Model-based Approaches to Demand Side Management of Continuous Industrial Processes

Modellgestützte Ansätze zur Laststeuerung kontinuierlicher industrieller Prozesse

Von der Fakultät für Maschinenwesen der
Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades eines
Doktors der Ingenieurwissenschaften genehmigte Dissertation

vorgelegt von

Pascal Schäfer

Berichter: Universitätsprofessor Alexander Mitsos, Ph. D.
Professor Ignacio E. Grossmann, Ph. D.

Tag der mündlichen Prüfung: 11. Dezember 2020

Titel: Model-based Approaches to Demand Side Management of Continuous Industrial Processes

Autor: Pascal Schäfer

Reihe: Aachener Verfahrenstechnik Series
AVT.SVT – Process Systems Engineering
Band 13 (2020)

Herausgeber: Aachener Verfahrenstechnik
Forckenbeckstraße 51
52074 Aachen
Tel.: +49 (0)241 80 97717
Fax.: +49 (0)241 80 92326
E-Mail: secretary.svt@avt.rwth-aachen.de
http://www.avt.rwth-aachen.de/AVT

Vorwort

Die vorliegende Arbeit entstand während meiner Zeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Systemverfahrenstechnik der RWTH Aachen. Der wichtigste Dank gilt meinem Doktorvater Prof. Alexander Mitsos für seine Unterstützung, die zahlreichen fachlichen Diskussionen und insbesondere die Freiheit, die er mir in meiner Forschung ließ und die letztendlich zu der vorliegenden Arbeit geführt hat. Daneben möchte ich Prof. Ignacio Grossmann für die Bereitschaft zur Übernahme des Koreferates danken.

Die Forschungsarbeiten wurden im Rahmen der ersten beiden Förderphasen des vom Bundesministerium für Bildung und Forschung geförderten Kopernikus-Projektes SynErgie durchgeführt. Mein herzlichster Dank gilt daher allen Beteiligten, die den reibungslosen Projektablauf ermöglicht haben. Insbesondere gilt hier mein Dank den Industriepartnern von Linde Engineering und Trimet SE sowie den Mitarbeitern des Lehrstuhls für Anlagen- und Prozesstechnik der TU München für die fachliche Unterstützung. Daneben möchte ich mich bei den Kolleginnen und Kollegen des operativen Lenkungskreises des Projektes für die stets freundliche und offene Atmosphäre in zahlreichen Meetings und Seminaren bedanken.

Ein ganz besonderer Dank geht natürlich an die Kolleginnen und Kollegen der SVT, die meine Zeit am Lehrstuhl nicht nur fachlich sondern auch menschlich extrem bereichert haben. Dies gilt insbesondere für meine beiden SynErgie-Mitstreiter, Adrian Caspari und Artur Schweidtmann, mit denen man sowohl offen jede neue Idee bis ins kleinste Detail diskutieren als auch auf diversen Dienstreisen die Münchner Brauhäuser erkunden konnte. Namentlich erwähnt werden sollen an dieser Stelle noch Wolfgang Huster als mein erster Kontakt zur SVT, Christian Redepenning als derjenige, der mich thematisch für den Lehrstuhl begeistert hat, Dominik Bongartz als mein Buddy der ersten Wochen, Luise Bering als beste Bürokollegin, die man sich wünschen kann, sowie Falco Jung als Reiseleitung zweier USA-Trips. Ein Dank geht auch an die nichtwissenschaftlichen Mitarbeiterinnen und Mitarbeiter des Lehrstuhls aus Sekretariat, Buchhaltung und IT sowie an unseren Oberingenieur Adel Mhamdi, die einen in jeder erdenklichen Situation unterstützt haben.

Schließlich möchte ich mich noch bei allen bedanken, die mich auf meinem akademischen Werdegang begleitet haben. Hier ist insbesondere meinen Kommilitonen der ersten Stunde für die schöne Zeit in den letzten bald zehn Jahren sowie meiner Familie für die fortwährende Unterstützung zu danken. Der letzte Dank ist aber der wichtigsten Person in meinem Leben vorbehalten, meiner Frau Verena, deren Unterstützung ich mir jederzeit sicher sein konnte und auch in Zukunft sein werde.

Aachen, Dezember 2020 *Pascal Schäfer*

Contents

Acronyms

ANN	Artificial neural network
ARMA	Autoregressive moving average
ASU	Air separation unit
CAES	Compressed air energy storage
CV	Controlled variable
DAE	Differential algebraic equation
DoF	Degree of freedom
DSM	Demand side management
eNMPC	Economic nonlinear model predictive control
FMU	Functional mock-up unit
FWT	Fast wavelet transform
GBM	Geometric Brownian motion
LMPC	Linear model predictive control
LIN	Liquid nitrogen
LOX	Liquid oxygen
LP	Linear program
MESH	Material, equilibrium, summation, heat
MIDO	Mixed-integer dynamic optimization
MILP	Mixed-integer linear program
MINLP	Mixed-integer nonlinear program
MV	Manipulated variable
NMPC	Nonlinear model predictive control
NLP	Nonlinear program
OCP	Optimal control problem
ODE	Ordinary differential equation
PRL	Primary balancing power (German: Primärregelleistung)
PSE	Process systems engineering
RMSE	Root mean square error
SQP	Sequential quadratic programming
SRL	Secondary balancing power (German: Sekundärregelleistung)
TSO	Transmission system operator

Abstract

Industrial demand side management is increasingly recognized as an important measure to align the network-wide electricity consumption with a volatile generation from intermittent renewable sources. At the same time, induced price fluctuations at liberalized electricity markets offer promising monetary incentives for flexible operation. Consequently, within the past decades, numerous works aimed at raising the economic potentials of adjusted production schedules by using model- and optimization-based approaches. Within this context, the contribution of the present thesis is fourfold.

First, the relevant literature currently mainly confines to scheduling optimization problems embedding (piecewise) linearized process models to maintain computational tractability. In contrast, we herein focus on the development of algorithms that allow for the consideration of long planning horizons with fine discretizations in the case of scheduling formulations with nonlinear process models. In particular, we show that feasible near-optimal schedules can thereby be furnished, although exposing only a small fraction of the degrees of freedom to the optimizer, which leads to substantial savings in computational times compared to solution approaches considering the full dimensionality.

Second, we build on existing works that argue to account for the transient process behavior during load changes in scheduling optimizations for processes with slow dynamics. In particular, we herein present a hybrid mechanistic/data-driven model reduction approach for distillation columns that allows for substantial reductions in computational times while providing similar prediction capabilities as full-order stage-to-stage models. Thereby, we enhance the real-time capability of economic nonlinear model predictive control schemes and enable participation in continuous electricity markets directly at the control layer.

Third, we address the optimal decision-making at pay-as-bid electricity markets, such as the German market for primary balancing power, where a severe trade-off between a higher probability of acceptance of a bid and a higher potential compensation payment exists. To account for this trade-off, we herein propose a two-stage stochastic problem formulation, where we optimize the bid at the balancing market auction in the first stage and market the remaining flexibility at the spot market in the second stage by a scheduling optimization.

Fourth, we critically question currently discussed measures increasing the flexibility potential of industrial processes from the perspective of the decarbonization of the electricity sector. For this purpose, we consider a generic process and evaluate whether measures that enhance the exploitation of time-variable electricity prices also result in a reduction of the consumption of fossil-generated electricity in favor of renewable electricity. Most importantly, we show that characteristic price fluctuation patterns prevent adequate monetary incentives for providing storage capacities that enable load shiftings on time scales in the order of one day, which would however be crucial for reducing the environmental impacts associated with the generation of the purchased electricity.

Kurzfassung

Industrielle Laststeuerung wird zunehmend als wichtige Maßnahme erkannt, den netzweiten Stromverbrauch an eine volatile Erzeugung aus erneuerbaren Quellen anzupassen. Gleichzeitig bieten die in liberalisierten Strommärkten verursachten Preisschwankungen vielversprechende monetäre Anreize für einen lastflexiblen Betrieb. Entsprechend versuchten in den vergangenen Jahrzehnten zahlreiche Arbeiten, wirtschaftliche Potenziale angepasster Produktionspläne mittels modell- und optimierungsbasierter Ansätzen zu heben. Vor diesem Hintergrund ist der Beitrag der vorliegenden Arbeit vierfacher Natur.

Erstens: Die relevante Literatur beschränkt sich derzeit primär auf Betriebsoptimierungsprobleme, die (abschnittsweise) linearisierte Prozessmodelle zwecks Einsatz effizienter numerischer Lösungsverfahren verwenden. Im Gegensatz dazu liegt der Fokus dieser Arbeit auf der Entwicklung von Algorithmen, die die Betrachtung langer Planungshorizonte mit feinen Diskretisierungen auch für Optimierungsprobleme mit nichtlinearen Prozessmodellen ermöglichen. Insbesondere zeigen wir, dass so zulässige, nah-optimale Produktionspläne identifiziert werden können, obwohl nur ein kleiner Anteil an Freiheitsgraden dem Optimierer zur Anpassung freigegeben wird. Dies führt zu erheblichen Reduktionen der Rechenzeit im Vergleich zu Lösungsansätzen, die die volle Dimensionalität berücksichtigen.

Zweitens erweitern wir bestehende Arbeiten, die für eine Berücksichtigung des transienten Prozessverhaltens während Lastwechseln bei Betriebsoptimierungen von Prozessen mit langsamer Dynamik plädieren. Hier stellen wir einen hybriden mechanistisch-datengetriebenen Modellreduktionsansatz für Rektifikationskolonnen vor, der signifikante Reduktionen der Rechenzeiten ermöglicht und gleichzeitig ähnliche Vorhersagegenauigkeiten aufweist wie rigorose Stufenmodelle. Dadurch verbessern wir die Echtzeitfähigkeit nichtlinearer modellprädiktiver Regelungsansätze mit ökonomischer Zielfunktion und erlauben eine Teilnahme an kontinuierlichen Strommärkten direkt auf Ebene der Regelung.

Drittens betrachten wir die optimale Entscheidungsfindung auf „Pay-as-bid“-Strommärkten, wie dem deutschen Markt für Primärregelleistung, wo die Abwägung zwischen einer höheren Wahrscheinlichkeit des Zuschlags für ein Angebot und einer höheren potenziellen Ausgleichszahlung erforderlich ist. Um dies zu berücksichtigen, wird hier eine zweistufige stochastische Problemformulierung vorgestellt, bei der wir in der ersten Stufe das Gebot für die Auktion auf dem Regelleistungsmarkt optimieren und in der zweiten Stufe die verbleibende Flexibilität für eine Betriebsoptimierung am Spotmarkt einsetzen.

Viertens hinterfragen wir derzeit diskutierte Maßnahmen zur Erhöhung des Flexibilitätspotenzials von industriellen Prozessen kritisch vor dem Hintergrund der Dekarbonisierung des Stromsektors. Dazu betrachten wir einen generischen Prozess und evaluieren, ob Maßnahmen, die die Ausnutzung variabler Strompreise erleichtern, auch zu einer Reduktion des Verbrauchs fossil erzeugten Stroms zugunsten von erneuerbarem führen. Insbesondere zeigen wir, dass typische Preisschwankungen adäquate monetäre Anreize für die Vorhaltung von Speicherkapazitäten, die Lastverschiebungen auf Zeitskalen in der Größenordnung eines Tages ermöglichen und daher entscheidend für die Verringerung der mit der Erzeugung des bezogenen Stroms verbundenen Umweltauswirkungen wären, verhindern.

