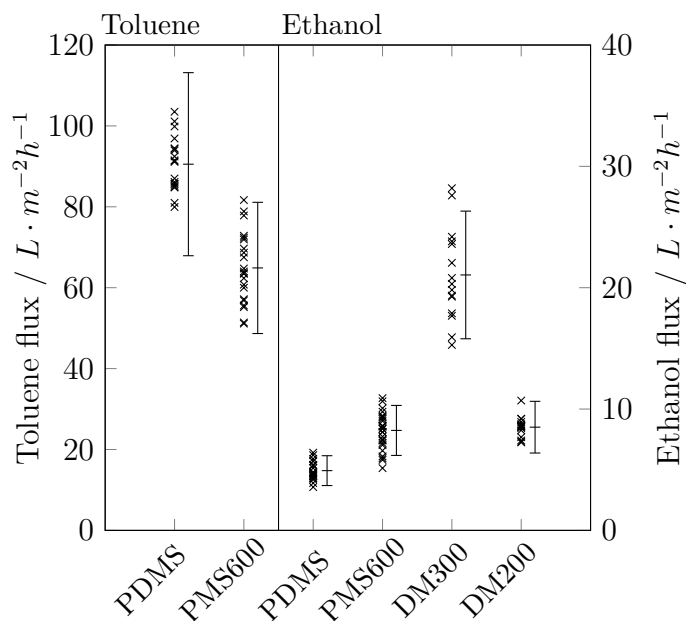
Figure 5.6: Flux measured with different membranes in toluene and ethanol (symbols), the mean of each data set (horizontal line) and error bars representing $\pm 25\%$ of the mean

only for the PDMS membrane used in toluene the percentage of confidence intervals that are smaller than $\pm 25\%$ is already 100% using four samples. For all other membrane/solvent combinations this is only achieved using 8 samples. The smaller variation of the measured fluxes for the PDMS membrane can be attributed to the different manufacturing processes. Considering the rather narrow distribution of the flux measurements for toluene and the PDMS and PuraMem® S600 and for ethanol and the DuraMem® 200 membrane the high portion of confidence intervals with elongation of less than $\pm 25\%$ is reasonable. The results for this additional set of measurements for the DuraMem® 200 membrane and ethanol, therefore, supports the hypothesis that the use of larger membrane samples directly results in a reduction of the required sample set for accurate flux estimates. However, the confidence interval for four samples still exceeds the $\pm 25\%$ limit in about 10% of the randomly selected samples. For the other membrane combinations, the ratio of sets with four samples that resulted in confidence intervals exceeding the $\pm 25\%$ limit was considerably larger. However, for all cases a sample size of eight suffices in keeping the limit of a maximum $\pm 25\%$ range of the confidence interval for each of the randomly selected sample sets, confirming the previous sample size recommendations.

Table 5.4: Percentage of confidence intervals smaller than $\pm 25\,\%$ derived from a series of 100 random drawn combinations of the available flux measurements for pure solvent flux from Böcking et al. [16], for a PDMS($2\,\mu$m), the PuraMem® S600 (PMS600), DuraMem® 300 (DM300) and DuraMem® 200 (DM200) membrane)

Sample size	PDMS		PMS600		DM300	DM200
	Toluene	Ethanol	Toluene	Ethanol	Ethanol	Ethanol
4	100%	71%	81%	43%	40%	91%
8	100%	100%	100%	100%	100%	100%

5.4.3 Uncertainties in rejection

To ensure the comparability of experimental results of the rejection as well, the same experiments using polystyrene oligomers measured with ethanol and toluene as solvents are performed at five different facilities following the same standardized experimental procedure as described in Section 5.3.1. Additionally, a verification of the analytical methods by cross validation the analytical results between at the different facilities excluded experimental errors resulting from the analysis. The standard deviation for the rejection is below $10\,\%$ [16] and comparability of experimental results measured in different facilities following the same standardized experimental procedure was successfully demonstrated in this cooperative study. Finally, all remaining experiments were conducted at least at two different experimental facilities. Figure 5.7 illustrates exemplary the rejection of the linear alkanes in three different solvents using the PuraMem® S600 membrane, as included in Figure 5.4 compared to results determined at RWTH Aachen.

The rejections of the different linear alkanes measured at the two facilities are similar and do not differ by more than $10\,\%$. Hence, the experimental results determined and presented in Section 5.4.1 contribute to an extended database for different commercially available membranes and can be transferred to different facilities, when using the same standardized experimental method.

5.4.4 Influence of uncertainty information on membrane selection and process analysis

In order investigate the effect of a variation in flux on the results of membrane screenings and the evaluation of process options for OSN, different literature studies
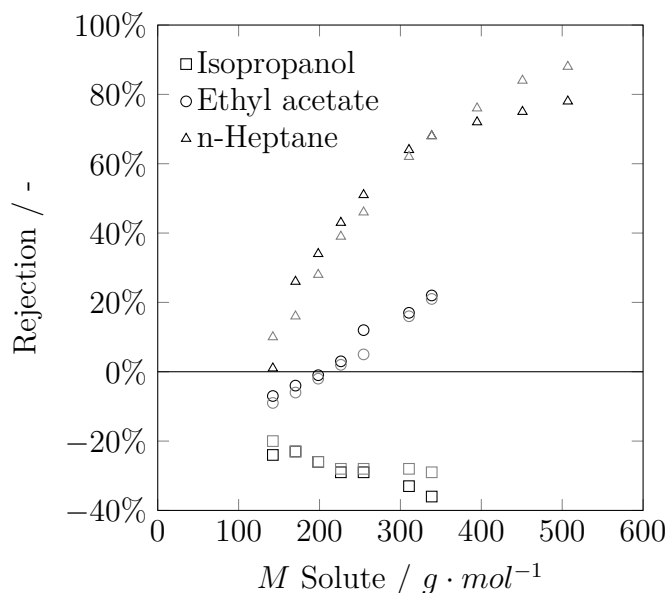
Figure 5.7: Rejections of linear alkanes in isopropanol, ethyl acetate and n-heptane using the PuraMem® S600 membrane measured in this study (black) and at RWTH Aachen (gray)

are discussed, reevaluating the therein reported data based on the knowledge gained in this study. For all studies, a variation in flux of $\pm 20\%$ was assumed, which already mandates a larger set of samples, as illustrated in the previous section. Any of the subsequent evaluations therefore represent a lower bound that would only increase for smaller sample sizes.

Membrane selection

Firman et al. [85] performed a membrane screening for the possible application of OSN for solvent recovery and removal of free fatty acids from an oil-hexane mixture. Flux and oil rejection were investigated for several tailor-made membranes and a commercially available membrane, performing lab-scale experiments with flat sheet membranes at different transmembrane pressures and temperatures. The active membrane area was $14.6\,\mathrm{cm^2}$, which is in the same range as the lab-scale experiments considered in the current study. Based on the results they concluded that the tailor-made poly(vinylideneflouride) membrane with an active layer of PDMS (PVDF-12SI), which was coated by a solution of hexane and 12% siloxane, is the best option. The flux through this membrane was $20.3\,\mathrm{L\,m^{-2}\,h^{-1}}$ at $20\,\mathrm{bar}$ and $30\,°\mathrm{C}$ and an oil rejection of 80% was determined for the oil/hexane mixture. The

PVDF-15SI membrane, which is manufactured in the same way but using a coating solution containing 15 % siloxane, provided a flux of $15.6\,\mathrm{L\,m^{-2}\,h^{-1}}$ and an even slightly higher rejection of (80.3 %). The authors further reported on standard deviations for the flux of both membranes (± 1.2 and $\pm 1.4\,\mathrm{L\,m^{-2}\,h^{-1}}$), based on three repeated experiments with constant samples of each membrane. Thus, these standard deviations do not reflect on the variance of different samples, but only the variance related to the experimental set-up and operation (intermediate depressurization, rinsing and cleaning), as well as the analytics. When assuming a variation of ± 20 % in flux for different samples, the PVDF-12SI can no longer be considered superior to the PVDF-15SI membrane with respect to the uncertainty related to the samples.

Process analysis

Micovic et al. [31] investigated a hybrid distillation-OSN process for the separation of a wide boiling mixture of n-decane, n-dodecanal and hexacosane. The optimized hybrid separation process was compared to a stand-alone distillation process based on costs, for which a cost factor of $120\,\mathrm{\text{\euro}\,m^{-2}\,a^{-1}}$ was considered for the OSN membrane. The estimated costs for the purification of one ton of product by the stand-alone distillation and the OSN-assisted process were determined as $5.87\,\mathrm{\text{\euro}\,a^{-1}}$ and $5.74\,\mathrm{\text{\euro}\,a^{-1}}$, respectively. Hence, a small, yet considerable cost benefit for an OSN-assisted process was determined. No variance information was considered in the design study. Considering a variation of ± 20 % in flux necessary membrane area for the same separation might either decrease for the higher flux, resulting in reduced costs of $5.68\,\mathrm{\text{\euro}\,a^{-1}}$, or they might increase and result in increased costs of $5.85\,\mathrm{\text{\euro}\,a^{-1}}$, becoming almost equal to the stand-alone distillation process.

In another study by Werth et al. [21, 22] the potential application for OSN membranes in solvent recovery and deacidification of non-edible and waste oils is investigated. After an initial membrane screening, appropriate flux models are identified and a comparative process design study is performed to determine the conditions under which an OSN-assisted evaporation process is competitive with a conventional two-stage evaporation process. Again, no information on the uncertainty with respect to variance of the membrane flux estimate was considered. A break-even price for a competitive OSN application of $140\,\mathrm{\text{\euro}\,m^{-2}}$ was determined. Given that the variance for the membrane rejection is negligible, the variance of the flux estimate

directly correlates with the determined break-even price, resulting in a potential need to decrease the price further to $116 \, €\, m^{-2}$.

Process design implications of flux and rejection variance

The variance analysis and the determination of the necessary sample size in Chapter 4 specifically focuses on pure solvent flux, in order to avoid any dependency on hydrodynamics. However, the characterization of the separation performance usually requires the evaluation of selectivity or rejection. While usually less pronounced, uncertainty related to rejection can have even more severe consequence on process design, as it was evaluated in two different Case Studies. For this, the experimental data and the uncertainties determined in Chapter 4 are used and a superstructure optimization is applied, which determines the optimal process configuration for a given separation task. The superstructure optimization was setup in GAMS and determines the permeate and retentate of each stage based on flux and rejection values but does not take any mass transport model into account. The process configuration is varied in order to minimize the necessary membrane area and the number of stages for the separation task. The superstructure optimization is described in detail in Chapter A in the Appendix and applied to two different Case Studies.

In Case Study 1 a concentration of one component with a molecular weight of $474 \, g \, mol^{-1}$ solved in ethanol is assessed, considering the experimentally determined performance metrics of the DuraMem$^®$ 300 membrane and the mixture of ethanol with styrene oligomers. The flux across the membrane at $30 \, bar$ was determined to $21.10 \, L \, m^{-2} \, h^{-1}$ while the confidence interval was $\pm 2.14 \, L \, m^{-2} \, h^{-1}$. The rejection for the styrene oligomer with a molecular weight of $474 \, g \, mol^{-1}$ was determined to $95.2 \, \% \, \pm 2.1 \, \%$.

For the optimization a feed flowrate of $100 \, kg \, h^{-1}$, a concentration from $0.05 \, g \, g^{-1}$ in the feed to $0.1 \, g \, g^{-1}$ in the final retentate and a solute recovery of at least $98 \, \%$ was assumed. Overall, nine different scenarios were evaluated, for which Table 5.5 provides in column 1 and 2 the constant values for flux and rejection, which were applied in each scenario. While in scenario 5 the expected values for flux and rejection obtained from the experiments are used, in the other scenarios all further combinations of the expected values or the minimum or maximum values of the respective confidence intervals of flux and rejection are used. In scenario 1, for example, the lower bounds of the confidence intervals for both values are used, and in

scenario 2 the minimum value of the confidence interval of the flux and the expected value for the rejection are used. Besides the values for flux and rejection, Table 5.5 additionally provides the performance metrics and the key structural results, in terms of number of required stages and overall membrane area. Figure 5.8 shows the optimized process structures for one of the case studies, for which an optimal process design was determined based on constant values for the estimated mean of flux and rejection as well as the upper and lower limits of the confidence intervals.

Table 5.5: Results for case study 1

Scenario	Flux $\mathrm{L\,m^{-2}\,h^{-1}}$	Rejection %	Nr. stages	Membrane area $\mathrm{m^2}$
1	18.96	93.1%	3	5.983
2	18.96	95.2%	2	5.157
3	18.96	97.3%	1	3.415
4	21.10	93.1%	2	5.467
5	21.10	95.2%	2	4.628
6	21.10	97.3%	1	3.068
7	23.24	93.1%	3	5.948
8	23.24	95.2%	2	4.973
9	23.24	97.3%	1	2.785

Based on the mean values of flux and rejection the optimized process needs two stages and a recycle stream of the retentate of the second stage to the feed stream of the first stage, while for the upper limits a single stage design without recycle is determined as optimal result and for the lower limits even a three-stage design with retentate recycles is required and determined as optimal design. The required membrane area varies by a considerable margin within these different designs from $2.8\,\mathrm{m^2}$ in the single stage design for the upper limits up to $5.9\,\mathrm{m^2}$ for the three-stage design for the lower limits, although the flux and rejection vary only by a relative margin of $\pm10\,\%$ and $\pm2.3\,\%$. The small variance results from a considerable number of measurements.

For the second case study a PDMS based membrane was used and an impurity was to be separated from a lower molecular weight molecule. The detailed results are presented in Chapter A.2 in the Appendix. However, the process structure again strongly depends on the flux and the rejection and different configurations were determined as optimized process structures ranging from a simple single stage set-up without recycle and a complex two stage set-up.
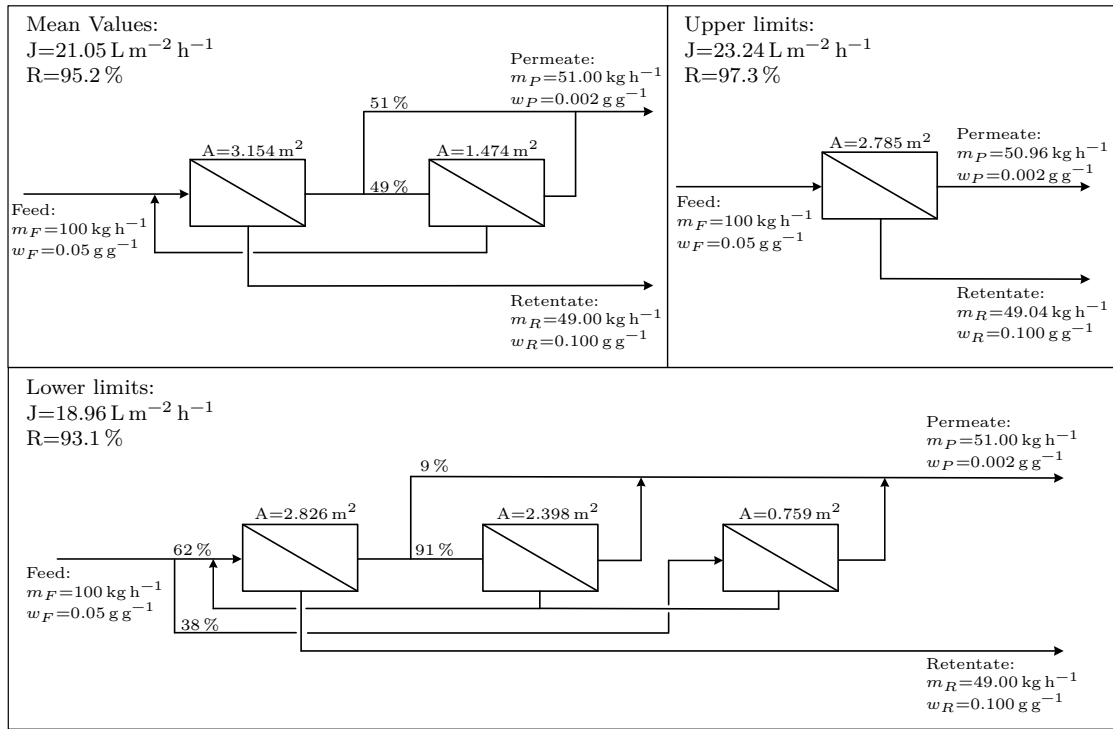
Figure 5.8: Process structure for three combinations in case study 1

For the considered case studies, even the small, experimentally determined confidence intervals for the rejection of $\pm 2.3\%$ resulted in variations of the optimal process configuration from a one to a three staged process. Therefore, further evaluation of the variance in rejection measurements should be conducted in future studies, considering that module geometry, hydrodynamic conditions and potential concentration polarization will affect the transferability of lab-scale measurements to the module scale.

5.5 Conclusion

In order to determine reliable flux and rejection of various components through OSN membranes a standardized method is reasonable for enabling the transfer of data to other applications and systems. Hence, the common experimental procedure has been revised in a cooperative study and was applied to different membrane-solvent-solute combinations for the purpose of creating a broad database on the one hand and for determining the comparability of the experimental results at different experimental facilities.

Therefore, the method of See Toh et al. [37] was adapted and extended in a cooperative study taking into account the suggestions of Marchetti et al. [39] and applied for the experiments presented in this chapter. Besides styrene oligomers, linear n-alkanes and further solutes to cover a wider range of different functional groups were used in the experiments performed in this study. Moreover, various solvents were used to cover a wide range of polarities as well as protic and aprotic solvents following the suggestions defined within the cooperative study [16]. Besides temperature and concentrations, Böcking et al. [16] proposed the measurement of different pressure levels in an increasing order as well as the cross-flow velocity in the membrane cells. Further, the experimental procedure is defined precisely starting with the implementation and rinsing of the fresh membrane sample and the definition of a minimum operation time. Experiments were performed applying this procedure at different laboratories using various lab-scale membrane plants. Following the experimental procedure, the results measured in the different facilities were comparable, but variations were still observed. The results of the collaborative study have been confirmed by a recent study of Merlet et al. [86]. However, these variations are within the range of the experimental error and the range of variations in flux resulting from fluctuations in the membrane material as determined in Chapter 4.

6

Uncertainties in performance modeling

In order to describe mass-transfer through organic solvent nanofiltration membranes, several models have been proposed and applied in different studies. Marchetti and Livingston [47] performed an extensive study on models for OSN and compared these models based on the model accuracy and complexity. However, the precision and reliability of model parameters has not been taken into account, although it is an essential aspect for the long-term objective of predictive performance models. In this chapter, the study of Marchetti and Livingston [47] is extended to include the aspect of parameter precision and the corresponding influence on the calculated results using different OSN models. Therefore, the results of parameter estimation are reproduced based on the available experimental permeation data and the models are evaluated based on an information criterion. Further, parameter uncertainty is investigated based on confidence intervals and an identifiability analysis, while the effect of parameter uncertainty on the model results is investigated in a subsequently performed Monte-Carlo simulation. The results confirm a significant influence of parameter precision on the reliability of model results and illustrate the inability to effectively discriminate the models based on the available experimental data. Consequently, consideration of the parameter precision should be considered crucial for model selection.

6.1 Introduction

To develop membrane processes for industrial applications, reliable models for different purposes are essential. Besides models for the mass-transfer through the membrane, the description of fluid dynamics in the membrane module and models for process scale to allow for an evaluation of the performance of the desired membrane process and thus an economical comparison with conventional processes [17, 87] are needed. In the scope of this chapter, models on the membrane scale are considered to describe mass-transfer of both solvent and solute depending on the operating conditions. Since mass-transfer in OSN is influenced by numerous interactions between solvent, solute and membrane, the transfer mechanism has not yet been identified completely and modeling is accordingly difficult [11, 17]. Thus, different approaches to model OSN exist which differ in their underlying transport mechanism and complexity. The approaches base on a physical understanding of the separation process simplified by idealized assumptions [17]. Several influencing factors are combined to parameter groups which can be determined by fitting to experimental permeation data [17]. This offers the advantage of a reduced number of experiments, which are necessary to determine the parameter values.

Among the semi-empirical models are the well-known solution-diffusion model (Section 2.2.2) and the pore-flow models (Section 2.2.1), that consider diffusional and convective transport through the membrane, respectively. A more complex variant of the solution-diffusion model is based on Maxwell-Stefan diffusion and takes frictional and convective coupling between solvent, solute and membrane into account. A combination of solution-diffusion and pore-flow models is incorporated in the solution-diffusion with imperfections model where the membrane is considered as a dense matrix with free volume elements called imperfections. The third group of OSN models originates from the thermodynamics of irreversible processes (Section 2.2.3). The membrane is described as a black box and thus membrane properties are not considered.

These models were investigated and compared in a previous work of Marchetti and Livingston [47] in order to determine the necessary degree of complexity of the models to describe experimental data for solvent and solute flux or rejection with a certain accuracy. For this, the model parameters were estimated by regression on a limited basis of experimental data, which were acquired using the common experimental procedure (Section 2.1.3). The evaluation of the models was con-

ducted regarding the accuracy of parameter estimation and complexity of the models. Marchetti and Livingston [47] used an information criterion which takes both the goodness of fit as well as the number of parameters of each model into account. The aim of this study was not to evaluate the physical correctness of the different models and the underlying theoretical considerations, but the ability of the models to represent experimental data.

However, parameter precision and its effect on model results was not considered by Marchetti and Livingston [47]. Model accuracy describes how well the experimental data is fitted by the model predictions, whereas parameter precision refers to the reciprocal of the variance of the parameter estimates [88]. For the design of new processes, it is crucial to predict the separation characteristics of a membrane for different process conditions in order to reduce the number of experiments. Thus, model parameters should be determined as precisely as possible and the uncertainty of the models should be evaluated before applying the models for process predictions. The objective of the current study is the extension of the work of Marchetti and Livingston [47] by an uncertainty analysis of the model parameters. Based on the experimental data used in the study of Marchetti and Livingston [47], the results for the parameter regression were reproduced since the parameter values were not published within the study. Subsequently, an information criterion as well as the Akaike weights [89] were used to evaluate the accuracy as well as the complexity of each model to determine the model with the highest probability of being the most suitable one for the description of present experimental data among the investigated models. The uncertainty of the model parameters was investigated based on confidence intervals of each model parameter and an identifiability analysis. Finally, a Monte-Carlo simulation is performed, in which the model parameters are varied within their confidence intervals in order to investigate the distinguishability of the different model approaches.

The different available mass-transfer models for OSN have already been described in Chapter 2. The different applied methods for parameter regression, evaluation of the information criterion and the Akaike weights as well as the methods for evaluation of the parameter uncertainty, the identifiability analysis and the Monte-Carlo simulation are presented in Section 6.2. Section 6.3 presents a discussion of the results of the applied methods followed by a conclusion in Section 6.4.

6.2 Methods

This section first introduces the methods used for parameter estimation for the OSN models described in Section 2.2. Subsequently, methods for model discrimination are introduced. Besides the information criterion already applied by Marchetti and Livingston [47], the parameter variance and identifiability were evaluated. Finally, the method used for the investigation of the influence of the model uncertainties is described.

6.2.1 Experimental data

Experimental data were taken from the publication of Marchetti and Livingston [47] and Postel et al. [66]. The flow and rejection from different membrane-solvent-solute systems are listed in both studies. Marchetti and Livingston [47] investigated the solvent flux and the rejection of different components by using four different commercially available polymeric membranes at four different pressure levels. They investigated the two polar membranes DuraMem® 200 and DuraMem® 500 in acetone and acetonitrile (ACN) as well as the two apolar membranes PuraMem® 280 and PuraMem® S600 in toluene using a mixture of styrene oligomers as solutes. Since the quantity of each oligomer solved in the solvent was only reported for α-methyl styrene, this solute is used as model substance in this study.

Postel et al. [66] investigated a non-commercial polymeric membrane with isopropanol, methanol, n-hexane and toluene as solvents and different linear alkanes as solutes. The results measured with the solute n-tetradecane listed by Postel et al. [66] were used to extend the study to a non-commercial membrane and to negative rejections. However, Postel et al. [66] performed rejection experiments at two pressure levels only resulting in a lower number of experimental data points compared to the data sets of Marchetti and Livingston [47].

For all data sets, further solutes present in the experimental investigations were neglected for the determination of all compositions, fluxes and rejections. Consequently, the systems were treated as binary systems composed of solvent and the specific solute. For the application of the models, permeate compositions and component-specific fluxes were calculated based on the reported experimental data for feed composition, rejection and flux from Marchetti and Livingston [47] and Postel et al. [66].

6.2.2 Parameter estimation

The models described in Section 2.2.1 to 2.2.3 were implemented in Matlab® and the nonlinear fitting function *nlinfit* was used for parameter regression. Besides the estimated values for the model parameters, the residuals between the model calculations and the experimental data, the Jacobian, the covariance matrix and the mean squared error of the obtained solution were determined based on a set of the experimentally determined fluxes and rejections at different pressure levels and initial values for the model parameters. The number of experimental data used for regression was determined by the limited number of available experimental data given by Marchetti and Livingston [47] and Postel et al. [66]. The former provided four pairs of values for solvent flux and rejection (at 10, 20, 30 and 40 bar), resulting in eight data points in total. The latter supplied two pairs of solvent flux and rejection (at 20 and 30 bar) which adds up to four data points. For parameter regression, the number of data points needs to be higher than the number of parameters. In this study, parameter regression was performed using all available data points. Nevertheless, it was not possible to fit the Maxwell-Stefan model to the data provided by Postel et al. [66] since the number of data points was too low.

For parameter regression, the rejections of the solutes were used instead of the solute fluxes since these are the commonly published values. In case the model itself does not include a specific equation to calculate the rejection, the rejection was determined based on the component specific fluxes. Since all values were determined in SI units, the values for solvent flux and rejection differ in several orders of magnitude. Hence, to avoid a stronger effect of the rejection compared to that of the smaller total flux values, the latter ones were scaled by multiplying with specific factors, which were set to 10^4 for the data of Marchetti and Livingston [47] and by 10^5 for the data of Postel et al. [66]. The objective function was provided in Equation 6.1 and includes the scaled total fluxes $J^*_{s,exp}$ and $J^*_{s,calc}$. The calculated values $J^*_{s,calc}$ and $R_{j,calc}$ depend on the model parameters $\vec{\theta}$ and other, experimental values y_i like pressure and concentrations.

$$obj(\vec{\theta}) = \sum_{i=1}^{N_{obs}} [J^*_{s,exp} - J^*_{s,calc}(\vec{\theta}, y_i)]^2 + \sum_{i=1}^{N_{obs}} [R_{j,exp} - R_{j,calc}(\vec{\theta}, y_i)]^2 \qquad (6.1)$$

For models, which are linear in parameters, the initial values for the model parameters were determined based on the order of magnitude of all values taken into account by the respective model equation. This method was used for the classical

(cf. Equations 2.16 and 2.17) and simple solution-diffusion model (cf. Equations 2.19 and 2.20) as well as for the solution-diffusion with imperfections model (cf. Equation 2.30 and 2.31). For the pore radius r_p in the Donnan Steric pore-flow model (cf. Equations 2.6 to 2.15) an initial value of $1\,\text{nm}$ was chosen, since OSN performs a separation on the molecular level. Moreover, the rejection calculated by the Donnan Steric pore-flow model was independent of the flux and thus, the two model parameters r_p and K_{HP} have been determined independently. Hence, the model parameter K_{HP} for the description of the flux (cf. Equation 2.4) was determined separately by using the function *mldivide*, which is applicable for linear equations. Moreover, both parameters were scaled by the factor 10^9 according to Equations 6.2 in order to avoid small values, which were determined by Matlab®.

$$r_p^* = 10^9 \cdot r_p; K_{HP}^* = 10^9 \cdot K_{HP} \qquad (6.2)$$

The initial values for the irreversible thermodynamics model (cf. Equations 2.32, 2.34 and 2.35) were varied several times for the regressions for the DuraMem® 200 membrane using acetone and α-methyl styrene. After reasonable results for flux and rejection have been achieved, the same initial values were used for different systems and were only changed if no reasonable results were achieved. Whereas for most of the models the same initial values for different membrane-solvent-solute systems yielded in reasonable parameter estimates, this was not the case for the Maxwell-Stefan model (cf. Equations 2.23 to 2.29). Hence, for the Maxwell-Stefan model the initial parameter values were varied several times for each membrane-solvent-solute system to achieve positive parameter values if possible.

In some specific cases, using the permeate compositions complying with the experimental results for parameter regression did not result in reasonable rejections. In these cases, a parabolic course was calculated for the rejection at different transmembrane pressures, including a maximum at mean transmembrane pressure. This does not comply with the experimental observations, where a constant increase of rejection with increasing transmembrane pressure was observed. Parabolic courses were calculated for the combination of linear models with only one parameter each for solvent and solute flux, like the simple solution-diffusion model, the classical solution-diffusion model and systems with low flux, as it is observed with the PuraMem® S600 in acetone and ACN. Therefore, these models were implicitly fitted, whereby the permeate composition was calculated. For this, the model was implemented in

GAMS and the parameters were determined with the CONOPT solver. The variable limits were chosen so that the parameters of the models were positive and the other variables, such as flux, rejection and composition, were within reasonable limits.

6.2.3 Information criterion for model comparison

The accuracy of the models was evaluated calculating the norm of the residuals ($resnorm$) for each pair of calculated $y_{i,calc}$ and measured value $y_{i,exp}$. The average of the experimental values $\bar{y}_{exp}$ was used to normalize the error.

$$resnorm = \sqrt{\sum_{i=1}^{N_{obs}} \left(\frac{y_{i,exp} - y_{i,calc}}{\bar{y}_{exp}} \right)^2} \tag{6.3}$$

Since *resnorm* only considers the accuracy of the model, but not the complexity, it was not suitable for a comparison of different models. Thus, the Akaike Information Criterion (AIC) was used, which decreases with increasing accuracy of a model. Marchetti and Livingston [47] use the following definition:

$$AIC = N_{obs} \cdot log \left(\frac{resnorm^2}{N_{obs}} \right) + 2 \cdot N_{param} \tag{6.4}$$

N_{param} denotes the number of parameters in the considered model. Lastly, the method of model discrimination based on Akaike weights was used. To compare the accuracy of N_{models} possible model candidates, the Akaike weights W (Equation 6.5) have been used to identify the model which is most likely the best one for the present experimental data among the investigated ones [89]. For this purpose, the difference of the AIC values ΔAIC_{model} of the currently evaluated model candidate and the lowest calculated AIC value of all investigated models was evaluated. Normalization with the sum of values calculated for all models leaded to the probability of the considered model to be the right one. Thus, the sum of all weights equals 1, whereas each weight represents the probability of the corresponding model to be the best one.

$$W_{model} = \frac{exp(-0.5 \cdot \Delta AIC_{model})}{\sum_{i=1}^{N_{models}} exp(-0.5 \cdot \Delta AIC_i)} \tag{6.5}$$

6.2.4 Analysis of Variance

An aspect, which was not evaluated by the information criterion, is the precision of the obtained parameters and the resulting model calculations. Whereas accuracy

refs to the distance between corresponding calculated and the experimental values, precision relates to the uncertainty of the model calculations [88], as illustrated in Figure 6.1.

After obtaining an estimation for a parameter by regression, the confidence intervals were calculated using the Matlab® function *nlparci*. This function calculates the confidence intervals of each parameter in a specific model using the estimated parameter values, the residuals of the regression and Jacobian matrix. For smaller levels of significance and for higher degree of freedom (DOF), the confidence intervals become narrower. If a confidence interval contains zero, the parameter is called non-significant [90]. This may be the case for very high levels of significance or very small DOFs. The latter can be caused either by few experimental values or by a large number of parameters in the respective model.
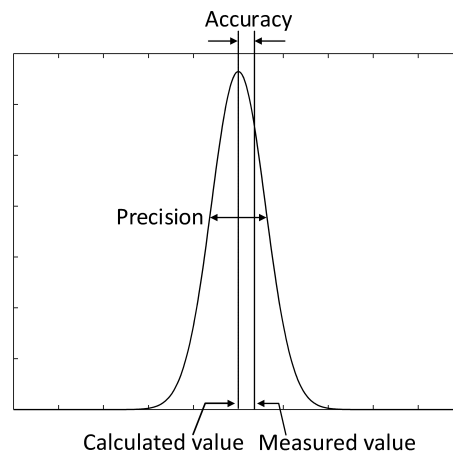
Figure 6.1: Distinction between accuracy and precision, adapted from [88], the line represents the probability distribution density of the calculated value

6.2.5 Identifiability analysis

Regarding the aspect of reliability of parameter estimations, an analysis of the parameter identifiability was performed. Thereby, the possibility of determining a parameter uniquely from a set of experimental data by regression was investigated [91]. In this context, an estimated set of parameters having no other parameter sets which yield the same results for the considered model, is considered to be unique [91, 92].

For the evaluation of the model's sensitivity towards the respective parameters, the eigenvalue method was applied. Therefore, the Hessian matrix of the objective func-

tion (Equation 6.1) was calculated [92]. In Gaussian approximation, Hessian matrix can be calculated from the Jacobian. By analyzing definiteness of the Hessian matrix, conclusions of the uniqueness of the model parameters are possible. If the identified model parameters correspond to a locally unique solution, the resulting Hessian matrix is positive definite. The definiteness of the Hessian matrix can be determined based on the eigenvalues of the Hessian matrix. For a positive definite matrix all eigenvalues are positive and hence, the model parameters are identifiable. However, Quaiser and Mönnigmann [92] state that in application of the eigenvalue method, all eigenvalues could be positive, but some are close to zero. If the eigenvalues differ in several orders of magnitude, a cut-off value can be defined under which an eigenvalue is effectively treated as zero [92]. The cut-off value was set to 10^{-10} in this study.

In order to determine the identifiability of the parameters in this study the following procedure was applied:

1) Calculate the Hessian matrix based on the Jacobian.

2) Determine eigenvalues and eigenvectors.

3) Sort eigenvalues by size: $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_{N_{param}}$

4) If λ_1 is greater than zero respectively the cut-off value: for this data set the model is at-a-point locally identifiable with respect to the parameters.

5) If λ_1 is smaller than zero respectively the cut-off value: Determine the position i of the greatest element of the corresponding eigenvector based on the absolute values of the elements: $u_{1,i} = max(|u_{1,1}|, |u_{1,2}|, ..., |u_{1,N_{param}}|)$

6) Determine the least identifiable parameter θ_i: The position of the last identifiable parameter in the vector of model parameters $\vec{\theta}$ is equal to the position of the greatest element in the eigenvector determined in the precious step.

6.2.6 Monte-Carlo simulation

The analysis of parameter uncertainty and identifiability leads to the question of their effects on the model results. Uncertainties of the parameters can either accumulate to even higher uncertainties of the model results or cancel each other out, leading to model results which seem to be less uncertain. Therefore, it was not

sufficient to calculate the model values at the lower and upper bounds of the parameters' confidence intervals. To evaluate these effects, a Monte-Carlo simulation was performed which is based on an algorithm using random numbers. The general method of Monte-Carlo simulation was primarily developed by Metropolis et al. [93] and poses an efficient method to evaluate large numerical problems [94]. Parameter values were sampled randomly from their respective confidence intervals and the resulting values of the considered model were evaluated. This process was repeated frequently to provide a high number of different combinations of randomly picked parameter values. Thus, deviations from the modeled values using the parameter estimates have been obtained which reflect the model uncertainty.

For this, random sampling of parameter values out of a uniform parameter distribution using the Matlab® function *makedist* and the limits of the confidence intervals of each parameter was performed 3000 times. The corresponding model was evaluated based on each of these 3000 parameter combinations. For the evaluation of the model uncertainty, the maximal lower and upper deviations of the solvent fluxes and rejections were determined. In order to discriminate the considered models for OSN, the respective uncertainty regions were compared for each membrane-solvent-solute system. In the case of overlaps between the uncertainty regions of different models, it was not possible to state which one of them provides the better regression results.

6.3 Results for model discrimination and discussion

The results of the different steps performed in order to discriminate between the different models, are described in the following Sections. For this, the focus was on the results for the four membranes DuraMem® 200, DuraMem® 500 used in acetone supplemented with α-methyl styrene as the solute as a representative example. Since the results of the parameter regression for the PuraMem® 280 and PuraMem® S600 membranes in the same solvent-solute mixture are significantly different to those of the DuraMem® membranes, these results are added to the discussion in Section 6.3.1. The results of the remaining systems are listed in Chapter C in the Appendix. First, the results of the parameter estimation reproducing the results of Marchetti and Livingston [47] are shortly presented in Section 6.3.1. Subsequently, the results of the analysis of variance and identifiability as well as for the Monte-Carlo simulation are discussed based on the exemplary systems in Section 6.3.2 and

Section 6.3.3. Finally, a summary of the results for all membrane-solvent-solute systems evaluated in this study is given in Section 6.3.4.

6.3.1 Model comparison based on Akaike information criterion

In Figure 6.2, the symbols represent the experimentally measured fluxes and rejections using the commercially available membranes DuraMem® 200, DuraMem® 500, PuraMem® 280 and PuraMem® S600 in acetone. The results obtained with the models are represented by the different lines connecting the calculated fluxes and rejections at the different pressure levels for a better overview.

For the two DuraMem® membranes (cf. Figure 6.2), all models are in good agreement with the experimental data, especially for the flux. Moreover, no significant difference in the accuracy of the models become apparent for the DuraMem® 200-acetone-α-methyl styrene system. This is also valid for the DuraMem® 500 membrane except that the fluxes calculated using the Maxwell-Stefan model differs from the results of the remaining models but are equally accurate. The difference between calculated and experimental results are in a range of $\pm 20\,\%$ of the experimental values, which was found to be a reasonable size of the experimental error in Chapter 4. Furthermore, the results obtained for both DuraMem® membranes are comparable to those obtained by Marchetti and Livingston [47].

The results for the PuraMem® 280 in acetone with α-methyl styrene as solutes qualitatively barely differ from those of the DuraMem® membranes. To the contrary, there are clear deviations between calculated and experimentally determined values for the rejection in the PuraMem® S600. Especially the simple solution-diffusion model and the classical solution-diffusion model differ significantly from the experimental data. The simple solution-diffusion model describes a linear relationship between the difference of the mole fractions in feed and permeate and the solute flux (cf. Equation 2.20). However, the experimental data rather show an exponential course as presented in Figure 6.3.

Since the experimentally determined permeate concentrations were used to adapt the model parameters, the simple solution-diffusion model is not capable of reproducing the experimental data due to the model structure. The same applies to the classical solution-diffusion model, which contains an exponential term in addition to the molar components in feed and permeate (cf. Equation 2.17). Although the
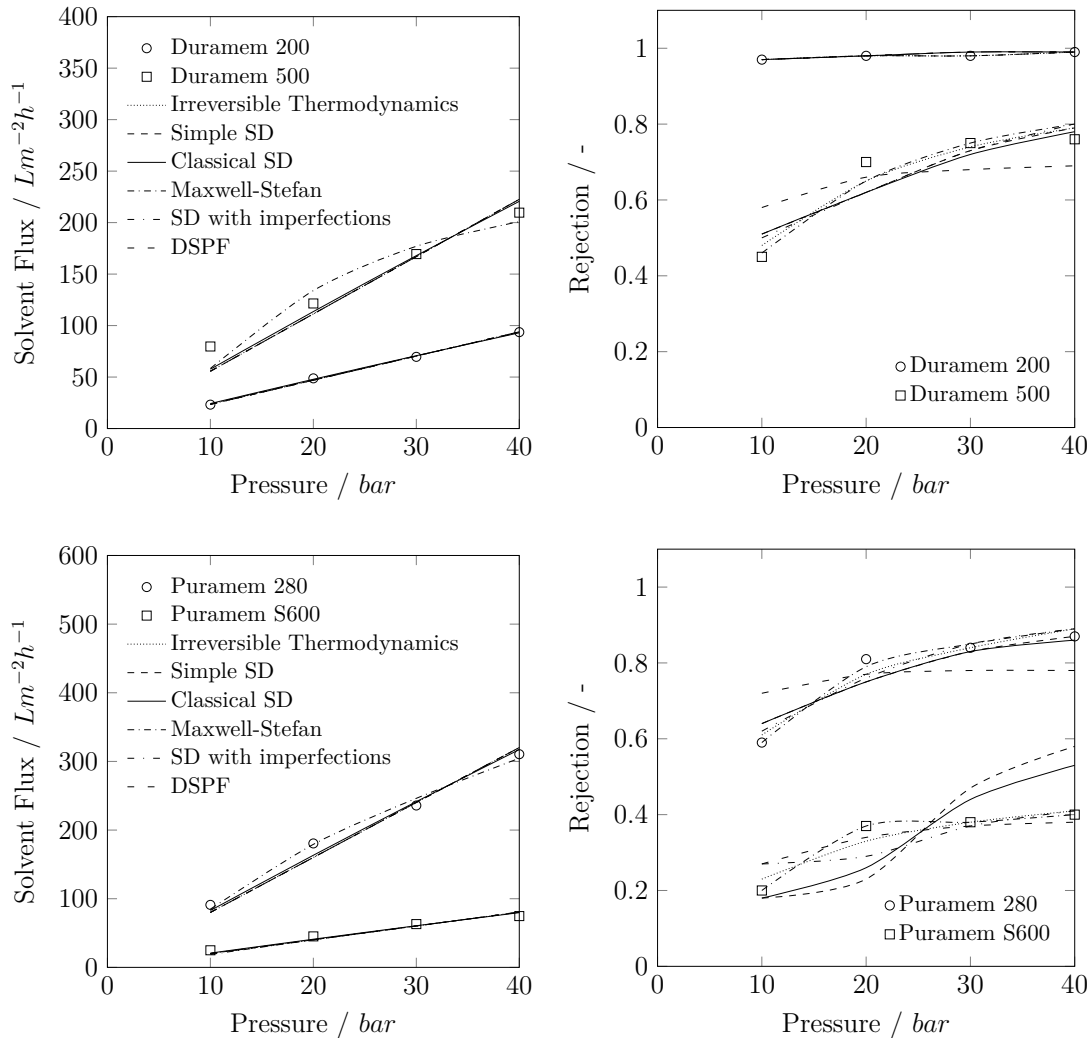
Figure 6.2: Experimental measured flux and rejection (provided by Marchetti and Livingston [47]) for the solute α-methyl styrene in acetone through DuraMem® (top) and PuraMem® (bottom) membranes (symbols) and calculated values for irreversible thermodynamics model, simple solution-diffusion (SD) model, classical solution-diffusion model, Maxwell-Stefan model, solution-diffusion with imperfections model and Donnan Steric pore-flow model (DSPF) connected by different lines
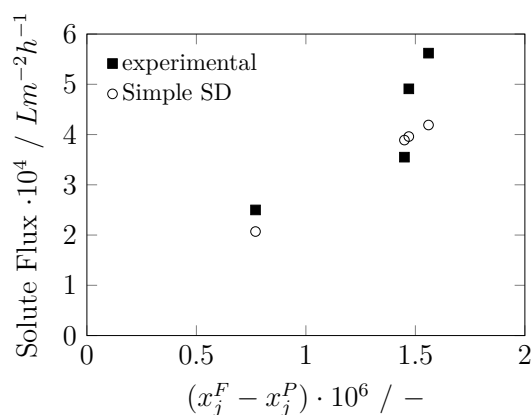
Figure 6.3: Solute flux plotted over difference in molar fraction of the solute between feed and permeate for experimental data provided by Marchetti and Livingston [47] (filled symbols) and simple solution-diffusion (SD) model (unfilled symbols) for PuraMem® S600 in acetone with α-methyl styrene as solute

exponential part reduces the deviations, even in this case the model is not able to describe the experimental data for rejection if the experimentally determined mole fractions of the solute in the permeate are used.

Therefore, both models were additionally fitted to the experimental data by solving an implicit system of equations and calculating the permeate concentration using GAMS as described in Section 6.2.2. The results shown in Figure 6.4 prove that by using this approach the simple solution-diffusion model and the classical solution-diffusion model represent the experimental data with similar accuracy as the other models. However, the permeate concentrations deviate from the experimentally determined concentrations, which would lead, for example, to an inaccurate design of a subsequent process step or the membrane separation step itself. Hence, this method of parameter regression is not meaningful and the experimental data should rather be critically questioned, since the measured rejection for α-methyl styrene stagnates between 20 and 40 bar, which is not to be expected.

The Akaike Information Criterion (AIC) and the Akaike weights (W) are listed in Table 6.1 for the DuraMem® membranes and in Table 6.2 for the PuraMem® membranes.

Since all models describe the experimental data for the DuraMem® 200 membrane with a similarly high accuracy, no model stands out with a particularly high probability of being the right one among the models examined. For the DuraMem® 500-
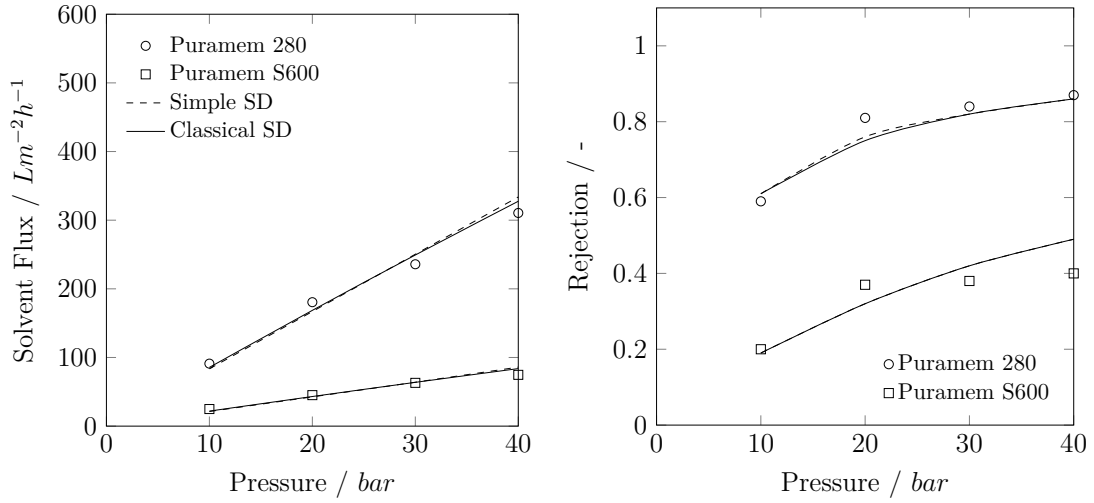
Figure 6.4: Experimental measured flux and rejection (provided by Marchetti and Livingston [47]) for the solute α-methyl styrene in acetone through PuraMem® S600 (symbols) and calculated values for simple solution-diffusion (SD) model and classical solution-diffusion model connected by different lines (implicit calculation)

Table 6.1: Resnorm, AIC and W of the simple Solution-diffusion (SSD) model, the solution-diffusion with imperfections (SDIM) model, the classical solution-diffusion (CSD) model, the Donnan Steric pore-flow (DSPF) model, the irreversible thermodynamics (IT) model and the Maxwell-Stefan (MS) model for DuraMem® 200 and DuraMem® 500 used in acetone and α-methyl styrene as solute

	DuraMem® 200			DuraMem® 500		
Model	Resnorm	AIC	W	Resnorm	AIC	W
SSD	0.02	-74.07	6.5%	0.34	-29.73	4.3%
SDIM	0.02	-76.98	27.8%	0.34	-27.80	1.6%
CSD	0.02	-74.63	8.6%	0.34	-27.80	1.6%
DSPF	0.02	-76.76	24.9%	0.52	-23.15	0.2%
IT	0.02	-76.96	27.5%	0.21	-35.45	75.5%
MS	0.02	-73.40	4.6%	0.21	-31.57	10.8%

Table 6.2: Resnorm, AIC and W of the simple Solution-diffusion (SSD) model, the solution-diffusion with imperfections (SDIM) model, the classical solution-diffusion (CSD) model, the Donnan Steric pore-flow (DSPF) model, the irreversible thermodynamics (IT) model and the Maxwell-Stefan (MS) model for PuraMem$^®$ 280 and PuraMem$^®$ S600 used in Acetone and with α-methyl styrene as solute

Model	PuraMem$^®$ 280			PuraMem$^®$ S600		
	Resnorm	AIC	W	Resnorm	AIC	W
SSD	0.19	−39.45	3.5%	1.52	−5.91	0.0%
SDIM	0.16	−39.47	3.5%	0.66	−17.21	0.0%
CSD	0.22	−37.12	3.1%	7.68	19.98	0.0%
DSPF	0.44	−25.85	0.0%	0.47	−24.67	0.0%
IT	0.12	−44.85	51.6%	0.32	−28.82	0.0%
MS	0.10	−44.23	37.8%	0.01	−83.89	100.0%

acetone-α-methyl styrene and the PuraMem$^®$ 280-acetone-α-methyl styrene systems the irreversible thermodynamics model (Equations 2.32, 2.34 and 2.35) is indicated to be the best model based on AIC and W. The Maxwell-Stefan model is evaluated as the most suitable model to describe the experimental data for the PuraMem$^®$ S600-acetone-α-methyl styrene system. These results differ from those obtained by Marchetti and Livingston [47], since the calculated norm of residuals already differ in both studies. Even small differences in the quality of the fit have a significant influence on the norm of residuals and thus on AIC and W. Hence, the difference in the indication of the best model can be attributed to even small differences in the approach to parameter regression. However, since the results for flux and rejection calculated using the different models are comparable are comparable to those of Marchetti and Livingston [47], the results of the parameter regression are a reasonable basis for the investigation of parameter accuracy and identifiability. That also holds for all the remaining systems listed in the Appendix.

6.3.2 Parameter variance and identifiability

Parameters determined by fitting the model to a limited set of experimental data are only estimates. However, the model may be able to describe these data points very accurately whereas the parameter values are not precisely identifiable. For a different set of parameter values, the model fit might still be as accurate. Thus,

the parameters cannot be determined uniquely. To evaluate the parameter precision 95 %-confidence intervals are determined. 95 % of the confidence intervals determined in this fashion will include the real value of the parameter. A narrow confidence interval corresponds to a high parameter precision, while a broad interval implies a high uncertainty and the model might not be accurate in predictions. Therefore, models with precisely estimated parameter values are to be favored for predictions.

Table 6.3 lists all parameter values for the different models applied for the data sets of Marchetti and Livingston [47] for the membranes DuraMem® 200 and DuraMem® 500 used in acetone with the solute α-methyl styrene. Confidence intervals marked with a star in the table are broader than ± 100 % of the parameter value. This applies to all five parameters of the Maxwell-Stefan model for both DuraMem® membranes in acetone. Hence, it is not possible to determine the model parameters of the Maxwell-Stefan model without a high uncertainty for the limited data set provided by Marchetti and Livingston [47]. Even though this data set already included flux and rejection at four different pressure levels. This does not imply that the Maxwell-Stefan model is physically incorrect, but rather that the respective model parameters cannot be determined from the available data. Using a higher number of data points might contribute to the parameter precision. While it is more likely that different dedicated measurements for diffusion or sorption may enable a more accurate estimation of these model parameters, as e.g. done by Hesse et al. [15].

The relative width of confidence intervals for the less complex solution-diffusion models (SSD and CSD) are narrower than ± 30 %. The same applies for the Donnan Steric pore-flow (DSPF) model. Hence, the parameters of these three models are considered as appropriately determined based on the data published by Marchetti and Livingston [47].

For the three model parameters of the solution-diffusion with imperfections model the broadness of the confidence intervals differs for both membranes in acetone. As shown in Table 6.3, the confidence interval for the first model parameter, which is the main parameter to describe the solvent flux, is narrow compared to the others. The second parameter describes not only the solvent flux but also the solute flux (cf. Equations 2.30 and 2.31). Considering the equation for the flux, the only difference between both summands is the osmotic pressure. Hence, the second model parameter L_{IM} does not have a significant influence on the flux since the effect of

transmembrane pressure and osmotic pressure are already accounted by the first term of the solvent flux equation. The same applies to the solute flux, where the first summand of the equation accounts for the influence of concentration difference and includes the third model parameter P_m, whereas the second summands including the parameter L_{IM} does not have a significant effect on the solute flux. Thus, the second parameter L_{IM} is almost zero and the confidence interval is large, since the effect of the parameter on the model results is not significant.

For the irreversible thermodynamics model, the confidence interval of the model parameter L_j is very broad in most cases. Due to the structure of the equation (cf. Equations 2.32 and 2.33) for the rejection, the effect of small changes of the parameter value on the accuracy of the model is very small. Hence, changing the parameter value in a broad range does not show significant effects on the model accuracy and the confidence intervals become broad.

Based on the evaluation of the confidence intervals, the parameters of the simple and classical solution-diffusion model and the Donnan Steric pore-flow model can be determined most accurately. The significance of the model parameters was investigated by the evaluation of the eigenvalues of the Hessian matrix, additionally. Table 6.4 lists the eigenvalues for all models for the two DuraMem® membranes in acetone. The listed values for the eigenvalues comply with the results for the confidence intervals. The majority of the eigenvalues for the Maxwell-Stefan model are very small or even negative. Based on the corresponding eigenvector of the smallest eigenvalue, the parameter b_1 is the least identifiable parameter for the system DuraMem® 200-acetone-α-methyl styrene, whereas K_{MS} is the least identifiable parameter for the system DuraMem® 500-acetone-α-methyl styrene. Since these parameters include different properties of the solvent and solutes, such as the diffusion coefficients in the membranes, further experiments might contribute to the identifiability of these parameters and increase the certainty of the model as already stated above. For the other models, the eigenvalues are larger. Hence, for the simple and classical solution-diffusion model, the solution-diffusion with imperfections model, the Donnan Steric pore-flow model and the irreversible thermodynamics model a model reduction is not reasonable. Since for some models broad confidence intervals where determined, the effect of this on the model results is investigated in a Monte-Carlo simulation in order to visualize the effects.

Table 6.3: Parameter values and confidence intervals for DuraMem® 200 and DuraMem® 500 used in acetone with α-methyl styrene as solute using the different models (simple Solution-diffusion (SSD) model, the solution-diffusion with imperfections (SDIM) model, the classical solution-diffusion (CSD) model, the Donnan Steric pore-flow (DSPF) model, the irreversible thermodynamics (IT) model and the Maxwell-Stefan (MS))

Model	Parameter	DuraMem® 200		DuraMem® 500	
		value	CI	value	CI
SSD	P_s^{SSD}	6.52e−8	2.17e−9	1.55e−7	2.55e−8
	P_j^{SSD}	7.39e−7	2.15e−7	5.42e−5	1.39e−5
SDIM	L_m	6.52e−8	2.00e−9	1.55e−7	2.93e−8
	L_{IM}	1.54e−13	2.37e−13	−8.89e−13	1.31e−11
	P_m	5.06e−7	4.09e−7	5.68e−5	4.17e−5
CSD	P_s^{CSD}	2.31	7.42e−2	2.31	7.42e−2
	P_j^{CSD}	7.59e−7	2.12e−7	7.59e−7	2.12e−7
DSPF	$K_{HP} \cdot 10^9$	1.96e−2	3.99e−4	4.64e−2	9.08e−4
	$r_p \cdot 10^9$	3.32e−1	3.53e−3	6.59e−1	1.15e−1
IT	L_p	6.52e−8	2.00e−9	1.55e−7	2.25e−8
	σ	9.89e−1	1.35e−2	1.00	5.92e−1
	L_j	1.68e−7	1.08e−7	1.67e−5	2.05e−5
MS	b_1	4.33e3	1.77e4	1.20e2	3.28e3
	b_2	−4.08e−4	1.22e2	5.33e−7	1.34e−5
	b_3	1.86e3	4.33e3	2.26e2	1.18e4
	K_s	8.74e−1	2.83	2.42	6.84e1
	$K_{MS,j}$	−8.76e−6	2.62	−1.06e1	3.19e2

6.3.3 Effect of parameter uncertainty

In a last step, the effect of the previously determined parameter uncertainty accounted for by the confidence intervals is investigated in a Monte-Carlo simulation by varying the parameter values within their confidence intervals and calculate the model results based on these new parameter values. Although a broad confidence interval of some parameters was determined, these do not necessarily result in broad variation in the model results for flux or rejection. On the contrary, a narrow confidence interval does not compulsorily result in only small variations in flux and rejection, since the variations of the parameters may cancel each other out or multiply.

Table 6.4: Eigenvalues for all models and the membranes DuraMem® 200 and DuraMem® 500 used in acetone with α-methyl styrene as solute (simple Solution-diffusion (SSD) model, the solution-diffusion with imperfections (SDIM) model, the classical solution-diffusion (CSD) model, the Donnan Steric pore-flow (DSPF) model, the irreversible thermodynamics (IT) model and the Maxwell-Stefan (MS))

Model	EV Nr.	DuraMem® 200	DuraMem® 500
SSD	1	3.06e9	1.01e8
	2	3.04e13	5.11e13
SDIM	1	7.19e8	1.48e7
	2	3.00e13	3.21e13
	3	9.19e21	1.63e21
CSD	1	2.38e−2	2.38e−2
	2	2.95e9	1.97e8
DSPF	1	2.70e6	2.70e6
	2	2.15e1	7.60
IT	1	6.66e−1	5.28e−2
	2	2.81e10	5.15e8
	3	3.02e13	3.76e13
MS	1	8.79e−13	−8.41e−9
	2	1.47e−11	3.79e−6
	3	5.24e−8	1.53e−4
	4	3.95e−2	2.59e1
	5	2.28e7	1.80e12

The uncertainty regions determined in the Monte-Carlo simulations for the two DuraMem® membranes in acetone with α-methyl styrene as solute are shown in Figure 6.5 as error bars. The origin model results which were calculated using the estimated parameters are shown as bars. Moreover, the experimental results of flux and rejection are added to the diagrams. The variation of the parameter values within their confidence intervals may result in unsymmetrical uncertainty regions. For the Maxwell-Stefan model, no uncertainty regions are shown since these were extremely broad for all investigated membrane-solvent-solute systems. The variations calculated for this model were always higher than $\pm 100\,\%$ of the model results calculated with the estimated parameters. The uncertainty regions for the Donnan Steric pore-flow model are displayed in the diagrams but not necessarily visible since these were very small.
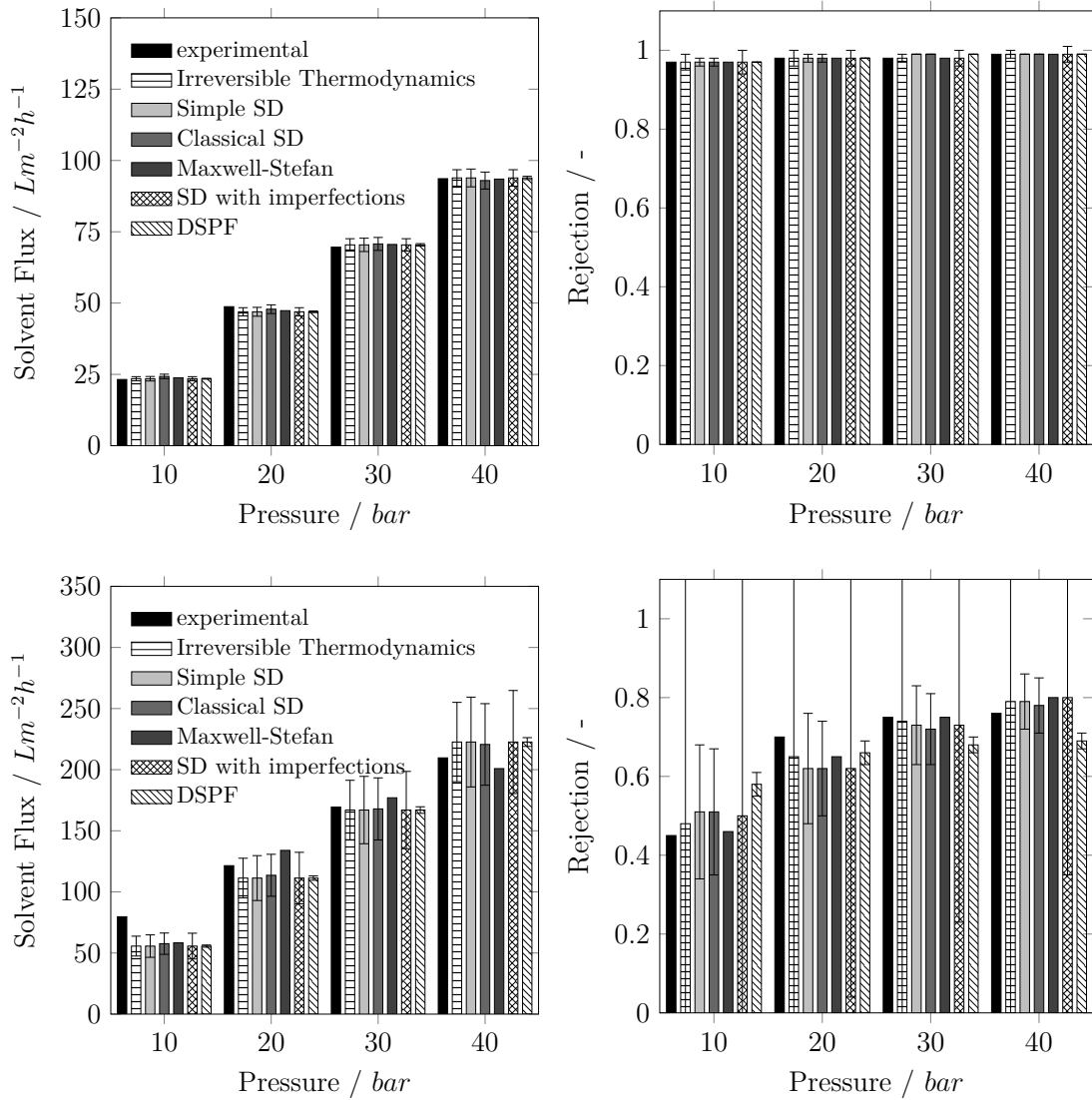
Figure 6.5: Experimental data (provided by Marchetti and Livingston [47]) and calculated values for flux and rejection for the mixture of acetone and α-methyl styrene using the DuraMem® 200 (top) and DuraMem® 500 (bottom), error bars represent the uncertainty regions determined by Monte-Carlo simulation (no bars are shown for Maxwell-Stefan model, due to extremely broad uncertainty regions)

The uncertainty regions for the flux of acetone through the DuraMem® 200 membrane as well as for the rejection of α-methyl styrene are relatively small for all models, except for the Maxwell-Stefan model. All uncertainty regions include the experimental value for each pressure level. Even though, the confidence interval determined for one model parameter of the solution-diffusion with imperfections model was broader than $\pm 100\,\%$ of the parameter value, the effect on the final model result is not as significant. In contrast to this, the broad confidence intervals for the model parameter of the Maxwell-Stefan model for this system results in very large fluctuations of the calculated flux and rejection and, thus, is highly uncertain. Since no significant differences of the broadness of the uncertainty regions for the other models is observed, it is not possible to discriminate between the different models for this system based on the available data.

The error bars for the models applied to the flux of acetone and the rejection of α-methyl styrene using the DuraMem® 500 membrane are broader Compared to the results for the DuraMem® 200 membrane. A reason for this might be the proportional increase of the flux with the increase of the transmembrane pressure for the DuraMem® 200 membrane, which is not as evident as for the DuraMem® 500. Moreover, the rejection using the DuraMem® 200 membrane was almost 1 for all pressure levels while only an intermediate rejection is observed for the DuraMem® 500 membrane.

For the irreversible thermodynamics model, the uncertainty resulting in broad error bars for the rejection is caused by the broad confidence interval for one of the model parameters. Equally large fluctuations were observed for the solution-diffusion with imperfections model. Hence, estimating the parameter values of these models based on the available data used in this study, no reliable predictions are possible due to the significant uncertainty. The uncertainty regions for the rejection observed for the simple and the classical solution-diffusion model as well as for the Donnan Steric pore-flow model are significantly narrower but still broader as for the DuraMem® 200 membrane. The Donnan Steric pore-flow model includes the slightest uncertainty but also the largest deviation from the experimental error.

6.3.4 Summary of results for various membrane-solvent-solute combinations

Summarizing the results of the previous subsections and for all investigated systems (cf. Appendix C), the Maxwell-Stefan model was often very accurate but also showed the highest uncertainties and should therefore not be considered as reliable for predictions based on the limited data from permeation experiments. Nevertheless, the precision of the model parameters can be increased by adding measurements of different sorption or diffusion experiments. The accuracy of the irreversible thermodynamics model was also high for many systems. Based on the norm of residuals, AIC and W, it was the best model for six out of 14 investigated systems. No other model performed as good as this model, with respect to the considered metrics. However, the subsequently performed Monte-Carlo simulations showed high uncertainties of this model especially for the rejection.

Evaluating the Donnan Steric pore-flow model based on AIC only, it was rated adequate to poor, but the model parameters were estimated precisely as the confidence intervals were narrow, which might be attributed to the independence of flux and rejection. On the other hand, the model was not able to describe negative rejections. In contrast, the solution-diffusion with imperfections model was always accurate and additionally able to describe negative rejections, but the confidence interval of one model parameter was also broad resulting in high fluctuations found in the Monte-Carlo simulation for many systems.

Almost no differences between the simple and the classical solution-diffusion models were found. For negative rejections, negative parameter values were necessary, but using these, the models were able to describe negative rejections in general. For a minor number of systems, both models were not able to describe the rejections when the permeate concentrations measured in the experiments were used. This was the case for systems with constant but low rejection or decreasing and negative for increasing transmembrane pressure. However, it was possible to describe these systems successfully when the permeate concentration was calculated as well. The variations found in Monte-Carlo simulations were always small for these two models.

6.4 Conclusion

In this chapter a systematic investigation of different models for OSN was presented. The study did not only evaluate the accuracy of the models in terms of goodness of fit but also included the investigation of the uncertainty of the model parameters and the influence of these on the model results. For this, different experimental data available in literature were used. Six different models commonly used to describe flux and rejection in OSN were fitted to the available data in order to reproduce the results of Marchetti and Livingston [47]. Subsequently, the study was extended by an analysis of variance and identifiability determining the confidence intervals of all model parameters and performing an eigenvalue analysis. Based on these results, a Monte-Carlo simulation was performed in order to determine the influence of the parameter uncertainty on the model results.

It was shown that no clear preference for any of the investigated models was possible. Determining the most suitable model only based on the goodness of fit neglects some further important aspects which are essential for identifying the best model for predictions. In order to identify the most suitable model for a specific membrane-solvent-solute systems, the measured experimental data should always be selected carefully and multiple repeated measurements should be performed in order to ensure reasonable experimental data. Based on these data, the parameters of different models can be estimated. For most of the systems, the simple or classical solution-diffusion model seem to be a good choice, but parameter uncertainties should always be investigated additionally. Different models, like the Maxwell-Stefan model, or further experimental measurements, like sorption or diffusion data, might be beneficial in order to increase the reliability of the models.

No clear recommendation for one of the models can be given, since the Monte-Carlo simulation showed that the models cannot be discriminated based on the available data. However, the current study clearly demonstrates that the parameter uncertainty has a strong influence on the model results. Thus, for reliable predictions precisely determined parameters are necessary.

7

Machine-based learning of predictive models: Flux

The complex interaction of different phenomena in the mass transfer through the membrane complicate the prediction of membrane performance severely, such that OSN is virtually not considered as an option in conceptual process design. Several attempts have been made to determine predictive models, which allow the determination of at least pure solvent flux through a given membrane. While these models correlate different important physical properties of the solvents and are derived from physical understanding, they provide a limited accuracy and not all of their parameters are identifiable based on available data. In contrast to previous approaches, this work presents a machine learning based approach for the identification of membrane-specific models for the prediction of solvent permeance. The data-driven approach, which is based on genetic programming, generates predictive models that show superior results in terms of accuracy and parameter precision when compared to previously proposed models. Applied to two respective sets of permeation data, the developed models were able to describe the permeance of various solvents with a mean percentage error below 9% and to predict different solvents with a mean percentage error of 15%. Further, the method was applied to solvent mixtures successfully.

7.1 Introduction

For OSN a various semi-empirical models considering different physical transport phenomena have been proposed (Section 2.2.1 to 2.2.3), including pore-flow, solution-diffusion and irreversible thermodynamic models. However, until now, only few efforts have been made to attempt the prediction of at least pure solvent flux through a specific membrane. To this extend phenomenological models have been proposed (Section 2.2.4), derived based on physical insight and correlating different physical properties of the solvents. The models differ in terms of general structure, as well in the type and number of solvent independent parameters, which have to be determined based on an available set of permeation data. While all these models show considerable differences, they all are intended to be generally applicable, requiring only the adjustment of membrane specific fitting parameters.

However, knowing that different transport models are superior to other depending on the type of membrane and membrane material [14] the applicability of a general model for all membranes and chemical systems is rather unlikely. This is supported by the differences in the previously proposed phenomenological models and the considered properties, as well as corresponding statements of the different authors. Geens et al. [55] indicated a necessary extension of their model by means of weighting factors in order to comply with the different behavior of polar and non-polar membranes, while Darvishmanesh et al. [1] indicated that the surface tension is less important for dense membranes but may be important for more porous membranes. However, due to the solvent-induced swelling of different membrane materials this differentiation is rather blurry, at least for polymeric OSN membranes. As an alternative approach to the previously proposed phenomenological models, the current study proposes the automatic generation of membrane-specific models, based on a tailored supervised machine learning approach. In specific, a combination of a genetic programming approach and a global deterministic optimization approach for nonlinear parameter regression is proposed, which is further extended by an additional post-processing for model reduction based on a parameter identifiability analysis. Based on the findings of an extensive collaborative experimental study [16] the available permeation data is divided into different classes, which are further considered for the generation of training, test and validation sets in the scope of a data pre-processing phase.

The existing phenomenological models for solvent flux and the considered membrane

and solvent properties were already discussed in in Section 2.2.4. Prior to presenting the developed machine learning approach, before a brief summary of available data-driven methods for model development is provided in Section 7.2. The novel approach proposed is further presented in detail in Section 7.3, followed by the application to different case studies for pure and mixed solvent flux predictions in Section 7.4. The comparison with the previously proposed models, highlight benefits with respect to accuracy of the predictions and identifiability of the model parameters, while an additional comparison of the introduced hybrid stochastic-deterministic optimization approach with a full enumeration elucidates the quality of the results obtained by the proposed algorithmic framework. Finally, Section 7.5 presents major conclusions and proposes further extensions.

7.2 Data-driven methods

While some authors of the studies on phenomenological models presented in Section 2.2.4 stated that the previously proposed models were able to describe solvent flux especially for non-polar membranes in good agreement with the considered experimental data, the proposed modification and categorization in polar and non-polar membranes, as presented by Darvishmanesh et al. [1], indicate a limited applicability of a fixed model structure independent of the membrane. Instead of trying to demystify the highly complex interactions between the different mass transfer phenomena, the derivation of membrane-specific models might provide a more suitable solution, for which appropriate correlations for solvent flux might be determined by a machine-based learning approach with a minimum effort. The current section briefly summarizes the most prominent methods, before the specific hybrid stochastic-deterministic approach proposed in the current chapter is presented in the subsequent section.

All of the previously presented phenomenological models represent a correlation of different properties of the solvents and membrane-specific parameters that are regressed to experimental data. In this sense, they can be interpreted as rather similar to quantitative structure-property relationships (QSPR) or quantitative structure-attribute relationships (QSAR) [95]

$$\text{Response=f(physical properties, chemical structure)},$$

which are usually applied in the context of thermodynamic or physical property

predictions, based on a data-driven correlation. For the phenomenological models, the desired response is either the flux or the permeance. The determination of such a correlation can be posed in a general form as a symbolic regression problem, which aims at the identification of a function that fits the given data points without making any assumptions about the structure of that function [96]. However, data-based approaches for the identification of such a correlation most often start with a pre-defined maximum model structure and evaluate the most appropriate one based on parameter regression. For a similar application Yangali-Quintanilla et al. [97] applied a combination of principal component analysis (PCA) and multiple linear regression (MLR) to derive QSAR models for the prediction of the rejection of emerging contaminants by aqueous nanofiltration. Shahmansouri and Bellona [98] applied MLR and artificial neural networks (ANN) to develop QSPRs for the rejection of nonionic organic compounds in aqueous nanofiltration. In general, a variety of data-driven models can be applied for the considered purpose, ranging from simple linear models, such as response surface models (RSM) [99], to highly nonlinear models, like multi-layer artificial neural networks [100]. Considering similar applications and the potential to derive nonlinear models especially ANN, the automated learning of algebraic models for optimization (ALAMO) approach, proposed by Cozad et al. [101], as well as the less popular genetic programming (GP) approach proposed by Koza [102], are interesting options for the considered purpose. A detailed discussion of the methods is provided in Chapter D.1 in the Appendix. The methods have been successfully applied to different membrane processes like aqueous nanofiltration, microfiltration and reverse osmosis in order to predict fouling or the flux based on operating parameters [103, 104] or the rejection of various organic components [105, 106]. These studies indicate the potential of data-driven methods. Yet none of them, investigated the application to the prediction of solvent flux for varying chemical species in organic solvent nanofiltration. However, since the ALAMO approach does not incorporate nonlinear combinations of basic functions and ANNs do not provide a transparent model equation, GP shows great potential to identify suitable compact models that accurately describe the performance of membrane processes. The subsequent section introduces the developed GP-based approach for this purpose.

7.3 A hybrid stochastic-deterministic approach for model development

In a first pre-processing step, the available permeation data is subdivided into training, test and validation data and different physical properties are selected as potential descriptors that are to be correlated in the symbolic regression. In order to determine accurate and solvent flux models with high parameter precision, a hybrid stochastic-deterministic approach was developed, which performs symbolic regression based on a modified version of the GP approach proposed by Madár et al. [107], while parameter regression is performed by means of a deterministic global optimization approach. The GP approach, which is implemented in Matlab® incorporates an orthogonal least squares (OLS) approach, in order to eliminate branches of the tree structure that do not contribute to a significant extent to the model result, as proposed by Madár et al. [107]. While linear regression is performed directly in Matlab®, making use of the *mldivide* operator, which takes advantage of symmetries in the problem by dispatching to an appropriate solver, nonlinear parameter regression is performed via an interface with the general algebraic modeling system (GAMS) making use of either BARON [108] or ANTIGONE [109] as global deterministic optimization algorithms. Finally, an additional identifiability analysis is performed as post-processing step to further evaluate the potential for model reduction. A schematic representation of the approach is illustrated in Figure 7.1, while the different steps are described in the following subsections.

7.3.1 Pre-processing of permeation data

In the first pre-processing step, the available permeation data is subdivided into training data, which is considered for symbolic and parameter regression, test data that is considered only for parameter regression and validation data, which is only considered for evaluation of the predictive power of the developed models, after symbolic and parameter regression have been performed. In order to avoid severe extrapolation each solvent group should be represented at least within the training data set, considering the molecules with the largest and the smallest molecular size, while remaining data is evenly distributed between test and validation data set.

As described in Section 2.2.4 all of the previously proposed models consider three types of parameters related to molecular size, solubility and solvent diffusivity. In
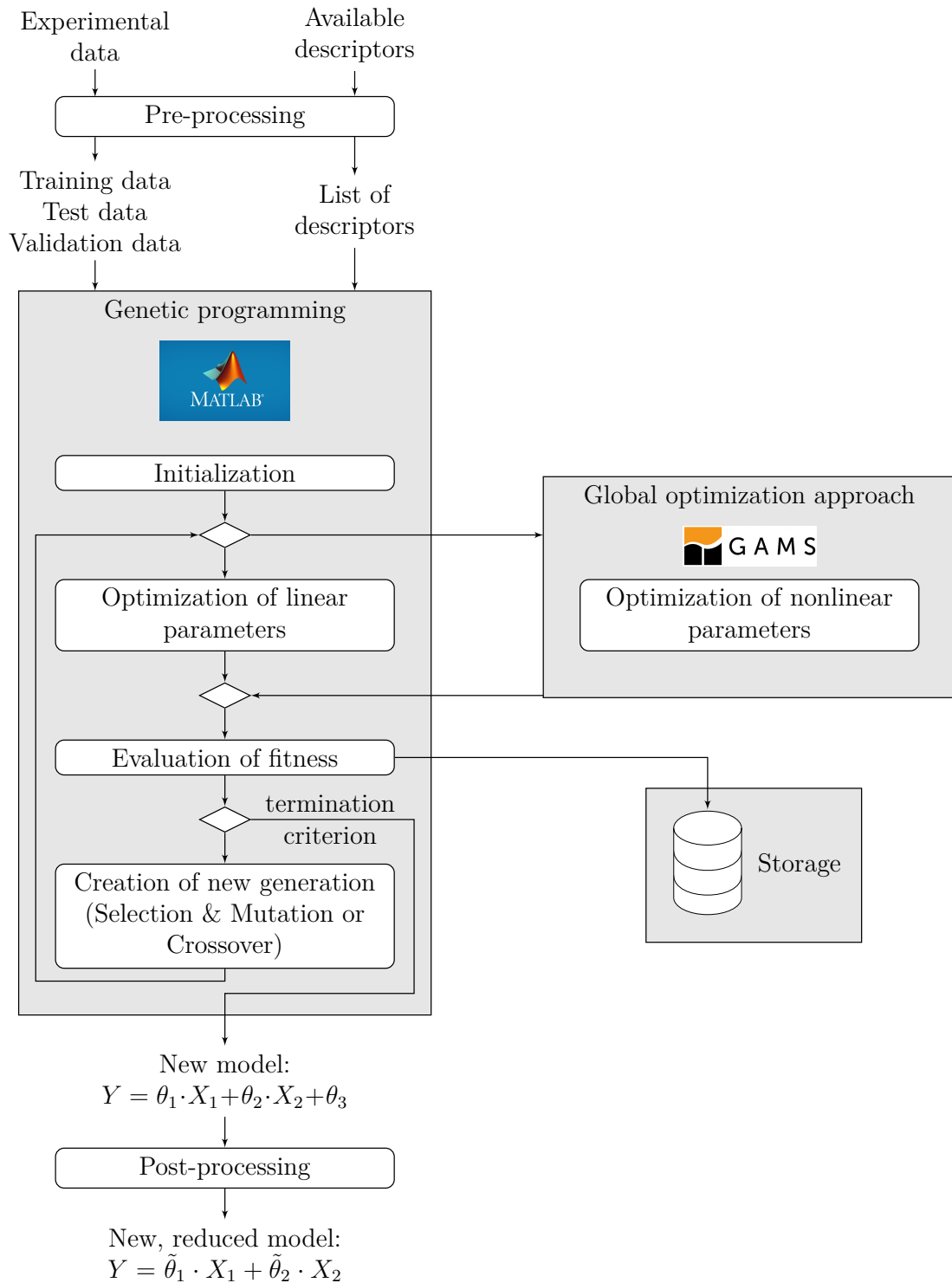
Figure 7.1: Schematic structure of the new approach

order to identify the most meaningful physical property parameters for the different categories, a range of widely available parameters is considered as potential descriptor for the symbolic regression. The group of molecular size parameters is formed by the molar mass M, the molar volume V_m, equivalent molecular diameter d_{eq} and the connectivity index $ConI$. While the molar mass is also related to the molecular weight cut-off, the molar volume was considered as important descriptor by several authors [51, 110, 111, 34] and included in the models of Bhanushali et al. [4] and Geens et al. [55]. The equivalent molecular diameter

$$d_{eq} = 2 \left(\frac{3V_m}{4\pi N_A} \right)^{1/3},$$ (7.1)

can be calculated based on the molar volume V_m and the Avogadro number N_A and was considered as important descriptor by Van der Bruggen et al. [112]. Finally, the connectivity index, as introduced by Randic [113] and extended by Kier and Hall [114], allows to differentiate between linear and branched molecules.

Since the solubility of the solvent was considered of high importance in previously reported models [1, 2, 55], the dielectric constant ϵ, dipole moment $\vec{d}$, surface tension γ, polarity Pol and the Hildebrand (δ_{HBP}) and Hansen solubility parameters (δ_d, δ_p and δ_h) are taken into consideration to describe the solubility of the solvent. While the dielectric constant, considered by Darvishmanesh et al. [1] and Machado et al. [3], is a bulk property, the dipole moment, considered by Marchetti et al. [2] is rather a molecular property. The surface tension was considered in various studies [1, 4, 55, 3] as representative of the polarity of the solvent and the membrane, whereas Blumenschein and Kätzel [33] used the Hildebrand solubility parameter to describe the interaction of solvent and membrane, similar to Robinson [115], and Tarleton et al. [116]. In the scope of different investigations, the Hansen solubility parameters, which are directly related to the Hildebrand solubility parameter

$$\delta_{HBP}^2 = \delta_d^2 + \delta_p^2 + \delta_h^2,$$ (7.2)

were further considered as representative for the swelling of the membrane material [117] or to explain solvent dependent membrane-solute sensitivity of OSN membranes [78]. Finally, the radius of the interaction sphere Ra

$$Ra^2 = 4(\delta_d - \delta_{d,M})^2 + (\delta_p - \delta_{p,M})^2 + (\delta_h - \delta_{h,M})^2,$$ (7.3)

as introduced by Hansen [118], can be used to evaluate the interaction between solvent and membrane, considering the Hansen solubility parameters $\delta_{d,M}$, $\delta_{p,M}$ and $\delta_{h,M}$ of the active layer of the membrane, which depending on the material may be evaluated based on the group contribution method of van Krevelen and Nijenhuis [119].

In accordance to the solution-diffusion model and as described by Reddy et al. [120] and Bhanushali et al. [4], the permeance depends also on the diffusivity of the solvent in the membrane, which depends on the reciprocal of the viscosity of the solvent. Thus, the viscosity η of the solvent is considered as additional descriptor and complemented with the density ρ as last descriptor for the symbolic regression. he descriptors are summarized in Table 7.1, while the respective solvent property data is provided in Table D.4 in Chapter D in the Appendix. If sufficient data is available, all descriptors are integrated in the terminal set of the GP approach.

Table 7.1: List of all descriptors and the corresponding symbols used in model development

Group	Descriptor	Symbol
Size	Molar mass	M
	Molar volume	V_m
	Equivalent molecular diameter	d_{eq}
	Connectivity index	$ConI$
Polarity	Dielectric constant	ϵ
	Surface tension	γ
	Hildebrand solubility parameter	δ_{HBP}
	Hansen solubility parameters (dispersive, polar, h-bonds)	$\delta_d, \delta_p, \delta_h$
	Difference of Hildebrand solubility parameters of membrane and solvent	$\Delta\delta_M$
	Interaction radius	Ra
	Dipole moment	$\vec{d}$
	Polarity	Pol
Further	Density	ρ
	Viscosity	η

7.3.2 Hybrid approach for automatic model development

The hybrid approach for automatic model development consists of the GP approach, that starts with the initialization of a random population of syntax trees, followed by a subsequent parameter regression, which may employ deterministic global optimization for nonlinear parameter regression, prior to the evaluation of the fitness and the creation of a subsequent generation by means of selection, mutation and crossover, until a specific termination criterion is met. The individual steps are explained in the subsequent subsections.

Initialization

Each model candidate is represented as a syntax tree (cf. Figure 7.2), which links different descriptors by means of additional functions and added parameters. The tree is modified in the different generations through the application of evolutionary operators.
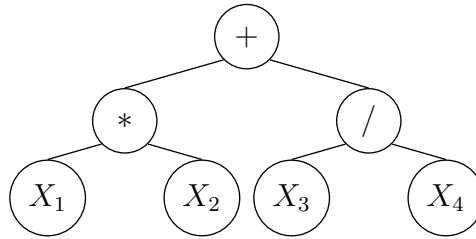
Figure 7.2: Tree with a tree depth of 3

In order to include the model candidates of Darvishmanesh et al. [1], Bhanushali et al. [4], Geens et al. [55], and Machado et al. [3] as well as a wide variety of alternative correlations, the previously described descriptors (Section 7.3.1) and a set of different functions $\{+, -, *, /, exp, ln, sqrt\}$ are considered as valid nodes. In order to avoid numerical problems, the

$$ln(x) = \begin{cases} ln(x) & , x \geq 0 \\ 1 & , x < 0 \end{cases} \tag{7.4}$$

and

$$sqrt(x) = \begin{cases} \sqrt{x} & , x \geq 0 \\ 1 & , x < 0 \end{cases} \tag{7.5}$$

functions are defined piecewise, avoiding non-defined and irrational values in case

of negative arguments. After the maximum tree depth and the population size are specified, the initial population is generated based on a random selection of descriptors and functions for each node of an individual syntax tree, except for the root node and the lowest level nodes. The root node has to be a function, while only descriptors are allowed for the lowest level nodes. While all intermediate level nodes may be either functions or descriptors, it is important to note that in the translation of the syntax tree to a functional program, all nodes on a branch below a descriptor are ignored. Thus, they do not contribute to the final model, but may become active again through mutation of the above descriptor node. In order to ensure a minimum number of trees with maximum tree depth in the initial population, 10% of the syntax trees in the initial population are generated by limiting the selection for nodes above the bottom layer to functions only.

Parameter optimization

Parameter estimation is performed with respect to the specified training data set applying either linear or nonlinear regression, depending on the functional representation of the syntax tree. The evaluation is made automatically, based on the arrangement of the functions in the syntax trees. If no function of the type $\{+, -\}$ is detected in the syntax tree below a function of the type $\{*, /, exp, ln, sqrt\}$, the tree represents a linear function and the linear regression is performed in Matlab$^{\circledR}$ using the *mldivide* operator.

Otherwise, the tree represents a nonlinear function with respect to the parameters and is first translated into a functional representation in which the parameters are inserted as optimization variables after each $\{+, -\}$ function. Other parameter locations would result in correlated parameters and are therefore neglected. The nonlinear parameter regression is performed in GAMS, based on an automatic model transfer and minimization of the corresponding least squares function as nonlinear programming problem, using either BARON [108] or ANTIGONE [109]. The results are transferred back to Matlab$^{\circledR}$ for further evaluation of the fitness of the corresponding syntax tree.

Evaluation of fitness

While parameter regression is performed based on the training data, both the training and test data are used to evaluate the fitness of the different syntax trees. In accordance with the parsimony principle the best model candidate is not necessary

the one that provides the highest accuracy with respect to the experimental data, but rather the one that provides the best approximation with the least number of parameters. Therefore, the Akaike information criterion (AIC) [121], which takes the goodness of fit and the number of parameters of each model candidate into account, is selected as fitness function. In case of least squares estimation with normally distributed errors the AIC is defined as

$$AIC = N_{obs} \cdot ln(\frac{\sum(\hat{\epsilon})^2}{N_{obs}}) + 2N_{param}, \tag{7.6}$$

for which N_{obs} is the number of observations, $\hat{\epsilon}$ are the residuals of the model and the experimental values and N_{param} is the number of parameters used in the model [121]. Since permeance values for different solvents can vary significantly, the absolute deviation in terms of the residuals $\hat{\epsilon}$ are substituted in Equation 7.6 for the relative errors (cf. Equation 7.7), in order to avoid emphasis on larger permeance values.

$$RE = \sum_{i=1}^{N_{exp}} \frac{P_{calc,i} - P_{exp,i}}{P_{exp,i}} \tag{7.7}$$

Each evaluated tree is stored in a separate file and hence are accessible after the run is finished. This allows for a detailed investigation of the evaluated trees and possible additional model candidates.

Creation of new generation

After the evaluation of the fitness, the best 10 % of the current generation are directly copied to the next generation, ensuring that the best solutions are preserved in subsequent generations. The remaining 90 % of the new generation are filled by syntax trees created by crossover or mutation, from syntax trees of the current generation, applying tournament selection [96]. Thereby, possible parent trees are selected as those with the best fitness of a subset selected by random drawing from a uniform distribution of the current generation, offering the possibility of constant selection pressure for all generations. Due to the repeated random drawing from the uniform distribution of all syntax trees in the current generation one tree can be considered several times, while others are not considered at all.
New syntax trees are generated from the selected parents, with additional application of either mutation or crossover, with a given probability for mutation, corresponding to the number of new trees created by mutation. In case of mutation a single

parent tree is selected, whilst for crossover two parent trees are selected, performing tournament selection with a subset size of two for the selection of each parent tree. The selected trees are copied to the new generation and subsequently modified. Both operators are only applied to active nodes, whereas node type and node content are handled separately. The node type is either 'function' or 'descriptor', while the content specifies the selected descriptor or function.

In case of mutation, one of the active nodes is selected by random drawing from a uniform distribution and type and content of this selected node are randomly modified, considering each option with equal probability (cf. Figure 7.3). Only for the root node and the lowest level nodes the node type is fixed, such that only the content is modified.
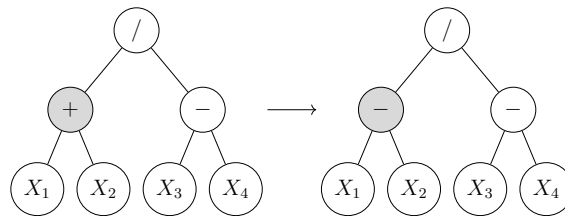
Figure 7.3: Mutation of GP tree [122]: Selected parent tree is copied and one selected node is modified by mutation to create the new syntax tree (the X_i represent different descriptors).

For crossover both selected trees are modified, based on the selection of an active node with similar node type in both syntax trees, exchanging the branch below this node between both of the selected syntax trees (cf. Figure 7.4). Thus, this one-point crossover operator requires two parent trees and generates two syntax trees for the new generation.
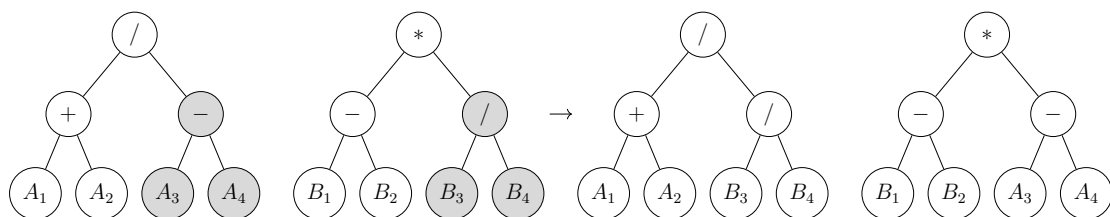
Figure 7.4: Crossover of GP trees [122]: Selected parent trees are copied and branches below a selected node are exchanged between both trees to create new syntax trees (the A_i and B_i represent different descriptors of the two parents).

In the first generations the probability of mutation is set to a high value to increase

diversity and ensure an adequate exploration phase [123]. In the subsequent generations, the probability of mutation p_m is gradually reduced from the initial value $p_{m,0}$ by

$$p_m = -0.1 \cdot ln(generation) + p_{m,0}, \qquad (7.8)$$

enforcing a focused exploitation phase in the end [123]. The change in mutation probability stops at a threshold value of 0.05 to uphold a small contribution of mutation.

Termination

Termination is either induced because a predefined maximum number of generations is reached or because of a maximum number of consecutive generations with a lack of improvement of the fitness. For all of the computations in the current study a maximum of 100 generations and 25 consecutive generations without improvement were defined.

7.3.3 Post-processing for potential model reduction

Since the fitness function (eq. Equation 7.6) already accounts for the number of parameters, models with fewer parameters are favored during the symbolic and parameter regression. Nevertheless, parameter precision is not accounted for in this process. Since models with accurately identifiable parameters are expected to be more reliable, the 95 % confidence intervals of the parameter estimates are calculated using the Matlab®-function *nlparci*. Furthermore, the eigenvalues of the Hessian matrix of the fitness function with respect to the parameters are evaluated, as suggested by Quaiser and Mönnigmann [92] in order to evaluate the precision of the parameter estimates. By canceling (setting to zero) parameters for which estimates are highly unprecise, the model complexity is systematically reduced. The reduced models are further refined by additional parameter regression (as in Section 7.3.2) and evaluation of the fitness function (as in Section 7.3.2).

7.3.4 Model Selection

While the specified objective function is sought to evaluate the best model and the developed hybrid approach allows for a global search, the genetic programming approach has an inherent stochastic character and does not guarantee to find the

global optimum. Hence, several runs should be performed to validate the quality of the determined solutions, while different measures, apart from the specified objective function, can be used to evaluate the most promising model candidates, which provide almost equal fitness in terms of the specified objective function.

Out of these runs, the most promising model candidates are further evaluated based on the mean percentage error

$$\bar{\varepsilon} = \frac{1}{N_{obs}} \sum_{i=1}^{N_{exp}} \frac{|P_{calc,i} - P_{exp,i}|}{P_{exp,i}}, \tag{7.9}$$

between the calculated $P_{calc,i}$ and the experimental permeance values $P_{exp,i}$ for the training and test data set, as well as the validation data set. Moreover, the best model candidate, according to the fitness function, is compared with the preceding literature models, based on Akaike weights and an analysis of parameter identifiability, as described in the post-processing step (Section 7.3.3). By using the Akaike weights

$$W = \frac{exp(\frac{-\Delta}{2})}{\sum_{i=1}^{N_{models}} exp(\frac{-\Delta_i}{2})} \ with \ \Delta = AIC - AIC_{best}, \tag{7.10}$$

the Akaike values are normalized [121]. The weights take on values between 0 and 1, which can be interpreted as the probability that one model is the best choice for the investigated data set.[121], making it possible to identify the most probable models in a set of various models [124].

7.3.5 Full Enumeration

In order to verify the quality of the developed hybrid approach for model development and to evaluate its effectiveness, a comparison is made with a full evaluation of all possible model candidates, resulting from the set of functions and operators. For this purpose, the same training set as for the hybrid approach was used. Subsequently, all possible syntax trees up to the predefined tree depth are generated by systematically changing the nodes. The translation of the models and parameter regression are performed as for the hybrid approach, while the evaluation is also based on the same criteria, as specified in Section 7.3.2. Based on the fitness, the best model is identified and the post-processing steps are applied as described in Section 7.3.3. The total number of possible combinations for each tree depth is

defined by

$$N_{comb} = \prod_{i=1}^{level} (n_i^{k_i}),$$ (7.11)

with n being the number of different options for filling a node (functions and/or descriptors), k being the number of different positions in each level of the tree i. Considering completely filled trees with a tree depth of 2, where only functions are allowed in the first level and both positions in level two can only be filled with descriptors, the number of possible combinations is 1792. However, for complete filled trees of a higher tree depth of 3 (only functions in level one and two and only descriptors in level tree) the number of possible combinations increases to 22.478.848. Since not completely filled trees with a tree depth of 3 are additionally possible, the number sums up to 22.882.048 trees in total. Further details can be found in the Appendix in Section D.3.

7.4 Case Studies for prediction of solvent flux

In order to evaluate the automatic approach for model development, it was applied for three case studies, comprising the prediction of pure solvent flux, based on permeation data for PDMS-based polymeric membranes published by Dijkstra et al. [51] and Machado et al. [62], and mixed solvent fluxes for permeation data for ceramic membranes, published by Marchetti et al. [2]. While all of the previously described descriptors (Section 7.3.1) were considered for the first two case studies, the Hildebrand solubility parameters of the solvent and membrane, as well as the interaction radius were neglected for the third case study on the mixed solvent fluxes for ceramic membranes. For each case study, the results of ten runs of the hybrid approach for model development were evaluated, and the most promising model candidates were compared based on the mean ($\bar{\varepsilon}$) and maximum percentage error (ε_{max}), while the best model was further compared with the existing literature models, based on Akaike weights. In accordance with suggestions by Poli et al. [96] a population size of 1000 individuals were considered for each run, which was initially verified by test runs with lower population sizes. If not stated otherwise, a maximum tree depth of three was considered.

7.4.1 Prediction of pure solvent permeance

Application of the approach is first illustrated for permeation data of polymeric membranes reported by Dijkstra et al. [51] and Machado et al. [3].

Permeance for PDMS-based membrane

The permeation data published by Dijkstra et al. [51] describes the pure solvent fluxes of a homologous series of alcohols, alkanes and ketones for a PDMS-based membrane with a support layer of Polyacrylonitrile (PAN). The permeation data were assigned to training, test and validation data set as described in Table 7.2, following the criteria described in Section 7.3.1.

Table 7.2: Allocation of training, test and validation data sets for Dijkstra et al. [51]

	Alkanes	Alcohols	Ketones
Training	n-pentane	ethanol	acetone
	n-octane	2-butanol	MIBK
	n-dodecane		
Test	n-hexane	1-propanol	MEK
	n-nonane	1-butanol	
Validation	n-heptane	isopropanol	diethyl ketone
	n-decane		

Table 7.3 provides an overview of the results of the ten consecutive runs and the resulting models, ordered according to their fitness, first limiting the models to linear parameter relationships. Besides the fitness, the mean ($\bar{\varepsilon}$) and maximum percentage error (ε_{max}) for training and test data as well as validation data are listed. While most runs determine the same model candidate, some determine a slightly inferior model, for which the mean percentage error of the training and test data only deviates by at most 2 %. Repeated execution of 100 runs validated that the best model candidate was found with a comparable frequency, confirming that the results of the ten runs are representative and that no further improvement of the objective function was obtained in the increased number of runs. While the third best model interestingly provides the lowest mean percentage error for the validation data set, the best model provides both, the lowest mean and maximum percentage error for the test and training data sets.

All runs terminated due to lack of improvement in 25 subsequent generations, within

Table 7.3: Results for the ten consecutive runs for models with linear parameter relationships for the data set of Dijkstra et al. [51].

Model Nr.	Run	Fitness	Training & test $\bar{\varepsilon}$	ε_{max}	Validation $\bar{\varepsilon}$
1	2-6,8	-51.33	7.0%	16.5%	11.5%
2	1,9	-47.71	8.8%	22.0%	14.2%
3	7	-46.76	8.5%	28.3%	7.3%
4	10	-46.04	9.0%	30.3%	14.4%

the maximum of 100 generations. As an example, Figure 7.5 illustrates the progress curve of the fitness plotted over the number of generations for the first model. While an initial plateau is passed around the 10th generation, the fitness further improves until generation 22, before staying constant until the final 47th generation. A single run converged within 2 min, while all ten runs were executed in 16 minutes.
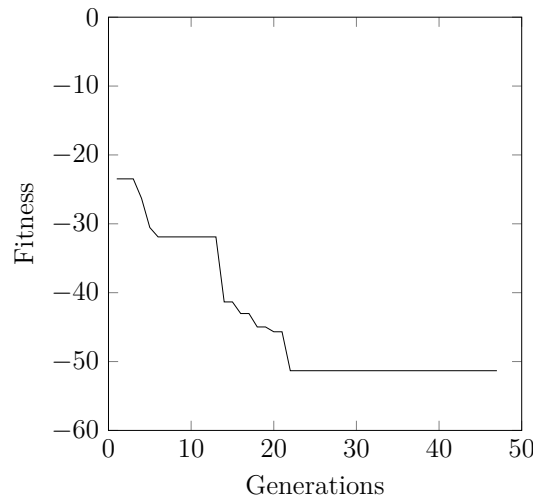
Figure 7.5: Progress curve for the first run for the data set of Dijkstra et al. [51].

The best model candidate

$$P_D^1 = 0.33 \cdot \frac{\delta_d}{\eta} + 0.87 \cdot \eta - 0.34 \cdot \delta_p - 1.04 \tag{7.12}$$

correlates the viscosity η, the Hansen solubility parameter for polar interactions δ_p and for dispersive interactions δ_d of the solvent and results in average mean percentage errors $\bar{\varepsilon}$ well below the expected experimental variation of $\pm 20\%$ (Chapter 4).

The second-best model

$$P_D^2 = 9.65 \cdot \frac{d_{eq}}{\eta} + 4.42 \cdot d_{eq} \cdot log(\eta) - 3.76 \tag{7.13}$$

correlates only the equivalent molecular diameter d_{eq} and the viscosity η of the solvent, while resulting in almost the same accuracy as the model with the best fitness (cf. Equation 7.12). Evaluation of the different models indicates the importance of the viscosity, as especially the inverse proportional relation, which fits the expectation and is in accordance with all previously proposed phenomenological models listed in Section 2.2.4. However, the models differ with respect to the other properties and the general model structure. Despite these differences, the permeance estimates derived from both models are generally within a range of $\pm 15\%$ with respect to the experimental data, as illustrated in Figure 7.6. Only for the lowest permeance values this limit is exceeded for three solvents, while the deviation is still in the range of $\pm 20\%$, highlighting the excellent agreement between model and experimental data.
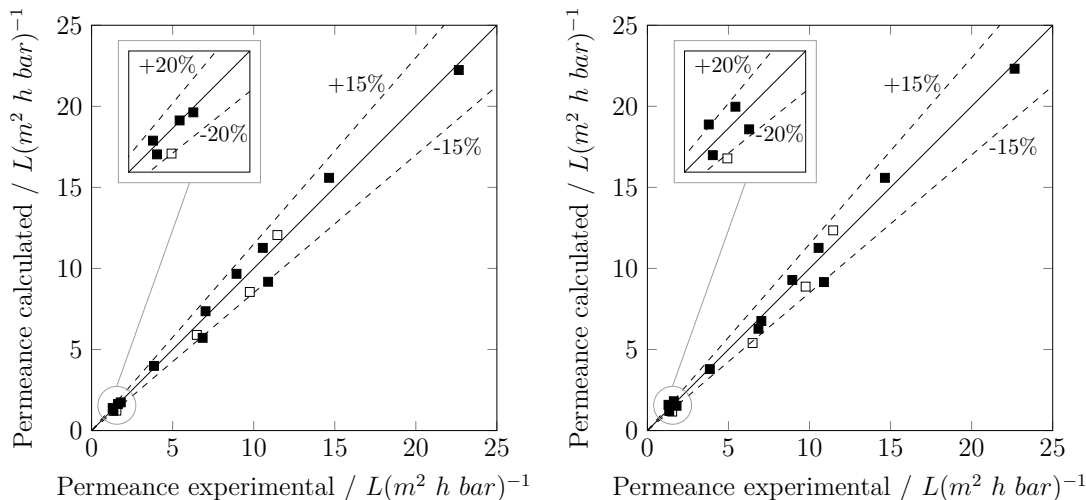
Figure 7.6: Calculated vs. experimental permeance values for training and test (filled) as well as validation (unfilled) data [51] for model P_D^1 (cf. Equation 7.12) (left) and model P_D^2 (cf. Equation 7.13) (right).

In order to evaluate the quality of the determined models further, they are compared with the results of a full enumeration of all models with linear parameter relationships, as well as the best model that was determined, when accounting for nonlinear parameter relationships. The results of this comparison are summarized

in Table 7.4. The full enumeration of all model candidates with linear parameter relationships validates that the previously determined best model candidate is in fact the global optimum of all models with a linear parameter relationship. Since this model was frequently determined in the ten (respective 100) runs, we can further assume that the approach with a set of ten individual runs provides the best model candidate with high probability, while the required computational effort (16 min) is much lower compared to the 12 h (factor of 45) required for the full enumeration.

Table 7.4: Comparison of results of run 1 (new method, linear parameters) with additional run with nonlinear models (new method, linear and nonlinear parameters) and the a full enumeration of all models with linear parameter relationships.

| Model Nr. | Fitness | Training & test | | Validation |
		$\bar{\varepsilon}$	ε_{max}	$\bar{\varepsilon}$
run 2-6,8	-51.33	7.0%	16.5%	11.5%
run 1,9	-47.71	8.8%	22.0%	14.2%
run (NL)	-53.64	7.2%	14.1%	8.7%
full	-51.33	7.0%	16.5%	11.5%

The considering of nonlinear parameter relationships does not only increase the number possible model candidates but mandates the more complicated global optimization for parameter regression. As a consequence, a single run with 1000 individuals per generation exceeds 48 h of computational time. The best model candidate, listed for run (NL) in Table 7.4, was determined after 45 h in the 29th generation of a single run and confirmed by 9 consecutive runs. Interestingly, this model

$$P_D^{NL} = \frac{2.62 \cdot \delta_{HBP} \cdot \sqrt{\delta_d}}{\delta_h + 2.02 \cdot ConI} - 10.62 \tag{7.14}$$

does not include the viscosity but the connectivity index $ConI$, as well as several properties in relation to the polarity of the solvent, such as the Hildebrand solubility parameter δ_{HBP} and the Hansen solubility parameters for dispersive interactions δ_d and hydrogen bonds δ_h. While it does neither improve $\bar{\varepsilon}$ nor ε_{max} much for the test and training data, it shows a considerable improvement in terms of $\bar{\varepsilon}$ for the test and validation data.

Prior to a final evaluation, the fitness values and the W of the best model candidates, as well as the 95% confidence intervals and eigenvalues of the model parameters are

evaluated. The results are listed in Table 7.5, together with the respective values of different literature models, for which parameter regression was performed based on the training and test data. While the fitness values already indicate that all of the derived models outperform the literature models in terms of accuracy, this is further highlighted by the Akaike weights, which indicate that model P_D^{NL} is the most likely the best candidate, while the probability for the literature models is practically zero. Since the size of the 95 % confidence intervals of the models proposed by Geens et al. [5] and Darvishmanesh et al. [1] exceed the parameter values, it is further concluded that these models lack parameter precision. This is however also the case for the second and fourth parameter of model P_D^1. Performing model reduction with respect to these parameters, canceling out unidentifiable parameters and re-regressing the remaining parameters, the fitness drops to considerable below that of P_D^2.

Table 7.5: Summary of fitness, mean percentage error ($\bar{\varepsilon}$), AIC weight (W), parameter value (θ), size of 95% confidence interval (CI) and eigenvalues (λ) for the different models, considering training and test set.