Publications and copyrights

The present thesis originates from the research performed by the author during his time at the Chair of Process Systems Engineering (AVT.SVT). Parts of the thesis have already been published in different journals. In particular, (parts of) the following publications listed in the order of their appearance are reused and integrated into the chapters of the thesis as stated below. Alexander Mitsos provided ideas and guidance as well as edits to all listed publications. Contributions of other co-authors are stated below the corresponding publications.

- P. Schäfer, A. M. Schweidtmann, P. H. A. Lenz, H. M. C. Markgraf, A. Mitsos. Wavelet-based grid-adaptation for nonlinear scheduling subject to time-variable electricity prices. *Computers & Chemical Engineering* **132**, 106598, 2020, `https://doi.org/10.1016/j.compchemeng.2019.106598`. Copyright © (2020) Elsevier.

  This publication describes the original algorithm and the first case study in Chapter 3.

  Co-authors' contributions: Artur M. Schweidtmann provided the framework for using ANNs within MAiNGO. Philipp H. A. Lenz implemented a first version of the wavelet-based grid adaptation algorithm during his bachelor's thesis. Hannah M. C. Markgraf performed the training of the ANNs for the case study and implemented a first version of the scheduling problem using the full time grid during her bachelor's thesis.

- P. Schäfer, A. M. Schweidtmann, A. Mitsos. Nonlinear scheduling with time-variable electricity prices using sensitivity-based truncations of wavelet transforms. *AIChE Journal* **66**(10), e16986, 2020, `https://doi.org/10.1002/aic.16986`. Copyright © (2020) John Wiley and Sons.

  This publication describes the further development of the algorithm and the second case study in Chapter 3.

  Co-authors' contributions: Artur M. Schweidtmann provided the framework for using ANNs within MAiNGO.

- P. Schäfer, A. Caspari, K. Kleinhans, A. Mhamdi, A. Mitsos. Reduced dynamic modeling approach for rectification columns based on compartmentalization and artificial neural networks. *AIChE Journal* **65**(5), e16568, 2019, `https://doi.org/10.1002/aic.16568`. Copyright © (2019) John Wiley and Sons.

  This publication presents the model formulation as well as the open-loop control case study in Chapter 4.

  Co-authors' contributions: Adrian Caspari provided the algorithmic framework for solving the open-loop dynamic optimization problems. Kerstin Kleinhans performed the data generation and the training of the ANNs for the model reduction

approach during her master's thesis. Adel Mhamdi provided advice during the development of the model and the formulation of the case study.

- P. Schäfer, A. Caspari, A. Mhamdi, A. Mitsos. Economic nonlinear model predictive control using hybrid mechanistic data-driven models for optimal operation in real-time electricity markets: In-silico application to air separation processes. *Journal of Process Control* **84**, 171–181, 2019, `https://doi.org/10.1016/j.jprocont.2019.10.008`. Copyright © (2019) Elsevier.

  This publication presents the closed-loop control case study in Chapter 4.

  Co-authors' contributions: Adrian Caspari provided the algorithmic framework for solving the open-loop dynamic optimization problems. Adel Mhamdi provided advice during the formulation of the case study.

- P. Schäfer, A. Caspari, A. M. Schweidtmann, Y. Vaupel, A. Mhamdi, A. Mitsos. The potential of hybrid mechanistic/data-driven approaches for reduced dynamic modeling: Application to distillation columns. *Chemie Ingenieur Technik* **92**(12), 1910–1920, 2020, `https://doi.org/10.1002/cite.202000048`. Copyright © (2020) John Wiley and Sons.

  Parts of this review/perspective paper are integrated in Chapter 4.

  Co-authors' contributions: Adrian Caspari, Artur M. Schweidtmann, Yannic Vaupel, and Adel Mhamdi contributed through numerous discussions about hybrid model structures, their application to chemical engineering problems, and their limitations.

- P. Schäfer, H. Westerholt, A. M. Schweidtmann, S. Ilieva, A. Mitsos. Model-based bidding strategies on the primary balancing market for energy-intense processes. *Computers & Chemical Engineering* **120**, 4–14, 2019, `https://doi.org/10.1016/j.compchemeng.2018.09.026`. Copyright © (2019) Elsevier.

  This publication forms the basis of Chapter 5.

  Co-authors' contributions: Hermann Westerholt implemented the two-stage stochastic problem formulation in his master's thesis. Artur M. Schweidtmann contributed through discussions about the problem formulation as well as the solution approach. Svetlina Ilieva provided industrial expertise for conducting the case study.

- P. Schäfer, T. M. Daun, A. Mitsos. Do investments in flexibility enhance sustainability? – A simulative study considering the German electricity sector. *AIChE Journal* **66**(11), e17010, 2020, `https://doi.org/10.1002/aic.17010`. Copyright © (2020) John Wiley and Sons.

  This publication forms the basis of Chapter 6.

  Co-authors' contributions: Torben M. Daun implemented the scheduling formulation using the generic process model in GAMS and performed the numerical optimizations during his bachelor's thesis.

Furthermore, during the time at AVT.SVT, the author supervised or co-supervised the student theses of R. Herding, L. Bering, L. Monigatti, H. Westerholt, H. Markgraf, P. Lenz, N. Hansmann, K. Kleinhans, T. Daun, L. Dederichs, and J. Lüthje. The work of all students is gratefully acknowledged. In particular, the bachelor's theses of H. Markgraf, P. Lenz, and T. Daun as well as the master's theses of H. Westerholt and K. Kleinhans formed the bases of some of the aforementioned publications. Their contribution is thus stated above.

In addition to that, the following publications, which are not part of this thesis, were authored or contributed to while being at AVT.SVT:

- A. Mitsos, N. Asprion, C. A. Floudas, M. Bortz, M. Baldea, D. Bonvin, A. Caspari, P. Schäfer. Challenges in process optimization for new feedstocks and energy sources. *Computers & Chemical Engineering* **113**, 209–221, 2018.
- P. Schäfer, L. F. Bering, A. Caspari, A. Mhamdi, A. Mitsos. Nonlinear dynamic optimization for improved load-shifting agility of cryogenic air separation plants. *Computer Aided Chemical Engineering* **44**, 547–552, 2018.
- A. Caspari, J. M. M. Faust, P. Schäfer, A. Mhamdi, A. Mitsos. Economic nonlinear model predictive control for flexible operation of air separation units. *IFAC-PapersOnLine* **51**(20), 295–300, 2018.
- A. Caspari, Y. Martin Pérez, C. Offermanns, P. Schäfer, A.-M. Ecker, A. Peschel, F. Schliebitz, G. Zapp, A. Mhamdi, A. Mitsos. Economic nonlinear model predictive control of multi-product air separation processes. *Computer Aided Chemical Engineering* **46**, 1–6, 2019.
- P. Schäfer, N. Hansmann, S. Ilieva, A. Mitsos. Model-based bidding strategies for simultaneous optimal participation in different balancing markets. *Computer Aided Chemical Engineering* **46**, 1639–1644, 2019.
- A. Caspari, C. Offermanns, P. Schäfer, A. Mhamdi, A. Mitsos. A flexible air separation process: 1. Design and steady-state optimizations. *AIChE Journal* **65**(11), e16705, 2019.
- A. Caspari, C. Offermanns, P. Schäfer, A. Mhamdi, A. Mitsos. A flexible air separation process: 2. Optimal operation using economic model predictive control. *AIChE Journal* **65**(11), e16721, 2019.
- A. Caspari, L. Lüken, P. Schäfer, Y. Vaupel, A. Mhamdi, L. T. Biegler, A. Mitsos. Dynamic optimization with complementarity constraints: Smoothing for direct shooting. *Computers & Chemical Engineering* **139**, 106891, 2020.
- P. Schäfer, A. Mitsos. Tailored time grids for nonlinear scheduling subject to time-variable electricity prices by wavelet-based analysis. *Computer Aided Chemical Engineering* **48**, 1123–1128, 2020.
- D. T. Doncevic, A. M. Schweidtmann, Y. Vaupel, P. Schäfer, A. Caspari, A. Mitsos. Deterministic global nonlinear model predictive control with neural networks embedded. *In press: IFAC-PapersOnLine*, 2020.

- J. T. Lüthje, J. C. Schulze, A. Caspari, A. Mhamdi, A. Mitsos, P. Schäfer. Adaptive learning of hybrid models for nonlinear model predictive control of distillation columns. `https://arxiv.org/abs/2011.12798`, 2020.

1. Introduction

Meeting nowadays energy demand is still primarily achieved by the depletion of fossil resources. The involved emissions of large quantities of greenhouse gases have already led to tremendous and potentially irreparable influences on the world's climate system. Consequently, the most crucial challenge for future societies and economies lies in the transformation of all value chains to substantially more sustainable ones while at the same time guaranteeing to provide the required amount of energy to all consumers.

Process systems engineering (PSE) can contribute to meeting this global challenge by developing new process designs and operating schemes that allow for the conversion of renewable energy sources and feedstocks into required products in a sustainable manner [1]. Among the variety of research topics that fit into this context, this thesis is concerned with the transformation of the electricity sector towards a high penetration of volatile generation from intermittent renewable sources and thereby aims at the reduction of the overall consumption of fossil-generated electricity in favor of low-carbon electricity. We acknowledge that our perspective is certainly shaped by the German energy system, which is characterized by large shares of non-dispatchable renewable generation by wind farms and photovoltaics. We however emphasize that in absence of opportunities for a base load capable renewable generation, e.g., from hydroelectric plants, similar circumstances will likely characterize many other energy systems as well - at least in the future.

In order to enable high shares of electricity generation by wind farms and photovoltaics, a certain degree of flexibility is required in the energy system, so that the currently mostly inelastic network-wide electricity demand can be aligned with the volatile nature of the generation [2]. To this end, there are two fundamental alternatives: first, the fluctuations of the electricity generation can be smoothened by storing excess generation and releasing it in times of shortages and second, the electricity consumption can be flexibilized to meet the fluctuating electricity supply by active load management, which is referred to as demand side management (DSM). Although measures that are taken on the supply side and those taken on the demand side act similarly from a network-wide perspective, i.e., load shedding of consumers acts similarly on the electricity grid level as discharging a battery storage, we herein primarily focus on measures of DSM. Being more specific, the main subject of this thesis lies in utilizing the load flexibility of existing continuous industrial processes by appropriate adjustments of their operation.

The provision of load flexibility by industrial processes is - in particular in the case of continuous processes - usually in a severe conflict with their primary purpose, that is, the production of desired goods [3]. More precisely, if either shedded load could not be caught up or a flexible operating strategy led to large amounts of products violating the specifications, i.e., if production targets were missed, the process operator would have certainly failed from an entrepreneurial perspective. Consequently, without any monetary incentives for a flexibilization of the electricity consumption, no rational entrepreneur would be willing to take such risks. Fortunately, monetary incentives are inherently set at liberalized electricity markets, where the fundamental principles of supply and demand lead to

fluctuating price signals depending on the instantaneous network-wide demand and the realized feed-in from intermittent renewable sources. Note that these circumstances have nowadays gained broad attention due to increasing numbers of periods with even negative prices. The exploitation of electricity market price spreads appearing on various time scales by shifting load from periods with high prices to those with low ones, i.e., by performing DSM, is hence currently considered an appealing measure to reduce the costs for the electricity purchase. Even more, this opportunity might ultimately cause a paradigm change. Namely, whereas the most efficient process, i.e., the process with the lowest energy consumption per output unit traditionally used to be the most competitive one, a higher degree of flexibility might eventually become a nonnegligible competitive edge.