Model	Fitness	$\bar{\varepsilon}$	W	θ	CI	λ
Model P_D^1	−51.33	7.0%	23.0%	0.34	0.07	0.50
				0.87	1.39	12.58
				−0.34	0.27	189.23
				−1.04	3.94	9.09×10^3
Model P_D^2	−47.71	8.8%	3.8%	9.65	1.55	0.22
				4.42	2.41	3.09
				−3.76	1.68	25.04
Model P_D^{NL}	−53.64	7.2%	73.2%	2.62	0.37	0.50
				2.02	0.19	38.53
				−10.62	1.81	820.44
Bhanushali	−24.95	25.6%	0.0%	0.37	0.07	7.95×10^5
Geens	−16.85	36.4%	0.0%	0.24	0.32	1.81×10^4
Darvishmanesh	−19.48	23.8%	0.0%	0.52	7.70	46.49
				6.37	5.07	386.90

Based on this post-processing step it can be concluded that with respect to the quality of fit and parameter precision P_D^2 is identified as best model with linear parameter relationship, which is slightly outperformed by model P_D^{NL}, with nonlinear parameter relationship. Note that the ranking and results of the identifiability analysis are based on the available data and that the identifiability of the parameters of model P_D^1 might be improved based on additional data, for which a model-based

experimental design approach could be used to determine suitable solvent candidates [125]. Yet, this is not part of the current study. While the results prove that an improved model candidate can be determined when considering nonlinear relationships, the rather small improvement does not necessarily justify the high computational load compared to the simpler linear relationship models, for which the possibility to determine the global has been validated by the full enumeration. Therefore, it is suggested to first derive model candidates with linear relationships and switch to the expensive derivation of nonlinear model candidates in case the derived models with linear relationships lack quality of fit or parameter precision. The subsequent examples indicate however that linear relationships result in sufficient model accuracy and parameter precision in most cases.

Permeance for MPF-50 Koch membrane

For the second case study, permeation data published by Machado et al. [3] for the PDMS/PAN membrane MPF-50, produced by Koch Membrane Systems U.S, is analyzed. Again, flux data of various solvents including alkanes, alcohols, and different acetates was reported. The division into the three data sets is indicated in Table 7.6.

Table 7.6: Training, test and validation data set for the data of Machado et al. [3]

	Alkanes	Alcohols	Acetates
Training	n-pentane	methanol	butyl acetate
	n-decane	ethanol	methyl acetate
	1-pentanol		
Test	n-heptane	1-butanol	ethyl acetate
Validation	n-octane	1-propanol	

Based on the results of the first case study only models with linear parameter relationships are derived in ten consecutive runs of the method. The results of these runs are listed in Table 7.7. The best model is determined in three of the ten runs, while nine of the ten runs result in model candidates with almost equal fitness and $\bar{\varepsilon}$. Only model 4, which was created in run 5 drops out to a recognizable extent. However, all models show mean percentage errors $\bar{\varepsilon}$ below the expected deviation of 20% (Chapter 4).

Figure 7.7 illustrates again a representative progress curve of the second run, indicating that the best model candidate is determined after nine generations and that

Table 7.7: Results of ten runs with maximum tree depth three for the data set of Machado et al. [62] without ketones sorted by fitness

Model Nr.	Run	Fitness	Training & test $\bar{\varepsilon}$	ε_{max}	Validation $\bar{\varepsilon}$
1	2,7,9	-52.33	3.2%	11.0%	17.4%
2	8	-51.94	3.0%	10.3%	14.8%
3	1,3,4,6,10	-50.37	3.6%	13.9%	9.2%
4	5	-47.11	6.9%	13.9%	3.8%

the algorithm terminates after 33 generations due to a lack of improvement in the fitness function.
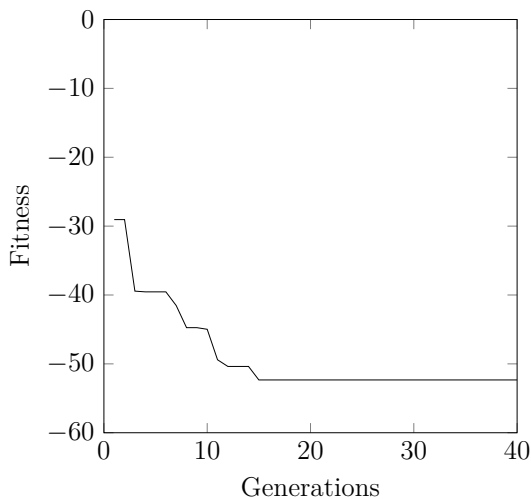
Figure 7.7: Progress curve for the second run for the data set of Machado et al. [62].

The determined model

$$P_M^1 = 0.02 \cdot \frac{Ra}{\Delta\delta_M} - 1.22 \cdot \delta_h - 76.70 \cdot d_{eq} + 71.6 \tag{7.15}$$

correlates the radius of the interaction sphere Ra and the difference of the Hildebrand solubility parameters $\Delta\delta_M$ of solvent and membrane, the equivalent molecular diameter d_{eq} and the Hansen solubility parameter for hydrogen bonds δ_h of the solvent. It does not only show the best fitness value, but also the lowest mean ($\bar{\varepsilon}$) and maximum percentage error (ε_{max}) for training and test data as well as validation

data. The second best model candidate

$$P_M^2 = 9.24 \cdot ConI - 0.05 \cdot \rho - 3.43 \cdot \Delta\delta_M - 243.35 \cdot d_{eq} + 204.22 \qquad (7.16)$$

correlates the connectivity index $ConI$, the density ρ, the equivalent molecular diameter ρ of the solvent and the difference of the Hildebrand solubility parameters of solvent and membrane $\Delta\delta_M$. Unlike the models created for the PDMS-based membrane in the first case study, none of the most promising models for the MPF-50 Koch membrane includes the viscosity. Rather, both models are based on different properties describing size and polarity of the solvent, indicating the benefit of membrane-specific models.

However, the third and fourth best model found for the data set of Machado et al. [62] include the reciprocal of the viscosity but were rated worse than the models that do not include the viscosity. The equations for these models as well as a figure illustrating the permeance plotted over the reciprocal of the viscosity of the solvents are provided in the Appendix (cf. Figure D.2). Comparing both data sets, the relation of the reciprocal of the viscosity and the permeance is stronger for the data set of Dijkstra et al. [51] such that the reciprocal of the viscosity is determined as significant factor for the most promising models. Nevertheless, the viscosity does have a similar influence on the permeance of solvents through the MPF-50 Koch membrane as well, but this is not as significant for the available data. If further data or other solvent groups are considered, this influence might appear more significant and could therefore be considered in the most promising models. Yet, based on the criteria of accuracy and parameter precision, other factors are deemed more significant for the permeance of solvents through the MPF-50 Koch membrane.

While the structure of model P_M^1 and P_M^2 is apparently different, there is no significant difference in the quality of the predicted permeance values. This is further illustrated in Figure 7.8 which provides parity plots of the calculated and experimentally measured permeance for the training and test data set (filled symbols) and the validation data set (unfilled symbols).

A deviation of $\pm15\,\%$ is exceeded only for one solvent from the validation data set in both cases. Since there are no significant difference between the quality of fit of both models, both are further evaluated in the post processing step, together with several
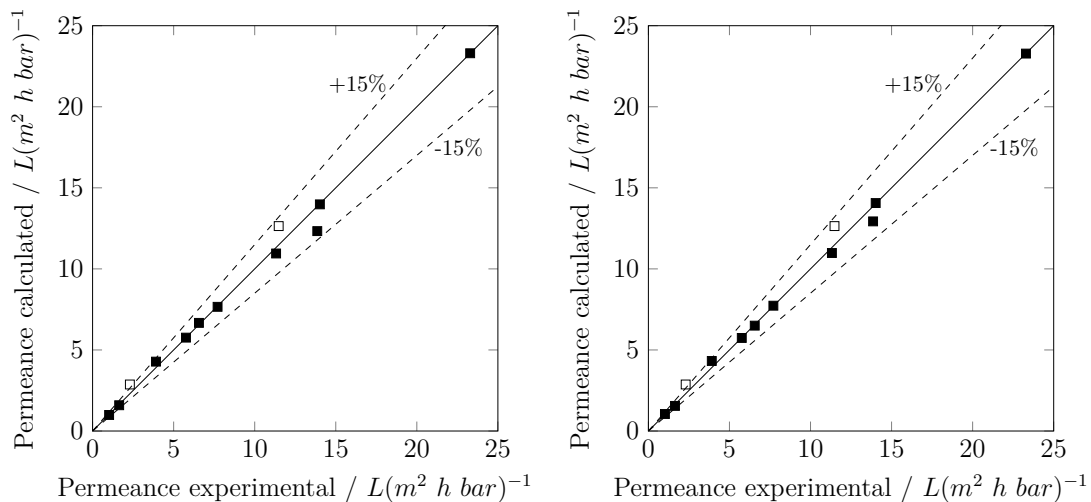
Figure 7.8: Calculated vs. experimental permeance values for training and test (filled) as well as validation (unfilled) data [62] for model P_M^1 (cf. Equation 7.15) (left) and model P_M^2 (cf. Equation 7.16) (right).

literature models with additional parameter regression. The results are summarized in Table 7.8.

According to the Akaike weights W both determined model candidates P_M^1 and P_M^2 are clearly favored over the respective literature models, while P_M^1 has a slightly higher probability compared to model P_M^2. While none of the confidence intervals includes zero for both models, parameter precision for model P_M^1 is extremely high, according to the tight confidence intervals. This is apparently not the case for the literature models of Geens et al. [55] and Darvishmanesh et al. [1], which also show a considerable worse fit of the experimental data according to the mean percentage error $\bar{\varepsilon}$. Only for the model of Bhanushali et al. [4] the single parameter can be identified accurately, while the quality of fit is still considerable worse, compared to the two developed model candidates P_M^1 and P_M^2.

7.4.2 Prediction of mixed solvent permeance

After illustrating the successful application of the method for the prediction of pure solvent permeance for polymeric membranes, the last case study illustrates the application to the prediction of pure and mixed solvent permeance for ceramic membranes, considering permeation data reported by Marchetti et al. [2]. The article reports data for eleven solvents, as well as different binary mixtures, for four ce-

Table 7.8: Summary of fitness, mean percentage error ($\bar{\varepsilon}$), AIC weight (W), parameter value (θ), size of 95 % confidence interval (CI) and eigenvalues (λ) for the different models, considering training and test set.

Model	Fitness	$\bar{\varepsilon}$	W	θ	CI	λ
Model P_M^1	-52.33	3.2%	54.7%	0.02	0.00	0.01
				-1.22	0.05	792.15
				-76.70	3.21	2.30
				71.60	2.70	$7.31{\times}10^4$
Model P_M^2	-51.94	3.0%	44.9%	9.24	3.09	0.00
				-0.05	0.01	0.10
				-3.43	0.37	5.91
				-243.35	48.59	141.94
				204.22	29.64	$3.86{\times}10^6$
Bhanushali	-30.27	23.0%	0.0%	0.49	0.06	$4.92{\times}10^3$
Geens	-13.13	50.2%	0.0%	0.26	0.40	$1.53{\times}10^4$
Darvishmanesh	-42.09	12.9%	3.3%	0.27	13.21	29.15
				7.13	7.65	391.03

ramic membranes (Inopor Nano 450 Da and 750 Da, Inopor Ultra 2000 Da and Sulzer 1000 Da). In order to compare the derived models with the advanced phenomenological model proposed by Marchetti et al. [2] (Section 2.2.4), the five solvents, which were considered for parameter estimation by Marchetti et al. [2], are allocated to the training data set, while all remaining data reported for a specific membrane was evenly distributed between the test and validation data set. Thereby, the test data set was composed of six pure solvents and 8 to 18 different binary mixtures, while the validation data set included data for 18 to 25 binary systems, depending on the specific membrane. Before summarizing the results for all four membranes, including a comparison with the phenomenological model proposed by Marchetti et al. [2], a more detailed description of the results for the Sulzer 1000 Da membrane is given first.

In order to apply the model development approach to the prediction of mixed solvent permeance, either a mixing rule can be applied or a combined permeance for the solvent mixture is determined based on properties of the mixtures. The latter was applied in the current case and the same literature data [126, 127, 128, 129, 130, 131] as considered by Marchetti et al. [2] was used for density, viscosity and surface tension of the mixtures. For those mixtures for which the compositions in literature differed from those investigated in this study linear interpolation was applied. The

remaining properties were determined by mixing rules based on the molar composition of the mixtures as it was done by Marchetti et al. [2] for the correction factor (Equation 2.45), except for molar volume and molecular diameter, which were determined based on the density and the molar mass, as well as the molar volume. Since no solubility parameters for ceramic membranes are available, the difference of the solubility parameters of the solvent and the membrane $\Delta\delta_M$ and the interaction radius Ra were not used for ceramic membranes.

Application of the model development approach, with similar settings as for the previous case studies, did provide the results summarized in Table 7.9 for ten consecutive runs.

Table 7.9: Results of ten runs with maximum tree depth three for the data set of Marchetti et al. [2] measured with the membrane Sulzer 1000 Da sorted by fitness

Model Nr.	Run	Fitness	Training & test $\bar{\varepsilon}$	ε_{max}	Validation $\bar{\varepsilon}$
1	1,3,5-7,9-10	-75.9	9.0%	36.9%	6.3%
2	8	-74.7	10.8%	30.0%	10.9%
3	2	-73.9	10.4%	33.1%	11.2%
4	4	-73.3	10.6%	31.7%	11.0%

While the best model candidates from each run show almost equivalent fitness, the best model is identified in seven of the ten runs. This model

$$P^1_{Mix} = 119.41 \cdot \frac{d_{eq}}{\eta} + 0.70 \cdot Pol \cdot \sqrt{\eta} - 57.56 \tag{7.17}$$

correlates the molecular diameter d_{eq}, the viscosity η and the polarity Pol of the solvent or solvent mixture and has three parameters, while the second model

$$P^2_{Mix} = 0.75 \cdot \frac{Pol}{\vec{d}} + 96.07 \cdot \frac{d_{eq}}{\eta} - 24.41 \tag{7.18}$$

correlates the same properties with the additional dipole moment $\vec{d}$, showing also three individual parameters. Since both models show almost equivalent fitness, the identifiability of the model parameters was further evaluated providing the data on confidence intervals and eigenvalues reported in Table 7.10. While the size of the 95 % confidence intervals for the parameters of both models are by far not as

narrow as in the previous case studies, they exceed the parameter values for two of the parameters of model P_{Mix}^2. Consequently, model candidate P_{Mix}^1, which is determined in the majority of runs is also considered favorable in terms of parameter precision.

Table 7.10: Summary of fitness, mean percentage error ($\bar{\varepsilon}$), parameter values (θ), size of the corresponding 95 % confidence interval (CI) and eigenvalues (λ) for the different models, considering training and test set from Marchetti et al. [2] for the Sulzer 1000 Da membrane.

Model	Fitness	$\bar{\varepsilon}$	θ	CI	λ
Model P_{Mix}^1	-75.9	9.0%	119.41	27.15	0.10
			0.70	0.38	3.82
			-57.52	41.27	1.68×10^4
Model P_{Mix}^2	-76.7	10.8%	0.75	2.48	0.14
			96.07	88.00	4.12
			-24.41	140.59	5339.14

In order to further illustrate the quality of fit of the developed model a parity plot of the model-based and experimentally measured permeances for the pure and mixed solvents is illustrated in Figure 7.9 for model P_{Mix}^1. As indicated by the dashed lines, only three of the overall 32 data points exceed a deviation of 20 %. These three data points correspond to binary mixtures, for which the approximation of the properties of the mixtures may result in slightly larger deviations, yet still within an acceptable error margin.

Through application of the proposed model generation approach, dedicated models with a comparable quality of fit and good parameter precision could be identified for all four membranes. The best model candidates for each of the membranes are listed in the Appendix in Table D.3. Table 7.11 further summarizes the results of a comparison of the newly developed models with the improved phenomenological model proposed by Marchetti et al. [2]. The evaluation of the latter model is based on the published parameters K_{HP} for each membrane, f_c of the solvents and the viscosity of the solvents or binary mixtures according to Equation 2.43. While a single model capable of accurately describing the different membranes would of course be favorable and the phenomenological model might provide higher accuracy for further extrapolations, parameter regression was performed on the basis of the same subset of the available data. The results indicate that for each of the membranes the individually derived model results in considerably lower mean per-
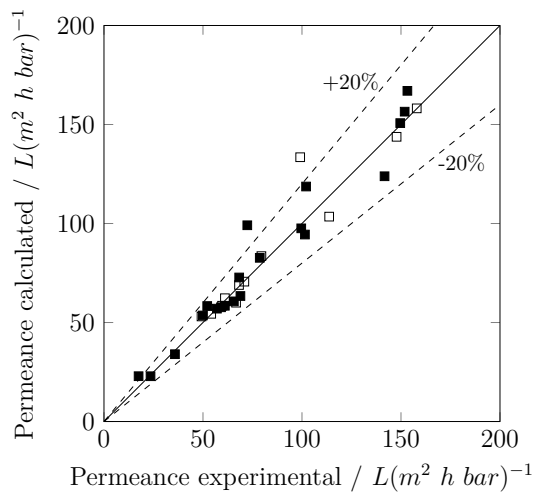
Figure 7.9: Calculated (filled) and predicted (unfilled) permeance using the model P_{Mix}^1 vs. experimental data reported for the Sulzer 1000 Da membrane by Marchetti et al. [2].

centage error $\bar{\varepsilon}$, while the Akaike weights W indicate clearly that the individual models are favored over the improved phenomenological model of Marchetti et al. [2], based on the considered data sets. Given the differences in model structure and enclosed descriptors, it might be concluded that different phenomena dominate the mass transfer for the different membranes, which indicates a potential limitation of a single membrane-independent model.

Table 7.11: Comparison of the newly developed individual models for each of the four membranes with the improved phenomenological model proposed by Marchetti et al. [2] considering training, test and validation data.

Membrane	New model		Marchetti et al. [2]	
	$\bar{\varepsilon}$	W	$\bar{\varepsilon}$	W
Innopor Nano 450 Da	25.9%	100.0%	35.8%	0.0%
Innopor Nano 750 Da	16.9%	100.0%	24.2%	0.0%
Innopor Ultra 2000 Da	15.6%	99.8%	15.0%	0.2%
Sulzer 1000 Da	9.0%	100.0%	11.9%	0.0%

Taking a closer look at the results it becomes obvious that interdependent of the model, the mean percentage error $\bar{\varepsilon}$ increases for the Inopor Ultra 2000 Da, Nano 750 Da and Nano 450 Da. One potential reason for the decreasing accuracy is related to the increasing number of binary data points for the latter membranes and the

larger deviation form a linear combination of the particular pure solvent permeance, which holds almost entirely for the Sulzer 1000 Da membrane. Nevertheless, even for most of the binary mixtures with minima in the permeance, the overall course of the permeance is represented more accurate with the newly developed model, compared to the phenomenological model of Marchetti et al. [2]. This is exemplarily illustrated for three binary mixtures in Figure 7.10 for the Inopor Nano 750 Da membrane.
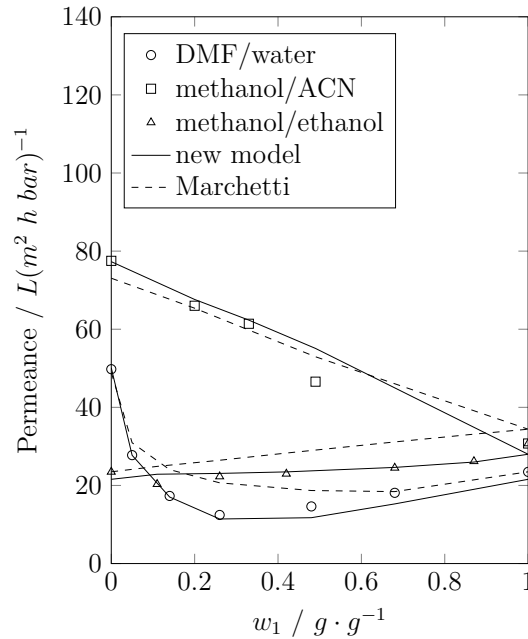
Figure 7.10: Experimentally measured (symbols) and calculated permeance values for the new individual model P^1_{Mix} (cf. Equation 7.17) (solid lines) and the phenomenological model of Marchetti et al. [2] (dashed lines) for the Inopor Nano 750 Da membrane.

7.5 Conclusion

Previous studies have primarily tried to develop a general phenomenological model for predicting the solvent flux for organic solvent nanofiltration membranes, based on the correlation of different thermodynamic and physical properties. However, the dominating mass transfer phenomena may vary for different types of membranes and the interaction with the different solvents, such that a single model structure can be insufficient for an accurate description of solvent flux. On the other hand, the limited availability of permeation data, as well as the variety of alternative descriptors for

molecular size and solubility of solvent molecules, limits the applicability of highly parameterized models, such as general response surface models and artificial neural networks, for the derivation of reliable quantitative structure-property relationships. In order to overcome the aforementioned limitations, a new hybrid stochastic-deterministic approach for model development was presented, which combines genetic programming with global deterministic optimization for parameter regression, as well as an analysis of parameter identifiability. The possibility to derive compact models which provide a good quality of fit with high parameter precision is demonstrated for three different case studies, covering polymeric membranes as well as ceramic membranes. For the latter also mixed solvent flux predictions are performed with high accuracy. The results for the different case studies indicate the potential for membrane-specific models, which in all cases outperformed the available phenomenological models, which were proposed in literature. Unlike the derived data-based models and with the sole exception of the correlation proposed by Bhanushali et al. [4], none of the available correlations showed sufficient identifiability of the regressed parameters, based on the available data sets.

The developed models represent an excellent tool to derive membrane rejection maps with significantly reduced experimental effort. Any membrane performance data available in literature can be used to derive possible data-based models using the presented approach, while membrane manufacturers are probably the most appropriate user of such a tool, building on in-house performance data, available for their membranes. Due to the consistent quality of commercially manufactured membranes, it is expected that membrane samples can be from the same or deviating batches, which follow the same production protocol and standards, as long as variations in membrane performance are within reasonable bounds (Chapter 4). However, the experiments should be conducted at the same conditions, to which the standardized measurement method described in Chapter 5 can be applied. The available amount of data should however at least reflect the diversity expected from the range of solvents that are to be evaluated, resulting in the challenge to produce meaningful interpolations that can be adapted to the desired scope of application. The more diverse data is included, the broader the potential range of applications for which the model can be applied. Yet the proposed method enables a reevaluation of a suitable model for an increasing amount of data at any time.

The most important next step is of course to extend the approach to the prediction

of permselectivity and rejection of dissolved solutes, in order to enable a prediction of the separation performance. These extensions are presented in Chapter 8.

8

Machine-based learning of predictive models: Rejection

For the prediction of solute rejection, a few phenomenological models have been developed, primarily for ceramic membranes, based on the assumption of a dominating size exclusion mechanism. However, especially for polymeric membranes the mutual interactions between membrane material, solvent and solute, need to be accounted for in order to accurately describe solvent flux and solute rejection. The dominating phenomena may strongly depend on the specific membrane, as well as the considered chemical systems. Building on the previous work on the automatic development of membrane-specific models for pure and mixed solvent flux, this chapter addresses the extension towards the prediction of solute rejection in pure and mixed solvents. For this purpose, automatically derived model candidates are evaluated with respect to accuracy and parameter precision for experimental data for the PuraMem® S600 membrane. Furthermore, the possibility to predict solute rejection in mixed solvent systems is evaluated based on the concept of membrane rejection maps, evaluating previously reported permeation data for a PuraMem® 280 membrane. The derived models show excellent accuracy and decent parameter precision, with deviations between measured and predicted solute rejection of less than 10% for most of the experimental data, while predictions for mixed solvents are in a similar range.

8.1 Introduction

In contrast to solvent flux through OSN membranes, as discussed in Section 2.2.4 and the previous Chapter 7, phenomenological models for solute rejection are rarely described in literature. Available studies [54, 132, 133] primarily focused on the analysis of the most important parameters influencing the rejection and the identification of interaction parameters between the membrane, solvent and solute as well as size parameters of the solutes affecting the separation (Section 2.2.5). Correlation of these parameters is consequently a logical step towards the prediction of solute rejection for OSN membranes. Yet, this correlation is not straightforward, since the number of potentially important parameters is high and the model structure is neither known a-priori, nor is it likely that a general model structure is valid for all membranes. Still, data-based methods for the development of solute rejection models in aqueous nanofiltration and reverse osmosis have shown promising results in different applications [47].

In the scope of this chapter, the method developed in Chapter 7 is extended in order to derive correlations for the prediction of the solute rejection in pure and mixed solvent systems. The investigations build on the determined rejection data for the PuraMem® S600 membrane, reported in Section 5.4 as well as previously reported permeation data of Schmidt et al. [65] for the PuraMem® 280 membrane. For the theoretical background, including available models and important parameters affecting solute rejection the reader is referred to Section 2.2.5, while the general approach for model development method is discussed in Section 8.2. The subsequent sections describe the application of the approach to the prediction of solute rejection (Section 8.2) and illustrate and discuss the results for the obtained models (Section 8.3). The resulting models allow for the prediction of the so-called membrane-rejection maps, which were experimentally derived by Schmidt et al. [65]. Finally, some conclusions and an outlook on future work are provided in Section 8.4.

8.2 Application of hybrid stochastic-deterministic approach for solute rejection

In order to evaluate the potential of an automatic model generation, two sets of solute rejection data were considered. The first set of data described in Section 5.4 considers a range of specifically selected solutes with different properties, solved

in six different solvents. The second set of data was extracted from the article of Schmidt et al. [65], who reported the rejection for five different solvents dissolved in the solvents toluene, hexane and isopropanol, as well binary and ternary mixtures of these solvents. The illustration of the so-called membrane rejection maps (MRM) allows for a graphical evaluation of the capability of the generated models to predict the rejection of the solutes in the ternary solvent mixtures. A brief summary of the experiments and the use of the proposed MRM of Schmidt et al. [65] is provided in Section 8.2.1. Finally, Sections 8.2.2 and 8.2.3 provide a summary of the application of the previously developed hybrid method for model development (Chapter 7).

8.2.1 Solute rejection data for ternary solvent mixture

The data set for ternary solvent mixtures used in this study was taken from Schmidt et al. [65], who determined the rejection of five different solutes (n-hexadecane (HD), 2,2,4,4,6,8,8-heptamethylnonane (HMN), phenyldodecane (PDD), 2,6-Diisopropylnaphthalene (DIPN), Triphenylphosphine(TPP)) solved in the solvents toluene, hexane and isopropanol, as well as binary and ternary mixtures of these solvents. Important solvent and solute properties are listed in table 1 and 2 of the article of Schmidt et al. [65]. The rejection experiments were conducted in the cross-flow system presented in Section 5.3.1.

For the investigated PuraMem$^®$ 280 (PM280) a MWCO of $280\,\mathrm{g\,mol^{-1}}$ is reported by the vendor determined under similar conditions as stated in Section 5.2 for the PuraMem$^®$ S600. Overall, 21 rejection experiments were reported for pressures of 20, 30 and 40 bar at 25 °C. While all five solutes were solved in the solvent for each of the measurements, the solvent was varied between the three pure solvents, binary mixtures at 25, 50 and 75 % of the primary solvent and nine compositions of ternary mixtures distributed in the Gibbs triangle, as illustrated in Figure 8.1.

Based on the measured solute rejections Schmidt et al. [65] proposed the so called MRM, which were obtained by smoothing and interpolating the measured rejections and which provide a useful insight of the rejection behavior of a solute in different solvent mixtures. In a subsequent publication, Schmidt et al. [134] illustrated how MRM can be used in order to identify solvent mixtures that enhance a specific separation task and applied them in a conceptual design work flow for the recycling of homogeneous catalysts in hydroformylation.

While the data set of Schmidt et al. [65] is by far not as diverse in terms of the considered solutes as that reported in Section 5.3 and 5.4, the available rejection
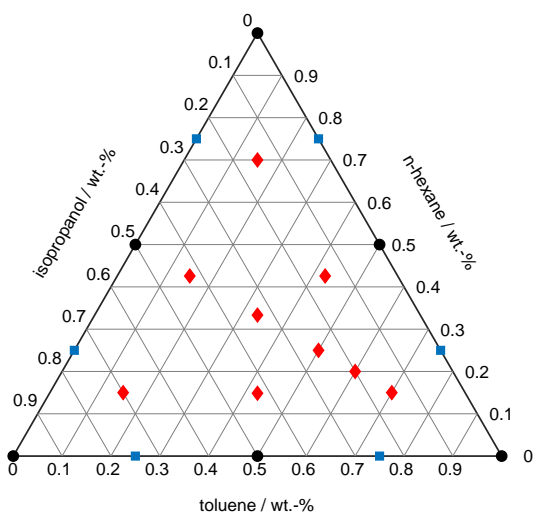
Figure 8.1: Composition of the different solvent mixtures for which solute rejection was evaluated (black circle: training data, blue square: test data, red diamond: validation data).

data for solvent mixtures allows for an evaluation of the potential prediction of solute rejection in mixed solvent systems and the possibility to predict the MRM proposed by Schmidt et al. [65] based on considerably less data.

8.2.2 Pre-processing of rejection data

The pre-processing for the rejection data was based on the steps presented in Section 7.3. The allocation of the rejection data of the solutes to the different data sets used for model development in the current study ensures that each functional group is considered within the training data (cf. Figure 8.2). The same allocation was used for all solvents. For the rejection data that was extracted from the article of Schmidt et al. [65] the evaluation was focused on the potential prediction of the solute rejection in the mixed solvent systems. Therefore, the training data set was based on the rejection data of all solutes solved in the pure solvents as well as the equimolar binary solvent mixtures. The remaining rejection data for the binary solvent mixtures was allocated to the test data, while the rejection data for all ternary mixtures was considered as validation data (cf. Figure 8.1).

As second preparation step, the potential descriptors for the model generation need to be defined. As previously described in Section 2.2.5, solute transport is influenced by solvent-solute-membrane interactions, for which Geens et al. [54] already pointed
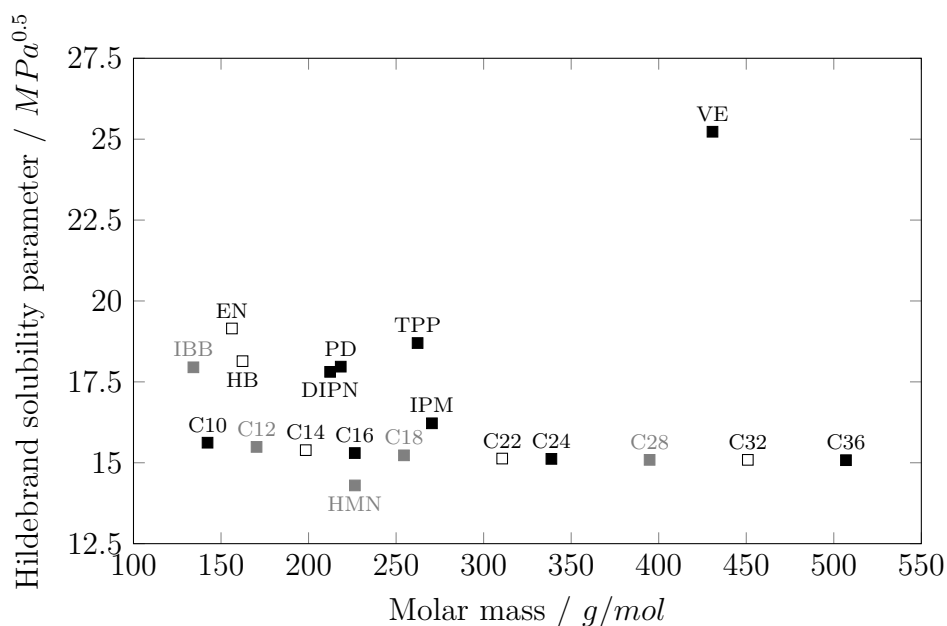
Figure 8.2: Allocations of the rejection data of the solutes solved in pure solvents to the different data sets. (black: training data, gray: test data, unfilled: validation data).

out important physical and chemical properties of solute, solvent and membrane, affecting solvent-solute, solute-membrane and solvent-membrane interactions. In a similar fashion Marchetti et al. [14] classified and highlighted important properties, as presented in Figure 8.3. Apart from the direct solute-membrane interactions, which according to Geens et al. [54] are expressed by polarity and charge effects, especially the interaction of the solvent with the solute and the membrane need to be accounted for, since solute rejection depends strongly on the solvent [135].

The most prominent property linked to these interactions is the solubility, for which mostly Hildebrand and Hansen solubility parameters are considered. A comparison between the difference between solute, solvent and membrane solubility parameters provide an indication of the preferred permeating component, in case solubility has a predominant effect on the mass transfer. E.g. Zeidler et al. [84] showed considerable variations in solute rejection for a GMT-oNF-2 (PDMS-PAN composite) membrane for three different solvents (THF, n-heptane, ethanol). They specifically indicated that negative rejections are linked to a smaller distance between the solubility parameters of the solute and the membrane compared to the solvent and the membrane. In a more recent study, Thiermeyer et al. [78] found an increasing
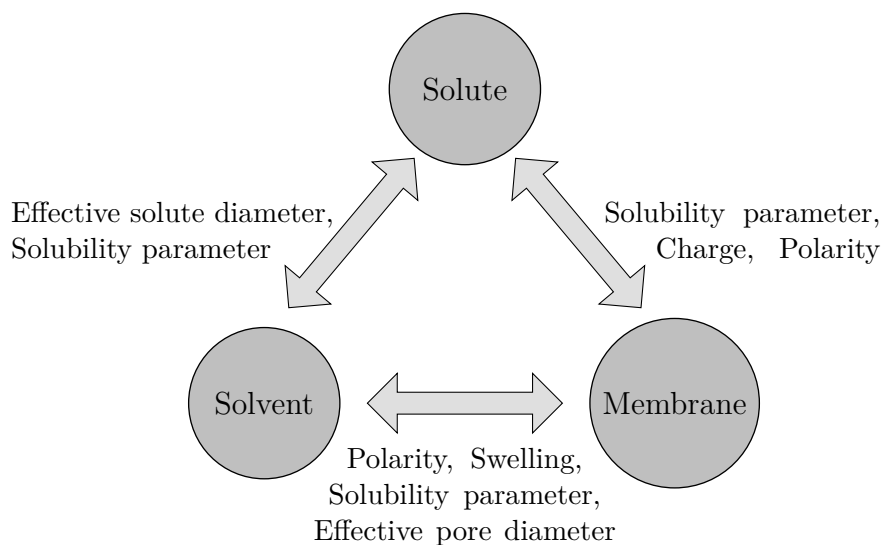
Figure 8.3: Properties affecting the interactions between Membrane, Solvent and Solutes [14]

influence of the solute properties on the rejection with increasing hydrogen bonding capacity of the solvent, characterized by the Hansen solubility parameter for hydrogen bonds $\delta_{h,i}$. Based on the experimental study of solute rejections of a PDMS and a PI-based membrane, for 5 solvents and 17 solutes, Thiermeyer et al. [78] evaluated three different solvent dependent solute sensitivity regions, in which the effect of functional groups of the solutes differs significantly. While the rejection of a solute mainly depended on the molecular size for the non-polar solvents heptane and toluene, solute rejection was strongly affected by the functional groups of the solutes in the polar solvents ethanol and isopropanol. In moderate polar solvents like ethyl acetate or methyl ethyl ketone these dependencies were still observable but much weaker.

In summary, the impact of solubilities on the solute rejection, characterized by the Hildebrand and Hansen solubility parameters, has been demonstrated by several studies and in multiple ways. Consequently, these parameters should be included in the development of solute rejection models, as already described in different publications [136, 84, 117].

Another important property that has been characterized as strongly affecting solute rejection is the dipole moment $\vec{d}$ of the solute [137]. Van der Bruggen et al. [133] already observed a decreasing rejection with increasing dipole moment of the solutes in their study of solute rejection for a variety of organic solutes in aqueous

nanofiltration. They attributed the dependency to the electrostatic interaction of the molecules, which results in different orientations of the solutes relative to the membrane. More recently [137] confirmed these results for OSN, also reporting a relationship between dipole moment and solute rejection.

Finally, the size of the solute is of crucial importance for the rejection. Commonly, as for aqueous nanofiltration, the molecular weight M is used as an indicator for the size of the solutes and molecular weight cut off (MWCO) values, referring to the molecular weight of a solute that is rejected by at least 90 %, are reported by manufactures to characterize the filtration performance of the membrane. However, as already stated by Geens et al. [55], the MWCO specified in water is not of much use for non-aqueous systems, while the statement holds generally true for varying solvents, since the effective solvent diameter differs due to salvation effects Geens et al. [54]. While different studies consequently promote the use of an effective solvent-dependent solute diameter [112, 54, 57], other reports indicate that the predicted effects on solute size were not reflected by the measured solute rejection [20, 56]. In any way, correlations for the calculation of solvent dependent solute size, require additional parameters, such as the association parameter of the solvent, which may be hard to retrieve, such that simple parameters for the molecular size are much more practical for application. Such parameters were evaluated in the study of Van der Bruggen et al. [133], evaluating the correlation between solute rejection, the molecular weight, the Stokes diameter, the equivalent molar diameter, as well as a diameter obtained with energy minimization calculations. They concluded that a good correlation with each of the parameters was feasible for the investigated organic solutes and three of four investigated membranes for aqueous nanofiltration. While the molecular weight is the easiest accessible parameter, the equivalent molar diameter d_{eq} provided a better correlation with solute rejection [112]. While Darvishmanesh et al. [137] also reported a good correlation between equivalent molar diameter d_{eq} and rejection. It needs to be noted that the resulting plots were slightly scattered. Yet, the equivalent molar diameter describes the size of a molecule, but it neglects the molecular structure [133].

In the scope of this study 29 different solvent and solute properties were selected based on the discussion above and the previous application for the generation of solvent flux models (Chapter 7). These descriptors are summarized in Table 8.1, while the respective solvent and solute property data is provided in Table D.4 in Chap-

ter D and Table E.21 in Chapter E in the Appendix. Properties of solvent mixtures were estimated by using a linear mixing rule and the pure solvent properties.

Table 8.1: List of all descriptors and the corresponding symbol used in model development (Polarity is only used for solvents)

Group	Descriptor	Symbol
Size	Molar mass	M
	Molar volume	V_m
	Equivalent molecular diameter	d_{eq}
Polarity	Dielectric constant	ϵ
	Surface tension	γ
	Hildebrand solubility parameter	δ
	Hansen solubility parameters (dispersive, polar, h-bonds)	$\delta_d,\delta_p,\delta_h$
	Difference of Hildebrand solubility parameters of membrane and solute/solvent	$\Delta\delta_M$
	Interaction radius	Ra
	Dipole moment	$\vec{d}$
	Polarity	Pol
Further	Density	ρ
	Viscosity	η

8.2.3 Application of hybrid method for model development

The hybrid method for automatic model development introduced in Section 7.3 was adapted by the previous modification of the pre-processing of the rejection data and further extended by dedication routines that allow for the prediction of the MRM for ternary systems, building solely on experimental data for pure and binary solvent systems.

8.3 Results and Discussion

The results of the model development approach are evaluated accounting for solute rejection in the different solvents, as well as in specific solvent groups. Finally, an evaluation of the model development approach for the solute rejection in mixed solvent systems according to the solute rejection data of Schmidt et al. [65] is presented in Section 8.3.2.

8.3.1 Prediction of solute rejection in pure solvents

The potential of the data-based model development approach is evaluated in two scenarios. The first scenario evaluates the application to solvent-group dependent rejection models, with respective division into training, test and validation data (cf. Figure 8.2) for each individual solvent group. In order to evaluate the possibility of deriving a model with higher predictive power also a solvent independent model was determined in the second scenario, considering rejection data from all six solvents.

Development of solvent-dependent solute rejection models

In order to develop solute rejection models for the specific types of solvents, the available rejection data was divided according to the three described solvent groups of polar, moderate polar and non-polar solvents. For each solvent group individual solute rejection models were determined based on repeated runs of the automatic model development approach, considering only linear parameters in a first step and a tree depth of tree. The tree depth was further increased in case a mean absolute deviation of 5 % was exceeded in the post-processing of the most promising candidate. The resulting statistics of the most promising model candidates are further listed in Section E.1.1-E.1.3 in Table E.1-E.6 of the Appendix, while parity plots for these models and the different solvent groups are illustrated in Figure 8.4.

Polar solvents

For polar solvents a maximum tree depth of three sufficed. In total, five different models were identified in the ten runs for polar solvents, which barely differ in their fitness (cf. Table E.2 in Section E.1.1 of the Appendix). Each model candidate includes at least the difference of the Hildebrand parameters of membrane and solute $\Delta\delta_{M,i}$, as well as another, varying polarity parameter of the solute. As a property of the solvent, the size or the polarity is considered in each of the models, however, in this case both properties are correlated with each other, since both solvents belong to the homologous series of alcohols. This leads to the fact that there is no clear indication to take only one of these properties into account and therefore, even if the size of the solvent molecule is included, the polarity is represented to a certain extent as well. Overall, it can be stated that the polarity of solute and solvent is crucial, which is consistent with the observations from the experiments (Section 5.4.1) and those of Thiermeyer et al. [78]. Furthermore, the Akaike weights for these models indicate that the best and the second-best model with a similarly high probability
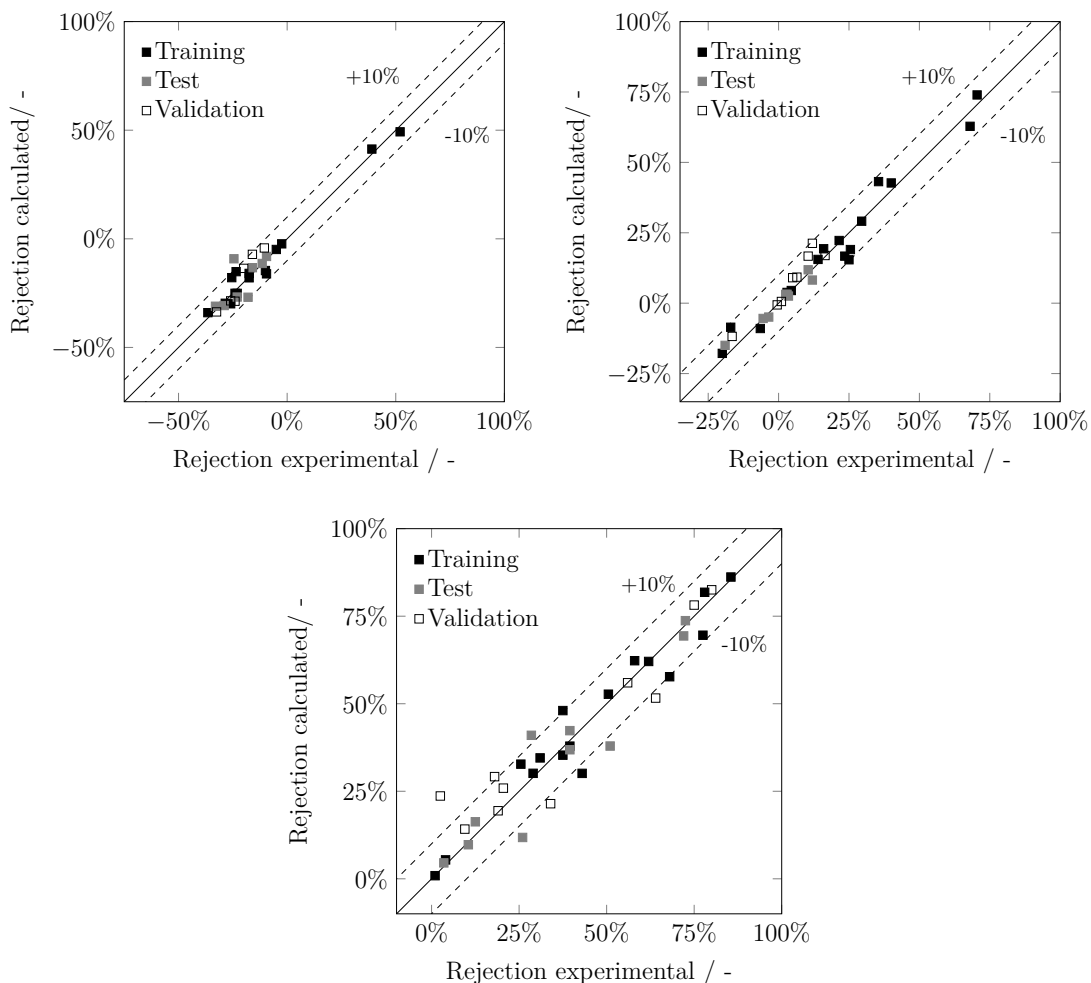
Figure 8.4: Calculated vs. experimental rejection of all solutes in the different solvent groups for the individual models: polar solvents and $R_{p,i}^3$ (top left), moderately polar solvents and $R_{m,i}^4$ (top right) and non-polar solvents and $R_{np,i}^3$ (bottom).

are the most suitable ones among the five models for the available data. While the parameter identifiability for the second-best model is limited, all parameters of the model with the best fitness are identifiable. Therefore, this model is examined in more detail in the following.

The parameter precision of the best model, according to AIC, is high, since the confidence intervals are small (cf. Table E.1 in Section E.1.1 of the Appendix). The best model in these metrics

$$R_{p,i}^3 = 1.067 \cdot \frac{\delta_{h,i}}{\gamma_i} + 1.750 \cdot \frac{\Delta\delta_{M,i}}{\delta_s} - 0.376, \qquad (8.1)$$

results in mean ($\hat{\varepsilon}$) and maximum absolute deviation (ε_{max}) of less than $4\,\%$ and $15\,\%$ respectively. The resulting model is rather simple, correlating the Hansen solubility parameter $\delta_{h,i}$ of the solute accounting for hydrogen bonds, the surface tension of the solute γ_i, as well as the difference in Hildebrand parameters of membrane and solute $\Delta\delta_{M,i}$ and the Hildebrand parameter of the solvent δ_s, using three fitting parameters. Hence, only polarity properties are taken into account in this model and the model allows for a considerably accurate representation of the experimental data, for which an absolute deviation of $\pm10\,\%$ is only exceeded for the solute heptamethylnonane in ethanol, which is allocated to the test data.

Moderately polar solvents

Because none of the model candidates with a tree depth of three did result in mean absolute deviation of less than $5\,\%$, model candidates with a tree depth of four were created for the group of moderately polar solvents. These models allowed for a significant improvement in terms of the AIC, as well as the mean and maximum absolute deviation. In the ten performed runs for a tree depth of four, ten different models with comparable accuracy were identified (cf. Table E.4 in Section E.1.2 of the Appendix). Each of these model candidates includes at least one polarity parameter of solvent and solute, as well as one size parameter of the solute, while the particularly selected property parameter of these groups differs. As indicated by the experimental data (Section 5.4.1) and mentioned by Thiermeyer et al. [78], for the group of moderately polar solvents, both size and polarity of the solute is crucial, since solutes with equal molecular weights as well as solutes with comparable polarity like the alkanes differ in their measured rejections. Moreover, a significant influence of the solvent is observed, which in this case is considered by polarity parameters, since the size of the solvents does not differ significantly. Taking the Akaike weights into account, the model with the highest fitness is indicated to be the most suitable one among these ten model candidates by an Akaike weight of $98.8\,\%$. Hence, this model is selected for more detailed discussion.

The parameters of the best model in terms of accuracy

$$R^4_{m,i} = -0.0173\cdot\delta_{h,i}\cdot\Delta\delta_{M,i}+0.003\cdot\Delta\delta_{M,i}\cdot Pol_s-0.181\cdot\frac{V_{m,i}}{M_s}+0.044\cdot\delta_{h,i}+0.004\cdot M_i-0.434,$$

$$(8.2)$$

are determined precisely (cf. Table E.3 in Section E.1.2 of the Appendix), which is another advantage of this model compared to the best model determined for a tree depth of three. Model $R^4_{m,i}$ shows a mean and maximum absolute deviations of less

than 4 % and 10 % respectively. The resulting model correlates the Hansen solubility parameter $\delta_{h,i}$ of the solute accounting for hydrogen bonds, the difference between the Hildebrand parameters of the membrane and the solute $\Delta\delta_{M,i}$ the molecular mass M_i and the molar volume $V_{m,i}$ of the solute, as well as the Polarity Pol_s and the molar mass M_s of the solvent and six fitting parameters. The model is more complex than $R_{p,i}^3$ and accounts for both size and polarity properties of the solute and the solvent.

Non-polar solvents

In ten consecutive runs for non-polar solvents with a tree depth of three only two different models with comparable fitness were identified (cf. Table E.6 in Section E.1.3 of the Appendix). Both models include the size of the solute as an influencing factor, but differ in the remaining incorporated properties, which complies again with the experimental results, revealing a strong correlation of solute size and rejection. However, the rejection of the same solute differs for different solvents, indicating an influence of the solvent as well. While the model with the best fitness includes the molecular weight of the solvent, the dispersive Hansen solubility parameter of the solvent is included in the second-best model. Since both properties account for the difference of the solvents, the method is not able to distinguish between these properties due to the empirical character and one out of these options is selected based on the effect on the fitness, although this does not necessarily have to correspond to the underlying physical effect. In this case, both the molecular weight and the dispersive Hansen solubility parameter of the solvent are indicated as useful properties to characterize the difference of the solvents.

The Akaike weights identify the model with the best fitness to be the most suitable one with a probability of 63.8 %. Moreover, the second-best model lacks in terms of parameter precision. Hence, the most promising model with a maximum tree depth of three

$$R_{np,i}^3 = 0.0021 \cdot \rho_i \cdot d_{eq,i} - 0.0004 \cdot V_{m,s} \cdot \delta_i - 0.744 \tag{8.3}$$

is discussed in detail. This model shows high parameter precision and mean and maximum absolute deviations of less than 5 % and 15 % respectively (cf. Table E.5 in Section E.1.3 in the Appendix). The model is of similar complexity than $R_{p,i}^3$, while correlating the density ρ_i, equivalent molecular diameter $d_{eq,i}$ and the Hildebrand solubility parameter δ_i of the solute with the molar volume $V_{m,s}$ of the solvent. Thus, both size and polarity of the solute are considered, while solvent size is accounted

for as well. As illustrated in Figure 8.4, the data is more scattered compared to the results of polar and moderately polar solvents. However, except for one data point, all deviation between $10\,\%$ and $14\,\%$ are allocated to the solvent heptane, indicating that the solvent properties might not be represented well enough. Although, increasing the tree depth to four did not result in a more accurate model (cf. Table E.5 in Section E.1.3 in the Appendix).

Evaluation of non-linear parameter models

The results for models with non-linear parameter relations are summarized in detail in Table E.7 and E.8 in Section E.1.4 in the Appendix. For polar and non-polar solvents the best identified models were exactly the same as determined considering only linear parameter relations (cf. $R_{p,i}^3$ and $R_{np,i}^3$). This was confirmed in ten and five consecutive runs for these solvent groups. For the groups of moderately polar solvents, the consideration of non-linear parameter relations did not result in an improved model, but rather resulted in an alternative model with similar accuracy to model $R_{m,i}^4$ in terms of mean and maximum absolute deviation ($3.8\,\%$ and $10.5\,\%$), including five linear parameters, rather than the six fitting parameters

$$R_{m,i}^{NL,4} = 0.011 \cdot M_i - 0.001 \cdot \frac{\gamma_i}{\delta_{p,i}} - 4.155e-4 \cdot \Delta\delta_{M,s} \cdot V_{m,i} - 1.441e-4 \cdot Ra_i \cdot M_s - 0.052$$

$$(8.4)$$

However, the confidence intervals of the different parameters indicate that at least two of the five parameters cannot be identified precisely from the available data. Besides the number of empirical parameters, the models differ in the incorporated solvent polarity parameter. Both models include the molar mass of the solute M_i and solvent M_s and the molar volume of the solute $V_{m,i}$. Additionally, in model $R_{m,i}^4$ the Hansen solubility parameter for hydrogen bonds of the solute, the difference between the Hildebrand parameter of the membrane and the solute and the polarity of the solvent are included, while model $R_{m,i}^{NL,4}$ incorporates the polar Hansen solubility parameter $\delta_{p,i}$, the interaction radius Ra_i and the surface tension γ_i of the solute, as well as the Hildebrand solubility parameter $\Delta\delta_{M,s}$ of the solvent. Hence, both models share a decisive dependency on polarity parameters, while the specific parameters differ for both models. Consequently, it may be argued that the experimental data is reflected approximately equally well by different model candidates, considering the same property groups, but different property parameters.

As already reported for model development of pure and mixed solvent flux in the previous chapter the evaluation of non-linear parameter relations results in a signif-

icant increase in computational effort, requiring 2 to 16 h depending on the data set and tree depth, instead of an average 2 min, required for the evaluation of models with only linear parameter relations. As the results of the exemplary runs for non-linear parameter relations do not indicate a considerable improvement in terms of accuracy and parameter precision, the subsequent investigations focused on models with linear parameter relations only.

Predictions for solvents not included in model development

Model $R_{m,i}^4$ is further evaluated for experimentally determined rejection data for the solvents acetone and n-butyl acetate, which were not considered at all for model development. Figure 8.5 illustrates the parity plot for the experimentally determined and predicted solute rejections in the two moderately polar solvents.
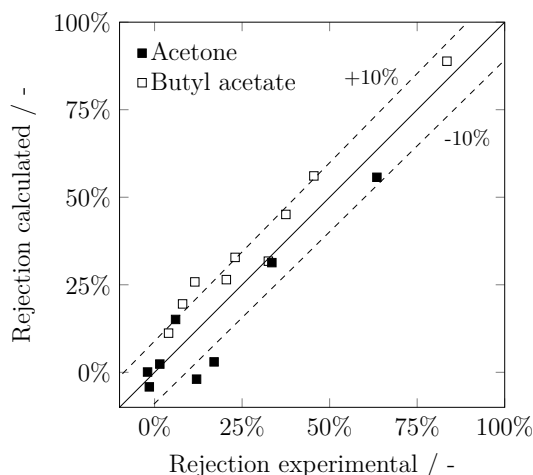
Figure 8.5: Calculated vs. experimental rejection of the solutes in the solvents acetone and n-butyl acetate using model $R_{m,i}^4$

While the deviations are slightly higher compared to the results for the solvents ethyl acetate and methyl ethyl ketone, which were considered for model development, the mean deviation is still below 8 % and a deviation of 15 % is never exceeded. Hence, the predicted rejections in different solvents, not included in the data sets for model development are in excellent agreement with the experimental data, proving the applicability of the model for predictions.

Development of a solvent independent solute rejection model

While the rejection models for the polarity-based solvent classes show very good performance, a solvent-independent model would of course allow for a broader ap-

plication. For this purpose, the model development methodology was applied to an extended data set for all considered solvents, performing ten runs with a maximum tree depth of four and five, allowing for more complex correlations. For both tree depth nine different models were identified in ten consecutive runs each (cf. Table E.10 and E.11 in Section E.1.5 in the Appendix). Regardless of the tree depth, each model includes at least parameters characterizing the size of the solute and the polarity of both solvent and solute, which matches the combined results for the three solvent groups. Considering only models with tree depth of four, the model with the best fitness is the most suitable one with a probability of 100%, while for a tree depth of five the model with the best and second-best fitness is rated with a probability of 71.9% and 28.1% to be the most suitable one respectively.

The best model candidates according to the AIC were further evaluated based on their mean and maximum absolute deviation, with respect to training and test data, as well as parameter precision. Although the best model in terms of accuracy for a tree depth of five did not provide sufficient parameter precision, the second-best model did provide considerable smaller intervals, while still resulting in comparable a mean and maximum absolute deviation (cf. Table E.9 in Section E.1.5 in the Appendix). Thus, the second best model is evaluated further. The most promising model candidates determined for a tree depth of four

$$R^4_{all,i} = -2.457e-7\cdot(\gamma_i - M_s)\cdot V_{m,s}\cdot M_i + 0.002\cdot\delta_{d,i}\cdot\gamma_s + 0.011\cdot\vec{d}_i\cdot\delta_{p,s} - 0.912 \quad (8.5)$$

and five

$$R^{5.2}_{all,i} = \frac{\vec{d}_s}{\Delta\delta_{M,s}}\cdot(0.085\cdot\epsilon_s + 2.085e-4\cdot\rho_i\cdot\vec{d}_i) + 0.606\cdot\frac{M_i}{V_{m,s}\cdot\epsilon_s} + 0.024\cdot\Delta\delta_{M,i} - 0.295$$
$$(8.6)$$

are rated almost equivalent in terms of mean ($\hat{\varepsilon}\approx 8.2\%$ and 8.7%) and maximum absolute deviation ($\epsilon_{max}\approx 35\%$ and 30%), whereas the more complex correlation enables an improved AIC value. The resulting model statistics are listed in more detail in Table E.9 in Section E.1.5 in the Appendix. Both models account for the molecular weight M_i of the solute, the dipole moment $\vec{d}_i$ of the solute and the molar volume $V_{m,s}$ of the solvent. Model $R^4_{all,i}$, however, also includes the surface tensions of the solute γ_i and solvent γ_s, the molar mass of the solvent M_s, as well as the dispersive Hansen parameter of the solute $\delta_{d,i}$ and the polar Hansen parameter of the solvent $\delta_{p,s}$. While considering the same number of property parameters, model

$R_{all,i}^{5.2}$ includes the difference of the Hildebrand parameter of solute and membrane $\Delta\delta_{M,i}$, as well as solvent and membrane $\Delta\delta_{M,s}$, the dielectric constant ϵ_s and the dipole moment $\vec{d}_s$ of the solvent, as well as the density of the solute ρ_i.

While presenting much more complex correlations, model $R_{all,i}^4$ and $R_{all,i}^{5.2}$ have a comparable number of fitting parameters as the solvent-dependent models $R_{p,i}^3$, $R_{m,i}^4$ and $R_{np,i}^3$.

However, as expected an extended applicability to a wider solvent space, comes at the cost of a reduced accuracy. As indicated by the parity plots in Figure 8.6, both models allow for good approximations for solute rejections. For both models $\approx 70\%$ of the model predictions deviate by less than 10% and more than 90% of the calculated rejection values do not differ more than 20% from the experimental values.
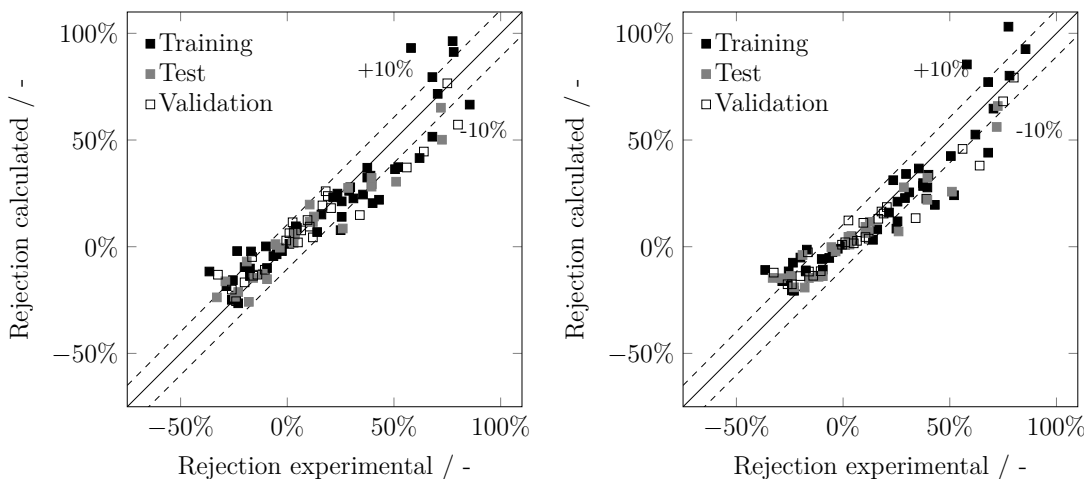
Figure 8.6: Calculated vs. experimental rejection of all solutes in each solvent using model $R_{all,i}^4$ (left) and model $R_{all,i}^{5.2}$ (right)

As both models do reflect the effect of solvent properties, the predicted solute rejections for the solvents acetone and n-butyl acetate, are expected to deviate from the previously derived model $R_{m,i}^4$. This deviation is illustrated in the parity plots in Figure 8.7 (compare with Figure 8.5). While there is a distinct tendency of the model $R_{all,i}^4$ to over- and underestimate the solute rejection depending on the solvent, the model predictions of model $R_{all,i}^{5.2}$ are considerably more accurate, showing a mean deviation of 7.2% and deviations of generally less than 15%. Hence, the predictions for the additional solvents using model $R_{all,i}^{5.2}$ are as accurate as using model $R_{m,i}^4$ derived based on only one solvent group.
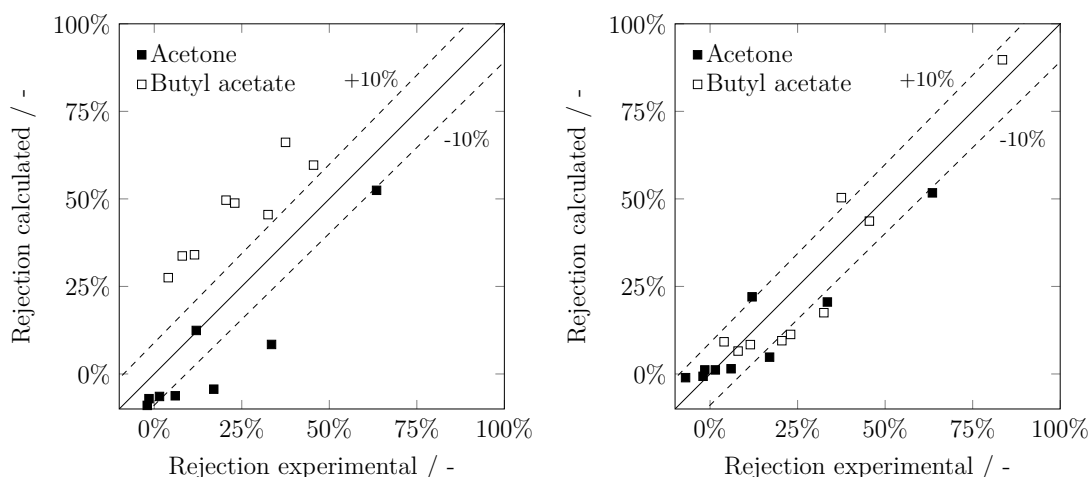
Figure 8.7: Calculated vs. experimental rejection of the solutes in the solvents acetone and n-butyl acetate using model $R_{all,i}^4$ (left) and model $R_{all,i}^{5.2}$ (right)

8.3.2 Prediction of solute rejection in mixed solvents

After showcasing the potential of the data-based modeling approach to derive models for solute rejection in pure solvents, the capability of deriving models for mixed solvents is investigated. The investigation is based on the rejection data published by Schmidt et al. [65] that covers solute rejection in isopropanol, toluene and hexane, which are considered as polar and non-polar solvents [78]. At first the generation of a more general solute-independent model is considered, before the potential of more accurate predictions by means of solute-specific models for the ternary solvent system is evaluated.

Development of solute-independent rejection models

In order to derive the data-based model, the available experimental rejection data is subdivided into training, test and validation data. It is important to note that only rejection data for pure solvents and equimolar binary solvent mixtures was considered for training data set, as illustrated in Figure 8.1. The remaining binary mixtures were allocated to the test data set. Based on the evaluation of different tree depths, a tree depth of four is deemed as preferred choice, resulting in significantly improved accuracy compared to model candidates derived for a tree depth of three. In ten consecutive runs for a tree depth of three, three different models were identi-

fied, all of which include the difference of the Hildebrand parameters of membrane and solvent $\Delta\delta_{M,i}$ and at least one additional solvent and solute polarity parameter each (cf. Table E.13 in Section E.2.1 in the Appendix). Hence, both solvent and solute polarity is taken into account in each model with a tree depth of three for the data set of mixed solvents. This complies with the model with a tree depth of four (cf. Table E.14 in Section E.2.1 in the Appendix). However, besides solvent and solute polarity, most of the models with a tree depth of four additionally incorporate a property parameter for the solvent size. Due to a higher tree depth and hence more available nodes in the tree, more complex models can be considered, in which additional effects like the solvent size can be included. Moreover, only half of these models include the difference of Hildebrand solubility parameters of the membrane and the solvent. The reason for this may be the statistical character of the method, but also the fact that, due to the higher complexity, the influence of solvent polarity is similarly well covered with other polarity parameters. On the other hand, trees with a higher depth also offer more possibilities for combining property parameters, which in this case led to ten different models with comparable fitness being found in ten runs. The considerable higher fitness of the models with a tree depth of four compared to the models with a tree depth of tree indicate the value of including these parameters in the models.

Moreover, the results are consistent with the results of Schmidt et al. [65] for the influence of the solvent and solute polarity, but not for the solute size. Only a few model candidates include the solute size as an influencing factor, while Schmidt et al. [65] observed a correlation between the solute size and the rejection, which was not completely independent of solvent polarity and of varying significance for the solutes and solvent mixtures. By combining the observations of Schmidt and the results for different model candidates, it is apparent that the influence of solvent and solute polarity either clearly predominates or that there is a correlation between solute polarity and size, thus no consideration of the solute size as an additional property is necessary.

While for a tree depth of three the Akaike weights do not clearly indicate one model to be favorable (probability of $47.1\,\%$ of the best model and $31.9\,\%$ for the second best model), the best model found with a tree depth of four is indicated to be the most suitable one among the models with the same tree depth with a probability of $99.9\,\%$. Table E.12 in Section E.2.1 in the Appendix presents a more elaborate overview of the according model statistics.

Overall, the model

$$R_{t,i}^4 = -2.582 \cdot \frac{\epsilon_s}{\delta_{h,i}} + 3.721 \cdot \frac{Ra_i}{Pol_s} + 178.395 \cdot \frac{\vec{d}_s}{\eta_s} + 686.443 \cdot d_{eq,s} - 0.083 \cdot V_{m,s} - 511.291$$

$$(8.7)$$

is evaluated as most promising. Model $R_{t,i}^4$ includes six fitting parameters and correlate a range of solute and solvent properties, including different solubility parameters as well as size related properties. While the correlation appears comparable complex, compared to the models derived in Section 8.3.1, the calculated rejections are in excellent agreement with the reported experimental data. As illustrated in the parity plot in Figure 8.8 the majority of the calculated rejections does not exceed an absolute deviation of 10 %, while even for the remaining data points deviations are always below 20 %.
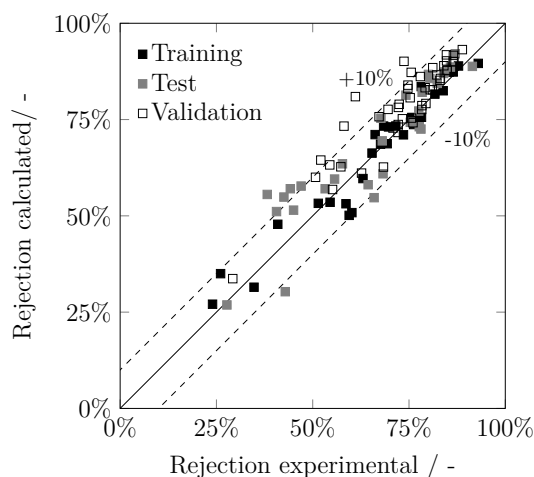
Figure 8.8: Calculated vs. experimental rejection of all solutes in ternary solvent mixtures of isopropanol, toluene and hexane using model $R_{t,i}^4$

The developed model further enables the prediction of the MRM introduced by Schmidt et al. [65]. This is illustrated exemplarily in Figure 8.9 for the rejection data of HMN. The top left diagram in the figure illustrates the MRM, based on smoothing and interpolating the measured rejections in the ternary system, while the top right diagram illustrates the predicted MRM, based on the developed model $R_{t,i}^4$, which is based solely on the rejection data obtained for pure and equimolar binary solvent mixtures. As highlighted by the third ternary graph on the bottom of Figure 8.9, the differences are mostly negligible. Similar illustrations of the

MRM for the remaining solutes are provided in Section E.2.1 in the Appendix. The derived data-based model does not only allow for remarkably accurate predictions of the experimentally determined rejections in the mixed solvent system, but also incorporates well identifiable parameters with respect to the confidence intervals.
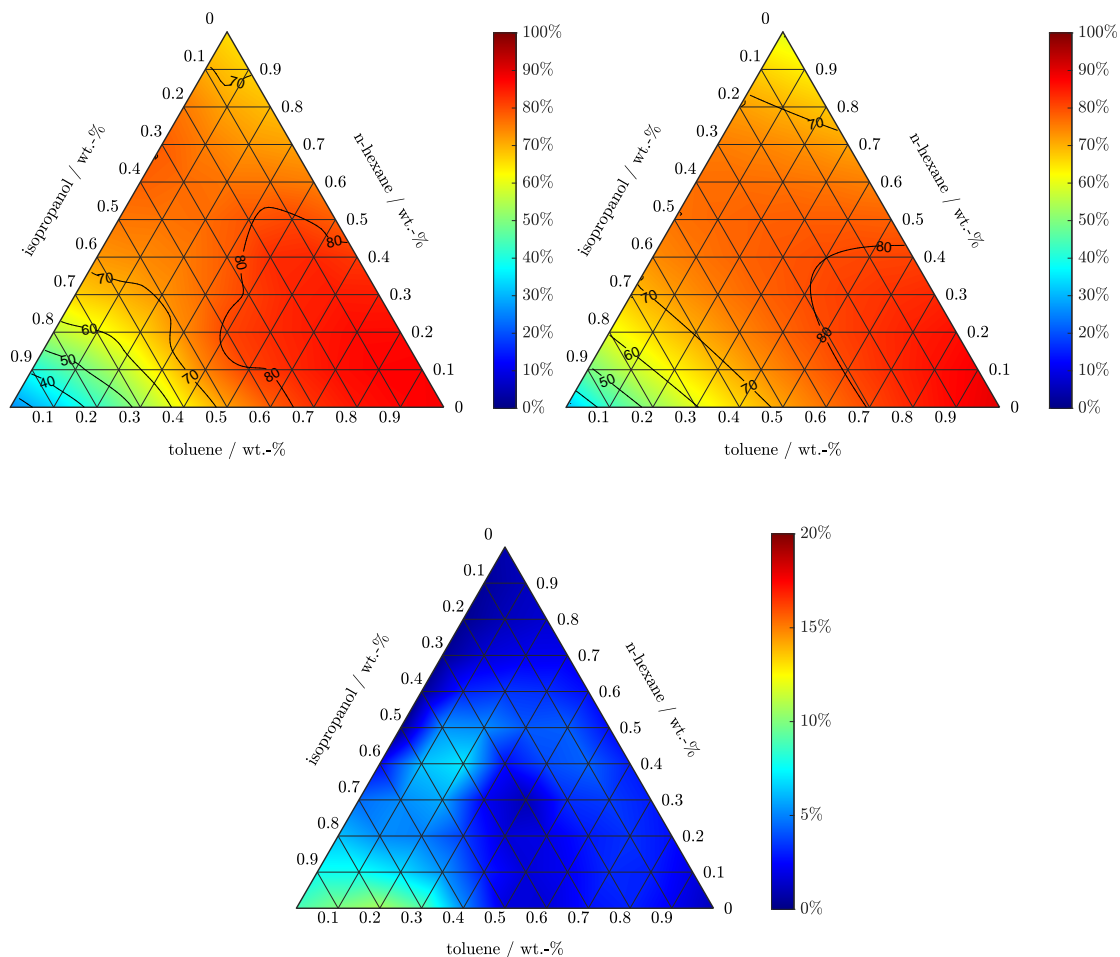
Figure 8.9: MRM for HMN, according to smoothing and interpolating of the measured rejections in the ternary system (top left) and model predictions of $R_{t,i}^4$ (top right), as well as absolute difference between both MRM (bottom).

Development of solute-specific rejection models

In order to evaluate the potential improvement in terms of model accuracy and parameter precision that may be achieved by a limitation of the validity range, solute-specific models were derived for the solute-specific experimental data. While this

limitation does of course reduce the range of applicability of the derived models, it offers the possibility to evaluate the potential improvement in terms of accuracy and potential simplicity, by considering a maximum tree depth of only three requiring fewer fitting parameters that are determined with narrow confidence intervals.

Including all identified models of ten runs for each solute in a subsequent evaluation of the included effects and parameters shows, that the main effect in these models is the solvent polarity and in some models the size of the solvent molecules is included as well (cf. Table E.16 to E.20 in Section E.2.2 in the Appendix). The specific property parameters for solvent polarity or size varies for the different models. Since the models are solute-specific, solute parameters are equivalent to constants in the generated models. While they are still considered as potential descriptors for model development, they do not have a considerable impact on the quality of the generated models.

While the best model for each solute is identified based on both fitness and parameter precision, the selected model always reflects the best fitness, except for TPP. For TPP the models with the best and second-best fitness show insufficient parameter precision and hence, the third best model is selected as the most promising one. The corresponding model candidates for the remaining solutes are summarized in Table E.15 in Section E.2.2 in the Appendix.

The results are further discussed based on the most promising model identified for HMN

$$R^3_{HMN} = -0.293 \cdot (\delta_{p,s} - \Delta\delta_{M,s}) \cdot \vec{d_s} \cdot \gamma_s + 61.884. \tag{8.8}$$

Model R^3_{HMN} includes only two fitting parameters and correlates the difference of the Hildebrand parameters of solvent and membrane $\Delta\delta_{M,s}$ and the polar Hansen solubility parameter $\delta_{p,s}$ of the solvent with the Dipole moment $\vec{d_s}$ and surface tension γ_s of the solvent. Comparing the effects included in model R^3_{HMN}, $R^4_{t,i}$ and the remaining model candidates for HMN (cf. Table E.16 in Section E.2.2 in the Appendix), it becomes apparent that each model includes at least the solvent polarity by incorporating at least to different solvent properties. Moreover, four model candidates of the solute-specific models include solute properties, which are similar to constants as discussed above. The solvent size is only included in the two solute-specific models with the lowest fitness and the model $R^4_{t,i}$, but not in model R^3_{HMN}, which leads to the conclusion that this effect is less prominent here.

While the model is much simpler compared to $R^4_{t,i}$, the potential improvement for the prediction of the MRM, illustrated in Figure 8.10, is limited, since the solute re-

jection predicted by model $R_{t,i}^4$ was already in good agreement with the experimental data.
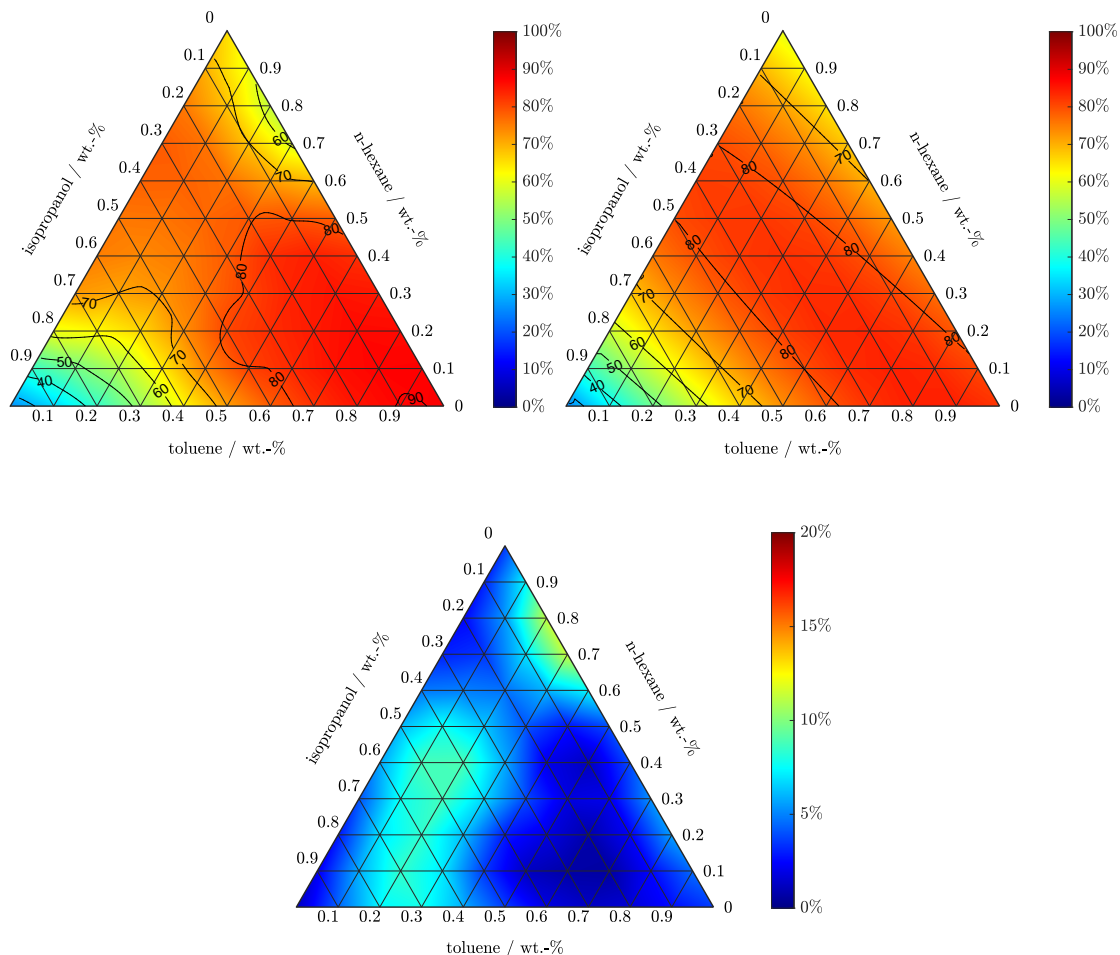
Figure 8.10: MRM for HMN, according to smoothing and interpolating of the measured rejections in the ternary system (top left) and model predictions of R_{HMN}^3 (top right), as well as absolute difference between both MRM (bottom).

As illustrated by the resulting statistics and the MRM illustrations in Table E.15 and Figure E.5-E.8 of Chapter E in the Appendix, similar results are obtained when generating solute-specific models for the other four solutes. The restriction to a specific solute allows for a slightly improved model accuracy, based on the reduced data set, as well as improved parameter precision, but at the cost of a limited validity range. In general, it can be concluded, that the presented approach allows for the derivation of either simplified solute-specific, as well as more general solute-

independent correlations for solute rejection, which enable an appropriate prediction of the solute rejection in the ternary mixed solvent systems, building only on the rejection data in pure solvents and equimolar binary solvent mixtures. Thus, the model development approach presents an excellent tool to derive the MRM presented by Schmidt et al. [65] with significantly reduced experimental effort.

8.4 Conclusion

While several correlations have been proposed for the prediction of solvent flux and various studies have addressed the analysis of prominent physicochemical properties that affect the rejection of solutes in organic solvent nanofiltration, no correlations for solute rejection modeling have been proposed so far. Building on the previous chapter for the automatic derivation of data-based models for pure and mixed solvent flux (Chapter 7), which outperformed previously proposed general phenomenological models in respect to model accuracy and parameter precision, the current chapter illustrated the possibility to derive appropriate correlations for solute rejections. The derived models accurately describe and predict solute rejections solved in pure solvents with various polarities. Constraining the range of application to solvent groups with consistent topology, as proposed by Thiermeyer et al. [78], allows for further increased model accuracy, while resulting in less complex correlations, indicating varying dominating transport mechanisms for the different solvent classes. The application of the model-development approach for solute rejection in mixed solvent systems further enables the prediction of membrane rejection maps [65], building solely on the rejection data in the pure solvents and equimolar binary solvent mixtures.

9

Conclusion and Outlook

To reduce the energy consumption in separation processes, it is essential to consider separation techniques that offer great potential for energy savings. These methods also include pressure driven membrane processes, such as organic solvent nanofiltration. However, in early stage of process design OSN is only rarely considered, due to still existing limitations. In order to overcome current limitations to some extend this thesis provided progress on the quantification and evaluation of the uncertainty related to experimental characterization of membrane performance in OSN and modeling of flux and rejection as well as the potential prediction of these quantities based on known experimental data for different chemical systems.

An elaborate evaluation of pure solvent flux measurements allowed for the first statistically valid quantification of the material-based uncertainty related to lab-scale measurements and the derivation of a minimal sample size for reliable quantification of solvent flux. The results indicated that the mean flux through the membrane estimated based on lab-scale data is an accurate approximation of the flux through an industrial scale module and provides additional information on the variance of the flux compared to an experimental investigation only based on the module. Based on a statistical evaluation of the experimental data a reasonable minimal sample size of four membrane samples was indicated to result in reliable estimations of the flux. While it has to be noted that these results are specific for the considered membranes and the size of the samples, the validation case studies support the derived recommendations. Although larger membrane samples can of course reduce the necessary sample size, the derived recommendations, based on the investigation of commercially produced membranes and appropriately selected solvents, should be considered as a best-case approximation. Considering the potential variations, the use of an appropriate number of samples in order to derive information on the stan-

dard deviation and allow for an estimation of the confidence intervals is generally recommended and easily possible in a single experimental run in common lab-scale set-ups offering multiple membrane cells. In future studies, the transferability of the minimum sample size found in this thesis to other membrane types and the comparability with commercial modules is of additional interest and may provide further insight into the influence of variations in membrane material during lab-scale experiments.

Further, the application of a standardized experimental procedure developed in a collaborative study with multiple partners enabled the comparability of experimental data within the range of the experimental error and the fluctuations caused by variations in the membrane material. This pointed out the transferability of experimental results to different facilities and hence, the possibility to combine the results in one large database. An expansion of the database in future studies is highly recommended in order to develop a global database for various membranes, solvents and solutes, which can be used as a basis for model development as well as for initial feasibility studies. The additionally performed evaluation of the influence of uncertainty information on process analysis and membrane selections demonstrated the importance of the consideration of this information.

Subsequently, the examination of the influence of the experimental uncertainties on modeling results of six different mass transfer models based on experimental data from literature, which were determined using common, non-standardized methods, revealed that none of the models was superior to the others taking into account the accuracy of the representation of experimental data and the uncertainties of the parameters. Hence, no clear recommendation for one of the models could be given. However, this clearly points out the effect of uncertainties in parameter regression on the model results and hence, in future studies, the precision and identifiability of the parameters should therefore always be taken into account, as these can strongly influence the model results and thus also predictions for other process conditions. An improvement in the accuracy by considering more complex models is conceivable in future studies if additional experimental data at further pressure levels are available. Moreover, the additional experimental data points as well as sorption or diffusion experiments could result in an increasing parameters precision, which should be investigated in future studies. However, model reduction should always be considered in cases of non-identifiable parameters. To enable predictions of flux and rejection, in the second part of this thesis a method to automatically

develop membrane-specific, predictive models was proposed. Comparing the newly developed models for solvent flux with existing models from literature based on accuracy and the parameter uncertainties, the results clearly indicated the potential of these membrane-specific models, which in all case studies outperformed the existing models. Extending the method to solute rejection, a prediction of the rejection in different solvents was possible but required a complex model structure, while solvent-dependent or solute-dependent models were less complex and more accurate. While the results showcase the potential of the developed method to automatically derive membrane-specific correlations for solvent flux and solute rejection, in pure and mixed solvent systems, the application should always be pursued with appropriate caution, as data-based models can only reflect dependencies that are depicted by the underlying data. Consequently, phenomenological models derived from physical reasoning, such as the model of Marchetti et al. [2], are more likely to provide reliable extrapolation for molecules that are not representing interpolations of the initially considered data set. However, the presented method is able to automatically derive different model structures and discriminate between potential molecular descriptors, based on the available data, covering effects that have not been included in the fixed model structure of the phenomenological models. Furthermore, the considered molecular descriptors are much easier retrieved, compared to properties like surface tension and polarizability considered by Marchetti et al. [2].

An improvement of the methodology is conceivable in the future by parallelizing the parameter optimization for the model candidates, especially with regard to nonlinear parameters. Further, the consideration of other parameters and descriptors based on quantum chemistry or thermodynamic models to describe the interactions between membrane, solvent and solutes might enable more accurate predictions. Furthermore, additional experimental data should be utilized in order to improve them and expand the applicability of the models. While a variety of literature studies report membrane performance data that can be exploited for model development, it can be assumed that even more data is available at membrane manufacturers, which frequently test the performance of their membranes for customer requests. Hence, membrane manufacturers are probably the most appropriate user of such a tool. Building a database of membrane-specific models enables a quick estimation of the feasibility of a membrane for a specific separation task and brings organic solvent nanofiltration membranes one step closer to a regular consideration in conceptual design studies.

Bibliography

[1] S. Darvishmanesh, A. Buekenhoudt, J. Degrève, and B. Van der Bruggen. "Coupled series–parallel resistance model for transport of solvent through inorganic nanofiltration membranes." *Separation and Purification Technology* 70.1 (2009), pp. 46–52.

[2] P. Marchetti, A. Butté, and A. G. Livingston. "An improved phenomenological model for prediction of solvent permeation through ceramic NF and UF membranes." *Journal of Membrane Science* 415-416 (2012), pp. 444–458.

[3] D. R. Machado, D. Hasson, and R. Semiat. "Effect of solvent properties on permeate flow through nanofiltration membranes." *Journal of Membrane Science* 166.1 (2000), pp. 63–69.

[4] D. Bhanushali, S. Kloos, C. Kurth, and D. Bhattacharyya. "Performance of solvent-resistant membranes for non-aqueous systems: solvent permeation results and modeling." *Journal of Membrane Science* 189.1 (2001), pp. 1–21.

[5] J. Geens, B. van der Bruggen, and C. Vandecasteele. "Transport model for solvent permeation through nanofiltration membranes." *Separation and Purification Technology* 48.3 (2006), pp. 255–263.

[6] V. Masson-Delmotte et al. *Global Warming of* 1.5 °C. *An IPCC Special Report on the impacts of global warming of* 1.5 °C *above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty.* 2018. URL: https://www.ipcc.ch/sr15/ (visited on 11/28/2019).

[7] United Nations Treaty Collection, ed. *Paris Agreement.* Paris, Dec. 12, 2015. URL: `https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-7-d&chapter=27&clang=_en` (visited on 11/28/2019).

[8] European Environmental Agency, ed. *2018 Industrial pollution country profiles, EEA-33.* 2018. URL: `https://www.eea.europa.eu/themes/industry/industrial-pollution/industrial-pollution-country-profiles-2018/2018-industrial-pollution-EEA33` (visited on 11/28/2019).

[9] U.S. Energy Information Administration, ed. *The United States uses a mix of energy sources.* 2018. URL: `https://www.eia.gov/energyexplained/us-energy-facts/` (visited on 11/28/2019).

[10] D. S. Sholl and R. P. Lively. "Seven chemical separations to change the world." *Nature* 532.7600 (2016), pp. 435–437.

[11] P. Lutze and A. Górak. *Reactive and Membrane-Assisted Separations.* De Gruyter Textbook. Berlin: De Gruyter, 2016.

[12] M. Priske, M. Lazar, C. Schnitzer, and G. Baumgarten. "Recent Applications of Organic Solvent Nanofiltration." *Chemie Ingenieur Technik* 88.1-2 (2016), pp. 39–49.

[13] L. S. White. "Development of large-scale applications in organic solvent nanofiltration and pervaporation for chemical and refining processes." *Journal of Membrane Science* 286.1-2 (2006), pp. 26–35.

[14] P. Marchetti, Jimenez Solomon, Maria F., G. Szekely, and A. G. Livingston. "Molecular Separation with Organic Solvent Nanofiltration: A Critical Review." *Chemical Reviews* 114.21 (2014), pp. 10735–10806.

[15] L. Hesse, J. Mićović, P. Schmidt, A. Górak, and G. Sadowski. "Modelling of organic-solvent flux through a polyimide membrane." *Journal of Membrane Science* 428 (2013), pp. 554–561.

[16] A. Böcking, V. Koleva, J. Wind, Y. Thiermeyer, S. Blumenschein, R. Goebel, M. Skiborowski, and M. Wessling. "Can the variance in membrane performance influence the design of organic solvent nanofiltration processes?" *Journal of Membrane Science* 575 (2019), pp. 217–228.

[17] T. Melin and R. Rautenbach. *Membranverfahren: Grundlagen der Modul- und Anlagenauslegung.* 3., aktualisierte und erw. Aufl. Chemische Technik/Verfahrenstechnik. Berlin and New York: Springer, 2007.

[18]   W. J. Koros, Y. H. Ma, and T. Shimidzu. "Terminology for membranes and membrane processes (IUPAC Recommendations 1996)." *Pure and Applied Chemistry* 68.7 (1996), pp. 1479–1489.

[19]   K. Ohlrogge and K. Ebert. *Membranen.* Wiley, 2006.

[20]   S. Darvishmanesh, T. Robberecht, P. Luis, J. Degrève, and B. Van der Bruggen. "Performance of Nanofiltration Membranes for Solvent Purification in the Oil Industry." *Journal of the American Oil Chemists' Society* 88.8 (2011), pp. 1255–1261.

[21]   K. Werth, P. Kaupenjohann, M. Knierbein, and M. Skiborowski. "Solvent recovery and deacidification by organic solvent nanofiltration: Experimental investigation and mass transfer modeling." *Journal of Membrane Science* 528 (2017), pp. 369–380.

[22]   K. Werth, P. Kaupenjohann, and M. Skiborowski. "The potential of organic solvent nanofiltration processes for oleochemical industry." *Separation and Purification Technology* 182 (2017), pp. 185–196.

[23]   D. Peshev, L. G. Peeva, G. Peev, I. Baptista, and A. T. Boam. "Application of organic solvent nanofiltration for concentration of antioxidant extracts of rosemary (Rosmarinus officiallis L.)" *Chemical Engineering Research and Design* 89.3 (2011), pp. 318–327.

[24]   I. Sereewatthanawut, I. Baptista, A. T. Boam, A. Hodgson, and A. G. Livingston. "Nanofiltration process for the nutritional enrichment and refining of rice bran oil." *Journal of Food Engineering* 102.1 (2011), pp. 16–24.

[25]   J. M. Dreimann, M. Skiborowski, A. Behr, and A. J. Vorholt. "Recycling Homogeneous Catalysts Simply by Organic Solvent Nanofiltration: New Ways to Efficient Catalysis." *ChemCatChem* 8.21 (2016), pp. 3330–3333.

[26]   W. L. Peddie, J. N. van Rensburg, H. C. Vosloo, and P. van der Gryp. "Technological evaluation of organic solvent nanofiltration for the recovery of homogeneous hydroformylation catalysts." *Chemical Engineering Research and Design* 121 (2017), pp. 219–232.

[27]   B. Bertleff, R. Goebel, J. Claußnitzer, W. Korth, M. Skiborowski, P. Wasserscheid, A. Jess, and J. Albert. "Investigations on Catalyst Stability and Product Isolation in the Extractive Oxidative Desulfurization of Fuels Us-

ing Polyoxometalates and Molecular Oxygen." *ChemCatChem* 10.20 (2018), pp. 4602–4609.

[28] G. Székely, J. Bandarra, W. Heggie, B. Sellergren, and F. C. Ferreira. "Organic solvent nanofiltration: A platform for removal of genotoxins from active pharmaceutical ingredients." *Journal of Membrane Science* 381.1-2 (2011), pp. 21–33.

[29] E. M. Rundquist, C. J. Pink, and A. G. Livingston. "Organic solvent nanofiltration: a potential alternative to distillation for solvent recovery from crystallisation mother liquors." *Green Chemistry* 14.8 (2012), pp. 2197–2205.

[30] A. Boam and A. Nozari. "Fine chemical: OSN – a lower energy alternative." *Filtration & Separation* 43.3 (2006), pp. 46–48.

[31] J. Micovic, K. Werth, and P. Lutze. "Hybrid separations combining distillation and organic solvent nanofiltration for separation of wide boiling mixtures." *Chemical Engineering Research and Design* 92.11 (2014), pp. 2131–2147.

[32] R. P. Lively and D. S. Sholl. "From water to organics in membrane separations." *Nature materials* 16.3 (2017), pp. 276–279.

[33] S. Blumenschein and U. Kätzel. "An heuristic-based selection process for organic solvent nanofiltration membranes." *Separation and Purification Technology* 183 (2017), pp. 83–95.

[34] L. White. "Transport properties of a polyimide solvent resistant nanofiltration membrane." *Journal of Membrane Science* 205.1-2 (2002), pp. 191–202.

[35] D. Bhanushali, S. Kloos, and D. Bhattacharyya. "Solute transport in solvent-resistant nanofiltration membranes for non-aqueous systems: experimental results and the role of solute–solvent coupling." *Journal of Membrane Science* 208.1 (2002), pp. 343–359.

[36] P. Vandezande, L. E. M. Gevers, J. Paul, I. F. J. Vankelecom, and P. Jacobs. "High throughput screening for rapid development of membranes and membrane processes." *Journal of Membrane Science* 250.1-2 (2005), pp. 305–310.

[37] Y. H. See Toh, X. X. Loh, K. Li, A. Bismarck, and A. G. Livingston. "In search of a standard method for the characterisation of organic solvent nanofiltration membranes." *Journal of Membrane Science* 291.1-2 (2007), pp. 120–125.

[38] S. M. Dutczak, M. Luiten-Olieman, H. J. Zwijnenberg, L. Bolhuis-Versteeg, L. Winnubst, M. A. Hempenius, N. E. Benes, M. Wessling, and D. Stamatialis. "Composite capillary membrane for solvent resistant nanofiltration." *Journal of Membrane Science* 372.1-2 (2011), pp. 182–190.

[39] P. Marchetti, L. Peeva, and A. Livingston. "The Selectivity Challenge in Organic Solvent Nanofiltration: Membrane and Process Solutions." *Annual review of chemical and biomolecular engineering* 8 (2017), pp. 473–497.

[40] P. Vandezande, L. E. M. Gevers, and I. F. J. Vankelecom. "Solvent resistant nanofiltration: separating on a molecular level." *Chemical Society reviews* 37.2 (2008), pp. 365–405.

[41] J. G. Wijmans and R. W. Baker. "The solution-diffusion model: a review." *Journal of Membrane Science* 107.1-2 (1995), pp. 1–21.

[42] W. Bowen, A. Mohammad, and N. Hilal. "Characterisation of nanofiltration membranes for predictive purposes — use of salts, uncharged solutes and atomic force microscopy." *Journal of Membrane Science* 126.1 (1997), pp. 91–105.

[43] W. Bowen and H. Mukhtar. "Characterisation and prediction of separation performance of nanofiltration membranes." *Journal of Membrane Science* 112.2 (1996), pp. 263–274.

[44] W. Bowen and J. S. Welfoot. "Modelling the performance of membrane nanofiltration—critical assessment and model development." *Chemical Engineering Science* 57.7 (2002), pp. 1121–1137.

[45] H. K. Lonsdale, U. Merten, and R. L. Riley. "Transport properties of cellulose acetate osmotic membranes." *Journal of Applied Polymer Science* 9.4 (1965), pp. 1341–1362.

[46] D. Paul. "Reformulation of the solution-diffusion theory of reverse osmosis." *Journal of Membrane Science* 241.2 (2004), pp. 371–386.

[47] P. Marchetti and A. G. Livingston. "Predictive membrane transport models for Organic Solvent Nanofiltration: How complex do we need to be?" *Journal of Membrane Science* 476 (2015), pp. 530–553.

[48] L. Hesse, S. Naeem, and G. Sadowski. "VOC sorption in glassy polyimides— Measurements and modeling." *Journal of Membrane Science* 415-416 (2012), pp. 596–607.

[49] T. K. Sherwood, P. L. T. Brian, and R. E. Fisher. "Desalination by Reverse Osmosis." *Industrial & Engineering Chemistry Fundamentals* 6.1 (1967), pp. 2–12.

[50] A. E. Yaroshchuk. "Solution-diffusion-imperfection model revised." *Journal of Membrane Science* 101.1-2 (1995), pp. 83–87.

[51] M. Dijkstra, S. Bach, and K. Ebert. "A transport model for organophilic nanofiltration." *Journal of Membrane Science* 286.1-2 (2006), pp. 60–68.

[52] O. Kedem and A. Katchalsky. "Thermodynamic analysis of the permeability of biological membranes to non-electrolytes." *Biochimica et Biophysica Acta* 27 (1958), pp. 229–246.

[53] K. S. Spiegler and O. Kedem. "Thermodynamics of hyperfiltration (reverse osmosis): criteria for efficient membranes." *Desalination* 1.4 (1966), pp. 311–326.

[54] J. Geens, A. Hillen, B. Bettens, B. Van der Bruggen, and C. Vandecasteele. "Solute transport in non-aqueous nanofiltration: effect of membrane material." *Journal of Chemical Technology & Biotechnology* 80.12 (2005), pp. 1371–1377.

[55] J. Geens, K. Boussu, C. Vandecasteele, and B. van der Bruggen. "Modelling of solute transport in non-aqueous nanofiltration." *Journal of Membrane Science* 281.1-2 (2006), pp. 139–148.

[56] S. Blumenschein, A. Böcking, U. Kätzel, S. Postel, and M. Wessling. "Rejection modeling of ceramic membranes in organic solvent nanofiltration." *Journal of Membrane Science* 510 (2016), pp. 191–200.

[57] S. Darvishmanesh and B. Van der Bruggen. "Mass Transport through Nanostructured Membranes: Towards a Predictive Tool." *Membranes* 6.49 (2016), pp. 1–15.

VI

[58] Y. A. Hussain, M. H. Al-Saleh, and S. S. Ar-Ratrout. "The effect of active layer non-uniformity on the flux and compaction of TFC membranes." *Desalination* 328 (2013), pp. 17–23.

[59] I. H. Tsibranska and B. Tylkowski. "Concentration of ethanolic extracts from Sideritis ssp. L. by nanofiltration: Comparison of dead-end and cross-flow modes." *Food and Bioproducts Processing* 91.2 (2013), pp. 169–174.

[60] B.-G. Wang, M. Ando, H. Yin, B. Hong, and Y. Peng. "Estimation of ultrafiltration membrane element flux from membrane sheet tests." *Desalination* 191.1-3 (2006), pp. 125–131.

[61] T. Schipolowski, A. Jeżowska, and G. Wozny. "Reliability of membrane test cell measurements." *Desalination* 189.1-3 (2006), pp. 71–80.

[62] D. R. Machado, D. Hasson, and R. Semiat. "Effect of solvent properties on permeate flow through nanofitration membranes, Part I." *Journal of Membrane Science* 163 (1999), pp. 93–102.

[63] B. Shi, D. Peshev, P. Marchetti, S. Zhang, and A. G. Livingston. "Multiscale modelling of OSN batch concentration with spiral-wound membrane modules using OSN Designer." *Chemical Engineering Research and Design* 109 (2016), pp. 385–396.

[64] I. Sereewatthanawut, F. W. Lim, Y. S. Bhole, D. Ormerod, A. Horvath, A. T. Boam, and A. G. Livingston. "Demonstration of Molecular Purification in Polar Aprotic Solvents by Organic Solvent Nanofiltration." *Organic Process Research & Development* 14.3 (2010), pp. 600–611.

[65] P. Schmidt, T. Köse, and P. Lutze. "Characterisation of organic solvent nanofiltration membranes in multi-component mixtures: Membrane rejection maps and membrane selectivity maps for conceptual process design." *Journal of Membrane Science* 429 (2013), pp. 103–120.

[66] S. Postel, G. Spalding, M. Chirnside, and M. Wessling. "On negative retentions in organic solvent nanofiltration." *Journal of Membrane Science* 447 (2013), pp. 57–65.

[67] T. C. Bailey and A. C. Gatrell. *Interactive spatial data analysis.* Longman Scientific & Technical Essex, 1995.

[68] R. H. Shumway. "Statistics and Data Analysis in Geology." *Technometrics* 29 (1987), p. 492.

[69] S. Keller, G. Korkmaz, M. Orr, A. Schroeder, and S. Shipp. "The Evolution of Data Quality: Understanding the Transdisciplinary Origins of Data Quality Concepts and Approaches." *Annual Review of Statistics and Its Application* 4.1 (2017), pp. 85–108.

[70] M. Razali, C. Didaskalou, J. F. Kim, M. Babaei, E. Drioli, Y. M. Lee, and G. Szekely. "Exploring and Exploiting the Effect of Solvent Treatment in Membrane Separations." *ACS applied materials & interfaces* 9.12 (2017), pp. 11279–11289.

[71] Evonik Resource Efficiency GmbH. *Technical Information - DuraMem®Products*. 2017. URL: https://duramem.evonik.com/product/duramem-puramem.

[72] K. Fischer and M. Wilken. "Experimental determination of oxygen and nitrogen solubility in organic solvents up to 10 MPa at temperatures between 298 K and 398 K." *The Journal of Chemical Thermodynamics* 33.10 (2001), pp. 1285–1308.

[73] Ž. R. Lazić. *Design of experiments in chemical engineering: A practical guide*. Weinheim: Wiley-VCH, 2004. URL: http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10300990.

[74] A. Sen and M. Srivastava. *Regression Analysis*. New York, NY: Springer New York, 1990.

[75] S. S. Shapiro and M. B. Wilk. "An Analysis of Variance Test for Normality (Complete Samples)." *Biometrika* 52.3/4 (1965), p. 591.

[76] F. Wilcoxon. "Individual Comparisons by Ranking Methods." *Biometrics Bulletin* 1.6 (1945), p. 80.

[77] S. Greenland, S. J. Senn, K. J. Rothman, J. B. Carlin, C. Poole, S. N. Goodman, and D. G. Altman. "Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations." *European journal of epidemiology* 31.4 (2016), pp. 337–350.

[78] Y. Thiermeyer, S. Blumenschein, and M. Skiborowski. "Solvent dependent membrane-solute sensitivity of OSN membranes." *Journal of Membrane Science* 567 (2018), pp. 7–17.

[79] C. Reichardt and T. Welton. *Solvents and solvent effects in organic chemistry*. Weinheim, Germany: Wiley-VCH, 2011.

[80] C. M. Hansen. *Hansen solubility parameters: A user's handbook.* 2nd ed. Boca Raton: CRC Press, 2007.

[81] Evonik Resource Efficiency GmbH. *Technical Information - DuraMem*® *Products.* 2017. URL: `https://duramem.evonik.com/product/duramem-puramem`.

[82] Evonik Resource Efficiency GmbH. *Technical Information - PuraMem*® *Products.* 2017. URL: `https://duramem.evonik.com/product/duramem-puramem`.

[83] K. Ebert, J. Koll, M. Dijkstra, and M. Eggers. "Fundamental studies on the performance of a hydrophobic solvent stable membrane in non-aqueous solutions." *Journal of Membrane Science* 285.1-2 (2006), pp. 75–80.

[84] S. Zeidler, U. Kätzel, and P. Kreis. "Systematic investigation on the influence of solutes on the separation behavior of a PDMS membrane in organic solvent nanofiltration." *Journal of Membrane Science* 429 (2013), pp. 295–303.

[85] L. R. Firman, N. A. Ochoa, J. Marchese, and C. L. Pagliero. "Deacidification and solvent recovery of soybean oil by nanofiltration membranes." *Journal of Membrane Science* 431 (2013), pp. 187–196.

[86] R. Merlet et al. "Comparing the Performance of Organic Solvent Nanofiltration Membranes in Non–Polar Solvents." *Chemie Ingenieur Technik* 93.9 (2021), pp. 1389–1395.

[87] D. Peshev and A. G. Livingston. "OSN Designer, a tool for predicting organic solvent nanofiltration technology performance using Aspen One, MATLAB and CAPE OPEN." *Chemical Engineering Science* 104 (2013), pp. 975–987.

[88] S. W. Smith. *The scientist and engineer's guide to digital signal processing.* 1. ed. San Diego, Calif.: California Technical Publ, 1997.

[89] K. P. Burnham and D. R. Anderson. *Model selection and multimodel inference: A practical information-theoretic approach.* Second edition. New York: Springer, 2002.

[90] L. Fahrmeir, R. Künstler, I. Pigeot, and G. Tutz. *Statistik: Der Weg zur Datenanalyse.* Fünfte, verbesserte Auflage. Springer-Lehrbuch. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, 2004.

[91] M. P. Little, W. F. Heidenreich, and G. Li. "Parameter identifiability and redundancy: theoretical considerations." *PloS one* 5.1 (2010), e8915.

[92] T. Quaiser and M. Mönnigmann. "Systematic identifiability testing for unambiguous mechanistic modeling–application to JAK-STAT, MAP kinase, and NF-kappa B signaling pathway models." *BMC systems biology* 3 (2009), p. 50.

[93] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. "Equation of State Calculations by Fast Computing Machines." *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092.

[94] W. K. Hastings. "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika* 57.1 (1970), pp. 97–109.

[95] K. Roy, S. Kar, and R. N. Das. *A Primer on QSAR/QSPR Modeling.* Cham: Springer International Publishing, 2015.

[96] R. Poli, W. B. Langdon, N. F. McPhee, and J. R. Koza. *A field guide to genetic programming.* Morrisville, NC: Lulu Press, 2008.

[97] V. Yangali-Quintanilla, A. Sadmani, M. McConville, M. Kennedy, and G. Amy. "A QSAR model for predicting rejection of emerging contaminants (pharmaceuticals, endocrine disruptors) by nanofiltration membranes." *Water research* 44.2 (2010), pp. 373–384.