Against the described backdrop, when approaching DSM from the perspective of a process systems engineer, the concept of model-based approaches gains a particular importance. That is, decisions about the process operation should preferably be made on a sound basis, having a model that predicts the behavior of the (abstracted) system depending on the selection of the values of the decision variables. Going one step further, such models then also allow for determining those values for the decision variables that lead to an optimal objective value, which is commonly achieved by performing numerical optimizations. Undeniably, the quality of the furnished results and thus the quality of the derived decisions relies on the quality of the applied model. Setting up a suitable model for a certain purpose is therefore maybe the most crucial part of the entire workflow and also the central aspect of the present thesis. As a generally applicable yet abstract rule, a suitable model always contains only the most relevant phenomena and discards all irrelevant ones. In the case of models in the context of DSM, the main challenge in modeling hence lies in the harmonization of three aspects: the addressed electricity market, the relevant decision-making layer, and the time scales and characteristics of process responses. The question of how to solve this challenge, i.e., how to harmonize the aspects by formulating a suitable model for our purposes, accompanies us through all chapters of the present thesis.

Certainly, this thesis is not the first work to discuss model-based approaches to DSM, as illustrated in the literature review in the following chapter. For instance, a comprehensive review paper from 2016 by Zhang and Grossmann [4] already lists more than 40 references since 1989 that fall under this category with numerous articles having been additionally published between 2016 and now. However, when revisiting the existing literature, one can easily figure out a predominant perspective on model-based approaches to DSM, which interprets DSM as a scheduling problem with time-variable input data and which sets up quasi-stationary linearized process models. Besides that, there are several aspects related to DSM that have been started to be discussed in recent years, such as implications due to the interaction of scheduling and control layers, economic potentials of partitioning the process flexibility among multiple markets, and a fair assessment of the reduction of environmental impacts when performing DSM. In the present thesis, we identify further relevant research perspectives beyond the state-of-the-art in DSM-related literature, for which we propose model-based solution approaches. We highlight that the main contribution lies in the development of methodologies. To this end, we make use of several well-established methods from various research fields within PSE and combine and tailor them for our specific purposes.

The remainder of the thesis is structured as follows: In Chapter 2, we first give an overview of the existing DSM-related literature clustered according to the discussion above. Based on this, we then derive so far mostly unnoticed but highly relevant research perspec-

tives and state the contributions of the present thesis in detail. In Chapter 3, we address the computational challenges when embedding nonlinear process models into discrete-time scheduling optimizations. We therein present an algorithm that allows for the consideration of long planning horizons with fine discretizations. The algorithm relies on a reduced-space scheduling formulation. In its original version, we assign one degree of freedom to multiple scheduling intervals and iteratively refine the assignment using a wavelet-based analysis. The algorithm is applied to the scheduling of a compressed air energy storage (CAES). Afterwards, we present an extension of the algorithm in the case of optimization problems with only continuous variables, where we directly expose the coefficients of the wavelet transform as variables to the optimizer, which allows for making decisions on refinements based on sensitivity information from Lagrangian multipliers. The extended algorithm is applied to the scheduling of an air separation unit (ASU). In Chapter 4, we propose a hybrid mechanistic/data-driven model reduction approach for distillation columns combining classical compartmentalization with surrogate modeling using artificial neural networks (ANNs). The reduction approach substantially enhances the real-time capability of model predictive control schemes. We thus apply the reduced hybrid model within a single-layer control framework of an ASU, which allows for performing DSM by participating in continuous electricity markets directly at the control layer. In Chapter 5, we introduce a two-stage stochastic problem formulation for optimizing the bids at pay-as-bid markets for balancing power. There, the selection of the offered capacities and ask prices corresponds to the first stage, whereas the second stage is a scheduling optimization subject to time-variable spot market prices. In a case study, we identify optimal bidding strategies for an industrial aluminum electrolysis cell that provides reserve capacities for balancing frequency fluctuations. In Chapter 6, we finally present a generic process model, which we use for assessing synergies and trade-offs between economic and environmental objectives of DSM. In particular, we therein discuss that measures increasing the flexibility potential of a process affect the two objectives unevenly due to different fluctuation characteristics in the underlying time series. We conclude the thesis in Chapter 7 with a summary of the insights and results from a high level and outline perspectives for future research.

2. Literature review and contributions of the thesis

Fluctuating spot electricity prices as well as high compensation payments for capacities that are reserved for the provision of ancillary services have motivated extensive research to assess the economic benefits from DSM for numerous energy-intense processes by using model-based approaches. Here, the majority of literature focuses on ASUs [5–18]. However, other processes are well-investigated as well, such as chlor-alkali electrolysis [19–23], electric arc steelmaking [24–27], aluminum electrolysis [28], cement [29, 30], seawater desalination [31–33], pulping [34] and electrically boosted glass furnaces [35]. Note that there are also several comparable efforts beyond DSM that address the operational optimization of processes on the supply side under the described circumstances, e.g., the scheduling of both conventional [36–39] and renewable thermal producers [40, 41] and energy storage systems [42, 43].

Approaching the referenced works from a methodological point of view, the applied models essentially fulfill two tasks. First, the actual process is modeled. The main purpose of this part of the model is to predict how changing the operation affects the key process variables, i.e., those variables which appear in the objective function or the constraints of the optimization problems, such as the power consumption, limiting product specifications, etc. Second, there is a need for quantifying the impacts of the adjusted operation. To this end, there is a corresponding part of the model that commonly computes the objective function for optimization as well as potential other performance indicators from the process variables.

In the following, we briefly review the existing literature in Section 2.1 against the backdrop of the discussion above. In particular, we first discuss the predominant perspective considering DSM as a scheduling problem with time-variable input data, i.e., using quasi-stationary (mostly linearized) process models and quantifying the benefits by considering a time-variable pricing of the consumed electricity that offers opportunities to reduce the electricity costs. Afterwards, we present recently addressed research fields beyond this perspective, such as the consideration of transient process behavior through the use of dynamic models, the partitioning of the available flexibility among multiple markets with individual rules, horizons, and pricing schemes, and the extension of the scheduling perspective towards the quantification of environmental impacts, particularly by using the reduction of the carbon footprint of the purchased electricity as an alternative objective function. Based on the clustered review, we then derive the research questions of the present thesis in Section 2.2 by identifying promising but so far mostly unnoticed topics.

2.1. Literature review

2.1.1. Demand side management as a scheduling problem with time-variable input data

Revisiting the referenced literature from the beginning of the chapter, one finds that the vast majority of works considers DSM as a scheduling problem with a time-variable pricing of one of the process inputs, namely of the purchased electricity. Thus, consumers are able to adapt to the given price signals retrieved from the market by adjusting their operation [4], making DSM a problem of enterprise-wide optimization [44]. Consequently, the majority of the DSM-related literature essentially builds on a sound framework, which provides numerous methods for process scheduling using discrete [45–48] and continuous-time formulations [49–53] as well as recent combinations thereof [54–56]. Advantages and disadvantages of the individual types of time representations have been widely reviewed [57–60], indicating no general superiority of one approach. In the case of scheduling subject to time-variable electricity prices, only a few authors however propose continuous-time formulations [24, 26, 61], as the input data is inherently discrete (e.g., hourly changing prices in day-ahead markets). Consequently, we limit the further review to the vast majority of literature in the context of DSM-oriented scheduling that relies on discrete-time formulations, mostly by applying the same discretization as the electricity prices.

In the light of the discussion above concerning the applied models, most of these works essentially use a similar objective function, namely the electricity cost added up over all intervals in the scheduling horizon (partially extended by accounting for additional raw materials and/or opportunity costs for shedded production). This objective function is minimized by assigning optimal loads to the individual intervals while respecting all technical (e.g., operating limits) and economic (e.g., production targets) constraints. In this setting, the process model simply represents the link between the load and those process variables, which are relevant for the constraints. For this purpose, one commonly relies on quasi-stationary models, i.e., by assuming that all scheduled set-point changes can be realized by the subordinated control with negligible delay. Note that often further constraints restrict the admissible set-point changes by introducing ramping limits, which will be discussed together with its implications in the following subsection.

The most basic version of such a process model in the context of DSM-oriented scheduling is provided in the early work of Daryanian et al. [62]. Therein, the authors introduce a storage-type customer that can vary its electricity consumption and store products up to a certain limit. As the production rate is proportional to the electricity consumption and there are no additional constraints beside operating/storage limits and a cumulative production target for the considered horizon, scheduling of this process corresponds to solving a linear program (LP), which is undeniably the least complicated type of optimization problems. The first major extension of this perspective is given by Ierapetritou et al. [5], who address the DSM-oriented scheduling of a real process, namely an ASU. Therein, the authors consider multiple admissible operating modes for the ASU (regular, assisted, and shut-down mode) by introducing binary variables as well as logical constraints on the switching between the modes by formulating sets of linear inequalities. As the authors apply a linearized power consumption function, the corresponding scheduling problem becomes a mixed-integer linear program (MILP), which is much more complicated to solve than an LP but can still be handled efficiently by state-of-the-art computer codes. Es-

sentially, this formed the basis of several subsequent works, such as the one by Karwan and Keblis [6] or the ones by Mitra et al. [30, 37], which either provide additional linear formulations for further logical constraints on mode switches or tighten existing constraint formulations to enhance the solvability. These MILP problem formulations found numerous applications in the literature, mostly in the context of ASUs (e.g., [7]) but for instance, also for modeling of a switchable chlor-alkali-electrolysis [21].

Real processes are however often not well represented by simplified linearized models that directly allow for (MI)LP formulations. Therefore, Zhang et al. [63] propose a framework to approximate nonconvex feasible sets with nonlinear objective functions by unions of polytopes - each of them with piecewise linear objective function - and thereby ultimately enable MILP formulations that can be handled efficiently in these case as well. The approach is referred to as convex region surrogate model and is successfully applied to the scheduling of ASUs in numerous case studies [8, 64–68]. Although such models relying on piecewise linearization might result in a large number of disjunctions reformulated using binary variables in case of strongly nonlinear multivariate relations, only a few authors tried a direct consideration of nonlinear process models in discrete-time scheduling so far, as this leads to (mixed-integer) nonlinear programs ((MI)NLPs) with potentially multiple suboptimal local optima. Consequently, solving these scheduling problems preferably requires global solution approaches to avoid furnishing suboptimal solutions [31], which however easily causes prohibitive problem sizes, i.e., numbers of variables and constraints. That is, state-of-the-art deterministic global solvers cannot cope with large numbers of intervals in discrete-time scheduling formulations that stem from long planning horizons with fine discretizations.

Particularly when embedding nonlinear relations, extending the scope of the scheduling calculations, i.e., towards longer horizons and/or finer temporal discretizations, is thus a crucial challenge. As an alternative to simplifications of the actual process model through piecewise linearizations, a second alternative lies in an approximation of the model for calculating the objective function in the scheduling optimization. To this end, methods for time series aggregation, which have been widely used in the context of energy systems engineering (cf. the overview given by Teichgraeber and Brandt [69] and the references therein), become highly appealing. Essentially, these approaches represent an entire horizon comprising a large number of intervals by only a few characteristic intervals and have mostly been used to consider operational decisions during the design optimization of energy systems by reducing the dimensionality of the corresponding MILP scheduling formulations [70–73]. The application of methods for time series aggregation to short-term scheduling subject to time-variable electricity prices using MINLP formulations is nevertheless in principle straight-forward. However, classical methods for time series aggregation do not consider the chronology of intervals. Therefore, their application to all systems with time-coupling constraints, such as storage or ramping limits, which characterize almost every DSM-relevant process, is de facto impossible, as furnished schedules would in general be infeasible due to violations of the time-coupling constraints. To overcome this restriction, recent literature proposes an aggregation method relying on the identification of characteristic periods, in which the chronology is preserved [74], finally enabling the consideration of long-term storage in design optimizations [75].