[98] A. Shahmansouri and C. Bellona. "Application of quantitative structure–property relationships (QSPRs) to predict the rejection of organic solutes by nanofiltration." *Separation and Purification Technology* 118 (2013), pp. 627–638.

[99] G. E. P. Box and N. R. Draper. *Empirical model-building and response surfaces.* Wiley series in probability and mathematical statistics: Applied probability and statistics. 1987.

[100] J. H. Lee, J. Shin, and M. J. Realff. "Machine learning: Overview of the recent progresses and implications for the process systems engineering field." *Computers & Chemical Engineering* 114 (2018), pp. 111–121.

[101] A. Cozad, N. V. Sahinidis, and D. C. Miller. "Learning surrogate models for simulation-based optimization." *AIChE Journal* 60.6 (2014), pp. 2211–2227.

[102] J. R. Koza. *Genetic programming: A paradigm for genetically breeding populations of computer programs to solve problems.* Stanford University, Department of Computer Science Stanford, CA, 1990.

[103] A. Fouladitajar, F. Zokaee Ashtiani, A. Okhovat, and B. Dabir. "Membrane fouling in microfiltration of oil-in-water emulsions; a comparison between constant pressure blocking laws and genetic programming (GP) model." *Desalination* 329 (2013), pp. 41–49.

[104] T.-M. Hwang, H. Oh, Y.-K. Choung, S. Oh, M. Jeon, J. H. Kim, S. H. Nam, and S. Lee. "Prediction of membrane fouling in the pilot-scale microfiltration system using genetic programming." *Desalination* 247.1-3 (2009), pp. 285–294.

[105] V. Yangali-Quintanilla, A. Verliefde, T.-U. Kim, A. Sadmani, M. Kennedy, and G. Amy. "Artificial neural network models based on QSAR for predicting rejection of neutral organic compounds by polyamide nanofiltration and reverse osmosis membranes." *Journal of Membrane Science* 342.1-2 (2009), pp. 251–262.

[106] Y. Ammi, L. Khaouane, and S. Hanini. "Prediction of the rejection of organic compounds (neutral and ionic) by nanofiltration and reverse osmosis membranes using neural networks." *Korean Journal of Chemical Engineering* 32.11 (2015), pp. 2300–2310.

[107] J. Madár, J. Abonyi, and F. Szeifert. "Genetic Programming for the Identification of Nonlinear Input−Output Models." *Industrial & Engineering Chemistry Research* 44.9 (2005), pp. 3178–3186.

[108] M. Tawarmalani and N. V. Sahinidis. "Global optimization of mixed-integer nonlinear programs: A theoretical and computational study." *Mathematical programming* 99 (2004), pp. 563–591.

[109] R. Misener and C. A. Floudas. "ANTIGONE: Algorithms for coNTinuous / Integer Global Optimization of Nonlinear Equations." *Journal of Global Optimization* 59.2-3 (2014), pp. 503–526.

[110] J. Santos, A. M. Hidalgo, R. Oliveira, S. Velizarov, and J. G. Crespo. "Analysis of solvent flux through nanofiltration membranes by mechanistic, chemometric and hybrid modelling." *Journal of Membrane Science* 300.1-2 (2007), pp. 191–204.

[111] P. Silva, S. Han, and A. G. Livingston. "Solvent transport in organic solvent nanofiltration membranes." *Journal of Membrane Science* 262.1-2 (2005), pp. 49–59.

[112] B. Van der Bruggen, J. Geens, and C. Vandecasteele. "Fluxes and rejections for nanofiltration with solvent stable polymeric membranes in water, ethanol and n-hexane." *Chemical Engineering Science* 57.13 (2002), pp. 2511–2518.

[113] M. Randic. "Characterization of molecular branching." *Journal of the American Chemical Society* 97.23 (1975), pp. 6609–6615.

[114] L. B. Kier and L. H. Hall. *Molecular connectivity in chemistry and drug research.* Medicinal chemistry, a series of monographs. New York: Academic Press, 1976.

[115] J. Robinson. "Solvent flux through dense polymeric nanofiltration membranes." *Journal of Membrane Science* 230.1-2 (2004), pp. 29–37.

[116] E. S. Tarleton, J. P. Robinson, and M. Salman. "Solvent-induced swelling of membranes — Measurements and influence in nanofiltration." *Journal of Membrane Science* 280.1-2 (2006), pp. 442–451.

[117] S. Postel, C. Schneider, and M. Wessling. "Solvent dependent solute solubility governs retention in silicone based organic solvent nanofiltration." *Journal of Membrane Science* 497 (2016), pp. 47–54.

[118] C. M. Hansen. "The three dimensional solubility parameter and solvent diffusion coefficient: their importance in surface coating formulation." PhD thesis. Copenhagen, Denmark, 1967.

[119] D. W. van Krevelen and K. t. Nijenhuis. *Properties of polymers: Their correlation with chemical structure their numerical estimation and prediction from additive group contributions.* 4th, completely rev. ed. Amsterdam: Elsevier, 2009.

[120] K. K. Reddy, T. Kawakatsu, J. B. Snape, and M. Nakajima. "Membrane Concentration and Separation of L-Aspartic Acid and L-Phenylalanine Derivatives in Organic Solvents." *Separation Science and Technology* 31.8 (1996), pp. 1161–1178.

[121] K. P. Burnham and D. R. Anderson. "Multimodel Inference." *Sociological Methods & Research* 33.2 (2004), pp. 261–304.

[122] S. Sette and L. Boullart. "Genetic programming: principles and applications." *Engineering Applications of Artificial Intelligence* 14.6 (2001), pp. 727–736.

[123]   M. Skiborowski, M. Rautenberg, and W. Marquardt. "A Hybrid Evolutionary–Deterministic Optimization Approach for Conceptual Design." *Industrial & Engineering Chemistry Research* 54.41 (2015), pp. 10054–10072.

[124]   C. Michalik, M. Stuckert, and W. Marquardt. "Optimal Experimental Design for Discriminating Numerous Model Candidates: The AWDC Criterion." *Industrial & Engineering Chemistry Research* 49.2 (2010), pp. 913–919.

[125]   M. Khayet, C. Cojocaru, and C. García-Payo. "Application of Response Surface Methodology and Experimental Design in Direct Contact Membrane Distillation." *Industrial & Engineering Chemistry Research* 46.17 (2007), pp. 5673–5685.

[126]   J. Geens, K. Peeters, B. van der Bruggen, and C. Vandecasteele. "Polymeric nanofiltration of binary water–alcohol mixtures: Influence of feed composition and membrane properties on permeability and rejection." *Journal of Membrane Science* 255.1-2 (2005), pp. 255–264.

[127]   M. Jelinska-Kazimierczuk and J. Szydlowski. "Physicochemical Properties of Solutions of Amides in H2O and in D2O." *Journal of Solution Chemistry* 30.7 (2001), pp. 623–640.

[128]   C. Moreau and G. Douhéret. "Thermodynamic behavior of water-acetonitrile mixtures excess volumes and viscosities." *Thermochimica Acta* 13.4 (1975), pp. 385–392.

[129]   N. Saha, B. Das, and D. K. Hazra. "Viscosities and Excess Molar Volumes for Acetonitrile + Methanol at 298.15, 308.15, and 318.15 K." *Journal of Chemical & Engineering Data* 40.6 (1995), pp. 1264–1266.

[130]   G. Ritzoulis, N. Papadopoulos, and D. Jannakoudakis. "Densities, viscosities, and dielectric constants of acetonitrile + toluene at 15, 25, and 35 .degree.C." *Journal of Chemical & Engineering Data* 31.2 (1986), pp. 146–148.

[131]   N. G. Tsierkezos and A. C. Filippou. "Thermodynamic investigation of N,N-dimethylformamide/toluene binary mixtures in the temperature range from 278.15 to 293.15K." *The Journal of Chemical Thermodynamics* 38.8 (2006), pp. 952–961.

[132]   Y. Kiso, Y. Sugiura, T. Kitao, and K. Nishimura. "Effects of hydrophobicity and molecular size on rejection of aromatic pesti-cides with nanofiltration membranes." *Journal of Membrane Science* 192 (2001), pp. 1–10.

[133] B. Van der Bruggen, J. Schaep, D. Wilms, and C. Vandecasteele. "Influence of molecular size, polarity and charge on the retention of organic molecules by nanofiltration." *Journal of Membrane Science* 156.1 (1999), pp. 29–41.

[134] P. Schmidt, E. L. Bednarz, P. Lutze, and A. Górak. "Characterisation of Organic Solvent Nanofiltration membranes in multi-component mixtures: Process design workflow for utilising targeted solvent modifications." *Chemical Engineering Science* 115 (2014), pp. 115–126.

[135] S. Darvishmanesh, J. Degrève, and B. Van der Bruggen. "Mechanisms of solute rejection in solvent resistant nanofiltration: the effect of solvent on solute rejection." *Physical Chemistry Chemical Physics* 12.40 (2010), pp. 13333–13342.

[136] E. S. Tarleton, J. P. Robinson, C. R. Millington, A. Nijmeijer, and M. L. Taylor. "The influence of polarity on flux and rejection behaviour in solvent resistant nanofiltration—Experimental observations." *Journal of Membrane Science* 278.1-2 (2006), pp. 318–327.

[137] S. Darvishmanesh, J. Vanneste, E. Tocci, J. C. Jansen, J. Jansen, F. Tasselli, F. Tasseli, J. Degrève, E. Drioli, and B. Van der Bruggen. "Physicochemical characterization of solute retention in solvent resistant nanofiltration: the effect of solute size, polarity, dipole moment, and solubility parameter." *The journal of physical chemistry. B* 115.49 (2011), pp. 14507–14517.

[138] M. J. Bagajewicz and V. Manousiouthakis. "Mass/heat-exchange network representation of distillation networks." *AIChE Journal* 38.11 (1992), pp. 1769–1800.

[139] M. Skiborowski, A. Mhamdi, K. Kraemer, and W. Marquardt. "Model-based structural optimization of seawater desalination plants." *Desalination* 292 (2012), pp. 30–44.

[140] M. Scholz, M. Alders, T. Lohaus, and M. Wessling. "Structural optimization of membrane-based biogas upgrading processes." *Journal of Membrane Science* 474 (2015), pp. 1–10.

[141] B. Ohs, J. Lohaus, and M. Wessling. "Optimization of membrane based nitrogen removal from natural gas." *Journal of Membrane Science* 498 (2016), pp. 291–301.

[142] Evonik Resource Efficiency GmbH. *Lab scale nanofiltration equipment -The METcell Range*. 2017. URL: https://duramem.evonik.com/product/duramem-puramem.

[143] V. Nissen. *Einführung in Evolutionäre Algorithmen: Optimierung nach dem Vorbild der Evolution*. Braunschweig, Wiesbaden: vieweg, 1997.

[144] S. Agatonovic-Kustrin, R. Beresford, and A. M. Yusof. "ANN modeling of the penetration across a polydimethylsiloxane membrane from theoretically derived molecular descriptors." *Journal of Pharmaceutical and Biomedical Analysis* 26.2 (2001), pp. 241–254.

[145] C. M. Hurvich and C.-L. Tsai. "A corrected Akaike information criterion for vector autoregressive model selection." *Journal of Time Series Analysis* 14.3 (1993), pp. 271–279.

[146] Z. T. Wilson and N. V. Sahinidis. "The ALAMO approach to machine learning." *Computers & Chemical Engineering* 106 (2017), pp. 785–795.

[147] J. R. Koza. *Genetic programming: on the programming of computers by means of natural selection*. Cambridge, Mass.: MIT Press, 1992.

[148] A. H. Gandomi, A. H. Alavi, and C. Ryan, eds. *Handbook of Genetic Programming Applications*. Cham: Springer International Publishing, 2015.

[149] H. Shokrkar, A. Salahi, N. Kasiri, and T. Mohammadi. "Prediction of permeation flux decline during MF of oily wastewater using genetic programming." *Chemical Engineering Research and Design* 90.6 (2012), pp. 846–853.

[150] S.-M. Park, J. Han, S. Lee, J. Sohn, Y.-M. Kim, J.-S. Choi, and S. Kim. "Analysis of reverse osmosis system performance using a genetic programming technique." *Desalination and Water Treatment* 43.1-3 (2012), pp. 281–290.

[151] A. Okhovat and S. M. Mousavi. "Modeling of arsenic, chromium and cadmium removal by nanofiltration process using genetic programming." *Applied Soft Computing* 12.2 (2012), pp. 793–799.

[152] D. R. Lide, ed. *CRC handbook of chemistry and physics: A ready-reference book of chemical and physical data*. 84th ed. Boca Raton: CRC Press, 2003.

[153] I. Smallwood. *Handbook of organic solvent properties*. Butterworth-Heinemann, 2012.

[154]   C. L. Yaws. *Thermophysical properties of chemicals and hydrocarbons.* Norwich, NY: Andrew, 2008. URL: http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10305680.

[155]   C. Reichardt and T. Welton. *Solvents and Solvent Effects in Organic Chemistry.* Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2010.

[156]   J. A. Riddick and W. B. Bunger. *Organic solvents: Physical properties and methods of purification.* 3. ed. On the basis of the 1. ed. by Arnold Weissberger and Eric S. Proskauer and the compl. rev. 2. ed. by John A. Riddick and Emory E. Toops, Jr. Vol. 2. Techniques of chemistry. New York: Wiley-Interscience, 1970.

[157]   M. D. Lechner, ed. *Surface Tension of Pure Liquids and Binary Liquid Mixtures.* Landolt-Börnstein - Group IV Physical Chemistry. Berlin/Heidelberg: Springer-Verlag, 1997.

[158]   H. Gilman and R. N. Meals. "Rearrangements in the Friedel-Crafts alkylation of benzene." *The Journal of Organic Chemistry* 08.2 (1943), pp. 126–146.

[159]   A. Et-Tahir, C. Boned, B. Lagourette, and P. Xans. "Determination of the viscosity of various hydrocarbons and mixtures of hydrocarbons versus temperature and pressure." *International Journal of Thermophysics* 16.6 (1995), pp. 1309–1334.

[160]   M. V. Forward, S. T. Bowden, and W. J. Jones. "S 26. Physical properties of triphenyl compounds of group V B elements." *Journal of the Chemical Society (Resumed)* (1949), S121.

[161]   Jiangping Liu. "Prediction of Fluid Dielectric Constants." Dissertation. Provo, UT, USA: Brigham Young University, 2011.

[162]   M. Fermeglia and G. Torriano. "Density, Viscosity, and Refractive Index for Binary Systems of n -C16 and Four Nonlinear Alkanes at 298.15 K." *Journal of Chemical & Engineering Data* 44.5 (1999), pp. 965–969.

[163]   A. F. Forziati and F. D. Rossini. "Physical properties of sixty API-NBS hydrocarbons." *Journal of Research of the National Bureau of Standards* 43.473-476 (1949).

[164] M. V. Rathnam, K. Jain, and M. S. S. Kumar. "Physical Properties of Binary Mixtures of Ethyl Formate with Benzene, Isopropyl Benzene, Isobutyl Benzene, and Butylbenzene at (303.15, 308.15, and 313.15) K." *Journal of Chemical & Engineering Data* 55.4 (2010), pp. 1722–1726.

[165] O. Mumm and J. Diederichsen. "Versuche zur Theorie der Allyl-Umlagerung (III. Mitteil.)" *Berichte der deutschen chemischen Gesellschaft (A and B Series)* 72.8 (1939), pp. 1523–1527.

[166] G. B. Arrowsmith, G. H. Jeffery, and A. I. Vogel. "369. Physical properties and chemical constitution. Part XLI. Naphthalene compounds." *Journal of the Chemical Society (Resumed)* (1965), p. 2072.

[167] M. L. Huber. *Models for Viscosity, Thermal Conductivity, and Surface Tension of Selected Pure Fluids as Implemented in REFPROP v10.0.* 2018. DOI: 10.6028/NIST.IR.8209.

[168] T. R. Lynch. *Process chemistry of lubricant base stocks.* Vol. 116. Chemical industries. Boca Raton, Fla.: CRC Press, 2008. URL: http://www.loc.gov/catdir/enhancements/fy0806/2007020175-d.html.

[169] C. W. Bonhorst, P. M. Althouse, and H. O. Triebold. "Esters of Naturally Occurring Fatty Acids - Physical Properties of Methyl, Propyl, and Isopropyl Esters of C 6 to C 18 Saturated Fatty Acids." *Industrial & Engineering Chemistry* 40.12 (1948), pp. 2379–2384.

[170] Thermo Fischer Scientific. *Safety data sheet - DL-α-Tocopherol.* URL: https://www.alfa.com/en/catalog/sds/A17039/.

[171] E. Szwajczak, J. Świergiel, R. Stagraczyński, and J. Jadżyn. "Viscous and dielectric properties of alpha -tocopherol and alpha -tocopherol acetate." *Physics and Chemistry of Liquids* 47.4 (2009), pp. 460–466.

[172] Thermo Fischer Scientific. *Safety data sheet - n-decane.* URL: https://www.alfa.com/de/catalog/sds/A14732/.

[173] Thermo Fischer Scientific. *Safety data sheet - dodecane.* URL: https://www.alfa.com/de/catalog/A14834/.

[174] A. Mejía, M. Cartes, and H. Segura. "Interfacial tensions of binary mixtures of ethanol with octane, decane, dodecane, and tetradecane." *The Journal of Chemical Thermodynamics* 43.9 (2011), pp. 1395–1400.

[175] Thermo Fischer Scientific. *Safety data sheet - n-tetradecane.* URL: https://www.alfa.com/de/catalog/A14672/.

[176] Thermo Fischer Scientific. *Safety data sheet - n-hexadecane.* URL: https://www.alfa.com/de/catalog/A10322/.

[177] J. Koefoed, J. V. Villadsen, L. G. Sillén, B. Zaar, and E. Diczfalusy. "Surface Tension of Liquid Mixtures. A Micro-Method Applied to the Systems: Chloroform-Carbon-tetrachloride, Benzene-Diphenylmethane, and Heptane-Hexadecane." *Acta Chemica Scandinavica* 12 (1958), pp. 1124–1135.

[178] Thermo Fischer Scientific. *Safety data sheet - n-octadecane.* URL: https://www.alfa.com/de/catalog/031954/.

[179] M. V. Dover and W. A. Hensley. "Properties of 1-Octadecene, n-Octadecane, and Di-m-tolylethane." *Industrial & Engineering Chemistry* 27.3 (1935), pp. 337–339.

[180] Thermo Fischer Scientific. *Safety data sheet - n-docosane.* URL: https://www.alfa.com/en/catalog/A18050/.

[181] A. J. Queimada, S. E. Quiñones-Cisneros, I. M. Marrucho, J. A. P. Coutinho, and E. H. Stenby. "Viscosity and Liquid Density of Asymmetric Hydrocarbon Mixtures." *International Journal of Thermophysics* 24.5 (2003), pp. 1221–1239.

[182] A. J. Queimada, F. A. Silva, A. I. Caço, I. M. Marrucho, and J. A. Coutinho. "Measurement and modeling of surface tensions of asymmetric systems: heptane, eicosane, docosane, tetracosane and their mixtures." *Fluid Phase Equilibria* 214.2 (2003), pp. 211–221.

[183] Thermo Fischer Scientific. *Safety data sheet - n-tetracosane.* URL: https://www.alfa.com/en/catalog/042462/.

[184] R. Anonymous. *Properties of hydrocarbon of high molecular weight.* Ed. by American Petroleum Institute Research Project 42. 1968.

[185] E. F. Meyer and K. S. Stec. "Evidence against energetically favored coiling of vapor-phase paraffins up to n-tetracosane." *Journal of the American Chemical Society* 93.21 (1971), pp. 5451–5454.

[186] Thermo Fischer Scientific. *Safety data sheet - n-octacosane.* URL: https://www.alfa.com/en/catalog/042460/.

[187] W. M. Mazee. "Some properties of hydrocarbons having more than twenty carbon atoms." *Recueil des Travaux Chimiques des Pays-Bas* 67.3 (1948), pp. 197–213.

[188] Thermo Fischer Scientific. *Safety data sheet - dotriacontane.* URL: `https://www.alfa.com/en/catalog/A17480/`.

[189] Thermo Fischer Scientific. *Safety data sheet - n-hexatriacontane.* URL: `https://www.alfa.com/en/catalog/031641/`.

<div style="text-align: right; font-size: 3em;">A</div>

Superstructure optimization of process concepts

A.1 Method

In order to elucidate the effect of uncertainties of flux and rejection measurement on subsequent process design, two different case studies were investigated using a superstructure optimization approach. The superstructure represents a so-called state-space approach [138] and builds on a previous implementation for the optimization of reverse osmosis processes for seawater desalination [139]. Similar superstructures have been used for the optimization of gas permeation processes for natural gas upgrading [140] and nitrogen removal from natural gas [141]. An illustration of the superstructure, which consists of a generic distribution network and a number of OSN membrane stages is provided in Figure A.1.

The distribution network allows distribution of the entering streams, which are the overall feed stream and all product streams from the membrane stages to all outgoing streams that are the feed streams of the OSN stages and the final permeate and retentate product streams. Thereby any kind of combination between the different membrane stages is possible. For splitters mass balances are used, while for mixers mass and component balances are utilized. Furthermore, a total mass and
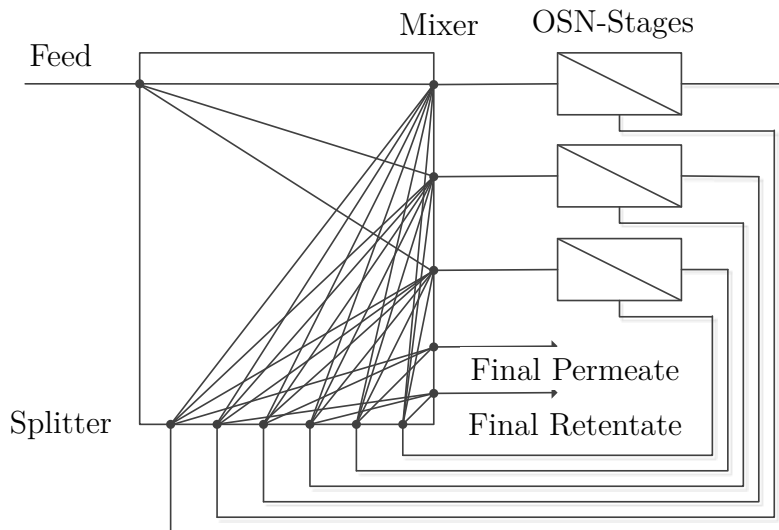
Figure A.1: Superstructure for membrane network optimization.

component balance between the feed and the final permeate and retentate is considered. The performance of each membrane stage is modeled based on the available performance metrics from the experimental data, which are an overall flux as well as solute-specific rejection. Furthermore, a concretization of the membrane stage in terms of 100 finite elements is performed to account for the composition changes along the membrane length. The permeate flux of each discrete is calculated as follows:

$$\dot{m}^{P}_{discrete} = \frac{A_{M,stage}}{N_{discretes}} \cdot J_{tot} \tag{A.1}$$

The retentate stream of each discrete is calculated based in mass balances. The concentration of each component in the retentate of the discrete is determined based on component balances, taking into account the rejection of the components. Further, the feed of each membrane discrete corresponds to the retentate of the previous discrete, except for the first discrete, where the feed of the discrete corresponds to the feed of the stage. The permeate of each stage is finally calculated based on the feed and the retentate of the stage, again by utilizing mass and component balances. However, a constant performance in terms of the experimentally determined values is assumed and concentration polarization and pressure drop are neglected, since no information on module geometry and hydrodynamics is considered. Operating

temperature and pressure are also fixed to the experimental conditions, such that only the membrane area remains as a degree of freedom for every membrane stage. For the current study, a maximum membrane area of $50\,\mathrm{m}^2$ per stage and a total of three possible stages was considered. Further constraints for optimization are that the membrane area decreases gradually in successive stages as well as the feed of each stage. The objective of the optimization is to perform a specific separation, in terms of purity and recovery constraints depending on the separation task, with a minimum number of membrane stages and membrane area. The objective function, which is minimized, is therefore as follows:

$$A_{total} = \sum_{i=1}^{N_{stage}} A_{M,stage} + N_{stages} \tag{A.2}$$

The superstructure optimization model is implemented in the general modeling framework GAMS as a mixed integer nonlinear programming problem (MINLP). The discrete decisions account for the existence of the different membrane stages, for which additional Big-M constraints enforce a minimum feed flowrate for existing membranes, while further constraints introduce additional cuts in order to eliminate topologically equivalent solutions by enforcing a specific order on the available membrane stages [139]. This is important to avoid multiplicity of equivalent optima in the solution space, which severely complicate global deterministic optimization. The resulting MINLP problem is solved using the global deterministic solver ANTIGONE in order to determine the best possible solution. In order to evaluate the effect of the extend of uncertainty related to the experimentally determined performance metrics, different scenarios are evaluated with the superstructure optimization approach considering a sampling within the ranges of the determined confidence intervals around the estimated mean values.

A.2 Case Study 2 - separation of impurity

As a second case study a separation step of two solutes solved in ethanol was chosen. The solute size equals 474 and $1098\,\mathrm{g\,mol^{-1}}$. The smaller solute is assumed to be the product while the bigger solute represents an impurity. The feed contains $0.05\,\mathrm{g\,g^{-1}}$ of the product and $0.003\,\mathrm{g\,g^{-1}}$ of the impurity while the flow rate is set to $100\,\mathrm{kg\,h^{-1}}$. The final permeate needs to contain at least $98\,\%$ of the product and the concentration of the impurity in the permeate needs to be lower than $0.001\,\mathrm{g\,g^{-1}}$.

For this case study the experimental values determined with the PDMS membrane and the mixture of ethanol with styrene oligomers are used. For this system the flux at $30\,\mathrm{bar}$ was experimentally determined to be $4.925\,\mathrm{L\,m^{-2}\,h^{-1}}$ $\pm 0.331\,\mathrm{L\,m^{-2}\,h^{-1}}$. The rejection of the styrene oligomer with a molecular weight of $474\,\mathrm{g\,mol^{-1}}$ equals $-11.8\,\%$ $\pm 6.6\,\%$ and the rejection of the styrene oligomer with a molecular weight of $1098\,\mathrm{g\,mol^{-1}}$ equals $88.2\,\%$ $\pm 6.2\,\%$. Compared to the first case study the flux is much smaller and additionally the experimentally determined confidence intervals were broader. The optimal process configuration was again determined using the means of the flux and the rejections but also for various combinations of the mean and the upper and lower limits of the confidence intervals. Overall, 27 different combinations of flux and rejections were evaluated. A summary of the results for these combinations are listed in Table A.1. Moreover, for three combinations the optimal process structure is shown in Figure A.2.
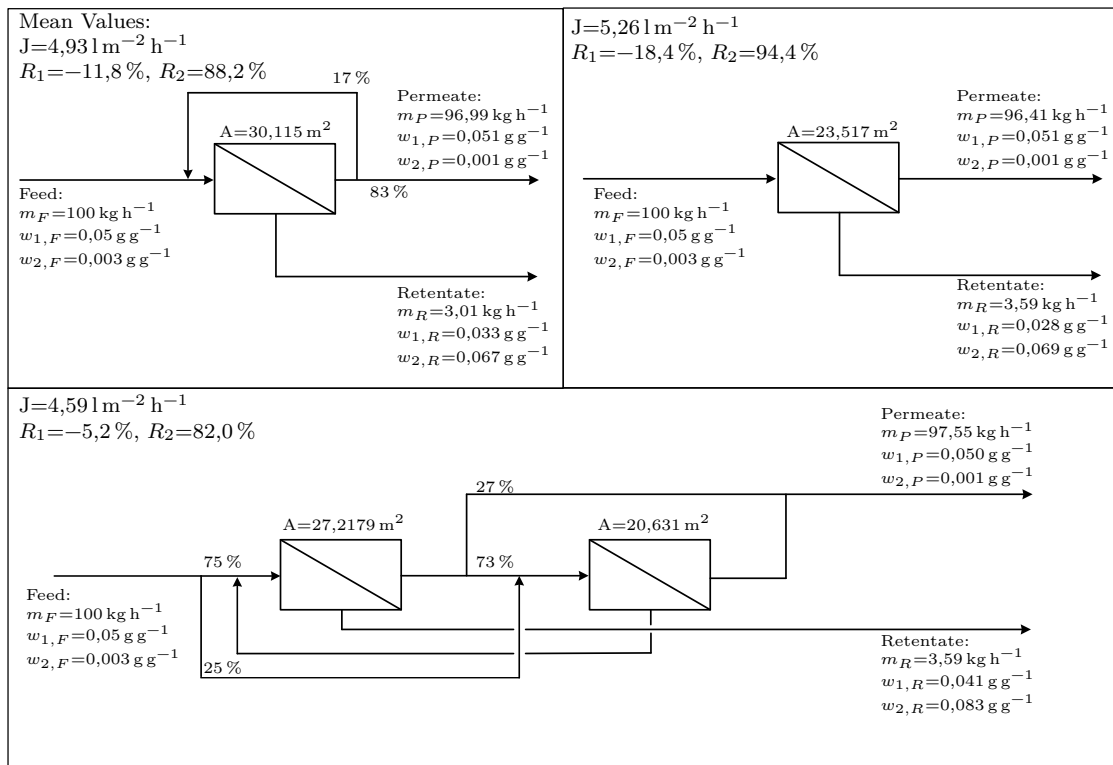
Figure A.2: Process structure for three combinations in case study 2

Considering the mean values of flux and rejection, the desired purity of the product is achieved using a single stage set-up with a recycle of $17\,\%$ of the permeate which is a rather simple set-up. Further the optimal process configurations for worst and

best combination of flux and rejection values are shown in Figure A.2 (top left). These combinations do not correspond to the combination of lower and upper limits of flux and rejection. The separation task is less complex for a high flux and high rejection of the impurity, whereas the rejection of the product is low. The optimal process configuration for this combination of flux and rejection values is shown in Figure A.2 (top right) and only a single stage without any recycle is sufficient in order to achieve the purity of the product. In comparison the process structure becomes very complex taking into account the worst combination of flux and rejection values, which are a low flux and a low rejection of the impurity while the rejection of the product is closer to zero (cf. Figure A.2 bottom). Moreover, the necessary membrane area ranges from 23.231 to 47.849 m^2. Hence, determining the feasibility of an OSN process is hard to determine taking into account the uncertainty of the experimental values.

Table A.1: Results for case study 2

Flux $\mathrm{L\,m^{-2}\,h^{-1}}$	Rejection product	impurity	Nr. stages	Membrane area $\mathrm{m^2}$
4.59	-18.4%	82.0%	2	47.342
4.59	-18.4%	88.2%	1	28.970
4.59	-18.4%	94.4%	1	26.559
4.59	-11.8%	82.0%	2	50.616
4.59	-11.8%	88.2%	1	32.272
4.59	-11.8%	94.4%	1	26.730
4.59	-5.2%	82.0%	2	47.849
4.59	-5.2%	88.2%	1	36.519
4.59	-5.2%	94.4%	1	26.886
4.93	-18.4%	82.0%	2	44.154
4.93	-18.4%	88.2%	1	27.034
4.93	-18.4%	94.4%	1	24.784
4.93	-11.8%	82.0%	2	30.801
4.93	-11.8%	88.2%	1	30.115
4.93	-11.8%	94.4%	1	24.943
4.93	-5.2%	82.0%	2	31.820
4.93	-5.2%	88.2%	1	34.078
4.93	-5.2%	94.4%	1	25.089
5.26	-18.4%	82.0%	2	41.407
5.26	-18.4%	88.2%	1	25.340
5.26	-18.4%	94.4%	1	23.231
5.26	-11.8%	82.0%	2	44.240
5.26	-11.8%	88.2%	1	28.228
5.26	-11.8%	94.4%	1	23.381
5.26	-5.2%	82.0%	3	48.400
5.26	-5.2%	88.2%	1	31.943
5.26	-5.2%	94.4%	1	23.517

B

Experimental investigation

B.1 Solvent mixtures

Additionally, the flux of solvents mixtures using one of the membranes were determined using a different experimental set-up and procedure as described in Section B.1.1. In experimental series with mixed solvents no solutes were used and hence, no rejection was determined. However, the composition of the permeate was determined using gas chromatography (Section B.1.2).

B.1.1 Experimental procedure

The permeate fluxes of the solvent mixtures are determined using the METcell CrossFlow System of Evonik Resource Efficiency GmbH [142]. Within a larger study, the plant was classified as comparable to a larger laboratory plant [16]. The experimental set-up is illustrated in Figure B.1.

The system includes a feed mix reservoir that holds a maximum volume of $800\,\mathrm{mL}$ [142]. The design of the METcell CrossFlow system allows batch experiments using up to six different membrane cells. Four METcells with a size of $2.5''$ and an active membrane area of $14\,\mathrm{cm^2}$ were used for measurements with solvent mixtures. A solvent-stable gear pump circulates the feed mixture with a constant volume flow
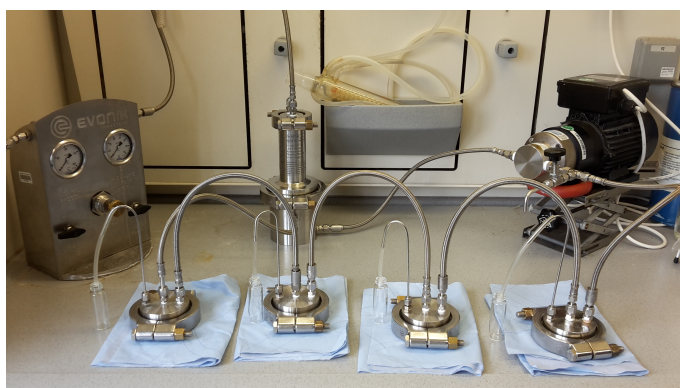
Figure B.1: Illustration of the METcell CrossFlow System

of $60\,\mathrm{L\,h^{-1}}$ (based on a solvent with a viscosity of $0.6\,\mathrm{mPa\,s}$). The pressure in the plant is applied by means of nitrogen, which is adjusted using a pressure regulator. A maximum pressure of $69\,\mathrm{bar}$ is permissible for the METcell CrossFlow system. The feed mixture flows across the membrane cells (METcells) arranged in series and is recirculated as retentate into the feed tank. Each METcell is equipped with a permeate outlet at which the permeate can be collected separately for each cell. In addition, there is a tap behind the pump which can be used to drain the feed mixture if required. Since the permeate is not recirculated, this design is less suitable for experiments with solutes, as the feed conditions change continuously without permeate recirculation. However, this system offers the advantage of investigating several membrane samples in a single experiment and, at the same time, the required quantities of solvent are lower than in the system used for the rejection measurements.

As solvents ethanol, ethyl acetate and n-heptane are used as pure solvents as well as binary and ternary mixtures. The compositions of the solvent mixtures used are shown in Figure B.2. A total of twelve binary and six ternary mixtures are used. The experiments are performed in the order given in Figure B.2, starting with pure ethanol.

Prior to the start of the experiment, samples of the PuraMem® S600 are cut to the appropriate size and built into the METcells. Subsequently, the solvent or solvent mixture is filled into the feed tank, the pump is started and the tank is closed. The system is supplied with nitrogen via the pressure control valve and a pressure of $30\,\mathrm{bar}$ is applied. The pressure is constantly monitored and readjusted during the experiment. The temperature in the system corresponds to the ambient tempera-
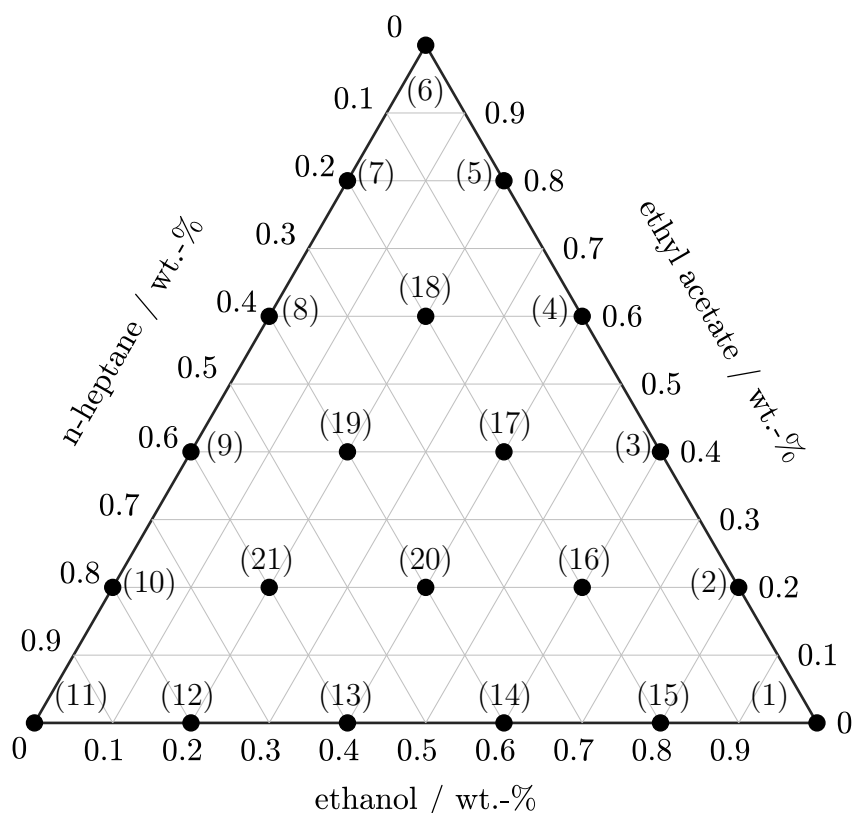
Figure B.2: Compositions of the solvent mixtures

ture of 25 °C. First, the membrane is rinsed until approximately 85 mL of permeate per cell is collected to remove conservatives. This permeate is discarded. After rinsing, the system is filled with 700 mL fresh solvent and the system is restarted. The permeate is collected over the entire duration of the test and the flux is determined gravimetrically. Once the flux in each cell is constant, the system reached steady state and samples are taken from all permeates and the feed. When changing to a new solvent mixture, the plant with the membrane samples of the previous measurement is first rinsed twice with 200 mL of the new mixture before the new membrane samples are installed.

B.1.2 Analytical methods for solvent composition

The weight fractions of the different solvents in a solvent mixture are determined by gas chromatography. The gas chromatograph GC-14B of the company Shimadzu

Europa GmbH is used for this purpose, which is equipped with a liquid auto sampler, flame ionization detector (FID) and a capillary column FS Supreme 5ms-HT. The column length was 30 m, the inner capillary diameter is 0.32 mm and the film thickness of the stationary phase is 0.25 μm. The mobile phase is formed by helium as carrier gas. Each sample was prepared for analysis in vials containing 0.8 g of the sample and 0.4 g of the internal standard dibuthyl ether. The injection volume of 0.5 μl was chosen and a split ration of 1:90 was applied. The temperature of the detector is constant at 280 °C and the injector is operated at 250 °C. During analysis a temperature program is applied. The column is first heated to a temperature of 70 °C, which is maintained for two minutes. The column is then heated to a temperature of 90 °C at a constant rate of 10 °C min^{-1}, which is kept constant for one minute. Each analysis is repeated three times and the concentration of the sample is evaluated based in the mean of the analyses and a calibration curve.

B.1.3 Experimental results
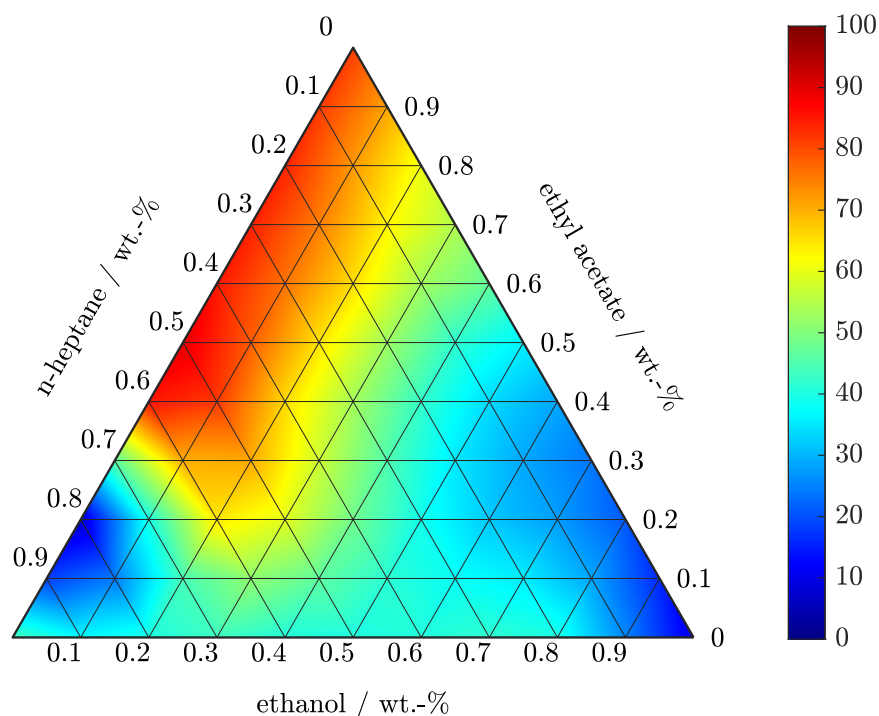
Figure B.3: Flux of the solvent mixtures of ethanol, ethyl acetate and n-heptane

XXX

Table B.1: Flux of ternary solvent mixtures of ethanol (EtOH), ethyl acetate (EtAc) and n-heptane (Hep) using the PuraMem® S600 membrane at 30 bar and the corresponding compositions of the feed and permeate as determined by gas chromatography

Mass fraction feed			Mass fraction permeate			Flux/$\mathrm{L\,m^{-2}\,h^{-1}}$
EtOH	EtAc	Hep	EtOH	EtAc	Hep	
1	0	0	1	0	0	10.07
0	1	0	0	1	0	81.45
0	0	1	0	0	1	44.39
0.79	0.21	0	0.76	0.24	0	20.87
0.60	0.40	0	0.56	0.44	0	28.72
0.40	0.60	0	0.37	0.63	0	47.13
0.21	0.79	0	0.19	0.81	0	60.35
0	0.78	0.22	0	0.77	0.23	83.94
0	0.63	0.37	0	0.63	0.37	85.23
0	0.38	0.62	0	0.41	0.59	85.16
0	0.20	0.80	0	0.21	0.79	70.59
0.20	0	0.80	0.17	0	0.83	40.15
0.47	0	0.53	0.29	0	0.71	40.54
0.68	0	0.34	0.49	0	0.51	40.72
0.83	0	0.17	0.78	0	0.22	39.94
0.61	0.20	0.18	0.54	0.22	0.24	32.01
0.43	0.39	0.18	0.36	0.41	0.23	44.62
0.22	0.58	0.20	0.19	0.59	0.21	65.46
0.23	0.40	0.37	0.19	0.41	0.40	65.81
0.42	0.21	0.37	0.32	0.21	0.47	48.41
0.22	0.19	0.59	0.18	0.19	0.62	64.34

B.2 Experimental Results for pure solvent flux and rejection

Table B.2: Flux and rejection of alkanes and SOA solved in Ethanol, Isopropanol and MEK using the DuraMem® 300 membrane

Membrane	Solute	Ethanol			Isopropanol			MEK		
		20 bar	30 bar	40 bar	20 bar	30 bar	40 bar	20 bar	30 bar	40 bar
DuraMem® 300 — Rejection	C10	0.77	0.77	0.82	0.41	0.31	0.57	0.72	0.76	0.82
	C12	0.83	0.83	0.86	0.47	0.37	0.64	0.85	0.90	0.90
	C14	0.87	0.87	0.88	0.51	0.42	0.69	0.91	0.98	0.96
	C16	0.90	0.90	0.90	0.57	0.48	0.73	0.92	0.93	0.92
	C18	0.89	0.89	0.89	0.56	0.47	0.70	0.93	1.00	0.94
	C22				0.61	0.53	0.74	0.88	1.00	1.00
	C24				0.62	0.54	0.74	1.00	1.00	1.00
	C28									
	C32									
	C36									
	SOA	0.93	0.93	0.92	0.70	0.63	0.81	1.00	1.00	0.99
	Flux / $\mathrm{L\,m^{-2}\,h^{-1}}$	11.50	17.52	24.06	3.23	3.96	2.96	15.77	10.65	21.05

Table B.3: Flux and rejection of alkanes and SOA solved in Ethanol, Isopropanol and MEK using the PuraMem® S600 and PDMS(2 µm) membranes

Membrane	Solute	Ethanol 20 bar	Ethanol 30 bar	Ethanol 40 bar	Isopropanol 20 bar	Isopropanol 30 bar	Isopropanol 40 bar	MEK 20 bar	MEK 30 bar	MEK 40 bar
PuraMem® S600 — Rejection	C10	−0.47	−0.23	−0.23	−0.12	−0.24	−0.16	−0.40	−0.20	−0.14
	C12	−0.17	−0.18	−0.36	−0.30	−0.23	−0.31	−0.40	−0.19	−0.13
	C14	−0.18	−0.24	−0.44	−0.39	−0.26	−0.39	−0.37	−0.17	−0.10
	C16	−0.31	−0.26	−0.49	−0.34	−0.29	−0.41	−0.35	−0.17	−0.05
	C18	−0.31	−0.33	−0.58	−0.10	−0.29	−0.44	−0.27	−0.06	0.01
	C22				−0.16	−0.33	−0.48	−0.16	0.01	0.11
	C24				−0.10	−0.37	−0.50	−0.10	0.05	0.16
	C28									
	C32									
	C36									
	SOA	0.83	0.84	0.82	0.72	0.79	0.82	0.93	0.93	0.96
	Flux / $L\,m^{-2}\,h^{-1}$	9.35	8.97	10.42	6.00	8.14	9.45	100.76	76.16	118.07
PDMS(2 µm) — Rejection	C10	−0.14	0.07	−0.16	0.38	−0.34	0.34	0.03	−0.15	−0.18
	C12	−0.28	−0.16	−0.43	−0.36	−0.22	−0.41	0.04	−0.15	−0.19
	C14	−0.40	−0.23	−0.54	−0.38	−0.30	−0.54	0.11	−0.13	−0.16
	C16	−0.41	−0.31	−0.59	−0.43	−0.24	−0.44	0.19	−0.09	−0.13
	C18	−0.53	−0.29	−0.76	−0.46	−0.36	−0.63	0.23	−0.03	−0.05
	C22				−0.43	−0.41	−0.56	0.30	0.10	−0.05
	C24				−0.46	−0.43	−0.60	0.35	0.18	0.15
	C28									
	C32									
	C36									
	SOA	0.90	0.87	0.93	0.78	0.75	0.78	1.00	1.00	1.00
	Flux / $L\,m^{-2}\,h^{-1}$	3.30	4.28	5.49	5.97	7.81	8.88	90.49	70.03	118.88

Table B.4: Flux and rejection of alkanes and SOA solved in ethyl acetate, toluene and n-heptane using the PuraMem® S600 and PDMS(2 μm) membranes

Membrane	Solute	Ethyl acetate			Toluene			n-Heptane		
		20 bar	30 bar	40 bar	20 bar	30 bar	40 bar	20 bar	30 bar	40 bar
PuraMem® S600 — Rejection	C10	−0.11	−0.07	−0.19	0.22	0.04	0.05	0.09	0.11	0.25
	C12	−0.10	−0.04	−0.17	0.27	0.13	0.14	0.15	0.59	0.22
	C14	−0.07	−0.01	−0.13	0.34	0.21	0.24	0.22	0.31	0.31
	C16	−0.02	0.03	−0.09	0.41	0.31	0.36	0.31	0.38	0.40
	C18	0.11	0.12	0.03	0.48	0.40	0.42	0.38	0.45	0.49
	C22	0.13	0.17	0.09	0.62	0.56	0.57	0.49	0.55	0.59
	C24	0.18	0.22	0.15	0.68	0.62	0.64	0.53	0.58	0.62
	C28				0.77	0.73	0.72	0.58	0.61	0.66
	C32				0.81	0.80	0.79	0.60	0.64	0.69
	C36				0.87	0.86	0.84	0.64	0.65	1.00
	SOA	0.97	0.97	0.97	0.94	0.92	0.90			
	Flux / L m^{-2} h^{-1}	47.71	61.81	83.28	45.98	73.23	89.14	6.69	12.73	22.03
PDMS(2 μm) — Rejection	C10	−0.10	0.01	−0.18	0.02	0.23	−0.01	0.13	0.24	0.17
	C12	−0.08	−0.01	−0.16	0.11	0.30	0.06	0.25	0.33	0.34
	C14	−0.05	0.02	−0.12	0.21	0.21	0.22	0.40	0.44	0.42
	C16	−0.02	0.08	−0.09	0.33	0.45	0.35	0.54	0.55	0.50
	C18	0.10	0.36	−0.01	0.47	0.55	0.45	0.65	0.65	0.62
	C22	0.14	−0.01	0.12	0.67	0.72	0.64	0.84	0.84	0.89
	C24	0.21	0.07	0.19	0.75	0.79	0.72	0.87	0.89	0.92
	C28				0.86	0.88	0.83	0.94	1.00	0.96
	C32				0.92	0.97	0.90	1.00	1.00	1.00
	C36				1.00	1.00	0.97	1.00	1.00	1.00
	SOA	1.00	1.00	1.00	1.00	1.00	0.97			
	Flux / L m^{-2} h^{-1}	61.71	86.53	108.31	64.26	89.36	106.53	62.34	90.40	107.16

Table B.5: Flux and rejection of the group of further solutes solved in Ethanol, Isopropanol and MEK using the PuraMem® S600 and PDMS(2 µm) membranes

Membrane	Solute	Ethanol			Isopropanol			MEK		
		20 bar	30 bar	40 bar	20 bar	30 bar	40 bar	20 bar	30 bar	40 bar
PuraMem® S600										
Rejection	IBB	−0.11	−0.16	−0.22	−0.09	−0.12	−0.24	0.04	0.03	0.03
	HB	−0.14	−0.20	−0.27	−0.11	−0.14	−0.30	0.06	0.05	0.06
	EN	−0.11	−0.16	−0.21	−0.10	−0.11	−0.26	0.11	0.12	0.15
	PD	−0.18	−0.26	−0.35	−0.14	−0.18	−0.36	0.12	0.14	0.16
	DIPN	−0.13	−0.18	−0.24	−0.09	−0.10	−0.24	0.22	0.25	0.29
	HMN	−0.17	−0.25	−0.33	−0.10	−0.10	−0.24	0.05	0.03	0.04
	IPM	−0.17	−0.24	−0.32	−0.16	−0.10	−0.23	0.21	0.24	0.27
	TPP	−0.04	−0.05	−0.06	−0.04	−0.03	−0.13	0.35	0.40	0.44
	VE	0.33	0.39	0.38	0.01	0.52	0.54	0.69	0.71	0.70
	Flux / L m^{-2} h^{-1}	6.54	8.32	10.53	7.31	9.81	12.09	64.58	88.85	108.37
PDMS(2 µm)										
Rejection	IBB	−0.13	−0.20	−0.28	−0.12	−0.19	−0.26	0.02	0.00	−0.01
	HB	−0.15	−0.25	−0.34	−0.15	−0.23	−0.30	0.04	0.03	0.03
	EN	−0.12	−0.19	−0.25	−0.12	−0.18	−0.23	0.12	0.14	0.16
	PD	−0.22	−0.35	−0.47	−0.20	−0.29	−0.38	0.12	0.12	0.14
	DIPN	−0.16	−0.26	−0.35	−0.15	−0.22	−0.28	0.26	0.29	0.32
	HMN	−0.25	−0.40	−0.55	−0.21	−0.31	−0.41	0.01	−0.02	−0.04
	IPM	−0.23	−0.37	−0.50	−0.17	−0.26	−0.32	0.23	0.25	0.27
	TPP	−0.06	−0.12	−0.14	−0.10	−0.12	−0.14	0.46	0.51	0.55
	VE	0.35	0.34	0.38	0.56	0.56	0.58	0.84	0.83	0.84
	Flux / L m^{-2} h^{-1}	4.37	5.29	6.34	8.95	11.41	13.43	70.72	90.36	107.86

Table B.6: Flux and rejection of the group of further solutes solved in ethyl acetate, toluene and n-heptane using the PuraMem® S600 and PDMS(2 µm) membranes

Membrane	Solute	Ethyl acetate			Toluene			n-Heptane		
		20 bar	30 bar	40 bar	20 bar	30 bar	40 bar	20 bar	30 bar	40 bar
PuraMem® S600 (Rejection)	IBB	0.04	0.04	0.04	0.10	0.11	0.13	0.01	0.04	0.04
	HB	0.06	0.07	0.07	0.17	0.19	0.23	0.06	0.10	0.11
	EN	0.09	0.11	0.12	0.16	0.18	0.22	0.01	0.03	0.03
	PD	0.15	0.16	0.18	0.35	0.38	0.43	0.22	0.29	0.32
	DIPN	0.23	0.26	0.30	0.36	0.40	0.45	0.18	0.26	0.28
	HMN	0.11	0.11	0.12	0.27	0.29	0.34	0.32	0.40	0.42
	IPM	0.27	0.30	0.33	0.49	0.51	0.56	0.30	0.38	0.40
	TPP	0.37	0.36	0.56						
	VE	0.72	0.68	0.75	0.82	0.78	0.81	0.53	0.58	0.60
	Flux / L m⁻² h⁻¹	59.39	81.23	96.19	39.43	60.10	76.34	9.35	15.84	25.43
PDMS(2 µm) (Rejection)	IBB	0.03	0.01	0.00	0.09	0.11	0.13	0.04	0.05	0.06
	HB	0.05	0.04	0.03	0.18	0.22	0.25	0.10	0.13	0.15
	EN	0.11	0.12	0.14	0.19	0.23	0.26	0.05	0.06	0.07
	PD	0.14	0.15	0.16	0.42	0.47	0.50	0.34	0.39	0.42
	DIPN	0.25	0.28	0.31	0.45	0.50	0.53	0.29	0.34	0.37
	HMN	0.06	0.05	0.03	0.33	0.36	0.38	0.49	0.54	0.55
	IPM	0.29	0.32	0.34	0.62	0.65	0.66	0.47	0.52	0.53
	TPP	0.55	0.67	0.72						
	VE	1.00	0.86	0.88	1.00	1.00	1.00	0.96	0.87	0.87
	Flux / L m⁻² h⁻¹	68.52	91.59	113.23	62.98	88.03	113.88	68.10	100.96	133.31

Table B.7: Flux and rejection of the group of further solutes solved in acetone and n-butyl acetate using the PuraMem® S600 and PDMS(2 µm) membranes

Membrane		Solute	Acetone			n-Butyl acetate		
			20 bar	30 bar	40 bar	20 bar	30 bar	40 bar
PuraMem® S600	Rejection	IBB	−0.02	−0.02	−0.02	0.03	0.04	0.04
		HB	−0.03	−0.02	−0.02	0.06	0.08	0.09
		EN	0.03	0.06	0.08	0.09	0.12	0.14
		PD	−0.01	0.02	0.02	0.19	0.23	0.26
		DIPN	0.12	0.17	0.21	0.27	0.33	0.36
		HMN	−0.07	−0.07	−0.07	0.17	0.21	0.23
		IPM	0.08	0.12	0.08	0.31	0.38	0.40
		TPP	0.27	0.34	0.38	0.39	0.46	0.49
		VE	0.00	0.64	0.65	0.79	0.84	0.83
	Flux [L m^{-2} h^{-1}]		49.08	64.85	77.06	28.10	40.68	53.17
PDMS(2 µm)	Rejection	IBB	−0.08	−0.10	−0.13	0.02	0.03	0.04
		HB	−0.09	−0.11	−0.16	0.06	0.07	0.10
		EN	0.01	0.04	0.03	0.11	0.14	0.17
		PD	−0.09	−0.11	−0.18	0.21	0.25	0.28
		DIPN	0.06	0.09	0.09	0.31	0.36	0.40
		HMN	−0.21	−0.29	−0.41	0.18	0.20	0.22
		IPM	−0.03	−0.04	−0.08	0.38	0.42	0.45
		TPP	0.30	0.37	0.41	0.52	0.57	0.59
		VE	0.60	0.67	0.69	1.00	1.00	1.00
	Flux [L m^{-2} h^{-1}]		43.20	54.75	65.02	47.05	64.29	80.40

Table B.8: Flux and rejection of styrene oligomers solved in ethanol using the DuraMem® 150, DuraMem® 200 and DuraMem® 300 membranes

Membrane		Mw / $\mathrm{g\,mol^{-1}}$	Ethanol		
			20 bar	30 bar	40 bar
DuraMem® 300	Rejection	266	0.86	0.86	0.79
		370	0.85	0.86	0.80
		474	0.85	0.85	0.79
		578	1.00	1.00	1.00
		682	1.00	1.00	1.00
		786	1.00	1.00	1.00
		890	1.00	1.00	1.00
		994	1.00	1.00	1.00
		1098	1.00	1.00	1.00
		1800			
	Flux / $\mathrm{L\,m^{-2}\,h^{-1}}$		0.76	0.83	1.07
DuraMem® 200	Rejection	266	0.88	0.83	0.88
		370	0.89	0.85	0.89
		474	0.89	0.85	0.89
		578	0.90	0.85	0.89
		682	1.00	1.00	1.00
		786	1.00	1.00	1.00
		890	1.00	1.00	1.00
		994	1.00	1.00	1.00
		1098	1.00	1.00	1.00
		1800			
	Flux / $\mathrm{L\,m^{-2}\,h^{-1}}$		6.19	9.08	12.10
DuraMem® 300	Rejection	266	0.97	0.96	0.96
		370	0.97	0.96	0.96
		474	0.96	0.96	0.97
		578	0.96	0.97	0.96
		682	0.96	0.96	0.97
		786	1.00	1.00	1.00
		890	1.00	1.00	1.00
		994	1.00	1.00	1.00
		1098	1.00	1.00	1.00
		1800			
	Flux [$\mathrm{L\,m^{-2}\,h^{-1}}$]		10.44	13.99	19.52

Table B.9: Flux and rejection of styrene oligomers solved in ethanol and toluene using the PuraMem® S600 and PDMS(2 μm) membranes (*=Peak was not clearly identifiable in analysis of chromatogram)

Membrane		Mw / g mol⁻¹	Ethanol			Toluene		
			20 bar	30 bar	40 bar	20 bar	30 bar	40 bar
PuraMem® S600	Rejection	266	−0.02	−0.13	−0.14	0.56	0.63	0.51
		370	0.01	−0.19	−0.01	0.82	0.84	0.78
		474	0.12	0.09	0.14	0.92	0.93	0.90
		578	0.28	0.29	0.33	0.90	0.90	0.88
		682	0.40	0.43	0.50	0.94	0.94	0.91
		786	0.53	0.56	0.59	0.92	0.93	0.89
		890	0.65	0.67	0.70	*	*	*
		994	0.72	0.74	0.74	*	*	*
		1098	0.69	0.76	1.00	*	*	*
		1800				1.00	1.00	1.00
	Flux / L m⁻² h⁻¹		5.65	7.74	9.95	46.55	68.75	90.38
PDMS(2 μm)	Rejection	266	−0.08	−0.32	−0.92	0.56	0.66	0.67
		370	−0.07	−0.33	−0.86	0.84	0.86	0.87
		474	0.02	−0.20	−0.61	0.93	0.93	0.93
		578	0.15	0.01	−0.26	0.98	1.00	1.00
		682	0.40	0.26	0.11	1.00	1.00	1.00
		786	0.48	0.34	0.35	1.00	1.00	1.00
		890	0.65	0.70	0.68	1.00	1.00	1.00
		994	0.73	0.80	0.69	1.00	1.00	1.00
		1098	0.89	0.84	0.82	1.00	1.00	1.00
		1800	0.97	0.97	0.97	1.00	1.00	1.00
	Flux / L m⁻² h⁻¹		3.14	4.30	5.48	67.01	94.12	116.94

C

Discrimination of OSN models

C.1 PuraMem membranes

Additionally to the results presented in Section 6.3, the parameter values of the models, the corresponding confidence intervals and the eigenvalues for the membranes PuraMem® 280 and PuraMem® S600 used in acetone with α-methyl styrene as solute are listed in Table C.1 and C.2. The results of the Monte-Carlo simulation for these data sets are illustrated in Figure C.1 and C.2.

Further, the membranes were used in ACN and toluene with α-methyl styrene as solute. The determined values for Resnorm, AIC and W of the different models fitted to these data sets are listed in Table C.3 to C.4. Moreover, the parameter values of the models, the corresponding confidence intervals and the eigenvalues are listed in Table C.5 to C.7 and the results of the Monte-Carlo simulation are displayed in Figure C.3 to C.6.

Table C.1: Parameter values and confidence intervals for PuraMem® 280 and PuraMem® S600 used in Acetone and with α-methyl styrene as solute

Model	Parameter	PuraMem® 280		PuraMem® S600	
		value	CI	value	CI
SSD	P_s^{SSD}	2.22e−7	1.85e−8	5.58e−8	4.68e−8
	P_j^{SSD}	4.35e−5	9.73e−6	7.48e−5	6.43e−5
SDIM	L_m	2.22e−7	2.01e−8	5.58e−8	2.40e−8
	L_{IM}	−3.41e−12	1.09e−11	7.97e−12	5.69e−12
	P_m	5.15e−5	2.76e−5	2.58e−5	3.10e−5
CSD	P_s^{CSD}	7.89	6.01e−1	1.98	1.22
	P_j^{CSD}	4.23e−5	8.64e−6	5.65e−5	3.57e−5
DSPF	$K_{HP} \cdot 10^9$	6.67e−2	1.36e−3	1.68e−2	3.43e−4
	$r_p \cdot 10^9$	5.85e−1	1.01e−1	1.02	1.42e−1
IT	L_p	2.22e−7	1.77e−8	5.58e−8	1.25e−8
	σ	1.10	4.92e−1	4.61e−1	1.50e−1
	L_j	1.66e−5	1.68e−5	7.13e−6	5.93e−6
MS	b_1	1.61e2	9.88e2	−1.59e7	3.10e10
	b_2	3.97e−8	3.41e−7	2.32e−6	7.78e−5
	b_3	1.46e3	4.43e3	2.39e3	5.33e3
	K_s	1.45	1.30	−3.38e−1	1.27e1
	$K_{MS,j}$	−8.38	2.10e1	1.81e−3	3.56

Table C.2: Eigenvalues for all models and the for membranes PuraMem® 280 and PuraMem® S600 used in Acetone and with α-methyl styrene as solute

Model	EV Nr.	PuraMem® 280	PuraMem® S600
SSD	1	1.09e8	1.59e7
	2	3.48e13	5.84e14
SDIM	1	1.58e7	1.79e7
	2	3.09e13	3.44e13
	3	7.90e20	1.25e22
CSD	1	2.38e$-$2	2.38e$-$2
	2	1.34e8	5.50e8
DSPF	1	2.70e6	2.70e6
	2	9.58	1.05
IT	1	3.73e$-$2	2.13e$-$1
	2	4.84e8	7.33e8
	3	3.30e13	4.99e13
MS	1	$-$2.03e$-$9	1.03e$-$24
	2	1.76e$-$8	2.00e$-$9
	3	3.04e$-$4	1.24e$-$3
	4	1.67e3	1.79e4
	5	1.55e14	3.32e11

Table C.3: Resnorm, AIC and W of the simple Solution-diffusion (SSD) model, the solution-diffusion with imperfections (SDIM) model, the classical solution-diffusion (CSD) model, the Donnan Steric pore-flow (DSPF) model, the irreversible thermodynamics (IT) model and the Maxwell-Stefan (MS) model for PuraMem® 280 and PuraMem® S600 used in ACN and α-methyl styrene as solute

	PuraMem® 280			PuraMem® S600		
Model	Resnorm	AIC	W	Resnorm	AIC	W
SSD	0.23	$-$36.24	2.0%	24.19	38.34	0.0%
SDIM	0.14	$-$42.03	36.0%	1.63	$-$2.79	0.0%
CSD	0.06	$-$56.38	0.0%	0.42	$-$26.44	0.4%
DSPF	0.22	$-$37.13	3.1%	1.61	$-$5.04	0.0%
IT	0.13	$-$42.90	55.6%	1.10	$-$9.17	0.0%
MS	0.21	$-$31.34	0.2%	0.18	$-$34.20	100.0%
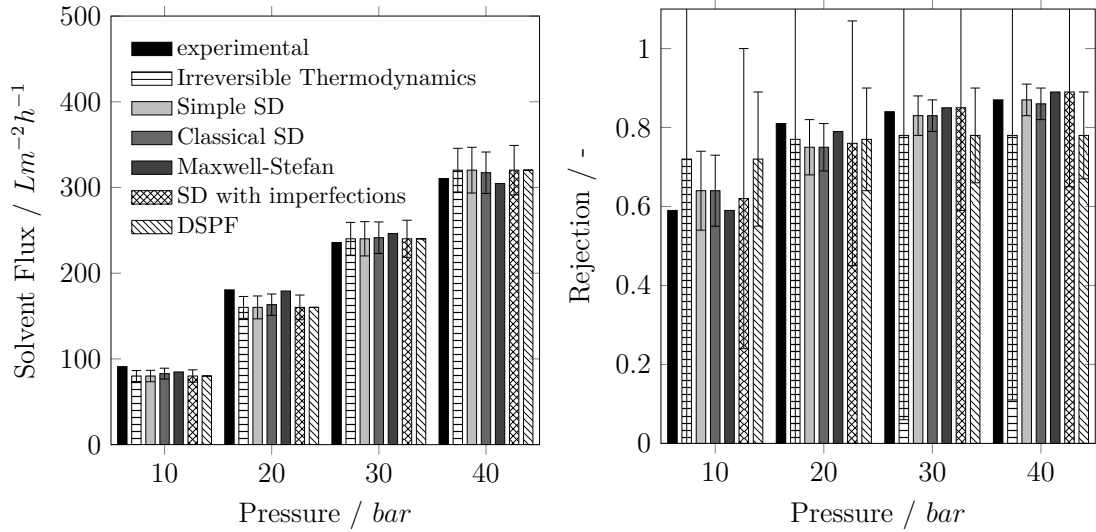
Figure C.1: Experimental data (provided by Marchetti and Livingston [47]) and calculated values for flux and rejection for the mixture of acetone and α-methyl styrene using the PuraMem® 280, error bars represent the uncertainty regions determined by Monte-Carlo simulation (no bars are shown for Maxwell-Stefan model, due to extremely broad uncertainty regions)

Table C.4: Resnorm, AIC and W of the simple Solution-diffusion (SSD) model, the solution-diffusion with imperfections (SDIM) model, the classical solution-diffusion (CSD) model, the Donnan Steric pore-flow (DSPF) model, the irreversible thermodynamics (IT) model and the Maxwell-Stefan (MS) model for PuraMem® 200 and PuraMem® 500 used in Toluene and α-methyl styrene as solute

	PuraMem® 280			PuraMem® S600		
Model	Resnorm	AIC	W	Resnorm	AIC	W
SSD	0.05	−59.37	0.0%	0.56	−21.83	0.0%
SDIM	0.02	−71.44	19.4%	0.43	−24.20	0.1%
CSD	1.12	−10.88	0.0%	0.00	0.00	0.0%
DSPF	0.31	−31.18	0.0%	0.66	−19.37	0.0%
IT	0.02	−74.26	79.5%	0.19	−37.29	99.4%
MS	0.03	−65.64	1.1%	0.81	−10.10	0.0%

Table C.5: Parameter values and confidence intervals for PuraMem® 280 and PuraMem® S600 used in ACN and with α-methyl styrene as solute

Model	Parameter	PuraMem® 280		PuraMem® S600	
		value	CI	value	CI
SSD	P_s^{SSD}	2.14e−7	3.60e−8	1.17e−8	1.30e−7
	P_j^{SSD}	3.76e−5	2.19e−5	1.63e−4	1.81e−3
SDIM	L_m	2.14e−7	3.84e−8	1.17e−8	1.01e−8
	L_{IM}	8.20e−12	2.32e−11	5.07e−12	4.40e−12
	P_m	2.19e−5	4.99e−5	1.16e−6	1.37e−5
CSD	P_s^{CSD}	1.08e1	1.72	5.92e−1	2.08
	P_j^{CSD}	3.73e−5	2.04e−5	4.77e−5	1.68e−4
DSPF	$K_{HP} \cdot 10^9$	7.27e−2	1.17e−3	3.99e−3	6.35e−5
	$r_p \cdot 10^9$	5.47e−1	5.39e−2	2.89	1.11
IT	L_p	2.14e−7	3.82e−8	1.17e−8	6.83e−9
	σ	8.95e−1	2.94e−1	1.73e1	1.00e4
	L_j	5.45e−6	1.11e−5	8.67e−4	5.04e−1
MS	b_1	2.40e1	1.09e2	−4.55	1.88
	b_2	4.16	1.83e1	1.88e3	1.38e3
	b_3	2.40e1	1.08e2	−4.59	1.91
	K_s	−4.45e8	1.48e15	1.31e8	6.58e12
	$K_{MS,j}$	3.83e−1	1.27e6	4.49e2	2.26e7

Table C.6: Parameter values and confidence intervals for PuraMem® 280 and PuraMem® S600 used in Toluene and with α-methyl styrene as solute

Model	Parameter	PuraMem® 280 value	CI	PuraMem® S600 value	CI
SSD	P_s^{SSD}	7.71e−8	6.27e−9	7.40e−8	2.26e−8
	P_j^{SSD}	5.95e−6	7.96e−7	4.76e−5	1.54e−5
SDIM	L_m	7.71e−8	5.71e−9	7.40e−8	2.01e−8
	L_{IM}	−4.37e−13	6.52e−13	2.59e−12	3.69e−12
	P_m	6.81e−6	1.49e−6	3.50e−5	2.05e−5
CSD	P_s^{CSD}	1.94	1.72e−1	1.86	4.32e−1
	P_j^{CSD}	5.98e−6	8.71e−7	3.92e−5	9.66e−6
DSPF	$K_{HP} \cdot 10^9$	4.55e−2	2.40e−4	4.37e−2	2.31e−4
	$r_p \cdot 10^9$	5.29e−1	8.43e−2	8.70e−1	1.71e−1
IT	L_p	7.71e−8	5.60e−9	7.40e−8	1.01e−8
	σ	1.06	1.18e−1	9.31e−1	5.60e−1
	L_j	3.17e−6	1.12e−6	2.00e−5	1.85e−5
MS	b_1	6.93e2	3.84e3	3.96e8	1.51e14
	b_2	3.13e−3	1.50e3	−3.97e5	5.90e11
	b_3	5.29e2	3.52e3	1.49e3	2.28e3
	K_s	8.04	1.89e2	4.34e7	1.67e17
	$K_{MS,j}$	−1.72e−3	8.27e2	−3.14e−6	1.20e4

Table C.7: Eigenvalues for all models and the for membranes PuraMem® 280 and PuraMem® S600 used in ACN and Toluene and with α-methyl styrene as solute

Model	EV Nr.	ACN PuraMem® 280	ACN PuraMem® S600	Toluene PuraMem® 280	Toluene PuraMem® S600
SSD	1	8.12e7	1.55e5	1.86e9	6.46e7
	2	3.27e13	2.21e16	4.76e13	2.73e14
SDIM	1	1.77e7	1.64e7	4.42e8	2.88e7
	2	3.02e13	3.02e13	3.39e13	3.82e13
	3	4.09e20	1.36e23	1.36e22	1.48e22
CSD	1	1.18e−2	1.18e−2	4.74e−2	4.74e−2
	2	9.11e7	1.54e9	2.92e9	8.70e8
DSPF	1	3.47e6	3.47e6	1.04e7	1.04e7
	2	9.24	6.11e−3	8.58	2.31
IT	1	5.10e−1	−1.37e−5	6.53e−2	1.65e−2
	2	1.25e9	4.20e3	5.57e9	2.27e8
	3	3.08e13	1.32e14	4.37e13	6.69e13
MS	1	5.33e−13	2.52e−13	8.80e−11	7.60e−30
	2	3.69e−6	1.74e−10	2.07e−9	1.27e−23
	3	1.99e−2	6.36e−8	3.21e−9	3.97e−23
	4	5.71e2	1.23e−2	2.57e−4	7.39e−8
	5	4.20e3	4.92e3	4.62e4	1.34e11
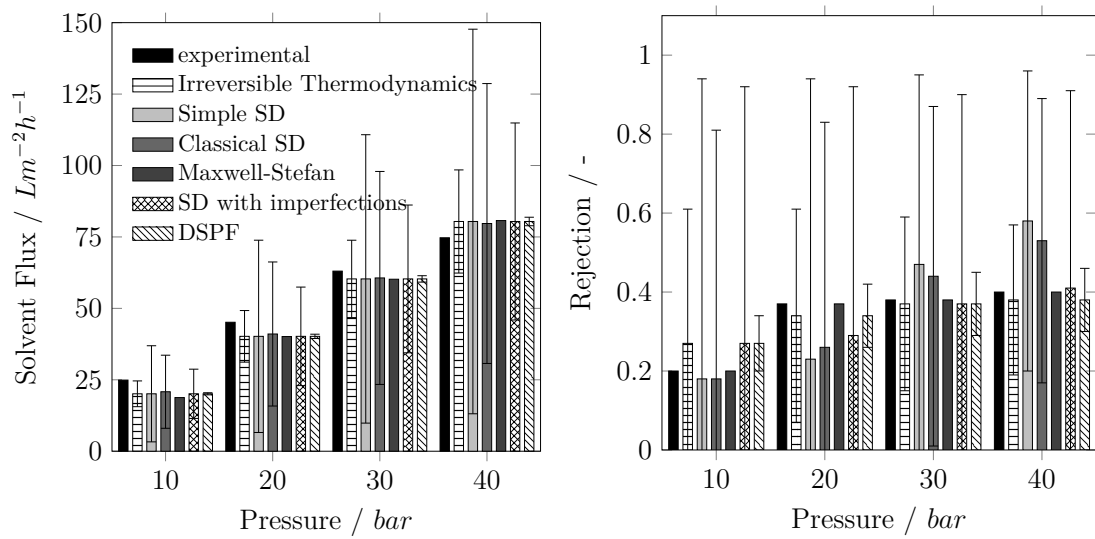
Figure C.2: Experimental data (provided by Marchetti and Livingston [47]) and calculated values for flux and rejection for the mixture of acetone and α-methyl styrene using the PuraMem® S600, error bars represent the uncertainty regions determined by Monte-Carlo simulation (no bars are shown for Maxwell-Stefan model, due to extremely broad uncertainty regions)
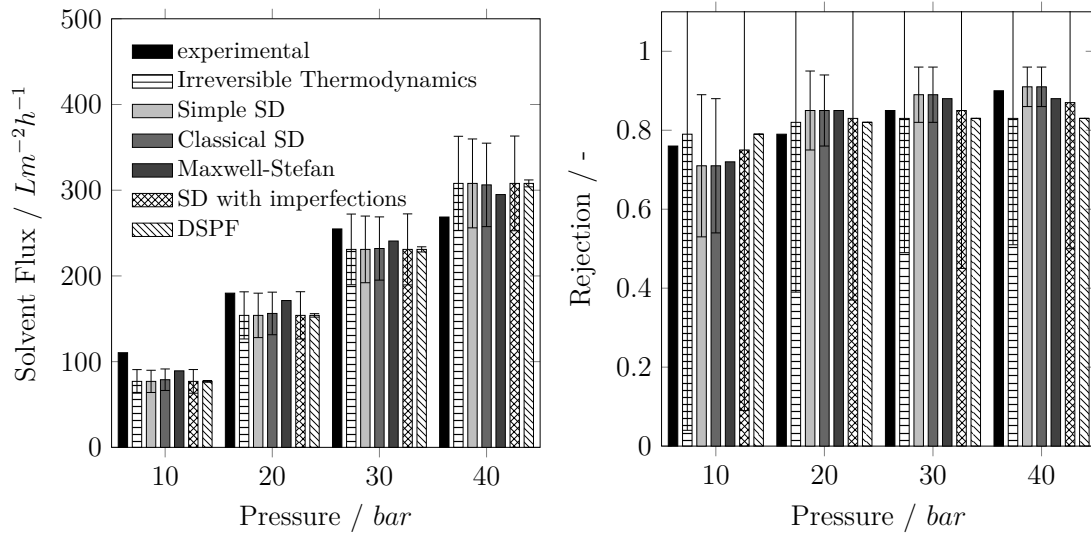
Figure C.3: Experimental data (provided by Marchetti and Livingston [47]) and calculated values for flux and rejection for the mixture of ACN and α-methyl styrene using the PuraMem® 280, error bars represent the uncertainty regions determined by Monte-Carlo simulation (no bars are shown for Maxwell-Stefan model, due to extremely broad uncertainty regions)
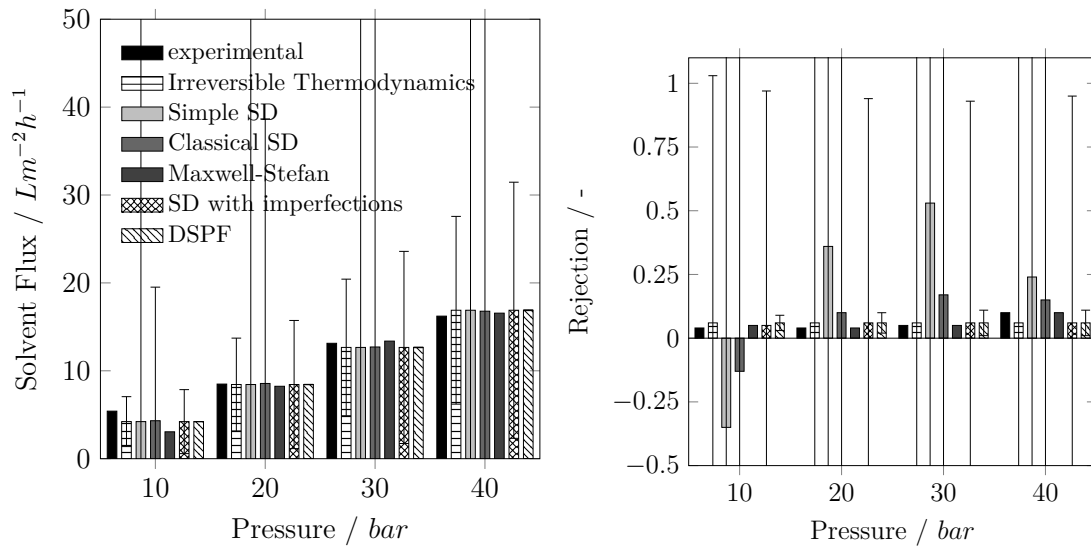
Figure C.4: Experimental data (provided by Marchetti and Livingston [47]) and calculated values for flux and rejection for the mixture of ACN and α-methyl styrene using the PuraMem® S600, error bars represent the uncertainty regions determined by Monte-Carlo simulation (no bars are shown for Maxwell-Stefan model, due to extremely broad uncertainty regions)
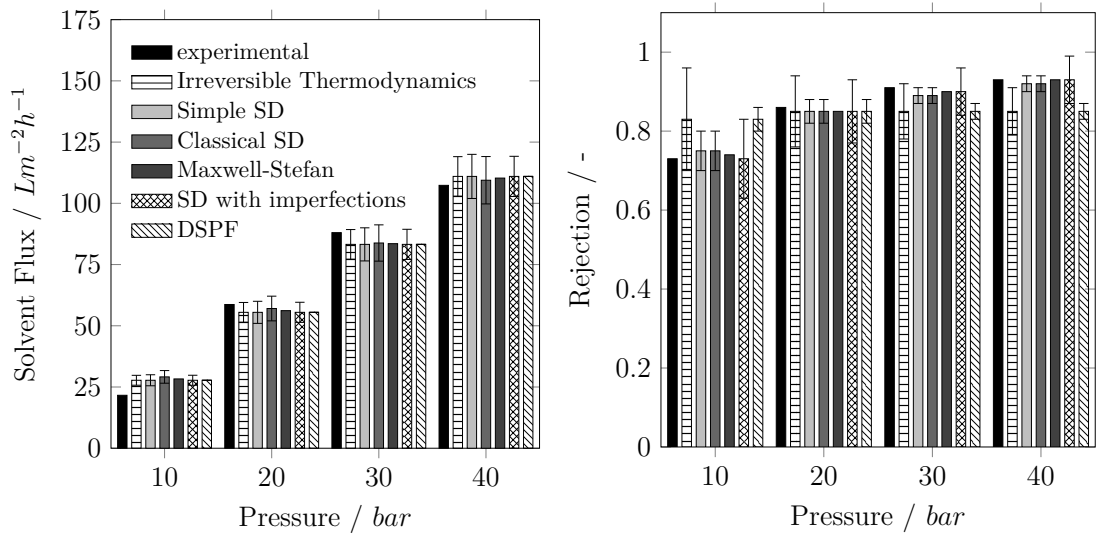
Figure C.5: Experimental data (provided by Marchetti and Livingston [47]) and calculated values for flux and rejection for the mixture of toluene and α-methyl styrene using the PuraMem® 280, error bars represent the uncertainty regions determined by Monte-Carlo simulation (no bars are shown for Maxwell-Stefan model, due to extremely broad uncertainty regions)
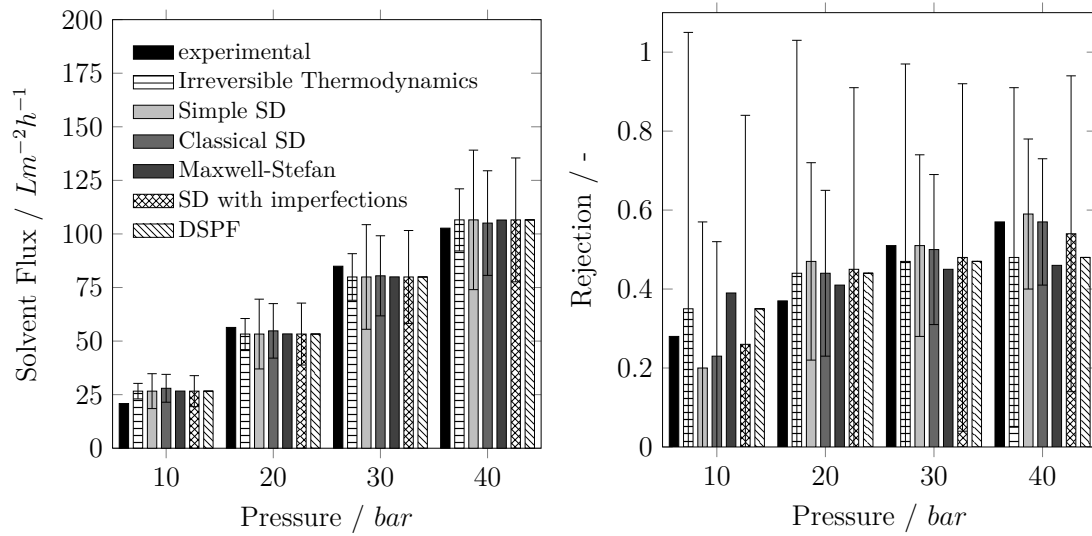
Figure C.6: Experimental data (provided by Marchetti and Livingston [47]) and calculated values for flux and rejection for the mixture of toluene and α-methyl styrene using the PuraMem® S600, error bars represent the uncertainty regions determined by Monte-Carlo simulation (no bars are shown for Maxwell-Stefan model, due to extremely broad uncertainty regions)

C.2 DuraMem membranes

The results of the parameter estimation, the confidence intervals and the eigen values for the membranes DuraMem® 200 and DuraMem® 500 used in ACN are listed in the following tables. Figure C.7 and C.8 illustrates the results of the corresponding Monte-Carlo simulations.

Table C.8: Resnorm, AIC and W of the simple Solution-diffusion (SSD) model, the solution-diffusion with imperfections (SDIM) model, the classical solution-diffusion (CSD) model, the Donnan Steric pore-flow (DSPF) model, the irreversible thermodynamics (IT) model and the Maxwell-Stefan (MS) model for DuraMem® 200 and DuraMem® 500 used in ACN and α-methyl styrene as solute

	DuraMem® 200			DuraMem® 500		
Model	Resnorm	AIC	W	Resnorm	AIC	W
SSD	0.01	-93.43	26.4%	0.15	-43.50	18.6%
SDIM	0.01	-91.45	9.8%	0.14	-41.54	7.0%
CSD	0.01	-93.44	26.5%	0.19	-39.55	3.6%
DSPF	0.01	-93.40	26.0%	0.39	-27.66	0.0%
IT	0.01	-91.47	9.9%	0.11	-45.45	49.3%
MS	0.01	-87.37	1.3%	0.15	-37.40	0.9%

Table C.9: Parameter values and confidence intervals for DuraMem® 200 and DuraMem® 500 used in ACN and with α-methyl styrene as solute

		DuraMem® 200		DuraMem® 500	
Model	Parameter	value	CI	value	CI
SSD	P_s^{SSD}	1.01e$-$7	3.58e$-$9	1.74e$-$7	2.72e$-$8
	P_j^{SSD}	1.42e$-$7	7.71e$-$7	4.33e$-$5	1.57e$-$5
SDIM	L_m	1.01e$-$7	4.12e$-$9	1.74e$-$7	3.12e$-$8
	L_{IM}	8.04e$-$15	1.08e$-$12	$-$4.14e$-$13	1.74e$-$11
	P_m	1.31e$-$7	1.82e$-$6	4.42e$-$5	4.22e$-$5
CSD	P_s^{CSD}	5.12	1.57e$-$1	8.79	1.30
	P_j^{CSD}	1.45e$-$7	6.83e$-$7	4.21e$-$5	1.45e$-$5
DSPF	$K_{HP} \cdot 10^9$	3.45e$-$2	5.48e$-$4	5.91e$-$2	9.49e$-$4
	$r_p \cdot 10^9$	3.47e$-$1	6.33e$-$3	5.73e$-$1	9.18e$-$2
IT	L_p	1.01e$-$7	4.12e$-$9	1.74e$-$7	3.03e$-$8
	σ	1.00	4.18e$-$2	1.04	6.56e$-$1
	L_j	2.92e$-$8	3.79e$-$7	1.05e$-$5	1.73e$-$5
MS	b_1	8.94e1	7.32e2	1.86e3	1.07e4
	b_2	$-$3.75	1.47e7	9.49e$-$8	1.13e$-$6
	b_3	8.74e1	6.99e2	5.19e2	1.16e4
	K_s	$-$8.03e9	1.45e17	3.80e4	6.27e9
	$K_{MS,j}$	$-$5.90	1.00e8	$-$6.01e5	9.92e10

Table C.10: Eigenvalues for all models and the for membranes DuraMem® 200 and DuraMem® 500 used in ACN and with α-methyl styrene as solute

| Model	EV Nr.	DuraMem® 200	DuraMem® 500
SSD	1	6.45e8	8.93e7
	2	3.00e13	3.68e13
SDIM	1	1.53e8	1.64e7
	2	3.00e13	3.11e13
	3	1.82e21	6.19e20
CSD	1	1.18e−2	1.18e−2
	2	6.22e8	1.16e8
DSPF	1	3.47e6	3.47e6
	2	8.65e−1	9.62
IT	1	2.87e−1	6.80e−2
	2	1.37e10	9.68e8
	3	3.00e13	3.40e13
MS	1	−8.96e−18	−4.39e−6
	2	4.65e−18	1.72e−13
	3	9.84e−10	3.65e−10
	4	3.93e−7	9.20e−7
	5	1.40e−1	2.28e133
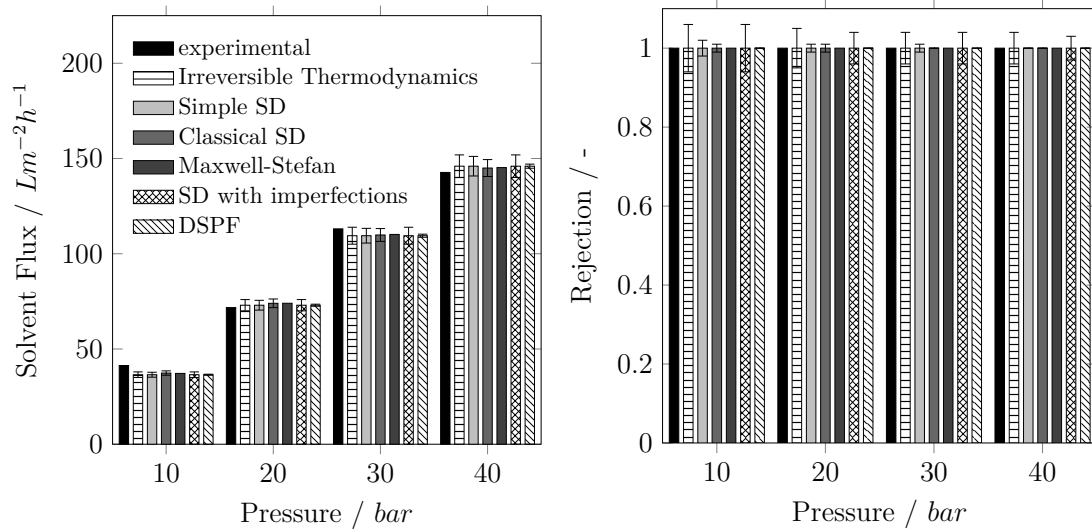
Figure C.7: Experimental data (provided by Marchetti and Livingston [47]) and calculated values for flux and rejection for the mixture of ACN and α-methyl styrene using the DuraMem® 200, error bars represent the uncertainty regions determined by Monte-Carlo simulation (no bars are shown for Maxwell-Stefan model, due to extremely broad uncertainty regions)
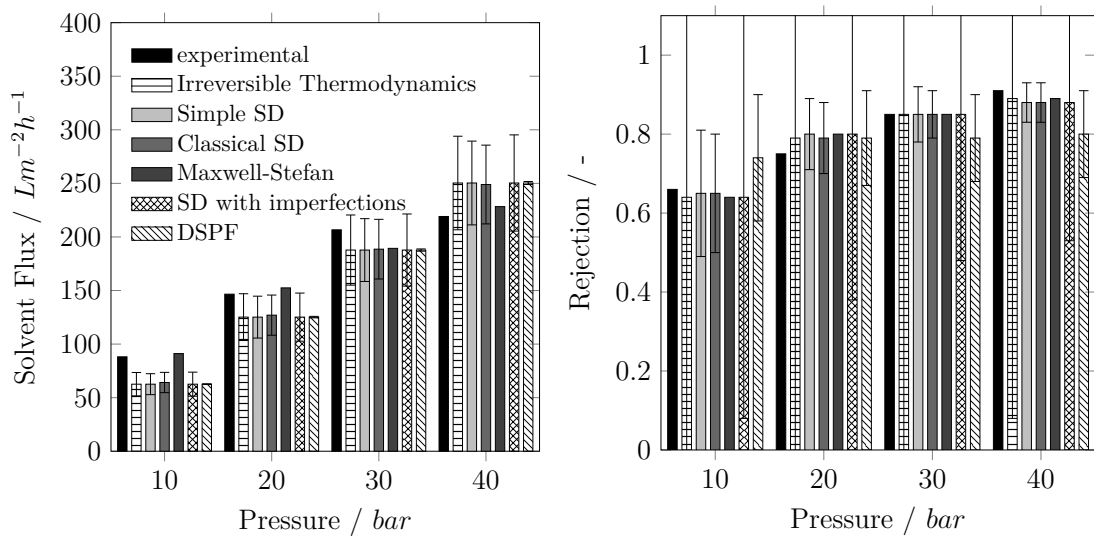
Figure C.8: Experimental data (provided by Marchetti and Livingston [47]) and calculated values for flux and rejection for the mixture of ACN and α-methyl styrene using the DuraMem® 500, error bars represent the uncertainty regions determined by Monte-Carlo simulation (no bars are shown for Maxwell-Stefan model, due to extremely broad uncertainty regions)

C.3 PDMS based membrane

The data set of Postel et al. [66] includes rejection and flux data for a PDMS based membrane used in Isopropanol (IPA), Methanol (MeOH), Toluene (Tol) and n-Hexane (Hex) with n-tetradecane as the solute. The results for the Resnorm, the AIC and W are listed in Table C.11. Estimated model parameters, confidence intervals and eigenvalues are listed in Table C.12 and C.14. Moreover, Figure C.9 to C.12 illustrate the results of the Monte-Carlo simulation. Due to the limited number of experimental data points, the Maxwell-Steffan model was not applicable for these data sets.

Table C.11: Resnorm, AIC and W of the simple Solution-diffusion (SSD) model, the solution-diffusion with imperfections (SDIM) model, the classical solution-diffusion (CSD) model, the Donnan Steric pore-flow (DSPF) model, the irreversible thermodynamics (IT) model and the Maxwell-Stefan (MS) model for Postel et al. [66] data set

Model	IPA			MeOH			Tol			Hex		
	Resnorm	AIC	W	Resnorm	AIC	W	Resnorm	AIC	W	Resnorm	AIC	W
SSD	0.64	−5.13	0.0%	1.22	0.05	0.0%	1.50	1.72	0.0%	1.08	−0.95	0.0%
SDIM	0.00	−107.40	100.0%	0.00	−101.26	100.0%	0.00	−95.20	100.0%	0.00	−99.72	100.0%
CSD	3.25	7.89	0.0%	10.40	17.19	0.0%	1.02	−1.39	0.0%	1.01	−1.50	0.0%
DSPF	2.94	7.07	0.0%	2.85	6.84	0.0%	0.17	−15.92	0.0%	0.04	26.86	0.0%
IT	0.00	−67.63	0.0%	0.04	−24.40	0.0%	0.00	−46.33	0.0%	0.00	−44.34	0.0%

Table C.12: Parameter values and confidence intervals for data set of Postel et al. [66] (Isopropanol and Methanol)

Model	Parameter	IPA value	IPA CI	MeOH value	MeOH CI
SSD	P_s^{SSD}	1.35e−7	5.66e−8	1.68e−7	1.59e−7
	P_j^{SSD}	−6.83e−5	2.98e−5	−1.11e−4	1.11e−4
SDIM	L_m	1.35e−7	5.04e−8	1.68e−7	1.58e−7
	L_{IM}	2.47e−12	6.91e−12	2.07e−11	6.76e−11
	P_m	−3.81e−5	8.54e−5	5.20e−5	5.14e−4
CSD	P_s^{CSD}	8.62	1.80e1	5.73	4.48e1
	P_j^{CSD}	1.86e−4	4.04e−4	−6.67e−4	6.14e−3
DSPF	$K_{HP} \cdot 10^9$	2.65e−1	8.76e−4	9.10e−2	3.98e−3
	$r_p \cdot 10^9$	6.05e1	2.15e8	6.06e1	3.12e8
IT	L_p	1.35e−7	5.03e−8	4.09e−9	4.09e−7
	σ	2.71e−1	3.49	4.94e4	5.07e6
	L_j	−6.65e−6	6.43e−5	−8.06e−1	8.26e1

Table C.13: Parameter values and confidence intervals for data set of Postel et al. [66] (Toluene and n-Hexane)

Model	Parameter	Tol value	Tol CI	Hex value	Hex CI
SSD	p_s^{SSD}	5.57e−7	1.33e−7	1.06e−6	1.59e−7
	p_j^{SSD}	6.37e−5	3.10e−5	4.08e−5	2.55e−5
SDIM	L_{rm}	5.57e−7	2.18e−7	1.06e−6	2.65e−7
	L_{IM}	7.57e−12	4.11e−11	1.60e−11	8.86e−11
	P_m	1.03e−5	2.92e−4	3.51e−6	2.10e−4
CSD	p_{CSD}^s	1.38e1	2.25	2.20e1	2.88
	p_{CSD}^j	4.50e−5	1.48e−5	4.12e−5	2.24e−5
DSPF	$K_{HP}\cdot10^9$	3.29e−1	1.20e−2	3.30e−1	4.36e−2
IT	$r_p\cdot10^9$	1.36	6.47e−1	9.79e−1	1.22e−1
	L_p	5.57e−7	2.17e−7	1.06e−6	2.57e−7
	σ	2.90e−1	7.06	3.46e−1	2.30
	L_j	1.09e−5	5.06e−4	7.43e−6	2.54e−4

Table C.14: Eigenvalues for all models and the data set of Postel et al. [66]

Model	EV Nr.	IPA	MeOH	Tol	Hex
SSD	1	4.73e7	2.67e7	2.39e8	5.05e8
	2	1.72e14	1.31e14	1.71e13	1.37e13
SDIM	1	4.56e6	1.24e6	7.22e6	2.06e7
	2	1.34e13	1.31e13	1.30e13	1.29e13
	3	9.40e22	1.68e22	1.61e22	3.01e21
CSD	1	3.21e−3	1.12e−2	2.10e−2	3.01e−2
	2	8.08e7	2.14e6	6.39e8	5.28e8
DSPF	1	5.03e7	3.77e6	4.51e6	1.24e6
	2	3.18e−16	3.04e−16	9.88e−2	4.97e−1
IT	1	5.06e−3	−1.53e−10	1.40e−2	1.61e−1
	2	2.51e9	2.23e−1	1.51e8	1.26e8
	3	2.44e13	3.64e16	1.30e13	1.30e13
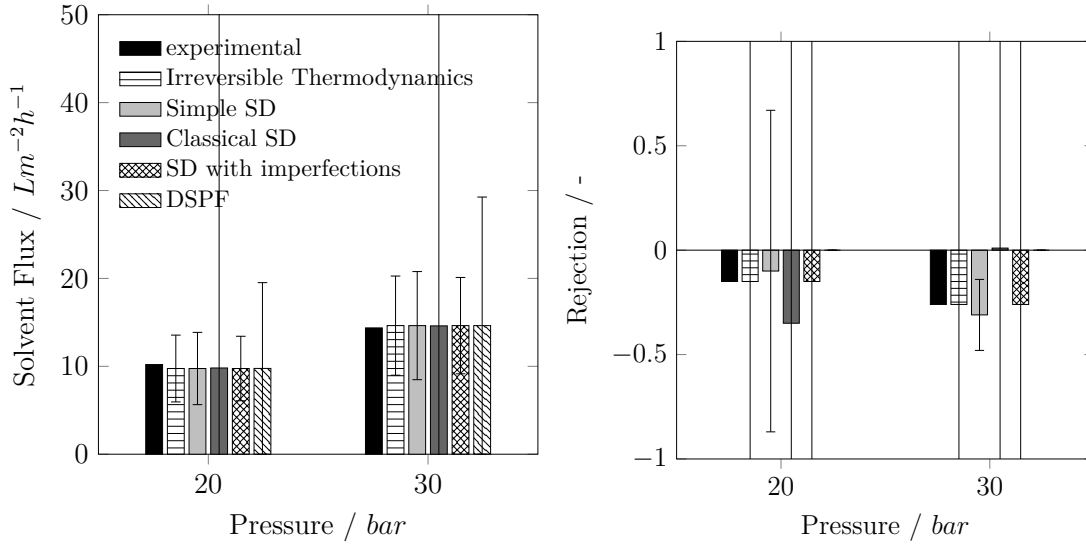
Figure C.9: Experimental data (provided by Postel et al. [66]) and calculated values for flux and rejection for the mixture of IPA and n-tetradecane of the data set of Postel et al. [66], error bars represent the uncertainty regions determined by Monte-Carlo simulation
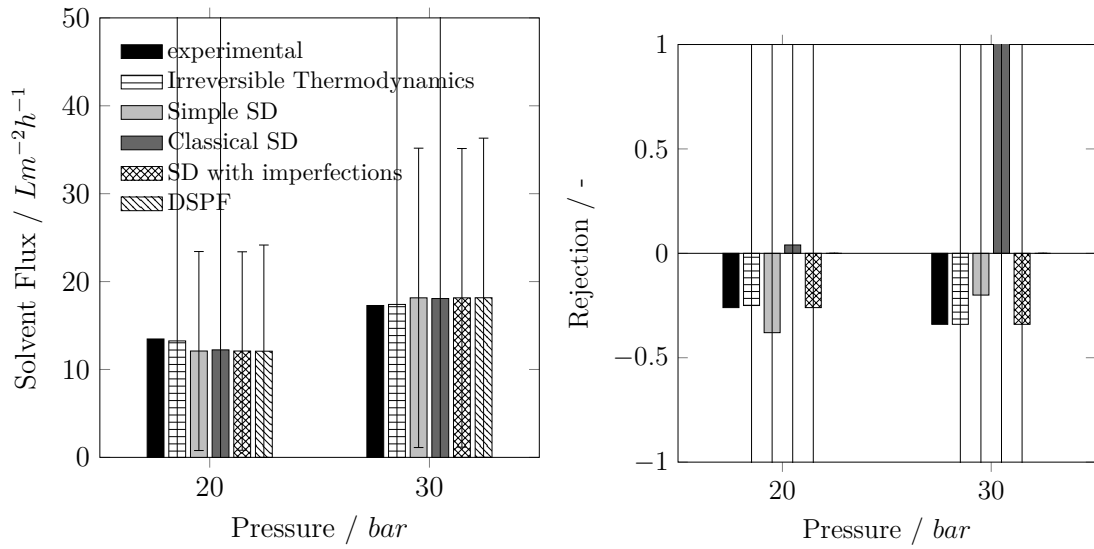
Figure C.10: Experimental data (provided by Postel et al. [66]) and calculated values for flux and rejection for the mixture of Methanol and n-tetradecane of the data set of Postel et al. [66], error bars represent the uncertainty regions determined by Monte-Carlo simulation
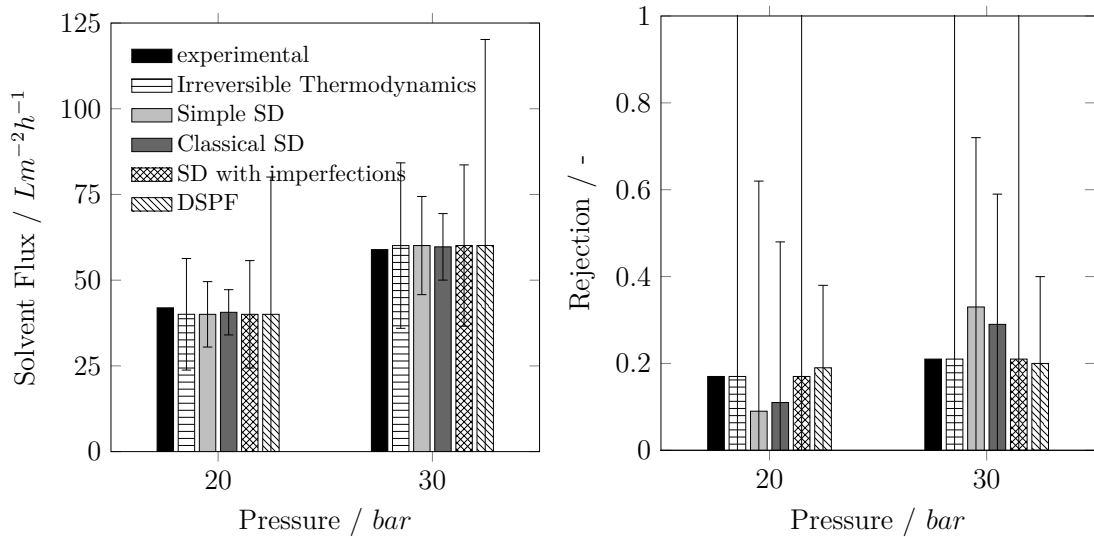
Figure C.11: Experimental data (provided by Postel et al. [66]) and calculated values for flux and rejection for the mixture of Toluene and n-tetradecane of the data set of Postel et al. [66], error bars represent the uncertainty regions determined by Monte-Carlo simulation
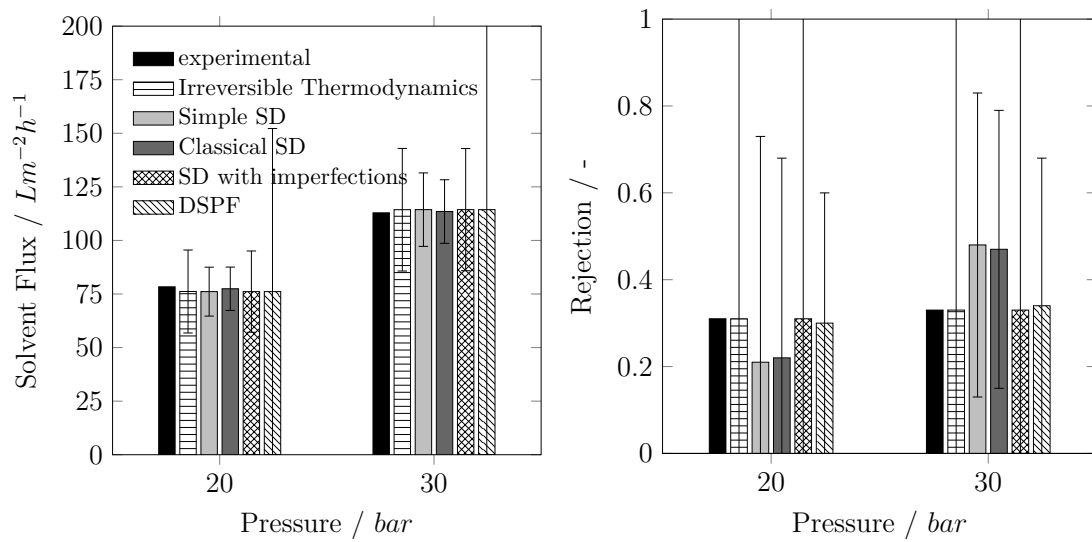
Figure C.12: Experimental data (provided by Postel et al. [66]) and calculated values for flux and rejection for the mixture of n-Hexane and n-tetradecane of the data set of Postel et al. [66], error bars represent the uncertainty regions determined by Monte-Carlo simulation

D

Additional information regarding model development for solvent flux

D.1 Data-driven methods

The following subsections provide a detailed discussion of the different data-driven methods. Moreover, successful applications of the methods to membrane processes are included.