In order to avoid the issues arising from time-coupling constraints, a second alternative has been proposed to allow for a low dimensionality of the optimization problem. Corresponding approaches make use of a re-assignment between degrees of freedom (DoFs), i.e.,

the actual decision variables, and the scheduling intervals, leading to tailored nonuniform time grids, essentially taking up an idea from the classical discrete-time batch scheduling [48]. More precisely, the dimensionality of the optimization problem is reduced by assigning one DoF to multiple intervals. As at the same time, all intervals of the original time series are explicitly considered further on, this allows for the consideration of time-coupling constraints, such as storage or ramping limits, and ultimately guarantees to furnish feasible schedules. For instance, both Pineda and Morales [76] as well as Palys and Daoutidis [77] apply such an approach by clustering multiple similar consecutive intervals in operational optimizations. To the best of our knowledge, there have however been so far no approaches to utilize the concept of tailored time grids in the context of DSM-oriented scheduling.

2.1.2. Interaction and integration of the scheduling and the control layer

The first extension of the perspective presented in the previous subsection solely addresses the process model, specifically the explicit consideration of the transient process behavior during set-point changes therein. This extension originates from the observation that process dynamics become important for scheduling decisions if set-point changes in response to market signals occur with period lengths that are in a similar order of magnitude as process time constants [78]. Neglecting the dynamics might then lead to suboptimal or even infeasible schedules [79], i.e., to scheduled set-point changes that cannot be tracked by the subordinated control without violating the specifications. As briefly stated above, most of the existing literature in the context of DSM-oriented scheduling only implicitly accounts for these circumstances by restricting the admissible set-point changes between two consecutive intervals. However, more recently, authors have started to argue to account for the transient closed-loop process behavior by using dynamic models and applied them in scheduling optimizations [10, 22, 80–82]. Along these lines, there are further approaches that aim at an explicit consideration of the subordinated control layer tracking the scheduled set-point changes either by scheduling calculations with sequential closed-loop process simulations [83] or by introducing the optimality conditions of the subordinated control problem into the formulation of the scheduling problem [84, 85].

In the context of DSM-oriented scheduling, the discussion above will mostly apply to the scheduling of ASUs, which are commonly characterized by large liquid hold-ups in the distillation columns and thus large time constants and dead-times [86, 87]. Consequently, most of the referenced approaches have been applied to ASUs. For instance, Pattison et al. [10] use low-dimensional Hammerstein-Wiener models with polynomial input and output nonlinearities for the closed-loop dynamics of scheduling-relevant variables of a single-product ASU and solve the dynamic optimization problem in a sequential manner. The same authors later extend the approach towards a moving horizon framework [88]. Kelley at al. [81] continue to use Hammerstein-Wiener models but with piecewise linear input and output nonlinearities, which allows for a simultaneous solution approach by formulating a large-scale MILP for scheduling of the same process [14]. Moreover, Tsay et al. [82] present a framework for learning of the scheduling-relevant dynamics and apply this to a large-scale multi-product ASU [18]. Finally, Dias at al. [83] also apply their proposed approach based on sequential scheduling calculations and closed-loop process simulations

to the scheduling of the aforementioned single-product ASU.

The previously mentioned works consider existing process control systems, mostly based on linear model predictive control (LMPC), and thus do not take the advantages of adaptations of the control systems into account. However, we have shown in a previous work, which has been published prior to the present thesis and is not part of it, that the use of detailed nonlinear models in a nonlinear model predictive control (NMPC) setting can substantially reduce load change periods of transiently operated distillation columns used in ASUs [89]. Having a suitable nonlinear model that is valid across the entire operating range then further allows for another promising alternative: integrating the economic perspective, i.e., the optimization subject to time-variable electricity prices in the context of DSM, directly into the controller [90]. This approach is commonly referred to as economic NMPC (eNMPC) [91]. Successful eNMPC applications in chemical engineering are reviewed by Ellis et al. [92] with a focus on batch optimizations. The sound theoretical framework concerning stability and optimality properties of eNMPC schemes, which is provided in the relevant literature, e.g., [93–95], is nevertheless still valid for continuous applications as well. The use of single-layer eNMPC schemes however requires that the nonlinear dynamic optimizations are conducted on time scales that are sufficient for both the economic performance and the traditional purpose of the control system, i.e., the rejection of disturbances. Ideally, the sampling rates for eNMPC should thus be in a similar order of magnitude as those for (regulatory) LMPC, which in contrast involves quadratic optimization problems that can be handled much more efficiently. A major challenge when applying eNMPC to chemical processes thus lies in solving nonlinear dynamic optimization problems in real-time, which particularly holds for large-scale distillation systems such as ASUs, where first-principle models easily lead to differential algebraic equation (DAE) system comprising hundreds of differential variables.

Existing applications of (e)NMPC schemes to ASUs therefore almost exclusively rely on fast update methods. For instance, Huang et al. [96] use sensitivity-based updates for regulatory NMPC of an uncoupled distillation section of an ASU producing both nitrogen and oxygen. Huang and Biegler [97] later extend this case study by an economic objective for optimizations subject to time-variable electricity prices. As an alternative to sensitivity-based updates, Caspari et al. [98] use suboptimal updates by performing only a limited number of iterations of the optimizer without interim line searches, thus restricting the number of function evaluations during the dynamic optimization, and focus on an integrated single-product nitrogen plant. For this particular application, the authors further identify promising improvements in terms of electricity costs enabled by the eNMPC when compared to a scheduling with low-dimensional dynamic models and a subordinated LMPC tracking controller [99]. The suboptimal framework based on limited function evaluations is also applied to a multi-product ASU comprising a double-column for producing both oxygen and nitrogen [100] as well as to a recently presented highly-flexible ASU configuration with an internal liquefaction cycle [101], which further allows to feed back stored product to maintain the internal columns flows during load decreases [16].

Although impressively large ASUs have already been considered in the literature, model sizes and complexities are still a bottleneck for the application of eNMPC in practice, such that fast control schemes alone do usually not allow for the desired high sampling rates. Thus, in order to enhance the real-time capability of the nonlinear dynamic optimizations, there is also a need to reduce the computational complexity of the process models - particularly of the distillation models - without impairing accuracy. Here, three fundamental

types of model reduction approaches for distillation columns have been proposed in the literature so far. They are based on (i) nonlinear wave theory [102–105], (ii) insertion of collocation points replacing single stages [106–108], and (iii) aggregation of multiple single stages to compartments [109–112]. Note that all three approaches have already been applied to modeling as well as dynamic optimization and/or control of ASUs.

Fu and Liu [113] present an advanced regulatory control scheme using a simplified wave model based on the earlier work of Liu et al. [114] on wave modeling for thermally coupled distillation columns. Moreover, Caspari et al. [105] extend the original nonlinear wave model by accounting for varying column hold-ups occurring in case of load changes and apply this model to closed-loop eNMPC of a single-product nitrogen plant subject to time-variable electricity prices. In contrast, Cao et al. [115] present a collocation-based reduction approach for the distillation columns used in ASUs. The authors then apply the reduced models to an open-loop operational optimization of a single-product nitrogen plant subject to time-variable electricity prices [116] as well as to the preemptive planning of economically optimal control actions in response to upcoming changes in the downstream demand [117]. Finally, Khowinij et al. [118] and Bian et al. [119] present compartmentalization approaches for dynamic modeling of the lower and the upper distillation column respectively in an integrated plant for producing highly pure oxygen and nitrogen. Later, Chen et al. [120] utilize these models in a regulatory NMPC scheme for the entire plant. In summary, out of the numerous works using reduced dynamic models for operational optimization of ASUs, only two, namely [116] and [105], consider the optimal operation in response to time-variable price signals. From this, only one, namely [105], focuses on closed-loop eNMPC, but still does not consider real-time relevant issues, such as time delays due to computational times as well as imperfect market forecasts.

2.1.3. Multi-level approaches to model the simultaneous participation in spot electricity and ancillary service markets

Essentially, all literature referenced in Subsections 2.1.1 and 2.1.2 relies on similar simple models for the purchase of the electricity when computing the objective functions. That is, the electricity is priced in a time-variable manner, i.e., the price per unit consumption differs over time, which de facto corresponds to a time-of-use tariff offered by the supplier to the consumer. Considering the actual participation in spot electricity markets, using such a pricing scheme still appears to be a reasonable approximation. For instance, at the German day-ahead market, which is the most relevant market for the short-term marketing of operational flexibility, prices almost solely depend on the instantaneously realized electricity generation from intermittent renewable sources as well as the nearly inelastic network-wide demand. Consequently, marginal day-ahead spot prices can be estimated quite well in the short-term. Furthermore, price time series show on average distinct repeating fluctuation patterns, which allows for sound estimations of long-term benefits of flexible operation. Note that even if considering continuous trading at intra-day markets in a pay-as-bid manner, applying a time-variable pricing appears to be a justifiable approach, as intra-day prices clearly show the trend to follow those from the day-ahead markets and usually only involve uncertainties up to some extent.

As a volatile electricity supply also threatens the grid stability, incentives are set by transmission system operators (TSOs) for both energy suppliers and consumers to provide

ancillary services, which thereby partially give up control of their energy consumption/generation. As the compensation payments for the provision of ancillary services are highly promising from an economic point of view, several works have started to take these into account in scheduling optimizations. For instance, Zhang and Hug [28] and Zhang et al. [25] consider an aluminum smelter and a steel plant respectively, which participate in both the spot electricity market as well as the market for spinning reserve, and use stochastic programming to account for uncertainties in price forecasts. Likewise, Zhang et al. [8, 65] focus on an ASU with cryogenic energy storage, which participates in the spot electricity market and simultaneously gets compensation payments for providing interruptible loads. In these works, the authors are primarily concerned with the uncertainty in the requests for the operating reserve. To this end, they first present a robust optimization approach guaranteeing the availability of the operating reserve at any point in time, which they later extent by an adjustable formulation allowing for recourse actions depending on previous realizations of the uncertainty. In addition to that, Otashu and Baldea [121] evaluate monetary benefits from the provision of frequency regulation capacities by chlor-alkali plants while accounting for process constraints and dynamics.

The main challenge in modeling the simultaneous participation in multiple markets and even more in solving these models lies in the fact that decisions have to be made in a multi-level manner. Note that this might originate from inherent uncertainties as in some of the previous references. In the general case, the multi-level perspective however already evolves from different temporal horizons of the decisions made at the various markets. In order to assess the economic benefits from such a multi-level market participation, Dowling et al. [122] provide a multi-scale optimization framework considering revenue potentials at both multiple spot markets with varying temporal discretizations and ancillary service markets. Therein, the authors identify promising incentives for additionally providing ancillary services that can substantially increase the overall revenues compared to a pure marketing of the operational flexibility at spot markets. Moreover, Otashu and Baldea [123] present a generic model for an electricity consumer, which exploits time-variable prices at a day-ahead spot electricity market in the upper level and offers remaining operational flexibility at a short-term ancillary service market. Along the same lines, Xenos et al. [9] consider an ASU and analyze remaining reserve capacities that can be marketed after optimally scheduling the production subject to time-variable day-ahead prices. Note that the literature referenced in this subsection exclusively considers exogenously determined prices, i.e., prices that are not affected by one's own market participation, which is a reasonable assumption if participating in markets applying marginal prices as done in the corresponding works. To the best of our knowledge, there have however been so far no approaches to consider the decision-making at other types of markets, particularly those where one's own participation will affect the market outcome.