D.1.1 Artificial Neural Networks (ANN)

Artificial neural networks (ANN) process inputs based on a specific network architecture, inspired by the biological neural networks. An ANN consists of different neurons arranged in one or multiple layers. Neurons are used for data processing and information is handed over to the next layer by a specific weighted connection, whereas there is no connection between the neurons of one layer. The network is further trained to map the inputs to outputs based on a training data set [143]. Nowadays, especially deep neural networks with a multitude of layers are experiencing an increasing popularity in current machine learning applications but require extensive data for an appropriate fitting [100].

ANNs have previously been applied in a similar context as covered by the current article. Yangali-Quintanilla et al. [105] did apply a three-step approach for the derivation of predictive models for the rejection of organic compounds in water by polyamide NF and RO membranes. The first determined important descriptors through the derivation of a QSAR model, based on PCA and MLR, before using the identified descriptors in order to determine a suitable ANN structure, evaluating different transfer functions and training methods. They concluded that the application of ANNs is highly promising for the prediction of the rejection of different compounds. Similarly, Ammi et al. [106] used ANNs to predict the rejection of organic compounds solved in water. However, other than Yangali-Quintanilla et al. [105], they did not apply PCA and MLR for a first determination of important descriptors and rather used 11 descriptors as input for a feed-forward ANN. By comparing the ANN with the previously proposed QSAR models developed by Yangali-Quintanilla et al. [97] and Yangali-Quintanilla et al. [105] the increased accuracy of the ANN models was illustrated. Agatonovic-Kustrin et al. [144] further used ANNs to estimate the permeability of drug components through polydimethylsiloxane (PDMS) membranes based on molecular descriptors. Other than the previous approaches they applied a genetic algorithm in order to determine the relevant set of descriptors, for which a feed-forward ANN with backpropagation rule and multilayer perception was trained and evaluated with respect to experimental data. Using this hybrid approach a 12-descriptor nonlinear ANN was developed, based on a data set of 254 compounds and 42 structural descriptors, considering 200 compounds for the ANN training.

Despite the success of using ANNs in related applications and the reported increased accuracy with respect to the PCA and MLE derived QSAR models it has to be considered that the ANN models are much larger in size and number of parameters and that despite relative importance evaluations of single descriptors [106] the relationships between input and output of the ANN is rather non-transparent.

D.1.2 Automated learning of algebraic models for optimization (ALAMO)

The ALAMO approach presented by Cozad et al. [101] was primarily developed to derive simple and accurate surrogate models with a functional form tailored for an optimization framework, considering the automated learning based on data derived

from available but rather complex simulation models. Yet, the approach can also be used to derive rather simple explicit models from experimental data, as opposed to ANN. The developed surrogate models represent a linear combination of nonlinear basis functions, which is optimized using a corrected Akaike information criterion [145] as objective function. The best solution is further determined by application of a global deterministic optimization approach using BARON [108]. The additional adaptive sampling methodology is able to make efficient use of small amounts of data and outperform space-filling models with larger training set [146], considering that the data is not available beforehand. While the approach is generally suited for a well-structured and efficient identification of high-quality correlations it does not allow for the identification of models that include a nonlinear combination of basic functions, such as those represented by the phenomenological models presented in Section 2.2.4. Yet, it would be of interest to include these models in the search space for the development of improved model candidates.

D.1.3 Genetic Programming (GP)

GP is an evolutionary computation technique, which evolves a population of computer programs, which are expressed as syntax trees rather than as lines of code [96]. The computer programs can represent the desired correlations that accordingly are stored in a syntax tree (cf. Figure D.1), which is modified in the different generations through the application of evolutionary operators. The tree-like representation made GP particularly suitable for symbolic regression problems starting from the earliest applications [147]. Compared to ANNs, models identified with GP are often more compact, use fewer parameters and are considered more useful for predictions [148]. As required for any metaheuristic, the application of GP requires the specification of various control parameters, including the maximum number of generations, population size, tree depth, probabilities for crossover and mutation and generation gap. Additionally, the terminal set containing the independent descriptors for the model, and the set of operators have to be defined initially [148]. However, unlike ANN and the ALAMO approach GP provides the capability to perform a direct symbolic regression.

GP was already successfully applied in the context of membrane processes. Shokrkar et al. [149] used GP for identifying a model for the prediction of flux decline in microfiltration (MF) process, identifying a correlation between the flux and different operating parameters, which was highly accurate (error of less than $5\,\%$).
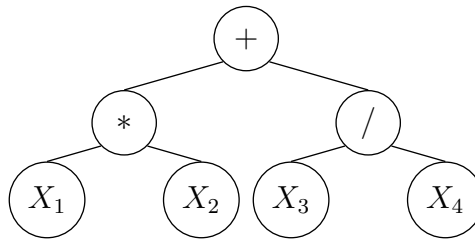
Figure D.1: Tree with a tree depth of 3

Fouladitajar et al. [103] identified a correlation between the operating conditions of a MF process and water flux and oil rejection, which outperformed other models in terms of accuracy. Hwang et al. [104] applied GP to model and predict membrane fouling in a MF-based drinking water production system, providing good agreement with experimental data, based on a correlation of the operating conditions with fouling. Park et al. [150] further demonstrated the ability of GP to identify a correlation of operating parameters and membrane permeability for a desalination plant, showing that the GP-based model was more accurate in representing the experimental data as the simple semi-empirical model, including two empirically determined parameters. Finally, Okhovat and Mousavi [151] investigated the application of GP for the identification of a model for ion rejection in aqueous nanofiltration, correlating feed concentration and transmembrane pressure with the rejection.

D.2 Translation of syntax trees

The following figures illustrate the translation of syntax trees including the functions $\exp, \sqrt{}$ and ln.
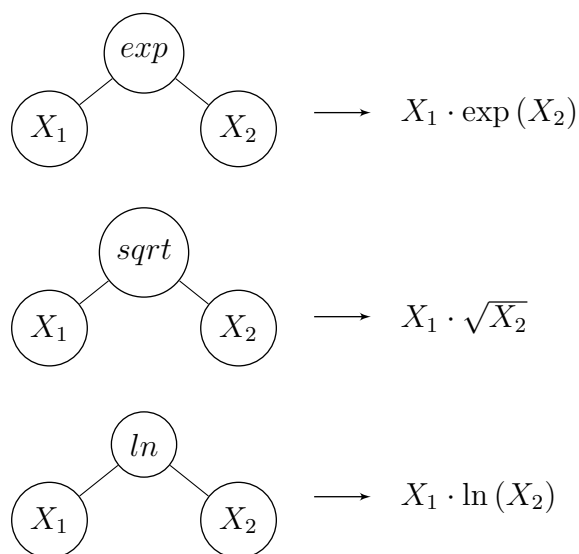
D.3 Calculation of possible combinations for full enumeration

For the determination of the number of possible combinations of 16 descriptors and 7 different functions in a syntax tree of the maximum tree depth of 3, the calculations are performed subsequently for the different options in the following. First, all complete filled trees with a tree depth of 2 and 3 are considered. Subsequently, trees with a tree depth of 3, which are not completely filled, are considered. The root of all trees can only be filled with one out of the seven functions and for the lowest nodes of the trees only descriptors are allowed. Moreover, for complete filled trees with a tree depth of 3 only functions are allowed in level 2, which is also valid for one out of two positions in level 2 in not completely filled trees with a tree depth of 3. For clarification, the nodes of the trees are marked with 'f' (function) or 'd' (descriptor), indicating the allowed type of the node.

D.4 Additional models for pure solvent permeance

For each data set different models were identified as the best models in each of the ten runs performed for the data sets. The third and fourth best model found for the

$$\longrightarrow \quad N_{comb,2,full} = f^1 \cdot d^2$$

$$= 7^1 \cdot 16^2 = 1.792$$

$$\longrightarrow \quad N_{comb,3,full} = f^1 \cdot f^2 \cdot d^4$$

$$= 7^1 \cdot 7^2 \cdot 16^4 = 22.478.848$$

$$\longrightarrow \quad N_{comb,3,r} = f^1 \cdot (f^1 \cdot d^1) \cdot d^2$$

$$= 7^1 \cdot (7^1 \cdot 16^1) \cdot 16^2 = 200.704$$

$$\longrightarrow \quad N_{comb,3,l} = f^1 \cdot (d^1 \cdot f^1) \cdot d^2$$

$$= 7^1 \cdot (16^1 \cdot 7^1) \cdot 16^2 = 200.704$$
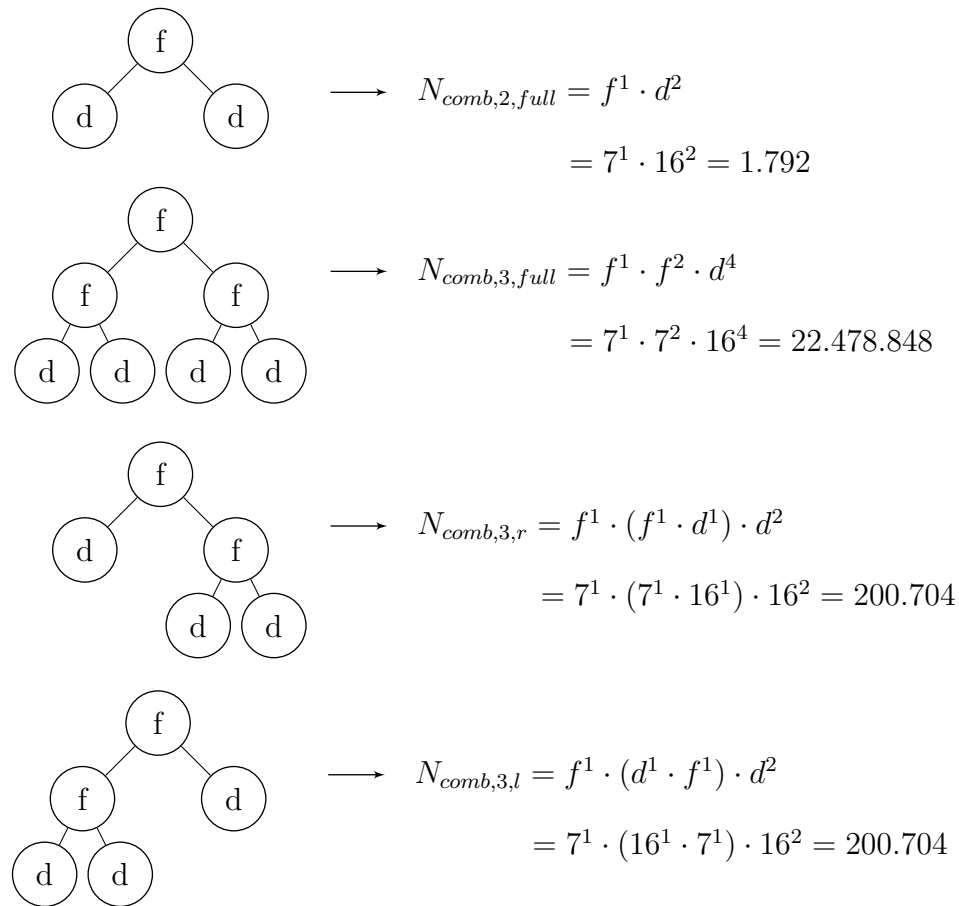
data set of Dijkstra et al. [51] are shown in Equation D.1 and D.2, respectively.

$$P_D^3 = -7.88 \cdot \frac{\Delta \delta_M - \delta_d}{\eta \sqrt{(\rho)}} + 0.60 \tag{D.1}$$

$$P_D^4 = 0.70 \cdot \frac{\eta}{d_{eq}} + 7.92 \cdot \frac{d_{eq}}{\eta} - 3.53 \tag{D.2}$$

Moreover, the method provides a large variety of different model candidates in each run, since each model candidate is evaluated and stored afterward. Table D.1 provides a list of the five best models found in run 1 of the method for the data set of Dijkstra et al. [51]. Table D.2 lists the results for run 2.

All model candidates listed include the viscosity of the solvent. Moreover, the main part includes the fraction of the molecular diameter d_{eq} and the viscosity η. The remaining model parts differ in both, structure and incorporated parameters. The top five models evaluated in run 1 include either the molecular weight M, the density

Table D.1: Five best models evaluated in run 1 for the data set of [51].

Fitness	Model equation
-47.71	$P_D = 9.65 \cdot \frac{d_{eq}}{\eta} + 4.42 \cdot d_{eq} \cdot \ln(\eta) - 3.76$
-46.03	$P_D = 8.89 \cdot \frac{d_{eq}}{\eta} + 0.03 \cdot M \cdot \ln(\eta) - 3.09$
-45.68	$P_D = 7.52 \cdot \frac{d_{eq}}{\eta} - 0.008 \cdot \rho + 1.09 \cdot \eta + 3.15$
-44.39	$P_D = 8.97 \cdot \frac{d_{eq}}{\eta} + 0.003 \cdot \rho \cdot \ln(\eta) - 3.17$
-44.28	$P_D = 8.03 \cdot \frac{d_{eq}}{\eta} + 0.42 \cdot \eta \cdot \ln(\delta_d) - 3.85$

Table D.2: Five best models evaluated in run 2 for the data set of [51].

Fitness	Model equation
-51.33	$P_D = 0.33 \cdot \frac{\delta_d}{\eta} - 0.34 \cdot \delta_p + 0.87 \cdot \eta - 1.04$
-43.89	$P_D = 8.11 \cdot \frac{d_{eq}}{\eta} + 1.70 \cdot \delta_d + 1.04 \cdot \eta - 30.42$
-41.34	$P_D = 7.85 \cdot \frac{d_{eq}}{\eta} - 0.08 \cdot \delta_p + 1.09 \cdot \eta - 3.16$
-37.64	$P_D = 7.85 \cdot \frac{d_{eq}}{\eta} - 0.04 \cdot \epsilon + 1.13 \cdot \eta - 3.14$
-34.15	$P_D = 0.32 \cdot \frac{\delta_{HBP}}{\eta} - 0.71 \cdot \delta_p + 1.15 \cdot \eta - 0.68$

ρ or the Hansen solubility parameter for dispersive interactions δ_d. In contrast to this, each of the top five models evaluated in run 2 include a parameter accounting for the polarity of the solvent, which is either the Hansen solubility parameter for dispersive interactions δ_d or the dielectric constant ϵ.

The third and fourth best model found for the data set of Machado et al. [62] are shown in Equation D.3 and D.4, respectively.

$$P_M^3 = 9.01 \cdot \frac{d_{eq}}{\eta} - 0.36 \cdot \gamma - 0.05 \cdot \delta_h + 9.93 \tag{D.3}$$

$$P_M^4 = 2.98 \frac{\delta_d \cdot d_{eq}}{\eta \cdot \ln(\rho)} - 0.16 \tag{D.4}$$

Figure D.2 illustrate the relationship between the reciprocal value of viscosity and the permeance for the data sets of Dijkstra et al. [51] and Machado et al. [62]. In addition, a trend line is added to both diagrams which correlates the reciprocal value

of viscosity directly with the permeance. The scattering for the data set of Machado et al. [62] is significantly greater, suggesting that the correlation between permeance and reciprocal of viscosity is not as clear as for the data set of Dijkstra et al. [51].
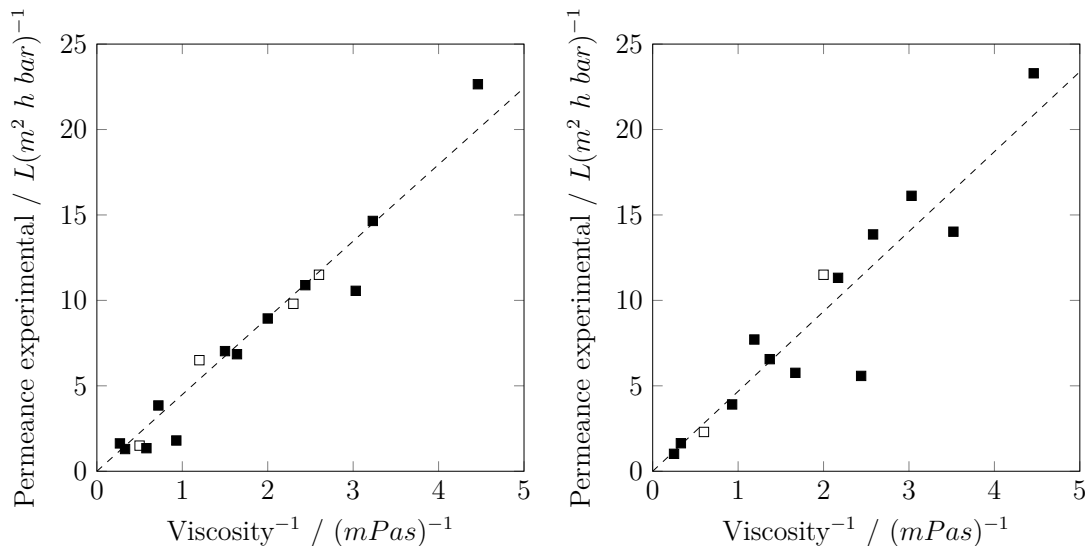
Figure D.2: Experimental permeance values vs. reciprocal of the viscosity of the solvent for training and test (filled) as well as validation (unfilled) data for the data set of Dijkstra et al. [51] (left) and Machado et al. [62] (right). The dashed lines represent trendlines.

D.5 Additional models for permeance of solvent mixtures

The fitness and the model equation of the best models found for the data sets for the membranes Inopor Nano 450 Da, Inopor Nano 750 Da and Inopor Ultra 2000 Da published by Marchetti et al. [2] are listed in Table D.3. The fitness of the models depends on the number of data points used for the application of the method, which differs for the different membranes. Hence, the fitness of the models is not comparable to each other.

Table D.3: Best models found in ten runs each for the data sets of [2].

Membrane	Fitness	Model equation
Inopor Nano 450 Da	-50.45	$P_{Mix,450} = 40.47 \cdot \frac{\delta_{HBP}}{V_m \cdot \eta \cdot \sqrt{(\delta_h)}} - 0.37$
Inopor Nano 750 Da	-87.04	$P_{Mix,750} = 501.29 \cdot \frac{\delta_d}{V_m \cdot \eta \cdot \sqrt{\gamma}} - 0.48$
Inopor Ultra 2000 Da	-58.95	$P_{Mix,Ultra} = -44.73 \cdot \frac{Pol \cdot \ln(\eta)}{\delta_h \sqrt{\gamma}} + 31.56$

D.6 Properties used for model development

The pure solvent properties used for model development are listed in Table D.4. Additionally, the molar volume V_m, the equivalent molecular diameter d_{eq}, the total Hildebrand solubility parameter δ_M, the difference of the solubility parameters of solvent and membrane $\Delta\delta_M$ and the interaction radius Ra are calculated using Equations D.5-D.9. For density, viscosity and surface tension of mixed solvents, the same literature data [126, 127, 128, 129, 130, 131] as considered by Marchetti et al. [2] was used. For different compositions as provided in the literature data, linear interpolation was applied. The remaining mixed solvent properties were determined based on mixing rules. The solubility parameters of PDMS were estimated using the group contribution method of van Krevelen and Nijenhuis [119] resulting in values of $\delta_{d,PDMS} = 13.9$ (MPa)$^{0.5}$, $\delta_{p,PDMS} = 0.3$ (MPa)$^{0.5}$, $\delta_{h,PDMS} = 6.9$ (MPa)$^{0.5}$ and $\delta_{HBP,PDMS} = 15.6$ (MPa)$^{0.5}$. Since no solubility parameters for ceramic membranes are available, the descriptors $\Delta\delta_M$ and Ra were not used for ceramic membranes.

$$V_m = \frac{M}{\rho} \tag{D.5}$$

$$d_{eq} = 2 \cdot \left(\frac{3 \cdot V_m}{4 \cdot \pi \cdot N_A} \right)^{\frac{1}{3}} \tag{D.6}$$

$$\delta_{HBP} = \sqrt{\delta_d^2 + \delta_p^2 + \delta_h^2} \tag{D.7}$$

$$\Delta\delta_M = |\delta_{HBP} - \delta_{HBP,PDM}| \tag{D.8}$$

$$Ra = \sqrt{4 \cdot (\delta_d - \delta_{d,PDMS})^2 + (\delta_p - \delta_{p,PDMS})^2 + (\delta_h - \delta_{h,PDMS})^2} \tag{D.9}$$

Table D.4: Properties of solvents used in model development

Solvent	M g mol⁻¹	ρ [152] kg m⁻³	ϵ [153] —	η [153] mPa s	γ [153] mN m⁻¹	δ_d [80] (MPa)⁰·⁵	δ_p [80] (MPa)⁰·⁵	δ_h [80] (MPa)⁰·⁵	$\vec{d}$ [153] debye	Pol [153] —
1-Butanol	74.1	809.5	18.2	3.0	24.6	16.0	5.7	15.8	1.7	60.2
1-Propanol	60.1	799.7	20.1	1.7	23.7	16.0	6.8	17.4	1.7	61.7
2-Butanol	74.1	806.3	16.6	3.7	23.0	15.8	5.7	14.5	1.7	50.6
Acetone	58.1	784.5	20.6	0.3	23.3	15.5	10.4	7.0	2.9	35.5
ACN	41.1	780.0 [2]	35.7 [2]	0.4 [2]	28.7 [2]	15.3	18.0	6.1	3.2	46.0
Butylacetate	116.2	882.5	5.0	0.7	25.1	15.8	3.7	6.3	1.8	24.1
DEK	86.1	809.8	17.0 [2]	0.4 [152]	24.7 [152]	15.8	7.6	4.7	1.7 [154]	26.5 [155]
DMF	73.1	944.5	37.0 [2]	0.9 [2]	36.3 [2]	17.4	13.7	11.3	3.8	40.4
DMSO	78.1	1095.0 [2]	46.4 [2]	2.0	43.0 [2]	18.4	16.4	10.2	4.0	44.4
Ethanol	46.1	789.3	22.4	1.1	22.3	15.8	8.8	19.4	1.7	65.4
Ethyl acetate	88.1	900.3	6.0	0.5	24.0	15.8	5.3	7.2	1.7	23.0
Isopropanol	60.1	780.9	18.3	2.0	21.7	15.8	6.1	16.4	1.7	54.6
MEK	72.1	799.9	18.5	0.4	24.6	16.0	9.0	5.1	2.8	32.7
Methanol	32.0	791.4	32.6	0.6	22.6	15.1	12.3	22.3	1.7	76.2
Methyl acetate	74.1	934.2	7.1 [152]	0.3 [152]	24.7 [152]	15.5	7.2	7.6	2.8 [154]	25.3 [155]
MIBK	100.2	796.5	13.1	0.6	23.6	15.3	6.1	4.1	2.8	27.0
NMP	99.1	1027.0 [2]	32.2 [2]	1.7 [2]	40.7 [2]	18.0	12.3	7.2	4.1 [152]	36.0
n-Decane	142.3	726.6	2.0	0.8 [152]	23.4 [152]	15.7	0.0	0.0	0.1 [156]	0.9
n-Dodecane	170.3	749.5	2.0 [152]	1.4 [152]	24.9 [152]	16.0	0.0	0.0	0.0 [154]	0.8
n-Heptane	100.2	679.5	1.9	0.4 [152]	19.7 [152]	15.3	0.0	0.0	0.0	1.2
n-Hexane	86.2	660.6	1.9	0.3	18.4	14.9	0.0	0.0	0.0	0.9
n-Nonane	128.3	719.2	2.0 [152]	0.7	22.4 [152]	15.7	0.0	0.0	0.0 [154]	0.9
n-Octane	114.2	698.6	1.9 [152]	0.5	21.7	15.5	0.0	0.0	0.0 [156]	1.2
n-Pentane	72.1	626.2	1.8	0.2 [152]	16.0	15.6	0.0	0.0	0.0	0.9
Pentanol	88.1	814.4	13.9	4.0	25.6	15.9	5.9	13.9	1.7	56.8 [155]
THF	72.1	883.3	7.5 [2]	0.5 [2]	25.0 [2]	16.8	5.7	8.0	1.8	21.0
Toluol	92.1	866.8	2.4	0.6 [2]	27.9 [2]	18.0	1.4	2.0	0.4	9.9
Water	18.0	1000.0 [2]	78.2 [2]	1.0 [152]	72.8	15.5	16.0	42.3	1.9	100.0

E

Additional information regarding model development for solute rejection

E.1 Resulting statistics of the developed model candidates for the different solute rejection data sets - Pure solvents

For each of the investigated cases in the article the results of the post-processing step are listed in the following subsections.

E.1.1 Solvent-dependent models - polar Solvents

Table E.1 provides a summary of the post-processing results of the most promising model candidates for solute rejection from the automatic model development process, considering the solute rejection data of the polar solvents. For polar solvents the maximum tree size of three is sufficient. The mean absolute deviation $\hat{\varepsilon}$ between experimental and calculated rejections of the training and test data set is $3.5\,\%$,

while an absolute deviation of 9 % was only exceeded for one solute within the test data set. All three parameters of this model are identifiable.

Table E.1: Summary of fitness, tree depth (TD), mean ($\hat{\varepsilon}$) and maximum absolute deviation(ε_{max}), parameter value (θ), size of 95 % confidence interval (CI) and eigenvalues (λ) for the best model with a maximum tree depth of 3 ($R_{p,i}^3$) for the group of polar solvents

Model	TD	Fitness	$\hat{\varepsilon}$	ϵ_{max}	θ	CI	λ
$R_{p,i}^3$	3.	-53.98	3.5%	15.3%	1.07	$5.01{\times}10^{-1}$	$3.64{\times}10^{-2}$
					1.75	$2.18{\times}10^{-1}$	$2.21{\times}10^{-1}$
					$-3.76{\times}10^{-1}$	$5.06{\times}10^{-2}$	$1.53{\times}10^{1}$

Table E.2 lists the fitness, the Akaike weight W and the model equations of all models identified in ten consecutive runs.

Table E.2: Fitness, Akaike weight W and the model found in ten runs for the group of polar solvents sorted by fitness

Run	Fitness	W	Model
5,10	-53.98	38.5%	$R_i = 1.067 \cdot \frac{\delta_{h,i}}{\gamma_i} + 1.750 \cdot \frac{\Delta\delta_{M,i}}{\delta_s} - 0.376$
4, 6, 7	-53.83	35.7%	$R_i = 3.872 \cdot \frac{d_{eq,s}}{Ra_i} - 4.369\text{E-}4 \cdot M_i + 0.116 \cdot \Delta\delta_{M,i} - 0.570$
1, 8	-51.42	10.7%	$R_i = 0.007 \cdot \Delta\delta_{M,i} \cdot \delta_{d,i} - 0.032 \cdot \frac{Ra_i}{d_{eq,s}} + 0.045$
2, 9	-51.41	10.7%	$R_i = 5.913 \cdot \frac{\Delta\delta_{M,i}}{Pol_s} + 0.146 \cdot \frac{\delta_{d,i}}{Ra_i} - 0.670$
3	-49.67	4.5%	$R_i = 0.006 \cdot \delta_i \cdot \Delta\delta_{M,i} - 1.168 \cdot \vec{d}_s - 0.051 \cdot Ra_i + 1.987$

E.1.2 Solvent-dependent models - moderately polar Solvents

Table E.3 provides a summary of the post-processing results of the most promising model candidates for solute rejection from the automatic model development process, considering the solute rejection data of the moderately polar solvents. Two of the parameters of the most promising model with a tree depth of tree $R_{m,i}^3$ were not identifiable according to the confidence intervals and the mean absolute deviation exceeded the limit of 5 %, such that further models with a tree depth of four were

generated. These allowed for a significant improvement in terms of fitness, $\hat{\varepsilon}$ and ε_{max} for the training and test data set. Moreover, model $R^4_{m,i}$ shows much better parameter precision compared to model $R^3_{m,i}$ although the number of parameters is increased.

Table E.3: Summary of fitness, tree depth (TD), mean ($\hat{\varepsilon}$) and maximum absolute deviation (ε_{max}), parameter value (θ), size of 95 % confidence interval (CI) and eigenvalues (λ) for the best models with a maximum tree depth of 3 ($R^3_{m,i}$ and $R^4_{m,i}$) for the group of moderately polar solvents

Model	TD	Fitness	$\hat{\varepsilon}$	ϵ_{max}	θ	CI	λ
$R^3_{m,i}$	3.	-25.35	5.2%	14.7%	-2.96×10^{-1}	1.71×10^{-1}	9.74×10^{-1}
					2.06×10^{-2}	3.08×10^{-2}	1.29
					4.90×10^{-3}	2.44×10^{-3}	9.57×10^{1}
					-1.70×10^{-1}	1.97×10^{-1}	1.21×10^{6}
$R^4_{m,i}$	4.	-59.83	3.1%	9.6%	-1.73×10^{-2}	9.89×10^{-3}	4.18×10^{-1}
					3.40×10^{-3}	1.47×10^{-3}	3.54
					-1.81×10^{-1}	7.30×10^{-2}	3.48×10^{1}
					4.37×10^{-2}	3.23×10^{-2}	4.35×10^{2}
					4.40×10^{-3}	1.32×10^{-3}	7.42×10^{4}
					-4.34×10^{-1}	2.12×10^{-1}	1.33×10^{6}

$$R^3_{m,i} = -0.296 \cdot \frac{M_i}{V_{m,s}} + 0.021 \cdot \Delta\delta_{M,i} + 0.005 \cdot M_i - 0.170 \qquad \text{(E.1)}$$

Table E.4 lists the fitness, the Akaike weight W and the model equations of all models identified in ten consecutive runs.

Table E.4: Fitness, Akaike weight W and the model found in ten runs for the group of moderately polar solvents sorted by fitness

Run	Fitness	W	Model
8	−59.83	98.8%	$R_i = -0.0173 \cdot \delta_{h,i} \cdot \Delta\delta_{M,i} + 0.003 \cdot \Delta\delta_{M,i} \cdot Pol_s - 0.181 \cdot \frac{V_{m,i}}{M_s} + 0.044 \cdot \delta_{h,i} + 0.004 \cdot M_i - 0.434$
5	−49.08	0.5%	$R_i = -0.490 \cdot \frac{V_{m,i}}{\rho_i} \cdot \delta_s + 1.347 \cdot \frac{\Delta\delta_{M,s}}{\delta_{d,i}} + 0.012 \cdot V_{m,i} - 0.001 \cdot \rho_i + 0.566$
6	−49.05	0.4%	$R_i = -433.027 \cdot \frac{V_{m,i}}{rho_i \cdot (\gamma_i + \Delta\delta_{M,i})} + 0.040 \cdot \frac{\delta_i}{Ra_i} + 0.311 \cdot \delta_{d,s} + 0.006 \cdot V_{m,i} - 5.154$
4	−48.01	0.3%	$R_i = -577.324 \cdot \frac{\frac{rho_i \cdot \gamma_i}{V_{m,i}}}{\rho_i \cdot (\delta_{h,i} + \gamma_i)} - 0.006 \cdot \frac{M_i}{Ra_i} + 0.158 \cdot \frac{V_{m,i}}{\delta_{d,s}} - 0.158$
7	−43.85	0.0%	$R_i = 4.342\text{E-05} \cdot \gamma_i \cdot M_i + 0.041 \cdot \frac{\vec{d_i}}{\Delta\delta_{M,i}} - 0.002 \cdot V_{m,i} \cdot deq_i + 0.015 \cdot \vec{d_s} \cdot \Delta\delta_{M,i} - 0.352$
9	−41.31	0.0%	$R_i = 0.039 \cdot \frac{\Delta\delta_{M,i}}{\vec{d_i}} + 0.005 \cdot M_i + 0.002 \cdot \Delta\delta_{M,i} \cdot \vec{d_s} - 0.001 \cdot \frac{V_{m,i}}{\eta_s} - 0.160$
2	−40.01	0.0%	$R_i = -2.130\text{E-4} \cdot \frac{\delta_{h,s} \cdot \gamma_i}{\delta_{p,i}} - 3.936\text{E-4} \cdot \delta_s \cdot V_{m,i} + 0.010 \cdot M_i - 0.007 \cdot Ra_i - 0.063$
10	−39.62	0.0%	$R_i = -0.048 \cdot \delta_{d,i} \cdot \eta_s - 0.003 \cdot \frac{\delta_s}{\delta_{p,i}} - 0.008 \cdot V_{m,i} + 0.198 \cdot \frac{M_i}{\delta_s} + 0.214$
1	−37.14	0.0%	$R_i = -0.303 \cdot \frac{V_{m,i}}{\gamma_i} + 0.001 \cdot \vec{d_s} \cdot \Delta\delta_{M,i} + 0.005 \cdot M_i + 0.018 \cdot \vec{d_i^2} - 0.136$
3	−33.35	0.0%	$R_i = 0.009 \cdot \delta_{p,i} \cdot \delta_{h,i} - 1.422 \cdot \frac{V_{m,i}}{\rho_s}{deq_i} + 0.005 \cdot \frac{M_i}{deq_{,i}} + 0.027 \cdot \Delta\delta_{M,i} - 0.616$

E.1.3 Solvent-dependent models - non-polar Solvents

Table E.5 provides a summary of the post-processing results of the most promising model candidates for solute rejection from the automatic model development process, considering the solute rejection data of the non-polar solvents. For non-polar solvents the maximum tree size of three is also sufficient. The mean absolute deviation $\hat{\varepsilon}$ between experimental and calculated rejections of training and test data set is 4.8% (cf. Table E.5), while an absolute deviation of 14.2% was determined. All four parameters of this model are identifiable. Although, the limit of 5% for the mean absolute deviation was not exceeded, models with a maximum tree depth of four have been evaluated in order to determine modes with a lower maximum absolute deviation. However, the best identified model with a tree depth of four ($R^4_{np,i}$) did show significant lower mean and maximum absolute deviations.

Table E.5: Summary of fitness, tree depth (TD), mean ($\hat{\varepsilon}$) and maximum absolute deviation (ε_{max}), parameter value (θ), size of 95% confidence interval (CI) and eigenvalues (λ) for the best model with a maximum tree depth of 3 ($R^3_{np,i}$) and 4 ($R^4_{np,i}$) for the group of non-polar solvents

Model	TD	Fitness	$\hat{\varepsilon}$	ϵ_{max}	θ	CI	λ
$R^3_{np,i}$	3.	-73.96	4.8%	14.2%	$2.12{\times}10^{-3}$	$3.25{\times}10^{-4}$	$3.41{\times}10^{-1}$
					$-3.58{\times}10^{-4}$	$1.35{\times}10^{-4}$	$2.47{\times}10^{5}$
					$-7.44{\times}10^{-1}$	$2.43{\times}10^{-1}$	$5.73{\times}10^{7}$
$R^4_{np,i}$	4.	-80.23	4.6%	14.3%	$2.26{\times}10^{-3}$	$3.96{\times}10^{-4}$	$3.63{\times}10^{-3}$
					$8.84{\times}10^{-1}$	2.55	$2.57{\times}10^{-1}$
					$-3.08{\times}10^{-4}$	$1.29{\times}10^{-4}$	$1.84{\times}10^{6}$
					-1.39	$3.35{\times}10^{-1}$	$1.34{\times}10^{6}$

$$R^4_{np,i} = 0.002 \cdot d_{eq,i} \cdot \rho_i + 0.884 \cdot \frac{Pol_s}{V_{m,i}} - 3.08\text{E-}4 \cdot \Delta\delta_{M,i} \cdot V_{m,s} - 1.392 \qquad (\text{E.2})$$

Table E.6 lists the fitness, the Akaike weight W and the model equations of all models identified in ten consecutive runs.

Table E.6: Fitness, Akaike weight W and the model found in ten runs for the group of non-polar solvents sorted by fitness

Run	Fitness	W	Model
1, 3, 8	-73.96	63.8%	$R_i = 0.00 \cdot \rho_i \cdot d_{eq,i} - 0.0004 \cdot V_{m,s} \cdot \delta_i - 0.744$
2, 4-7, 9, 10	-72.82	36.2%	$R_i = -264.648 \cdot \frac{d_{eq,i}}{M_i} + 0.009 \cdot \frac{\delta_{d,s}}{\eta_i} + 1.436$

E.1.4 Solvent-dependent models - including non-linear parameters

The following table provides a summary of the results of runs including nonlinear parameter relationships for the different data sets of solute rejections in pure solvents used for model development.

Table E.7: Summary of fitness, tree depth (TD) and Runtime for the runs including nonlinear parameter relationships

Data set	TD	Fitness	Runtime
polar solvents	3	-53.98	2 h
moderate polar solvents	4	-44.79	14 h
non-polar solvents	3	-73.96	3.75 h

Table E.8: Summary of fitness, tree depth (TD), mean ($\hat{\varepsilon}$) and maximum absolute deviation (ε_{max}), parameter value (θ), size of 95 % confidence interval (CI) and eigenvalues (λ) for the best model with a maximum tree depth of 4 ($R_{np,i}^4$) for the group of moderately polar solvents including non-linear parameter relationships

Model	TD	Fitness	$\hat{\varepsilon}$	ϵ_{max}	θ	CI	λ
$R_{m,i}^{NL,4}$	4.	-44.79	3.8%	10.5%	$1.06{\times}10^{-2}$	$3.39{\times}10^{-3}$	$7.52{\times}10^{-1}$
					$-1.41{\times}10^{-3}$	$8.73{\times}10^{-4}$	$2.47{\times}10^{3}$
					$-4.15{\times}10^{-4}$	$1.47{\times}10^{-4}$	$1.35{\times}10^{5}$
					$-1.44{\times}10^{-4}$	$2.02{\times}10^{-4}$	$1.84{\times}10^{6}$
					$-5.20{\times}10^{-2}$	$1.66{\times}10^{-1}$	$5.59{\times}10^{8}$

E.1.5 Solvent-independent models

Table E.9 provides a summary of the post-processing results of the most promising model candidates for solute rejection from the automatic model development process, considering the solute rejection data of all six solvents simultaneously.

Table E.9: Summary of fitness, mean ($\hat{\varepsilon}$) and maximum absolute deviation (ε_{max}), parameter value (θ), size of 95 % confidence interval (CI) and eigenvalues (λ) for the best models with a maximum tree depth of 4 ($R^4_{all,i}$) and 5 ($R^{5.1}_{all,i}$ and $R^{5.2}_{all,i}$)

Model	TD	Fitness	$\hat{\varepsilon}$	ϵ_{max}	θ	CI	λ
$R^4_{all,i}$	4.	-96.30	8.2%	35.1%	$-2.46{\times}10^{-7}$	$3.52{\times}10^{-8}$	1.74
					$1.78{\times}10^{-3}$	$4.83{\times}10^{-4}$	$1.54{\times}10^{3}$
					$1.06{\times}10^{-2}$	$6.65{\times}10^{-3}$	$3.36{\times}10^{6}$
					$-9.12{\times}10^{-1}$	$1.94{\times}10^{-1}$	$1.48{\times}10^{14}$
$R^{5.1}_{all,i}$	5.	-116.68	8.4%	30.0%	$6.04{\times}10^{-6}$	$1.06{\times}10^{-6}$	2.97
					$5.00{\times}10^{-3}$	$5.03{\times}10^{-3}$	$7.85{\times}10^{1}$
					$1.05{\times}10^{-1}$	$6.72{\times}10^{-2}$	$5.61{\times}10^{2}$
					$5.68{\times}10^{-2}$	$1.30{\times}10^{-2}$	$3.70{\times}10^{3}$
					$-4.16{\times}10^{-1}$	$1.41{\times}10^{-1}$	$2.79{\times}10^{11}$
$R^{5.2}_{all,i}$	5.	-114.80	8.7%	27.9%	$8.47{\times}10^{-2}$	$6.08{\times}10^{-2}$	4.63
					$2.09{\times}10^{-4}$	$1.09{\times}10^{-4}$	$2.26{\times}10^{1}$
					$6.06{\times}10^{-1}$	$9.03{\times}10^{-2}$	$5.04{\times}10^{1}$
					$2.37{\times}10^{-2}$	$1.48{\times}10^{-2}$	$4.40{\times}10^{2}$
					$-2.95{\times}10^{-1}$	$8.44{\times}10^{-2}$	$1.25{\times}10^{7}$

$$R^{5.1}_{all,i} = 6.037\text{E-}6{\cdot}V_{m,i}{\cdot}\frac{\rho_s - \eta_i}{\epsilon_s} + 0.005{\cdot}\delta_{p,i}{\cdot}\vec{d_i}{\cdot}\vec{d_s^2} + 0.105{\cdot}d_{eq,i}{\cdot}\vec{d_s} + 0.057{\cdot}\Delta\delta_{M,i} - 0.416$$

$$\text{(E.3)}$$

Table E.10 and E.11 lists the fitness, the Akaike weight W and the model equations of all models identified in ten consecutive runs for both tree sizes.

Table E.10: Fitness, Akaike weight W and the model found in ten runs for all solvents and a tree depth of four sorted by fitness

Run	Fitness	W	Model
2	−96.30	100.0%	$R_i = -2.457\text{E-}7 \cdot (\gamma_i - M_s) \cdot V_{m,s} \cdot M_i + 0.002 \cdot \delta_{d,i} \cdot \gamma_s + 0.011 \cdot \vec{d_i} \cdot \delta_{p,s} - 0.912$
4, 8	−79.49	0.0%	$R_i = -6.130\text{E-}6 \cdot \frac{V_{m,i}}{\varepsilon_s} \cdot (\eta_i - \rho_s) + 0.057 \cdot \Delta\delta_{M,i} + 0.126 \cdot d_{eq,i} \cdot \vec{d_s} - 0.433$
10	−71.75	0.0%	$R_i = 0.004 \cdot \frac{V_{m,i}}{\varepsilon_s} + 1.826\text{E-}4 \cdot \rho_s + 0.013 \cdot d_{eq,i} \cdot \gamma_s + 0.071 \cdot \vec{d_s} - 0.102 \cdot \Delta\delta_{M,i} - 0.559$
9	−63.04	0.0%	$R_i = -2.009\text{E-}7 \cdot M_s \cdot M_i \cdot (\gamma_i - V_{m,s}) + 0.002 \cdot \gamma_s \cdot \delta_i + 1.319\text{E-}4 \cdot M_i - 0.968$
1	−61.96	0.0%	$R_i = 0.006 \cdot \delta_{p,s} \cdot \vec{d_s} + 0.002 \cdot \delta_i \cdot \gamma_s - 9.517\text{E-}5 \cdot M_i \cdot \vec{d_s} - 0.001 \cdot \rho_s + 0.002 \cdot M_i - 0.232$
7	−50.31	0.0%	$R_i = -0.090 \cdot \frac{M_i}{\delta_{d,s}} + 0.001 \cdot \delta_i \cdot \delta_s + 0.005 \cdot \delta_{p,s} \cdot \delta_{h,i} + 0.130 \cdot \frac{M_i}{\delta_s} - 0.685$
5	−28.68	0.0%	$R_i = 6.380\text{E-}6 \cdot (M_i \cdot \delta_{d,s} - (V_{m,s} - \rho_i)) \cdot (\delta_i - \delta_i) - 0.257$
6	−26.15	0.0%	$R_i = -3.731\text{E-}8 \cdot M_i \cdot (M_s + Ra_s) \cdot (\Delta\delta_{M,s} - \delta_i) \cdot (Ra_s + \gamma_s) - 0.223$
3	−17.42	0.0%	$R_i = -6.714\text{E-}5 \cdot \frac{\delta_i + \Delta\delta_{M,i}}{\delta_s} \cdot \frac{V_{m,i}}{d_{eq,i}} \cdot (\vec{d_s} - \gamma_s) - 0.298$

Table E.11: Fitness, Akaike weight W and the model found in ten runs for all solvents and a tree depth of five sorted by fitness

Run	Fitness	W	Model
2, 3	−116.68	71.9%	$R_i = 6.037\text{E-}6 \cdot V_{m,i} \cdot \frac{\rho_s - \eta_i}{\epsilon_s} + 0.005 \cdot \delta_{p,i}\vec{d_i}\vec{d_s}^2 + 0.105 \cdot d_{eq,i}\vec{d_s} + 0.057 \cdot \Delta\delta_{M,i} - 0.416$
4	−114.80	28.1%	$R_i = \frac{\vec{d_s}}{\Delta\delta_{M,s}} \cdot (0.085 \cdot \epsilon_s + 2.085\text{E-}4 \cdot \rho_i\vec{d_i}) + 0.606 \cdot \frac{M_i}{V_{m,s}\epsilon_s} + 0.024 \cdot \Delta\delta_{M,i} - 0.295$
5	−84.62	0.0%	$R_i = -3.459\text{E-}5 \cdot \frac{(\eta_i - M_i + Pol_s + \gamma_i)(M_i + V_{m,s})}{\epsilon_s + Ra_i} - 4.1069\text{E-}5 \cdot (\delta_i d_{eq,i} - \Delta\delta_{M,s} + \delta_i)(2\gamma_i - \delta_{d,i}\gamma_s) - 0.443$
1	−80.04	0.0%	$R_i = 3.442\text{E-}4 \cdot \frac{(M_i - \gamma_s)(M_s - \delta_{p,s})}{\delta_i} - 0.008 \cdot \frac{(\eta_s - d_{eq,i})\delta_{d,s}}{\Delta\delta_{M,s}} + 0.585 \cdot \frac{\delta_{h,i}}{\gamma_i} + 0.002 \cdot \gamma_s\delta_{d,i} - 1.196$
6	−77.45	0.0%	$R_i = -4.192\text{E-}5 \cdot \delta_{d,s}(Ra_s + \rho_s + \eta_i + Ra_i) + 0.002 \cdot \gamma_s\delta_{d,i} - 3.223 \cdot \frac{\gamma_i}{\rho_s} + 0.001 \cdot \frac{V_{m,i}}{\epsilon_s}$ $+ 1.578\text{E-}5 \cdot M_s M_i - 0.524$
8	−66.70	0.0%	$R_i = -3.452\text{E-}4 \cdot \frac{V_{m,i}}{d_{eq,i} + \Delta\delta_{M,i}} + 0.005 \cdot \vec{d_i}\vec{d_s}\delta_{p,s} + 0.002 \cdot M_i + 7.606\text{E-}4 \cdot \Delta\delta_{M,i}\gamma_s$ $- 3.523\text{E-}5 \cdot Pol_s M_i - 0.483$
7	−65.65	0.0%	$R_i = -4.018\text{E-}5 \cdot (\gamma_s + Pol_s - \vec{d_i})(M_i - \Delta\delta_{M,i} + \gamma_s) + 0.003 \cdot M_i + 9.300\text{E-}4 \cdot \delta_i(Ra_s + \gamma_s) - 0.762$
10	−62.63	0.0%	$R_i = -15593.369 \cdot \frac{Ra_i - \Delta\delta_{M,s}}{\rho_i\rho_s} - 0.008 \cdot \frac{\gamma_s}{\eta_i} + 0.071 \cdot \delta_i + 0.067\delta_{d,s} + 0.003 \cdot \frac{V_{m,i}}{\epsilon_s} + 0.029 \cdot \delta_{p,s}$ $- 0.064 \cdot \Delta\delta_{M,s} - 2.048$
9	−58.82	0.0%	$R_i = -1.568\text{E-}6 \cdot (\delta_{d,i}\gamma_i - \gamma_s V_{m,s})M_i + 0.001 \cdot \Delta\delta_{M,i}(\delta_i + \delta_s) - 0.002 \cdot M_i - 0.868$

E.2 Resulting statistics of the developed model candidates for the different solute rejection data sets - Mixed solvents

E.2.1 Solute-independent models

The models found for differing maximum tree depth vary in their complexity, the accuracy and the number of necessary parameters as listed in Table E.12. Using a maximum tree depth of three, the best model found describes the rejection for training and test data set with a mean absolute deviation of $\pm 12\,\%$. The mean absolute deviation decreases and the fitness and number of parameters increase with increasing maximum three depth. The most promising model candidate ($R_{t,i}^4$) describes the rejection for training and test data accurate, while providing a comparable fitness and low mean and maximum absolute deviation, while showing a much lower maximum absolute deviation between experimental and calculated rejection of $\pm 17.4\,\%$. Furthermore, the confidence intervals of the parameters are narrow and the parameters are well identifiable.

Table E.12: Summary of fitness, tree depth (TD), mean ($\hat{\varepsilon}$) and maximum absolute deviation (ε_{max}), parameter value (θ), size of $95\,\%$ confidence interval (CI) and eigenvalues (λ) for the best models with a maximum tree depth of 3 ($R_{t,i}^3$) and 4 ($R_{t,i}^4$) for ternary solvent mixtures

Model	TD	Fitness	$\hat{\varepsilon}$	ϵ_{max}	θ	CI	λ
$R_{t,i}^3$	3.	-153.88	10.1%	36.9%	-1.66×10^2	3.92×10^1	3.61×10^{-1}
					7.62×10^1	5.17	3.02×10^1
$R_{t.i}^4$	4.	-233.22	4.7%	17.4%	-2.58	4.60×10^{-1}	2.95×10^{-3}
					3.72	1.31	1.21×10^{-1}
					1.78×10^2	3.61×10^1	1.70×10^1
					6.86×10^2	1.32×10^2	7.02×10^2
					-8.32×10^{-2}	4.73×10^{-2}	1.60×10^3
					-5.11×10^2	1.11×10^2	1.73×10^5

$$R_{t,i}^3 = 1.067 \cdot \frac{\delta_{h,i}}{\gamma_i} + 1.750 \cdot \frac{\Delta\delta_{M,i}}{\delta_s} - 0.376 \qquad \text{(E.4)}$$

Table E.13 and E.14 lists the fitness, the Akaike weight W and the model equations of all models identified in ten consecutive runs for both tree sizes.

Table E.13: Fitness, Akaike weight W and the model found in ten runs for the mixed solvents and a tree depth of three sorted by fitness

Run	Fitness	W	Model
2, 10	-153.88	47.1%	$R_i = 1.067 \cdot \frac{\delta_{h,i}}{\gamma_i} + 1.750 \cdot \frac{\Delta\delta_{M,i}}{\delta_s} - 0.376$
1, 3-7, 9	-153.10	31.9%	$R_i = -1.055 \cdot \frac{\epsilon_s \cdot \Delta\delta_{M,i}}{\Delta\delta_{M,s} \cdot \delta_{h,i}} + 76.238$
8	-153.26	20.9%	$R_i = -0.65114 \cdot \frac{\delta_{h,s} \cdot Ra_i}{\Delta\delta_{M,s} \cdot \delta_{h,i}} + 75.413$

Table E.14: Fitness, Akaike weight W and the model found in ten runs for the mixed solvents and a tree depth of four sorted by fitness

Run	Fitness	W	Model
2	−233.22	99.9%	$R_i = -2.582 \cdot \frac{\epsilon_s}{\delta_{h,i}} + 3.721 \cdot \frac{Ra_i}{Pol_s} + 178.395 \cdot \frac{\vec{d_s}}{\eta_s} + 686.443 \cdot d_{eq,s} - 0.083 \cdot V_{m,s} - 511.291$
6	−216.67	0.0%	$R_i = -71.359 \cdot \eta_s^2 - 218.887 \cdot \frac{M_s}{Ra_s} - 2.564 \cdot \frac{\epsilon_s}{\delta_{h,i}} + 0.193 \cdot V_{m,s} + 2.001 \cdot \rho_s - 339.509$
7	−216.28	0.0%	$R_i = -9990.443 \cdot d_{eq,s}^2 + 1.564 \cdot \frac{\Delta\delta_{M,i}}{Pol_s} - 2.552 \cdot \frac{\epsilon_s}{\delta_{h,i}} - 2680.828 \cdot \gamma_s + 8019.745 \cdot \delta_{d,s} - 64576.684$
3	−213.76	0.0%	$R_i = -16394.210 \cdot \frac{\Delta\delta_{M,i}}{\rho_s \cdot \delta_{h,i} \cdot \Delta\delta_{M,s}} + 0.282 \cdot Ra_s \cdot Pol_s + 0.210 \cdot V_{m,s} - 13.135 \cdot \epsilon_s + 33.312$
10	−213.12	0.0%	$R_i = -2.282 \cdot \frac{\epsilon_s}{\frac{\delta_{h,i}}{Ra_i \cdot \delta_s}} - 0.159 \cdot \frac{\eta_i}{V_{m,s}} + 79.822 \cdot \frac{\epsilon_s}{d_{eq,i}} - 3.960 \cdot \delta_{p,s}^2 + 63.301$
1	−210.43	0.0%	$R_i = -0.406 \cdot \frac{\Delta\delta_{M,s} \cdot \delta_{h,i}}{\frac{\delta_{h,s}}{\eta_i}} - 0.475 \cdot Ra_s \cdot M_s - 680.419 \cdot \gamma_s + 2055.089 \cdot \delta_{d,s} - 17216.986$
4	−209.62	0.0%	$R_i = -2.797 \cdot \frac{\delta_{h,s}}{\delta_{h,i}} + 0.0185 \cdot V_{m,s} \cdot \delta_{h,i} + 23.127 \cdot \delta_s + 1.865 \cdot \Delta\delta_{M,s} - 612.455$
9	−209.48	0.0%	$R_i = -0.094 \cdot \frac{\eta_s}{V_{m,s}} - 2.598 \cdot \frac{\epsilon_s}{\delta_{h,i}} - 0.191 \cdot Pol_s \cdot \delta_s + 63.697 \cdot \delta_{p,s} \cdot d_{eq,s} + 63.647$
5	−208.90	0.0%	$R_i = -20.8117 \cdot \frac{\eta_s}{\delta_{h,i}} - 1.143 \cdot \frac{\Delta\delta_{M,i}}{V_{m,i}} + 196.254 \cdot \frac{\delta_s}{\gamma_s} + 24.652 \cdot \delta_{d,s} + 0.218 \cdot V_{m,i} - 496.039$
8	−205.80	0.0%	$R_i = -244.572 \cdot \frac{\epsilon_s}{\delta_{h,i} \cdot (\gamma_i + M_s)} + -1814.677 \cdot \frac{\gamma_s}{M_s} - 23279.247 \cdot d_{eq,s} - 291.398 \cdot \delta_s + 22136.462$
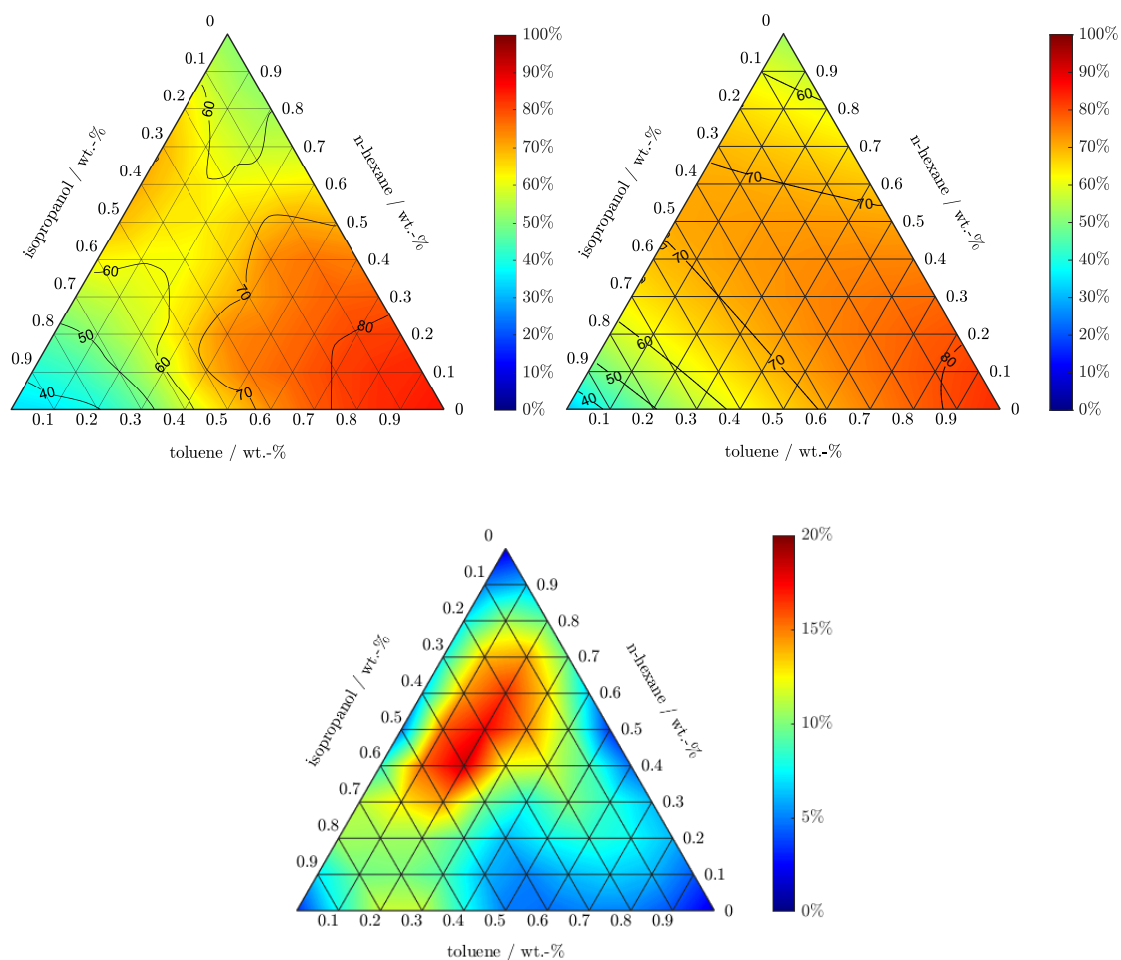
Figure E.1: MRM for DIPN, according to smoothing and interpolating of the measured rejections in the ternary system (top left) and model predictions of $R_{t,i}^4$ (top right), as well as absolute difference between both MRM (bottom).
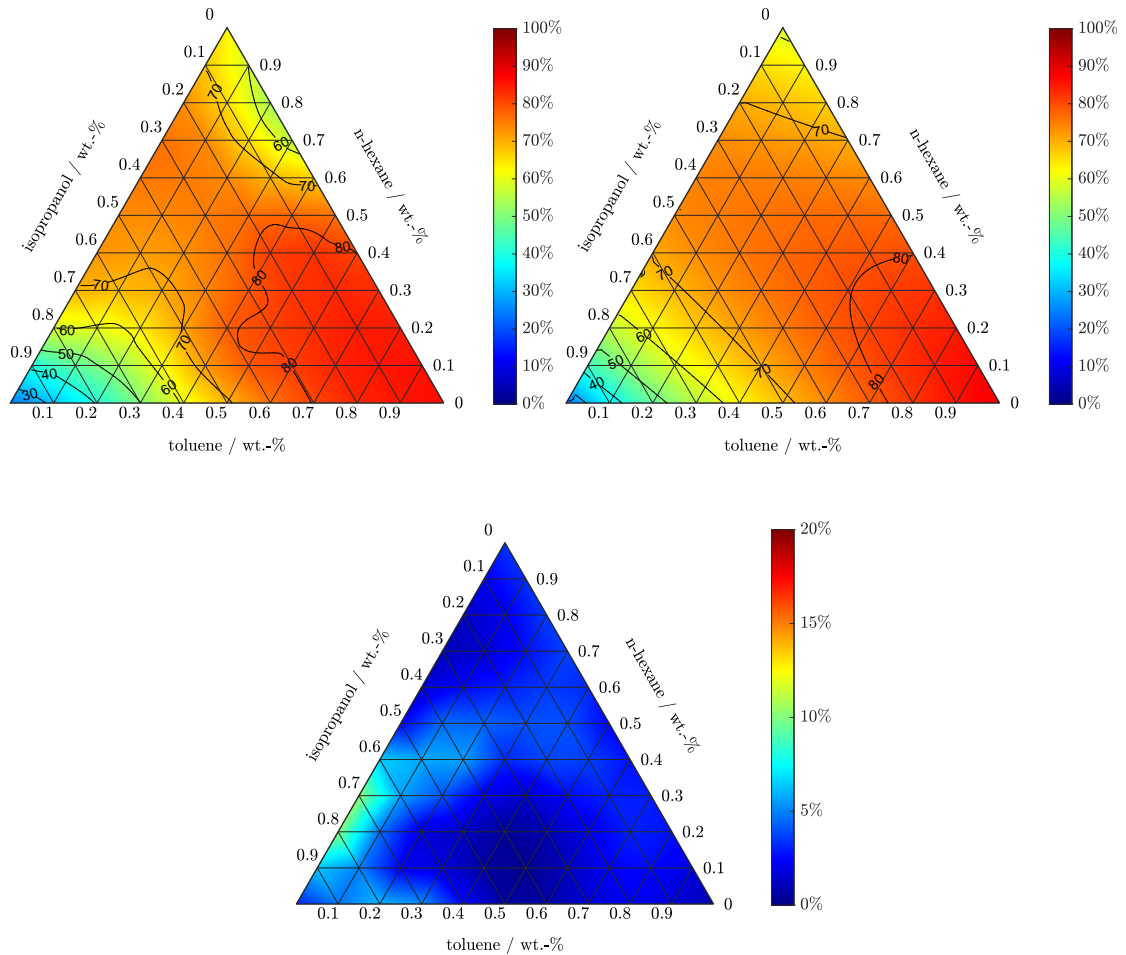
Figure E.2: MRM for HD, according to smoothing and interpolating of the measured rejections in the ternary system (top left) and model predictions of $R_{t,i}^4$ (top right), as well as absolute difference between both MRM (bottom).
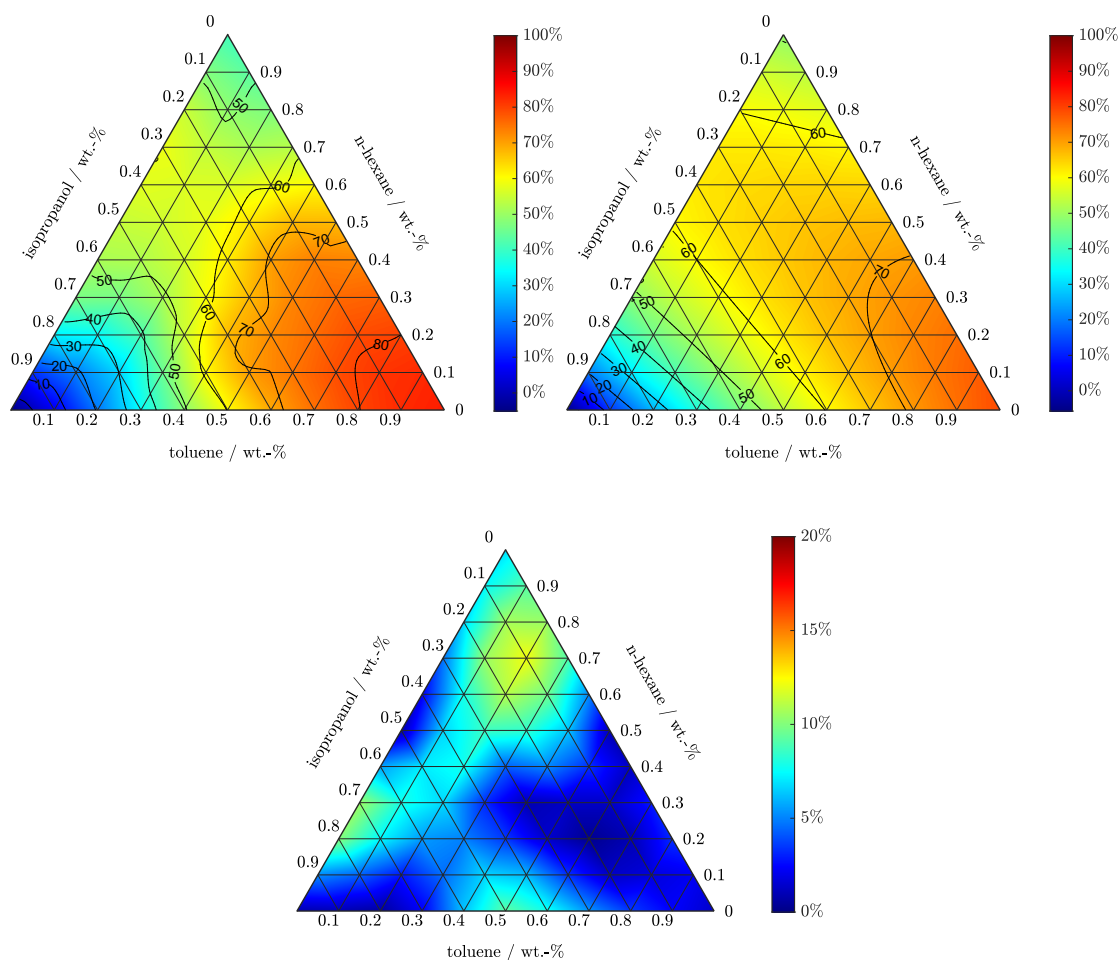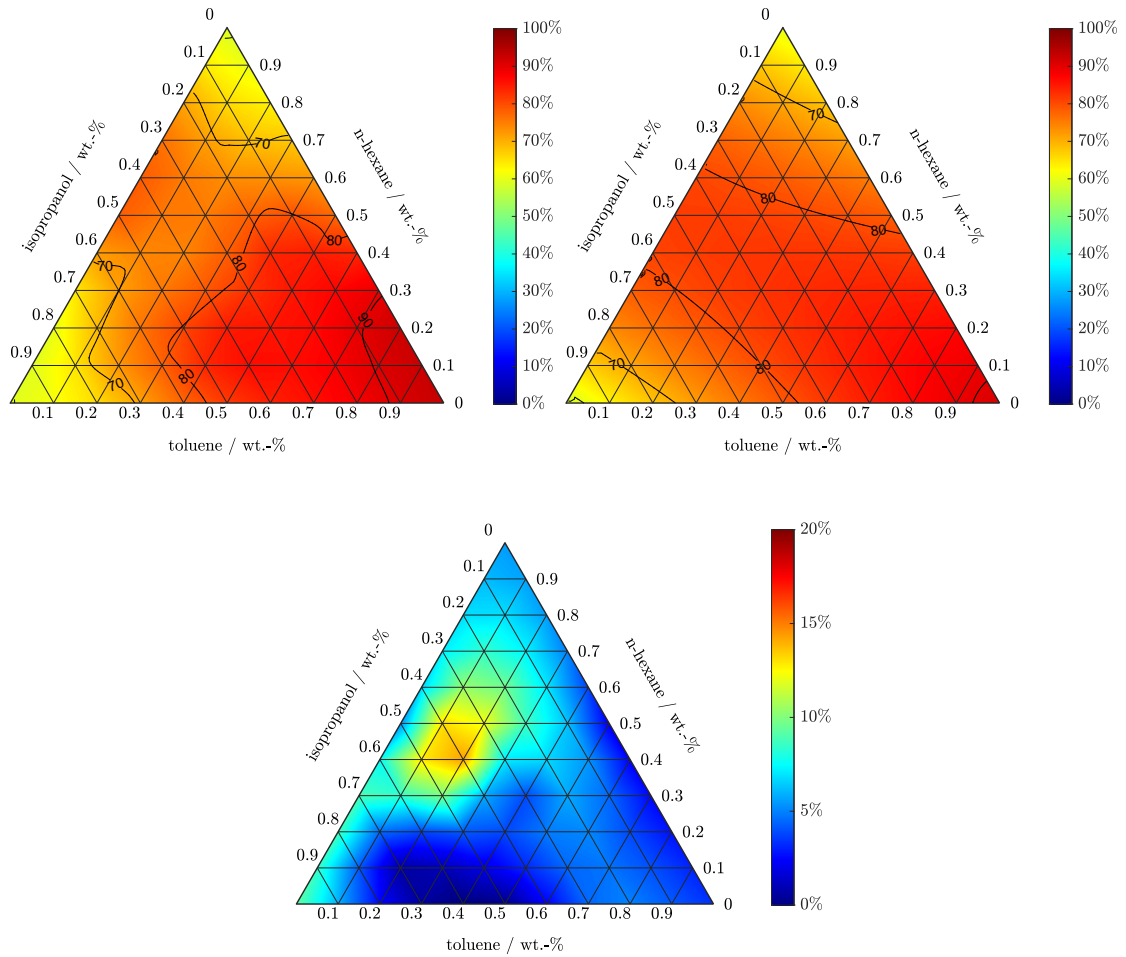
Figure E.3: MRM for PDD, according to smoothing and interpolating of the measured rejections in the ternary system (top left) and model predictions of $R_{t,i}^4$ (top right), as well as absolute difference between both MRM (bottom).

Figure E.4: MRM for TPP, according to smoothing and interpolating of the measured rejections in the ternary system (top left) and model predictions of $R_{t,i}^4$ (top right), as well as absolute difference between both MRM (bottom).

E.2.2 Solute-specific models

For the identification of solute specific models the distribution of the solvent mixtures as described in Section 8.2 is used, but only the rejection value of one solvent at a time was included. The solute specific models were already comparably accurate as the model presented in Section 8.3 using a maximum tree depth of three (Table E.15). All solute specific models incorporate less parameters of which the confidence intervals are narrow. The results are further discussed based on HMN as an example.

Table E.15: Summary of fitness, tree depth (TD), mean ($\hat{\varepsilon}$) and maximum absolute deviation(ε_{max}), parameter value (θ), size of 95 % confidence interval (CI) and eigenvalues (λ) for the best solute specific models with a maximum tree depth of 3 for DIPN (R^3_{DIPN}), HD (R^3_{HD}), HMN (R^3_{HMN}), PDD (R^3_{DIPN}) and TPP (R^3_{TPP}) for ternary solvent mixtures

Model	TD	Fitness	$\hat{\varepsilon}$	ϵ_{max}	θ	CI	λ
R^3_{DIPN}	3.	-36.90	6.4%	18.7%	-7.07×10^{-1}	2.58×10^{-1}	2.87
					4.66×10^1	8.12	5.94×10^3
R^3_{HD}	3.	-51.14	7.9%	16.5%	2.94×10^{-1}	2.58×10^{-1}	2.87
					5.98×10^1	8.12	5.94×10^3
R^3_{HMN}	3.	-52.69	4.7%	12.5%	-2.93×10^{-1}	8.31×10^{-2}	5.53
					6.19×10^1	5.70	2.83×10^4
R^3_{PDD}	3.	-47.26	3.7%	12.5%	-3.34×10^{-1}	1.08×10^{-1}	7.26×10^{-3}
					-2.29	6.15×10^{-1}	2.03×10^4
					6.19×10^2	1.76×10^2	1.13×10^7
R^3_{TPP}	3.	-41.65	5.9%	19.9%	2.18×10^{-2}	1.19×10^{-2}	5.71×10^{-1}
					3.97×10^1	2.01×10^1	1.71×10^7

$$R^3_{DIPN} = -0.707 \cdot (\epsilon_s - \delta_{d,s}) \cdot \vec{d}_s \cdot \Delta\delta_{M,s} + 46.584 \tag{E.5}$$

$$R^3_{HD} = 0.294 \cdot \gamma_s \cdot \vec{d}_s \cdot (\Delta\delta_{M,s} - \delta_{p,s}) + 59.806 \tag{E.6}$$

$$R^3_{PDD} = -0.334 \cdot M_s \cdot Ra_s - 2.286 \cdot \delta_s \cdot \delta_{p,s} + 619.385 \tag{E.7}$$

$$R^3_{TPP} = 0.022 \cdot (\delta_s - \Delta\delta_{M,s}) \cdot \gamma_s \cdot \Delta\delta_{M,s} + 39.726 \tag{E.8}$$

Table E.16 to E.20 lists the fitness, the Akaike weight W and the model equations of all models identified in ten consecutive runs for each solute.

Table E.16: Fitness, Akaike weight W and the model found in ten runs for HMN in mixed solvents and a tree depth of three sorted by fitness

Run	Fitness	W	Model
3, 4, 6	-52.69	29.4%	$R_i = -0.293 \cdot (\delta_{p,s} - \Delta\delta_{M,s}) \cdot \vec{d_s} \cdot \gamma_s + 61.884$
8	-52.17	22.6%	$R_i = -1.630 \cdot (\Delta\delta_{M,s} - \eta_s) \cdot (\Delta\delta_{M,s} - \delta_{d,i}) + 12.66$
9	-51.53	16.5%	$R_i = 0.068 \cdot (\delta_{p,i} + \Delta\delta_{M,s}) \cdot \delta_s^2 + -132.331$
2	-50.87	11.8%	$R_i = 0.968 \cdot \delta_s \cdot \epsilon_i \cdot (\gamma_i + \Delta\delta_{M,s}) - 556.666$
7, 10	-50.86	11.7%	$R_i = -240.829 \cdot \frac{\delta_i}{d_{eq,s}} + 0.044 \cdot V_{m,s}$ $-7090.752 \cdot d_{eq,s} + 9965.139$
1, 5	-50.09	8.0%	$R_i = -238.575 \cdot \frac{\epsilon_s}{Ra_s} + 230.880 \cdot d_{eq,s} \cdot \eta_s + 34.913$

Table E.17: Fitness, Akaike weight W and the model found in ten runs for DIPN in mixed solvents and a tree depth of three sorted by fitness

Run	Fitness	W	Model
3, 9	-36.90	21.5%	$R_i = -0.707 \cdot (\epsilon_s - \delta_{d,s}) \cdot \vec{d_s} \cdot \Delta\delta_{M,s} + 46.584$
4, 6	-36.76	20.1%	$R_i = -2087.2067 \cdot \frac{\delta_{h,i}}{d_{eq,s}} - 7173.204 \cdot d_{eq,s}$ $+0.079 \cdot V_{m,s} + 9954.813$
2	-36.57	18.2%	$R_i = -6.105 \cdot \delta_{d,s} \cdot \Delta\delta_{M,s} - 0.062 \cdot \delta_{h,s} \cdot \rho_s + 1102.007$
5, 7	-36.51	17.7%	$R_i = -2.036 \cdot \delta_s^2 + 0.077 \cdot V_{m,s} + 76.454 \cdot \delta_s - 642.772$
1, 10	-36.15	14.8%	$R_i = -908.034 \cdot \frac{\delta_{p,s}}{Ra_s} - 5.319 \cdot Ra_s \cdot \delta_{d,s} + 1649.587$
8	-34.84	7.7%	$R_i = -321.4367 \cdot \frac{\epsilon_s}{Ra_s} + 3704.961 \cdot \frac{\eta_s}{\delta_i} + 17.857$

Table E.18: Fitness, Akaike weight W and the model found in ten runs for HD in mixed solvents and a tree depth of three sorted by fitness

Run	Fitness	W	Model
5	-51.14	43.6%	$R_i = 0.294 \cdot \gamma_s \cdot \vec{d_s} \cdot (\Delta\delta_{M,s} - \delta_{p,s}) + 59.806$
9, 10	-50.05	25.3%	$R_i = -1712.202 \cdot \frac{\epsilon_i}{d_{eq,s}} - 7220.566 \cdot d_{eq,s}$ $+0.049 \cdot V_{m,s} + 10138.936$
2, 4, 6-8	-49.31	17.4%	$R_i = 2.074 \cdot \delta_s \cdot \Delta\delta_{M,s} + 0.045 \cdot V_{m,s}$ $-20.820 \cdot \Delta\delta_{M,s} - 58.1967$
1, 3	-48.82	13.7%	$R_i = -220.883 \cdot \frac{\epsilon_s}{Ra_s} + 6.498 \cdot Pol_s \cdot d_{eq,s} + 78.389$

Table E.19: Fitness, Akaike weight W and the model found in ten runs for PDD in mixed solvents and a tree depth of three sorted by fitness

Run	Fitness	W	Model
5, 10	-47.26	49.1%	$R_i = -0.334 \cdot M_s \cdot Ra_s - 2.286 \cdot \delta_s \cdot \delta_{p,s} + 619.385$
3, 4, 6-9	-45.60	21.4%	$R_i = -53.975 \cdot \frac{\delta_s}{d_{eq,s}} + 27.491 \cdot M_i - 7589.604 \cdot d_{eq,s}$
1	-44.86	14.8%	$R_i = -50.056 \cdot \delta_{h,s} + 23.159 \cdot Pol_s \cdot d_{eq,s} + 25.851$
2	-44.86	14.8%	$R_i = 23.159 \cdot d_{eq,s} \cdot Pol_s - 94.607 \cdot \frac{\delta_{h,s}}{\delta_{p,s}} + 25.8506$

Table E.20: Fitness, Akaike weight W and the model found in ten runs for TPP in mixed solvents and a tree depth of three sorted by fitness

| Run	Fitness	W	Model
1-3, 9	-44.73	57.3%	$R_i = 0.009 \cdot \frac{\gamma_s \cdot Pol_s}{\epsilon_s - \delta_{p,i}} + 66.1664$
10	-42.82	13.4%	$R_i = -0.231 \cdot \frac{\delta_{h,s} \cdot \gamma_s}{d_{eq,s} - \delta_{h,s}} + 67.126$
8	-41.65	12.3%	$R_i = 0.022 \cdot (\delta_s - \Delta\delta_{M,s}) \cdot \gamma_s \cdot \Delta\delta_{M,s} + 39.726$
4	-41.45	11.1%	$R_i = 0.365 \cdot \epsilon_s \cdot \Delta\delta_{M,s} \cdot (\gamma_s - \epsilon_s) + 57.471$
5-7	-40.22	6.0%	$R_i = 0.214 \cdot M_s \cdot \delta_s + 11.650 \cdot \frac{V_{m,i}}{\gamma_s} - 364.358$
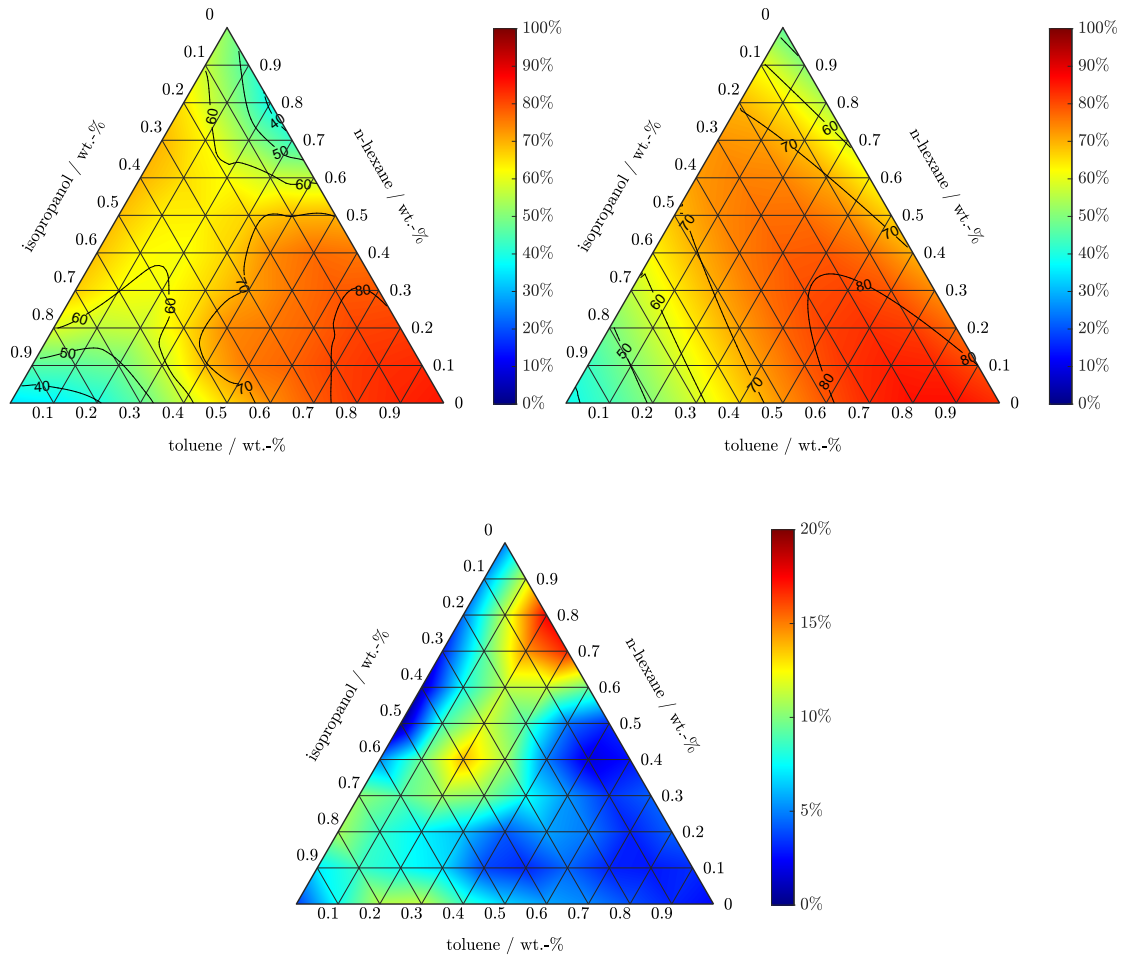
Figure E.5: MRM for DIPN, according to smoothing and interpolating of the measured rejections in the ternary system (top left) and model predictions of R^3_{DIPM} (top right), as well as absolute difference between both MRM (bottom).
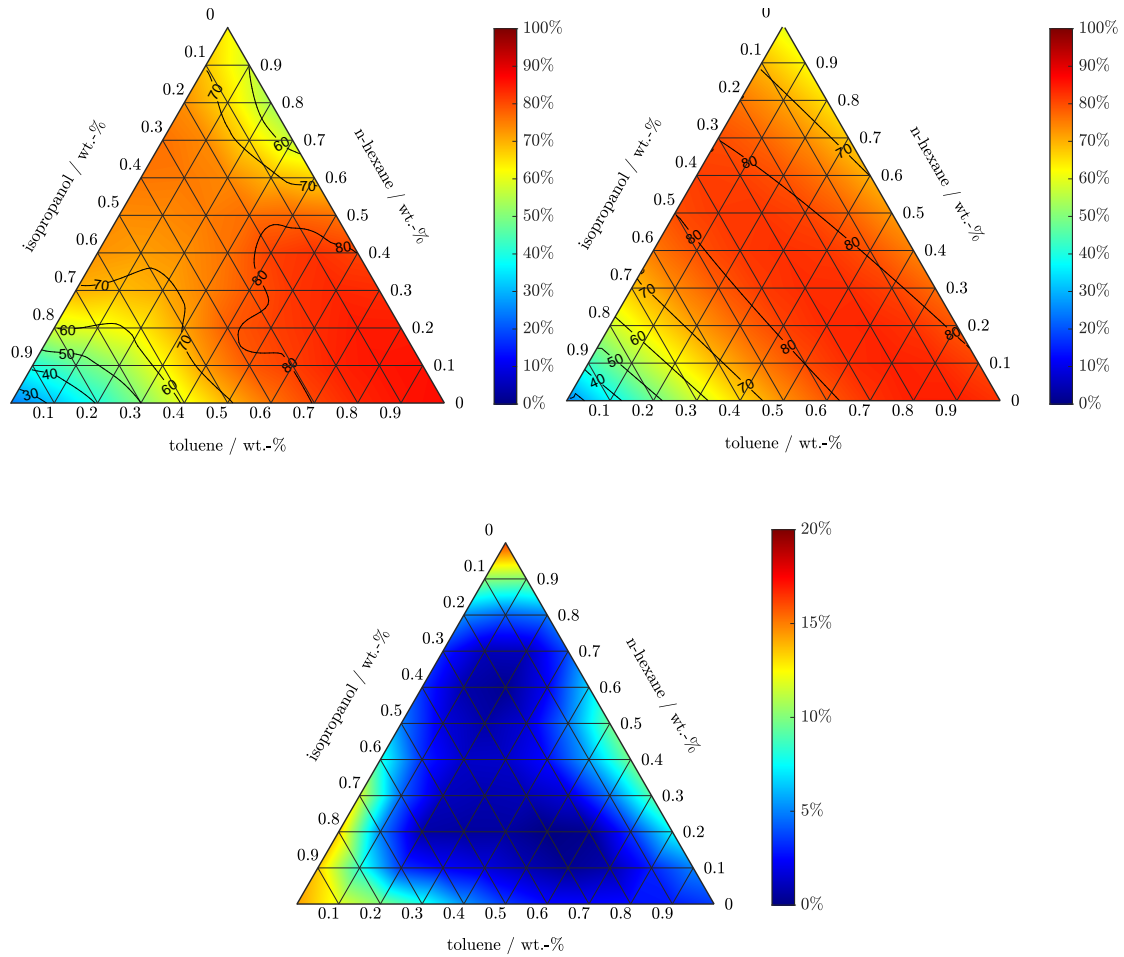
Figure E.6: MRM for HD, according to smoothing and interpolating of the measured rejections in the ternary system (top left) and model predictions of R_{HD}^3 (top right), as well as absolute difference between both MRM (bottom).
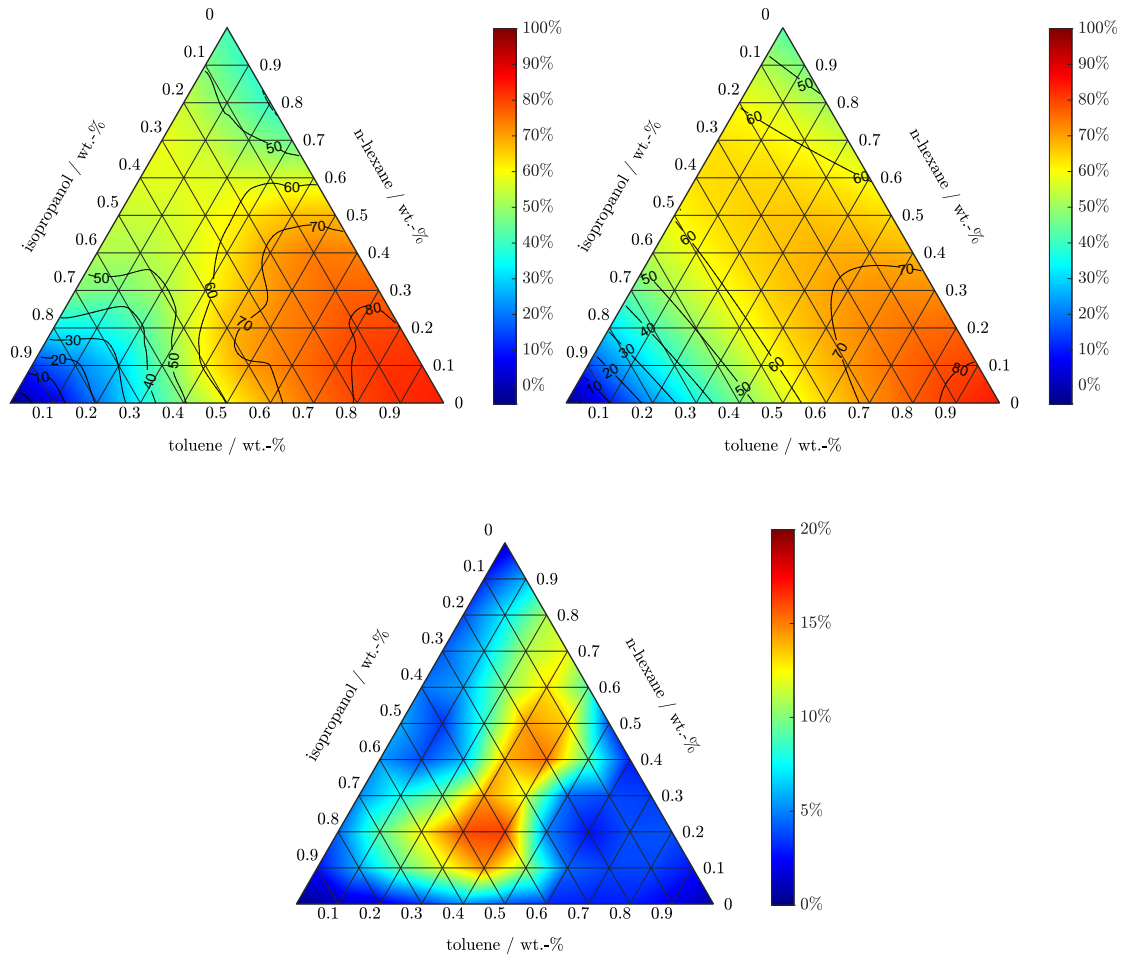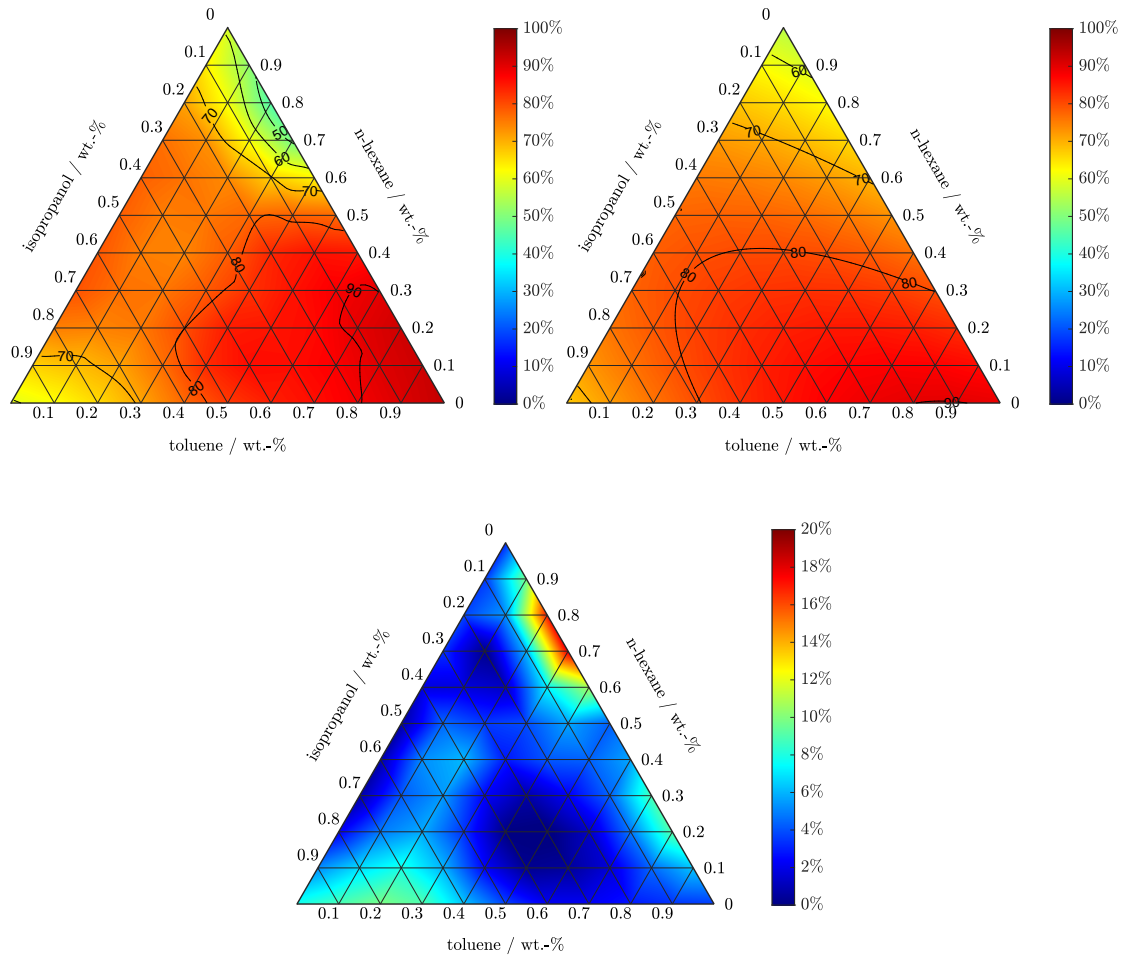
Figure E.7: MRM for PDD, according to smoothing and interpolating of the measured rejections in the ternary system (top left) and model predictions of R_{PDD}^3 (top right), as well as absolute difference between both MRM (bottom).

Figure E.8: MRM for TPP, according to smoothing and interpolating of the measured rejections in the ternary system (top left) and model predictions of R^3_{TPP} (top right), as well as absolute difference between both MRM (bottom).

E.3 Properties used for model development

The solvent and solute properties used for model development are listed in Table D.4 in Chapter D and Table E.21. Additionally, the molar volume V_m, the equivalent molecular diameter d_{eq}, the total Hildebrand solubility parameter δ, the difference of the Solubility parameters of solvent/solute and membrane $\Delta\delta_{M,s}$ and the interaction radius Ra are calculated using Equations D.5-D.9. The solubility parameters of PDMS and PI were estimated using the group contribution method of van Krevelen and Nijenhuis [119] resulting in values of $\delta_{d,PDMS} = 13.9$ (MPa)$^{0.5}$, $\delta_{p,PDMS} = 0.3$ (MPa)$^{0.5}$, $\delta_{h,PDMS} = 6.9$ (MPa)$^{0.5}$, $\delta_{d,PI}20.2$ (MPa)$^{0.5}$, $\delta_{p,PI} = 14.8$ (MPa)$^{0.5}$ and $\delta_{h,PI} = 8.4$ (MPa)$^{0.5}$.

Table E.21: Properties of solutes used in model development and predictions

Solute	M g mol^{-1}	ρ kg m^{-3}	ϵ −	η mPa s	γ mN m^{-1}	δ_d[1] (MPa)$^{0.5}$	δ_p[1] (MPa)$^{0.5}$	δ_h[1] (MPa)$^{0.5}$	$\vec{d}$ [154] debye
DIPN	212.3	883.9[1]	9.8[1]	7.5[1]	34.5 [157]	17.7	1.0	1.6	0.4 [1]
PDD	246.4	856.0 [158]	4.3[1]	4.8 [159]	29.4 [1]	17.8	1.9	0.7	0.0
TPP	262.3	1105.0 [160]	2.9 [161]	4.6 [2]	40.7 [157]	18.2 [119]	0.1 [119]	4.6 [119]	1.5
HMN	226.4	793.0 [162]	2.0 [161]	3.3 [162]	9.9 [1]	14.1	1.4	1.8	0.0
IBB	134.2	853.0 [163]	2.3 [152]	0.9 [164]	27.0 [157]	17.6	2.6	2.2	0.5
HB	162.3	861.0 [152]	2.3 [152]	1.4 [165]	29.8 [157]	17.8	2.8	2.0	0.0
EN	156.2	992.0 [166]	2.7 [152]	3.8 [167]	36.1 [157]	19.0	1.0	2.4	0.0
PD	218.4	856.0 [152]	4.4[1]	2.5 [168]	30.5 [157]	17.8	2.2	1.2	0.0
IPM	270.5	850.0 [169]	3.1 [152]	5.6 [169]	24.3 [1]	15.4	3.0	4.2	2.1
VE	430.7	950.0 [170]	NA	224.2 [171]	37.3 [2]	24.6	0.8	5.4	2.3
C10	142.3	726.0 [172]	2.0 [152]	0.8 [152]	23.4 [152]	15.1	2.8	2.7	0.0
C12	170.5	745.0 [173]	2.0 [152]	1.4 [152]	24.2 [174]	15.1	2.5	2.3	0.0
C14	198.4	759.0 [175]	2.0 [152]	2.1 [152]	26.3 [157]	15.1	2.2	1.8	0.0
C16	226.4	770.0 [176]	2.0 [152]	3.0 [152]	26.6 [177]	15.1	1.9	1.4	0.0
C18	254.5	778.5 [178]	2.1 [161]	3.6 [179]	25.6 [1]	15.1	1.6	1.0	0.0
C22	310.6	791.0 [180]	2.1 [152]	6.9 [181]	26.5 [182]	15.1	1.0	0.1	0.0
C24	338.7	795.8 [183]	2.1 [161]	8.9 [184]	26.9 [185]	15.1	0.7	0.0	0.0
C28	394.8	806.7 [186]	2.1 [161]	15.4 [187]	27.8 [1]	15.1	0.1	0.0	0.0
C32	450.9	805.8 [188]	2.1 [161]	16.5 [184]	29.8 [157]	15.1	0.0	0.0	0.0
C36	507.0	780.0 [189]	2.1 [161]	27.8 [187]	29.6 [1]	15.1	0.0	0.0	0.0

[1] calculated by ICAS 15–ProPred Version 4.3 using Marrero and Gani method,
[2] calculated by Aspen Properties Version 8.8 using NRTL method based on available experimental data

Publications

Journal Articles

R. Goebel, T. Glaser, M. Skiborowski, Machine-based learning of predictive models in organic solvent nanofiltration: Solute rejection in pure and mixed solvents, Separation & Purification Technology 248 (2020), pp. 117046

R. Goebel, M. Skiborowski, Machine-based learning of predictive models in organic solvent nanofiltration: Pure and mixed solvent flux, Separation & Purification Technology 237 (2020), pp. 116363.

R. Goebel, M. Schreiber, V. Koleva, M. Horn, A. Górak, M. Skiborowski, On the reliability of lab-scale experiments for the determination of membrane specific flux measurements in organic solvent nanofiltration, Chemical Engineering Research and Design 148 (2019), pp. 271–279.

A. Böcking, V. Koleva, J. Wind, Y. Thiermeyer, S. Blumenschein, R. Goebel, M. Skiborowski, W. Wessling, Can the variance in membrane performance influence the design of organic solvent nanofiltration processes?, Journal of Membrane Science 575 (2019), pp. 2017–2228.

B. Bertleff, R. Goebel, J. Claußnitzer, W. Korth, M. Skiborowski, P. Wasserscheid, A. Jess, J. Albert, Investigations on Catalyst Stability and Product Isolation in the Extractive Oxidative Desulfurization of Fuels Using Polyoxometalates and Molecular Oxygen, ChemCatChem 10 (20), 2018, pp. 4602–4609.

R. Goebel, T. Glaser, I. Niederkleine, M. Skiborowski, Towards predictive models for organic solvent nanofiltration, Computer Aided Chemical Engineering 43 (2018), pp. 115–120.

T. Goetsch, P. Zimmermann, R. van den Bongard, S. Enders, T. Zeiner, Superposition of Liquid–Liquid and Solid–Liquid Equilibria of Linear and Branched Molecules: Ternary Systems, Industrial & Engineering Chemistry Research 56 (1),

(2017), pp. 417–423.

T. Goetsch, P. Zimmermann, R. van den Bongard, S. Enders, T. Zeiner, Superposition of Liquid–Liquid and Solid–Liquid Equilibria of Linear and Branched Molecules: Binary Systems, Industrial & Engineering Chemistry Research 55 (42), (2016), pp. 11167–11174.

R. Schulz, R. van den Bongard, J. Islam, T. Zeiner, Purification of Terpenyl Amine by Reactive Extraction, Industrial & Engineering Chemisty Research 55 (19), (2016), pp. 5763–5769.

Conference Contributions (Selection)

R. Goebel, M. Skiborowski, Machine-based learning of predictive solvent flux models for ceramic membranes, Presentation at 17. Aachener Membran Kolloquium (2018), Aachen, Germany.

R. Goebel, T. Glaser, I. Niederkleine, M. Skiborowski, Development of predictive models for organic solvent nanofiltration based on a data-driven approach, Presentation at 28th European Symposium on Computer-Aided Process Engineering (ESCAPE) (2018), Graz, Austria.

R. Goebel, M. Skiborowski, Automatic model identification for organic solvent nanofiltration by means of machine learning, Poster at 16th Membrane poster day (2018), Leuven, Belgium.

R. Goebel, I. Niederkleine, T. Glaser, M. Skiborowski, Vorhersage des Trennverhaltens organophiler Nanofiltrationsmembranen, Presentation at Jahrestreffen der ProcessNet Fachgruppen Fluidverfahrenstechnik und Membrantechnik (2018), Munich, Germany.

R. van den Bongard, L. Lingemann, M. Skiborowski, Systematic identification of suitable models for describing separation efficiency in organic solvent nanofiltration, Presentation at 10th World Congress of Chemical Engineering (2017), Barcelona, Spain.

R. van den Bongard, P. Larysz, M. Skiborowski, A new tool for prediction of

solvent permeance, Presentation at 6th international Conference on organic solvent nanofiltration (2017), St. Petersburg, Russia.

R. van den Bongard, L. Lingemann, M. Skiborowski, Systematische Modellidentifikation für die Beschreibung der Trennleistung bei der organophilen Nanofiltrationsolutions, Presentation at Jahrestreffen der ProcessNet Fachgruppen Fluidverfahrenstechnik und Membrantechnik (2017), Cologne, Germany.

R. van den Bongard, K. Werth, M. Skiborowski, Untersuchung der quantitativen Vorhersagbarkeit des Permeatflusses in der organophilen Nanofiltration, Presentation at Jahrestreffen der ProcessNet-Fachgruppen Mechanische Flüssigkeitsabtrennung und Membrantechnik (2016), Kassel, Germany.

Supervised Theses

Master Theses

L. Lingemann, Systematic investigation of suitable models for organic solvent nanofiltration, 2016.

P. Larysz, Tool-development for the prediction of solvent permeance using genetic programming, 2017.

I. Niederkleine, Erweiterung einer Methodik zur Vorhersage des Flusses von Lösungsmittelgemischen in der organophilen Nanofiltration, 2017.

T. Glaser, Extension of a Method for Prediction of Rejection in Organic Solvent Nanofiltration, 2017.

Bachelor Theses

R. Wilmshöfer, Experimentelle Untersuchung zur Lösungsmittelpolarität in der organophilen Nanofiltration, 2016.

M. Peterwitz, Experimentelle Untersuchung des Einflusses der Lösungsmittelviskösität in der organophilen Nanofiltration, 2016.

L. Pietzka, Experimentelle Untersuchung zur Trenncharakteristik organophiler Nanofiltrationsmembranen, 2016.

M. Schreiber, Experimentelle und theoretische Untersuchung der Vorhersagbarkeit von Lösungsmittelflüssen durch organophile Nanofiltrationsmembranen, 2017.

M. Imhausen, Untersuchung zur Übertragbarkeit des Trennverhaltens auf unterschiedliche Lösungsmittel in der organophilen Nanofiltration, 2017.

S. Krick, Experimentelle Untersuchung des Trennverhaltens einer organophilen Nanofiltrationsmembran in Lösungsmittelgemischen, 2018.

B. Wintz, Entwicklung einer Modelldatenbank für die organophile Nanofiltration, 2018.

P. Lotz, Automatische Entwicklung von Druckverlust-Modellen für Packungskolonnen durch maschinelles Lernen, 2018.

Declaration

Contents of this thesis were acquired within the research project "Energieeffiziente Stofftrennung in der chemischen und pharmazeutischen Industrie durch Membranverfahren" (ESIMEM) funded by the Federal Ministry for Economic Affairs and Energy under project number 03ET1279F. Parts of this work have already been published by the author or contain measured data evolved from supervised student theses at the Laboratory of Fluid Separations, Faculty of Biochemical and Chemical Engineering, TU Dortmund University. Detailed information is given below:

Chapter 2	2.2.4	Modified from	A
	2.2.5	Modified from	B
Chapter 4		Published in	C
Chapter 5	5.2, 5.3	Modified from	B
	5.4.1	Published in	B
		Experimental data from	a
	5.4.2, 5.4.4	Published in	C
	5.4.3	Experimental data from	D
Chapter 6		Method partly from	b
Chapter 7		Published in	A
Chapter 8		Published in	B
		Experimental data from	a
Chapter A		Published in	D
Chapter B		Partly published in	B
	B.1	Experimental data partly from	d
	B.2	Experimental data from	a
	B.4	Experimental data from	c
Chapter D		Published in	A (SI)
Chapter E		Published in	B (SI)

Publications

A R. Goebel, M. Skiborowski, Machine-based learning of predictive models in organic solvent nanofiltration: Pure and mixed solvent flux, Separation & Purification Technology 237 (2020), pp. 116363

B R. Goebel, T. Glaser, M. Skiborowski, Machine-based learning of predictive models in organic solvent nanofiltration: Solute rejection in pure and mixed solvents, Separation & Purification Technology 248 (2020), pp. 117046

C R. Goebel, M. Schreiber, V. Koleva, M. Horn, A. Górak, M. Skiborowski, On the reliability of lab-scale experiments for the determination of membrane specific flux measurements in organic solvent nanofiltration, Chemical and Engineering Research and Design 148 (2019), pp. 271-279

D A. Böcking, V. Koleva, J. Wind, Y. Thiermeyer, S. Blumenschein, R. Goebel, M. Skiborowski, M. Wessling, Can the variance in membrane performance influence the design of organic solvent nanofiltration processes?, Journal of Membrane Science 575 (2019), pp. 217-228

Student theses

a Masterarbeit T. Glaser, Extension of a Method for Prediction of Rejection in Organic Solvent Nanofiltration, 2017

b Masterarbeit L. Lingemann, Systematic investigation of suitable models for organic solvent nanofiltration, 2016

c Masterarbeit I. Niederkleine, Erweiterung einer Methodik zur Vorhersage des Flusses von Lösungsmittelgemischen in der organophilen Nanofiltration, 2017

d Bachelorarbeit L. Pietzka, Experimentelle Untersuchung zur Trenncharakteristik organophiler Nanofiltrationsmembranen, 2016