2.1.4. Quantification of the environmental benefits of performing demand side management

When considering the alternative options to provide the required flexibility to the electricity grid, particularly the storage of surplus energy and the production of e-fuels, one finds that extensive studies have been carried out to quantify the environmental impacts and identify the most sustainable of various technological alternatives [124–127]. In contrast,

when considering industrial DSM, the research focus shifts to an almost exclusively entrepreneurial perspective by solely using economic objective functions (cf. the references at the beginning of this chapter). Nevertheless, first few approaches exist that try to assess the environmental impacts of performing DSM in response to price signals.

Finn and Fitzpatrick [128] consider an industrial manufacturer in Ireland and identify a positive correlation between cost savings and wind power consumption. In particular, the authors find that relative monthly electricity cost savings are on average accompanied by relative increases in the consumption of electricity generated by wind farms. Going one step further, hourly changing electricity mixes also result in hourly-changing emission factors and thereby allow for the use of an environmental objective for the scheduling optimizations as well. To this end, Kopsakangas-Savolainen et al. [129] optimize the electricity purchase of two finish industrial consumers with regard to a minimization of the greenhouse gas emissions that are associated with the generation of the electricity. Along the same lines, Summerbell et al. [130] focus on a cement plant in the UK and optimize the schedules subject to time-variable carbon footprints of the electricity and compare them to the economic optima leading to the lowest electricity costs. The authors find synergetic effects between the two objectives, i.e., the optimization for cost savings also reduces the CO_2 emissions associated with the generation of the purchased electricity and vice versa. At the same time, the authors also identify possibilities for balancing between the objectives, as optimizations for one objective leave a substantial reduction potential of approximately 50% in the other one unexploited. Following a similar approach, Kelley et al. [131] use an existing model for an ASU from their previous work [14] and compare economically and environmentally optimized scheduling alternatives for this process using electricity data for California. Therein, the authors also find mostly synergetic effects. However, in specific periods, the environmentally optimized schedules lead to substantially increased electricity costs by several percentage points compared to the stationary reference operation. Noticeably, the economically optimized schedules always leave an environmental optimization potential unexploited, but they never lead to higher CO_2 emissions compared to the stationary reference operation. Trade-offs like the ones observed in the previous references can be systematically taken into account by multi-objective optimizations. For instance, Baumgärtner et al. [132] use an ϵ-constraint method and present Pareto curves for the optimal integrated design and operation of a utility system. The authors consider total annualized costs and global warming impacts as objectives and illustrate significant potentials for balancing between these two objectives. Comparable approaches have however not yet been taken to systematically assess the impacts of measures increasing the flexibility potential on the demand side.

2.2. Contributions of the thesis

As can be seen from the literature review in the previous section, numerous model-based approaches have been applied in the literature to optimally raise the benefits of actively using the operational flexibility of energy-intense industrial processes by performing DSM. However, promising further research perspectives can be identified that have not yet been considered in the literature. These are thus addressed in the present thesis using model-based methods as well. As in the remainder of the thesis, the focus is set on the formulation of suitable models for the intended purposes further on, we continue to use the order of

Section 2.1 and present and discuss the contributions of the thesis in accordance with that. That is, we first present algorithms that enhance the solvability of scheduling formulations when embedding nonlinear process models. Second, we propose model formulations that allow for performing DSM directly by a single-layer eNMPC. Third, we propose a two-stage stochastic problem formulation for optimizing the participation at pay-as-bid balancing markets. Finally, we present a generic process model that enables an environmental assessment of monetarily incentivized measures for DSM.

2.2.1. Computationally tractable approximations of nonlinear scheduling problems

A stated in Subsection 2.1.1, an accurate capturing of nonlinear process characteristics by using nonlinear models typically prohibits long planning horizons with fine discretizations in discrete-time scheduling. Having algorithms that enable more efficient yet accurate solutions to nonlinear scheduling problems would then also allow for making use of recent advancements in formulating sophisticated data-driven surrogate models that are capable of capturing highly complex nonlinear chemical engineering relations (e.g., [133, 134]). To this end, we herein propose an algorithm relying on a reduced-space scheduling formulation exposing only the actual decision variables to the optimizer (see, e.g., [135–137]). In the original version of the algorithm, we assign one DoF to multiple generally nonconsecutive intervals with similar electricity prices, thereby furnishing a tailored time grid using fewer decision variables. The time grid is iteratively refined by conducting a wavelet-based analysis of the solution from the previous iteration. This analysis identifies both parts of the time grid that need a finer discretization and parts where a coarser one can be used, thus allowing for a systematic introduction of additional and deletion of insignificant decision variables from the optimization problem. We also present a further development of the original algorithm, enhancing its performance in case of scheduling problems comprising only continuous DoFs. Therein, we perform optimizations directly in the space of coefficients of the wavelet transform, making the explicit derivation of a time grid superfluous. Thereby, dimensionality reductions do no longer correspond to assigning one DoF to multiple intervals but rather turn into truncations of the wavelet transform. This procedure allows for calculating the marginal effect of relaxing the constraints truncating the wavelet transform, i.e., the Lagrangian multipliers, and thus provides a quantitative measure for the improvements in the objective function from the introduction of additional decision variables.

2.2.2. Real-time capable control strategies for continuous market participation

Following the discussion in Subsection 2.1.2, there is still a continued demand for highly accurate dynamic models for distillation columns that substantially improve the computational efficiency in dynamic optimizations and thus allow for deciding on energy trading in real-time also in case of processes with large time constants. Therefore, we herein present a hybrid mechanistic/data-driven modeling approach, enhancing the classical compartmentalization [109–112] by integrating machine learning. In particular, due to numerous recent successful applications in PSE to replace complex nonlinear thermodynamic calcu-

lations [138–140], we therein substitute the stage-to-stage calculations inside the compartments by using ANNs as explicit surrogate models for the intra-compartment input/output relations. The proposed ANN-based compartment inherits the advantages of classical compartment models, i.e., the compliance with integral balance relations and the possibilities for substantial reductions in the number of differential variables without significantly influencing the quality of dynamic responses. At the same time, the proposed model avoids the classical drawbacks by omitting time-consuming flash calculations on each stage. The reduced hybrid model then forms the basis of a real-time capable control strategy that allows for the participation in continuous intra-day electricity markets directly at the control layer. Therein, the ANN-based modeling approach is combined with a suboptimal eNMPC scheme that relies on a restricted number of function evaluations in the dynamic optimization [98], which further enhances the real-time capability of the control strategy.

2.2.3. Optimal participation in pay-as-bid markets for ancillary services

Several existing works have discussed potential benefits from the simultaneous participation in multiple energy markets. In particular, authors have found very promising revenue potentials when partitioning the available operational flexibility among both spot electricity and ancillary service markets (cf. Subsection 2.1.3). However, there is still a substantial research gap concerning the optimal decision-making at pay-as-bid markets, where ask prices have to balance between higher probabilities for acceptance of the bids and higher potential compensation payments from the TSO. Closing this gap is desired, as several ancillary service markets, particularly the European markets for balancing power that are highly promising for marketing the operational flexibility of energy-intense industrial processes [3, 141], apply a pay-as-bid pricing. Therefore, we herein present a framework for model-based bidding strategies based on two-stage stochastic programming. As an example, we consider the German market for primary balancing power (PRL), where participants can receive highly valuable compensation payments for offering additional reserve capacities that are automatically retrieved to control the grid frequency. In our framework, the first stage is formulated as an MINLP and corresponds to the identification of the optimal bidding strategy at the PRL market, i.e., what amounts to offer and what prices to ask to maximize the expected revenues from marketing the operational flexibility given a forecast of the development of the maximum compensation payments at the PRL market. Depending on the results of the PRL auction, the remaining flexibility is then used in the second stage to exploit price spreads at the spot electricity markets, namely the day-ahead market, via a subordinated MILP scheduling optimization. The two-stage problem is solved using a tailored enumeration approach by decomposing the MINLP formulation of the deterministic equivalent into NLPs for identifying the optimal ask prices and MILPs for determining the corresponding production schedules for each element within the finite set of realizations of the offered capacity.

2.2.4. Environmental assessment of monetarily incentivized measures for demand side management

A stated in Subsection 2.1.4, some authors have started to extend the mostly economic perspective on DSM by assessing the environmental impacts of performing DSM in response to price signals. However, there have not yet been systematic approaches towards a bi-objective assessment of measures increasing the flexibility potential on the demand side that target generalizable conclusions not limited to a specific application and provide thorough explanations for their observations. In particular, this includes the question if economically-driven measures, i.e., measures that enhance the exploitation of fluctuating electricity prices, promote the penetration of renewables by allowing to consume electricity primarily in periods with a high renewable share. To this end, we assess if the current pricing at the electricity markets sets incentives for investments in increased flexibility that are suitable for reducing the environmental impacts of the electricity consumption, which we measure by the contribution to the integral residual load. We herein provide a computational study focusing on the German electricity sector using both historic electricity time series data as well as projected time series data for the future. In a first step, we conduct a quantitative characterization of the time-variable input data sets using frequency and correlation analysis. In a second step, we investigate the effects of the different characteristics of the time series on process scheduling. For this purpose, we perform single- and bi-objective optimizations of the production schedule with regard to both electricity costs and residual load contributions using a generic process model. In this model, we introduce a general formulation for DSM activities allowing for load shiftings, including temporary shut-downs. We investigate numerous parametrizations of the generic process model considering varying average utilization rates, storage capacities, ramping limits, and off-design efficiency losses. Thereby, the vast majority of DSM-relevant processes is covered, which enables the desired generalizable conclusions not limited to a specific process. Finally, by analyzing the results of the scheduling optimizations, we systematically identify measures increasing the flexibility potential on the demand side, which are driven by the economic perspective and those driven by the environmental perspective, which allows for the discussion of contradictions in the current market mechanisms.

3. Low-dimensional approximations of nonlinear scheduling problems with time-variable electricity prices

This chapter is concerned with the development of algorithms that allow for long planning horizons with fine discretizations in discrete-time nonlinear scheduling problems with time-variable electricity prices. As the proposed approach relies on a reduced-space formulation of the optimization problem to directly translate reductions in the number of decision variables into reductions of the problem size, we first present a general reduced-space scheduling formulation in Section 3.1. Afterwards, we describe the proposed algorithm in Section 3.2. Therein, we initially describe the concept of the linear mapping procedure for reducing the number of decision variables by assigning one DoF to multiple generally nonconsecutive intervals with similar electricity prices, furnishing a tailored nonuniform time grid. Second, we briefly present the idea behind the wavelet-based analysis and describe how this can be used for iterative refinement of the time grid. Third, we discuss the properties of the algorithm, particularly its guarantee for furnishing feasible schedules. In Section 3.3, we introduce the case study for assessing the performance of the algorithm, namely the scheduling of a CAES formulated as an MINLP, where we use ANNs as sophisticated surrogate models for the highly nonlinear efficiency characteristics of the turbomachines. The performance of the algorithm in furnishing feasible near-optimal schedules in the first case study is then evaluated in Section 3.4. Therein, a distinct focus is set on both the convergence behavior of the algorithm with increasing lengths of the planning horizon as well as on the computational performance, i.e., the scaling of CPU times. The further development of the proposed algorithm in case of scheduling problems comprising only continuous DoFs is described in Section 3.5. Therein, we first discuss the remaining shortcomings of the original algorithm. Second, we introduce the approach for dimensionality reductions by exposing the coefficients of truncated wavelet transforms as only decision variables to the optimizer, which makes the explicit derivation of a time grid superfluous. Third, we describe the proposed further development of the refinement procedure by using sensitivity information from Lagrangian multipliers, which provides a quantitative measure for the improvements in the objective function from the introduction of additional decision variables. The second case study, which we use for assessing the performance of the extended algorithm, is presented in Section 3.6 and corresponds to the scheduling of a flexible ASU subject to time-variable electricity prices formulated as an NLP. Note that we therein also apply ANNs as surrogate models, representing the nonlinear power consumption characteristic of the process. The results of the scheduling optimizations are presented in Section 3.7. Again, we therein set a distinct focus on the discussion of the convergence behavior of the algorithm with increasing lengths of the planning horizon as well as on the computational performance. Furthermore, we demonstrate the advantages

compared to the original version of the algorithm in terms of a reduced number of iterations. Finally, the chapter is concluded and further research perspectives are sketched in Section 3.8.

3.1. General reduced-space scheduling formulation

The algorithm introduced in Section 3.2 targets the solution of the general scheduling problem given in (3.1a)-(3.1e). This corresponds to a reduced-space scheduling formulation. Note that a significant share of all discrete-time scheduling problems with time-variable electricity prices that have been considered in literature so far can be formulated equivalently. We do not claim that the reduced-space formulation is always beneficial. Nevertheless, the reduced-space approach is highly suited for exploiting the advantages of a reduced number of decision variables when applying the proposed algorithms.

In (3.1a)-(3.1e), the production schedule is optimized by minimizing an objective function $\Phi(\cdot)$ (3.1a), mostly corresponding to operational costs. Note that alternative objectives might be possible, such as maximizing profitability or minimizing environmental impacts (cf. Chapter 6) and would require only minor changes. In most cases, the objective function can then be written as a sum over interval-specific objective functions $\phi_t(\cdot)$ for each time interval $t = 1 \dots T$. For instance, when minimizing electricity costs, this will simply correspond to the product of the instantaneous electricity price and the current electricity consumption. d_i denote the continuous process inputs (such as mass flows, power consumptions, etc.) and y_j the binary ones (e.g., operating modes). The values of the process inputs in each time interval t are the DoFs for optimization, referred to as $d_{i,t}$ and $y_{j,t}$ and limited by respective bounds (3.1d)-(3.1e). If the DoFs are specified, the entire schedule can be calculated. Moreover, two kinds of constraints are imposed denoted by $g_{l,t}(\cdot)$ and $h_m(\cdot)$: one that needs to be fulfilled in every interval (3.1b), e.g., operating bounds, and one for the last interval (3.1c), which allows for the consideration of production targets. Other types of constraints (e.g., targets at intermediate time points, minimum requirements at final time, etc.) are possible as well and could be added in a similar way as (3.1b)-(3.1c), but are omitted for readability. The (in)equality signs in (3.1b)-(3.1c) are chosen to match the case studies, however different ones could be applied as well. Furthermore, we assume that the function $g_{l,t}(\cdot)$ in (3.1b) depends on all DoFs up to time t, as this is typically the case for real processes. However, the assumption is not necessarily required. Finally note that in a reduced-space formulation, $\Phi(\cdot)$, $\phi_t(\cdot)$, $g_{l,t}(\cdot)$, and $h_m(\cdot)$ are explicit functions of the optimization variables, i.e., the DoFs $d_{i,t}$ and $y_{j,t}$.

$$\min_{d_{i,t}, y_{j,t}} \quad \Phi(d_{i,1}, \dots, d_{i,T}, y_{j,1}, \dots, y_{j,T}) = \sum_{t=1}^{T} \phi_t(d_{i,t}, y_{j,t}) \tag{3.1a}$$

$$\text{s.t.} \quad 0 \leq g_{l,t}(d_{i,1}, \dots, d_{i,t}, y_{j,1}, \dots, y_{j,t}), \quad \forall l, t \tag{3.1b}$$

$$0 = h_m(d_{i,1}, \dots, d_{i,T}, y_{j,1}, \dots, y_{j,T}), \quad \forall m \tag{3.1c}$$

$$d_i^l \leq d_{i,t} \leq d_i^u, \quad \forall i, t \tag{3.1d}$$

$$y_{j,t} \in \{0, 1\}, \quad \forall j, t \tag{3.1e}$$

In the general problem formulation, we also allow for further intermediate variables $x_{k,t}$, that are however not exposed to the optimizer and thus not presented as part of the

reduced-space optimization problem (3.1a)-(3.1e). Instead, they are calculated successively and in an explicit manner from the DoFs and their initial conditions $x_{k,1}$:

$$x_{k,t+1} = f_k(x_{k,t}, d_{i,t}, y_{j,t}), \quad \forall k, t. \tag{3.2}$$

Constraints on $x_{k,t}$ can also be accounted for in the reduced-space formulation above by including them into (3.1b)-(3.1c). We thereby explicitly allow for time-coupling constraints that enable the modeling of storage devices with limited capacities. Note that the other common type of time-coupling constraints, i.e., ramping limits, can also be formulated equivalently to (3.1c).

3.2. Original algorithm: Tailored time grids through wavelet-based grid adaptation

The proposed algorithm comprises two key features to exploit the reduced-space formulation (i.e., hiding (3.2) from the optimizer): (i) an assignment of one DoF to multiple generally nonconsecutive intervals with similar electricity prices by a linear mapping furnishing a tailored nonuniform time grid and (ii) a wavelet-based grid adaptation procedure for iterative refinements of the assignment, for which we use the algorithm from Schlegel et al. [142] that was originally proposed for refining the control vector parametrization in dynamic optimization. Both key features are explained in the following. Furthermore, the properties of the algorithm are discussed.

3.2.1. Linear mapping for reducing the number of optimization variables

We substitute the original chronologically ordered DoFs $(d_{i,1}, \dots, d_{i,T})^T$ and $(y_{j,1}, \dots, y_{j,T})^T$, i.e., the original optimization variables in the reduced-space formulation (3.1a)-(3.1e), by new decision variables $(\hat{d}_{i,1}, \dots, \hat{d}_{i,\hat{T}})^T$ and $(\hat{y}_{j,1}, \dots, \hat{y}_{j,\hat{T}})^T$ with reduced dimensionality. Formally, these new optimization variables are mapped to the original DoFs by a linear relation using the matrix $\boldsymbol{A} \in \mathbb{R}^{T \times \hat{T}}$ (cf. (3.3a)-(3.3b)). Note that their index $\hat{t} = 1 \dots \hat{T}$ does not necessarily have a chronological meaning. In fact, we also cluster disjoint time intervals. A significant reduction of the number of optimization variables is achieved by choosing $\hat{T} \ll T$.

$$(d_{i,1}, \dots, d_{i,T})^T = \boldsymbol{A} \cdot (\hat{d}_{i,1}, \dots, \hat{d}_{i,\hat{T}})^T, \quad \forall i \tag{3.3a}$$

$$(y_{j,1}, \dots, y_{j,T})^T = \boldsymbol{A} \cdot (\hat{y}_{j,1}, \dots, \hat{y}_{j,\hat{T}})^T, \quad \forall j \tag{3.3b}$$

The use of a common matrix for all inputs is not mandatory and only selected to simplify the notation. As the algorithm targets the solution of scheduling problems with time-variable electricity prices, we consider a mapping that assigns one DoF to multiple intervals of the scheduling problem that comprise similar electricity prices. Thus, each row of $\boldsymbol{A}$ has $\hat{T} - 1$ zero entries and one entry equal to one.

Inserting the right-hand side of (3.3a)-(3.3b) into the optimization problem (3.1a)-(3.1e), we obtain a new reduced-space formulation with lower dimensionality:

$$\min_{\hat{d}_{i,\hat{t}},\hat{y}_{j,\hat{t}}} \quad \hat{\Phi}(\hat{d}_{i,1},\ldots,\hat{d}_{i,\hat{T}},\hat{y}_{j,1},\ldots,\hat{y}_{j,\hat{T}}) = \sum_{\hat{t}=1}^{\hat{T}} \hat{\phi}_{\hat{t}}(\hat{d}_{i,\hat{t}},\hat{y}_{j,\hat{t}}) \tag{3.4a}$$

$$\text{s.t.} \quad 0 \leq \hat{g}_{l,t}(\hat{d}_{i,1},\ldots,\hat{d}_{i,\hat{T}},\hat{y}_{j,1},\ldots,\hat{y}_{j,\hat{T}}), \quad \forall l,t \tag{3.4b}$$

$$0 = \hat{h}_{m}(\hat{d}_{i,1},\ldots,\hat{d}_{i,\hat{T}},\hat{y}_{j,1},\ldots,\hat{y}_{j,\hat{T}}), \quad \forall m \tag{3.4c}$$

$$d_i^l \leq \hat{d}_{i,\hat{t}} \leq d_i^u, \quad \forall i,\hat{t} \tag{3.4d}$$

$$\hat{y}_{j,\hat{t}} \in \{0,1\}, \quad \forall j,\hat{t}. \tag{3.4e}$$

In comparison to problem (3.1a)-(3.1e), there are fewer optimization variables in the formulation (3.4a)-(3.4e). The objective still remains a summation over single functions, but now with index $\hat{t}$, i.e., with fewer summands. The number of constraints (3.4b) in contrast remains unchanged, i.e., constraints still have to hold for every t. Moreover, note that internal calculations of intermediate variables (3.2) also remain unaffected, such that the system (3.4a)-(3.4e) still respects the chronology of the horizon. Thus, dependencies between model variables in time steps t and $t+1$ can still be accounted for, allowing for the consideration of storage relations.

A direct analysis of the left-hand side of (3.3a)-(3.3b) for an iterative adaptation of $\boldsymbol{A}$ would typically involve the consideration of a strongly fluctuating time series, such that the application of the algorithm from Schlegel et al. [142] might show poor performance. Thus, we propose the following decomposition of $\boldsymbol{A}$ introducing matrices $\boldsymbol{P}$ and $\boldsymbol{B}$:

$$\boldsymbol{A} = \boldsymbol{P}^{-1} \cdot \boldsymbol{B}.$$

Having electricity prices $\boldsymbol{c} = (c_1,\ldots,c_T)^T$, the permutation matrix $\boldsymbol{P} \in \mathbb{R}^{T\times T}$ is defined in a way that $\boldsymbol{P}\cdot\boldsymbol{c}$ furnishes a vector of electricity prices in descending order. As $\boldsymbol{P}$ is constant during the iterations, the remaining matrix $\boldsymbol{B} \in \mathbb{R}^{T\times\hat{T}}$ is to be adapted over the iterations. In $\boldsymbol{B}$, each row also has $\hat{T}-1$ zero entries and one entry equal to one. However, the rows are arranged in a way that blocks of column vectors with all entries equal to one are formed. The blocks are located from top left to bottom right. The sizes of the blocks thus determine how many of the original DoFs, i.e., how many intervals with similar prices, are represented by the same decision variables. Compared to the multiplication with $\boldsymbol{A}$ (i.e., the left-hand side of (3.3a)-(3.3b)), the series after multiplication with $\boldsymbol{B}$ will likely show substantially less fluctuations and is thus expected more suitable for the wavelet-based grid adaptation. The entire mapping procedure is illustrated in Figure 3.1 for a small example.

3.2.2. Wavelet-based analysis and iterative refinement of the time grid

We apply the algorithm from Schlegel et al. [142] (Figure 3.2) to iteratively refine the time grid, i.e., the first mapping in Figure 3.1 using the matrix $\boldsymbol{B}$. More precisely, this corresponds to deciding how many decision variables to use for representing the full time grid. Hence, starting in iteration 0 from an initial assignment matrix $\boldsymbol{B}^0$, the algorithm

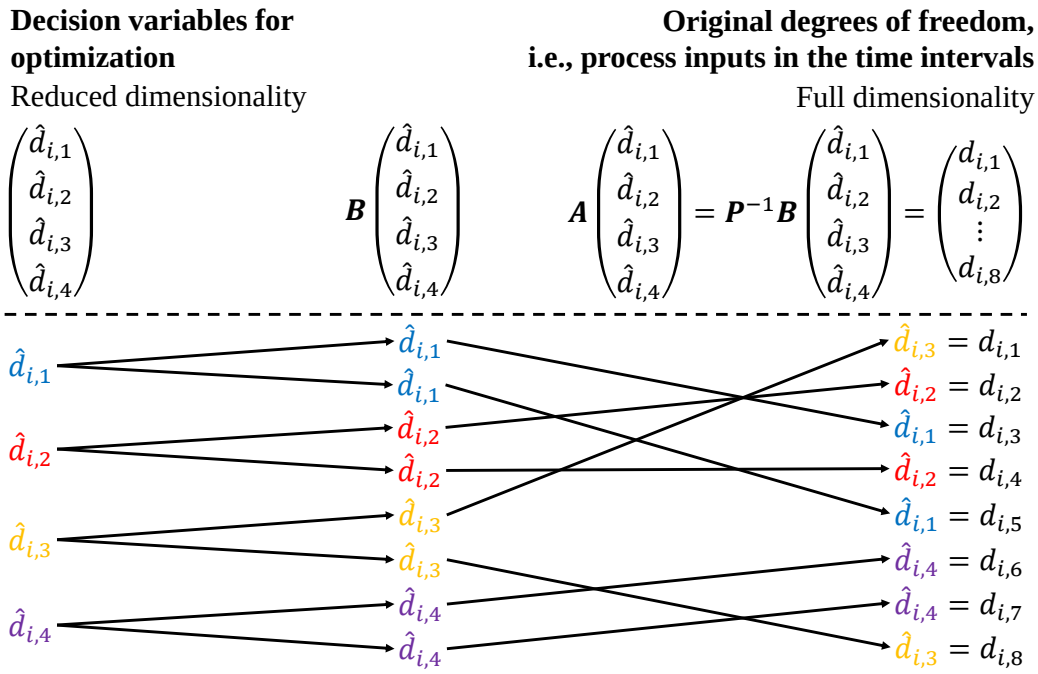

Figure 3.1.: Illustration of the assignment of one degree of freedom to multiple nonconsecutive intervals by the linear mapping procedure with $T = 8$ and $\hat{T} = 4$. The first multiplication with the matrix $\boldsymbol{B}$ is a mapping of a small number of decision variables to a larger number of degrees of freedom. The following multiplication with $\boldsymbol{P}^{-1}$ is a rearrangement. The first decision variable (blue) is used for the intervals with the highest and second highest electricity price ($t = 3, 5$). The second variable (red) is used for the intervals with the third and fourth highest price ($t = 2, 4$). Likewise, the third (yellow) and fourth (purple) variable are also assigned to two intervals ($t = 1, 8$ and $t = 6, 7$).

iteratively generates a sequence $\boldsymbol{B}^0, \boldsymbol{B}^1, \ldots, \boldsymbol{B}^s$ by solving (3.4a)-(3.4e). The corresponding solutions $(\hat{d}^{s*}_{i,1}, \ldots, \hat{d}^{s*}_{i,\hat{T}})^T$ and $(\hat{y}^{s*}_{j,1}, \ldots, \hat{y}^{s*}_{j,\hat{T}})^T$ are analyzed to determine the refined assignment matrix $\boldsymbol{B}^{s+1}$ for the next iteration until a desired stopping criterion is reached. The adaptations of the matrix $\boldsymbol{B}$ correspond to adding additional decision variables to the optimization problem (3.4a)-(3.4e), i.e., to inserting additional grid points, or to removing irrelevant ones, i.e., to deleting grid points.

The algorithm decides on whether to increase or decrease the number of optimization variables by analyzing the differences in the solution from the previous iteration, i.e., by analyzing $|\Delta\hat{d}^{s*}_{i,\hat{t}}| = |\hat{d}^{s*}_{i,\hat{t}} - \hat{d}^{s*}_{i,\hat{t}-1}|$. Easily spoken, if $|\Delta\hat{d}^{s*}_{i,\hat{t}}|$ is small, continuing to use individual optimization variables $\hat{d}^{s+1}_{i,\hat{t}}$ and $\hat{d}^{s+1}_{i,\hat{t}-1}$ in the next iteration is insignificant. In contrast, if $|\Delta\hat{d}^{s*}_{i,\hat{t}}|$ is large, it is worth investigating benefits from increasing the number of optimization variables to represent this part of the time grid.

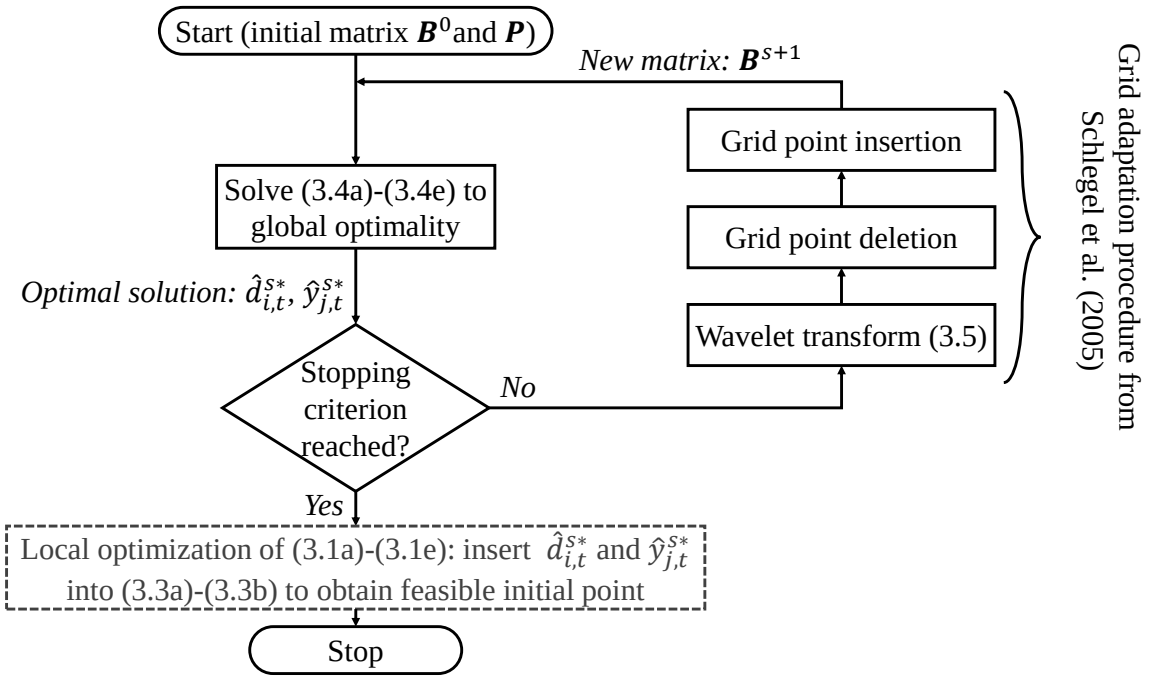

Figure 3.2.: Flowchart of the wavelet-based grid adaptation procedure. The last gray box corresponds to an optional a posteriori search.

In order to enable systematic decision criteria for grid point insertion and deletion, a wavelet transform in the so-called Haar basis is conducted, which we denote by $\boldsymbol{\mathcal{W}}_{\psi_{Haar}}(\cdot)$. For instance, if the continuous optimization variable $\hat{d}_{i,\hat{t}}$ is used to decide on refinements of the time grid, the corresponding wavelet transform of the optimized time series reads:

$$\boldsymbol{\mathcal{W}}_{\psi_{Haar}}\left(\boldsymbol{B}^s \cdot (\hat{d}_{i,1}^{s*}, \ldots, \hat{d}_{i,\hat{T}}^{s*})^T\right) = \boldsymbol{\lambda}_i^{s*}. \tag{3.5}$$

Thereby, we represent the entire optimized time series after the first linear mapping (cf. Figure 3.1) unambiguously by T coefficients $\lambda_{i,a,b}^{s*}$ with $a \in \{-1, \ldots, N-1\}$ and $b \in \{0, \ldots, 2^a - 1\}$, which we summarize in the vector $\boldsymbol{\lambda}_i^{s*}$. Formally, these coefficients allow to express a piecewise linear signal by a linear combination of wavelet basis functions $\psi_{a,b}$. An efficient procedure to obtain $\lambda_{i,a,b}^{s*}$, which is referred to as fast wavelet transform (FWT), can be found in the relevant literature [143].

The set of basis functions $\psi_{a,b}$ with nonzero coefficients $\lambda_{i,a,b}^{s*}$ is strongly related to the matrix $\boldsymbol{B}^s$ (cf. Appendix A.2). E.g., assuming pairwise different $\hat{d}_{i,\hat{t}}^{s*}$ in Figure 3.1, the wavelet representation would in general use four basis functions $\psi_{-1,0}$, $\psi_{0,0}$, $\psi_{1,0}$, and $\psi_{1,1}$ with nonzero coefficients. The coefficients of the wavelet transform (3.5) are analyzed to decide on grid point deletions and insertions. To this end, we appply the algorithm proposed by Schlegel et al. [142]. A detailed treatment can be found in their work. Consequently, we only review the key ideas. The required formulas as well as further informations are given in Appendix A.1.

The aforementioned decision whether to increase or decrease the number of optimization variables to represent a specific part of the time grid can now be made based on the absolutes of the wavelet coefficients $|\lambda_{a,b}^{s*}|$ instead of the differences $|\Delta\hat{d}_{i,\hat{t}}^{s*}|$. That is, comparably low $|\lambda_{a,b}^{s*}|$ correspond to the presence of optimization variables with low infor-

mational content that can be deleted from the optimization problem with minor influences on the objective value to be expected. In contrast, comparably high $|\lambda_{a,b}^{s*}|$ indicate that using additional optimization variables might enable significant benefits. Consequently, threshold values ϵ^d for deletion and ϵ^i for insertion are used and can be tuned to increase the performance. The so-called norm equality of the wavelet transform (cf. [144]) further allows for proper scaling, such that $0 \leq \epsilon^d, \epsilon^i \leq 1$ holds.

By performing grid point insertions and deletions, the relevant set of basis functions $\psi_{a,b}$ for the next iteration is identified. Based on this, the new mapping can be derived using the procedure given in Appendix A.2. Assuming that in the wavelet representation of the example in Figure 3.1, the coefficient $\lambda_{1,0}^{s*}$ would fulfill the criterion for insertion and $\lambda_{1,1}^{s*}$ the criterion for deletion, in the next iteration, there would be individual decision variables for the interval with highest ($t = 3$), with second highest ($t = 5$), with third highest ($t = 2$), and with fourth highest electricity price ($t = 4$). The remaining fifth decision variable would be assigned to all other intervals. This would correspond to using basis functions $\psi_{-1,0}$, $\psi_{0,0}$, $\psi_{1,0}$, $\psi_{2,0}$, and $\psi_{2,1}$. Note that in contrast to dynamic optimization problems, for which this procedure was originally proposed, there is a finest resolution for the optimization variables that gives exactly a one-to-one assignment between the decision variables for optimization and the DoFs in (3.3a)-(3.3b) with $T = \hat{T}$. Consequently, inserting grid points that correspond to using basis functions with an assignment of multiple optimization variables to the same interval (i.e., basis functions $\psi_{N,b}$, $\psi_{N+1,b}$, etc.) is not allowed.

Further note that the wavelet transform (3.5) requires $T = 2^N$ with $N \in \{1, 2, 3, \ldots\}$, i.e., scheduling horizons of $2, 4, 8, 16, \ldots$ intervals (e.g., hours). In the computational study, we confine to these cases. For different scheduling horizons, e.g., $T = 24$ (one day with hourly discretization), one could decompose the horizon into multiple subhorizons - in this case $T_1 = 16$ and $T_2 = 8$.

3.2.3. Properties of the algorithm

If the deletion of grid points is prohibited, the objective value in iteration $s+1$ is guaranteed to be at least as good as the objective value in iteration s. Thus, using $\epsilon^d = 0$ gives a nonincreasing sequence of objective values for any ϵ^i. If $\epsilon^i = 1$ is additionally chosen (i.e., selecting all possible grid points for insertion), the algorithm ultimately results in a one-to-one assignment in (3.3a)-(3.3b) with $\boldsymbol{B} \in \mathbb{R}^{T \times T}$. Consequently, the solution of (3.4a)-(3.4e) then becomes equivalent to the solution (3.1a)-(3.1e). If the subproblems in each iteration are solved to global optimality, the sequence of solution points converges to the global minimizer of (3.1a)-(3.1e) in this case.

Due to the computational complexity of the original problem, convergence of the optimization problems to solve to (3.1a)-(3.1e) is in general not desired. A reasonable selection of ϵ^d and ϵ^i thus should balance between exploration through increasing the number of optimization variables for improved objective values and maintenance of low dimensionality in the optimization problem to limit the computational burden. Furthermore, appropriate stopping criteria should be applied. A suitable criterion indicating near-optimal solution points furnished by the grid adaptation is the relative improvement in the objective function from one iteration to the next:

$$\Delta\Phi^{s*} = \frac{|\Phi^{s-1*} - \Phi^{s*}|}{|\Phi^{s*}|}.$$

If $\Delta\Phi^{s*}$ falls below a threshold, the algorithm is terminated. As a reasonable selection for the threshold value, we herein apply the relative optimality tolerance *optcr* for the solution of (3.4a)-(3.4e) for this purpose as well, i.e., the algorithm is finally terminated if $\Delta\Phi^{s*} < optcr$. Moreover, if the insertion step only proposes refinements that would lead to a finer resolution than a one-to-one assignment, the algorithm is also terminated.

In each iteration s, inserting the optimal solution $(\hat{d}^{s*}_{i,1}, \ldots, \hat{d}^{s*}_{i,\hat{T}})^T$ and $(\hat{y}^{s*}_{j,1}, \ldots, \hat{y}^{s*}_{j,\hat{T}})^T$ into (3.3a)-(3.3b) gives a feasible point for the optimization problem (3.1a)-(3.1e). Note that this point will in general not satisfy any optimality conditions. However, as the point is feasible and expected near-optimal if sufficiently many iterations have been performed, it represents a suitable initial point for a local search considering the full time grid of optimization problem (3.1a)-(3.1e). The search does not necessarily need to be performed among all optimization variables. For instance, performing the local search only on continuous variables that can be handled very efficiently in state-of-the-art local NLP solvers seems to be an appealing alternative. We acknowledge that there is no guarantee that the local solver will converge to the actual global minimum. However, a feasible point satisfying local optimality conditions and likely further improving the objective value can thereby be furnished requiring comparably low additional computational effort. Alternatively, the feasible point generated through (3.3a)-(3.3b) could also be used as initial point for a global solution of (3.1a)-(3.1e); this would be beneficial for global algorithms whose computational requirements strongly depend on having good initial guesses.

Various works demonstrate the successful practical application of the grid adaptation to dynamic optimization problems. In particular, authors have shown that the algorithm can identify high-quality solutions by using tailored discretizations that are comparable to the solutions when using very fine equidistant discretizations, but require substantially less computational time [142, 145, 146]. In particular, authors have also shown that an adequate balance between maintaining a low number of optimization variables and improving the objective function can be achieved by tuning the parameters and that few adaptation steps starting from coarse discretizations commonly suffice for obtaining near-optimal solutions. Nevertheless, the performance of the algorithm will certainly depend on the actual problem considered. As our algorithm essentially exploits the strong relation between instantaneous electricity prices and optimal values of the DoFs in the corresponding intervals, we expect the performance to be best if this relation is very distinct.

3.3. Case study 1: Scheduling of a compressed air energy storage

Although the present thesis is primarily concerned with industrial DSM, we herein consider the operational optimization of an energy storage system, which however shows strong similarities from a methodological point of view. Among the variety of opportunities for energy storage, CAES belongs to the few technologies that have been commercialized and are readily available at large scale [147]. In a CAES plant (Figure 3.3), electricity is bought at low prices. The energy is used to pressurize and store air in a cavern. In times of high

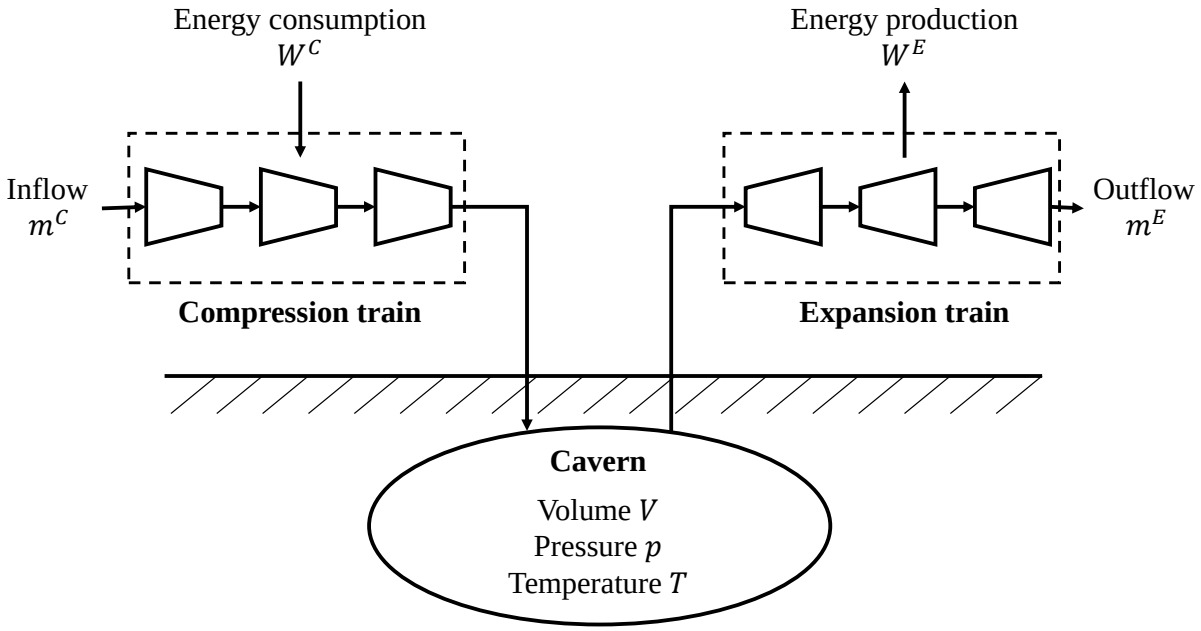

Figure 3.3.: Simplified flowsheet of a compressed air energy storage.

electricity prices, the cavern is discharged via turbines that power a generator. The concept relies on large price spreads to overcome the efficiency losses of the turbomachines [148]. Optimal scheduling is thus crucial for the economic performance of the plant [42].

Scheduling of a CAES plant represents a computationally challenging optimization problem, as the efficiency characteristics of turbomachines commonly show highly nonlinear behavior. Moreover, scheduling of CAES plants involves strong coupling between time intervals, in a sense that the pressure of the cavern in the interval t and thus also the necessary work for storing additional gas/the released work by expanding stored gas depends on the cavern pressure of the previous interval $t-1$.

In the following, we first present the considered set of equations for modeling a CAES plant with a distinct focus on the relations inside the turbomachines to highlight the capabilities of the presented algorithm to solve nonlinear scheduling problems involving time-coupling constraints. Afterwards, we briefly discuss how a reduced-space scheduling formulation in alignment with (3.1a)-(3.1e) arises from the modeling equations.

3.3.1. Mixed-integer nonlinear scheduling model

We aim at maximizing the profit, which corresponds to the difference between the revenues from electricity disposal and the costs for electricity purchase. The objective $\Phi(\cdot)$ to be minimized reads:

$$\Phi(W_t^E, W_t^C) = -\sum_{t=1}^{T} c_t \cdot (W_t^E - W_t^C),$$

where c_t is the instantaneous spot electricity price in hour t (without distinction of selling vs. purchasing price). Note that we omit the multiplication with the interval length $\Delta t =$

1 h for readability when presenting the equations. We use historic German day-ahead spot electricity prices from the end of January 2018, which are retrieved from EPEX SPOT SE (`www.epexspot.com`). We assume the prices to be known for the entire planning horizon. W_t^E and W_t^C denote the electricity production/consumption of the expansion/compression train in the corresponding hour. We use w_t^E and w_t^C to denote specific energies and m_t^E and m_t^C as masses that are processed by the expansion/compression train, i.e.,

$$W_t^E = m_t^E \cdot w_t^E, \quad \forall t$$
$$W_t^C = m_t^C \cdot w_t^C, \quad \forall t.$$

We further introduce binary variables y_t^E and y_t^C that indicate if the cavern is charged or discharged in the current interval. There is no sense in simultaneously charging and discharging if energy is bought and sold at the same price. However, there is the possibility of neither charging nor discharging during intervals with intermediate electricity prices. We thus require

$$y_t^E + y_t^C \leq 1, \quad \forall t.$$

Furthermore, we set limits for the mass flows only in case that the operating mode is active. Otherwise we force them to be zero:

$$y_t^E \cdot m_{min}^E \leq m_t^E \leq y_t^E \cdot m_{max}^E, \quad \forall t$$
$$y_t^C \cdot m_{min}^C \leq m_t^C \leq y_t^C \cdot m_{max}^C, \quad \forall t.$$

Moreover, a linear balance is used for modeling the accumulation of mass storage within the cavern

$$m_{t+1} = m_t + (m_t^C - m_t^E), \quad \forall t,$$

with m_t as the stored mass. Assuming air to be an ideal gas, the cavern pressure p_t can be calculated via

$$p_t = \frac{m_t \cdot R \cdot T}{M \cdot V}, \quad \forall t,$$

using parameter values $R = 8.314 \, \frac{\mathrm{J}}{\mathrm{mol \cdot K}}$ (ideal gas constant), $M = 0.028949 \, \frac{\mathrm{kg}}{\mathrm{mol}}$ (molar mass of air), $T = 323.15$ K (constant cavern temperature), and $V = 40,000$ m^3 (cavern volume). The cavern pressure is used to determine the pressure ratios of the expansion/compression train Π_t^E and Π_t^C. We account for pressure losses in the pipes through a constant pressure difference $\Delta p = 50,000$ Pa. $p_a = 100,000$ Pa further corresponds to the ambient pressure.

$$\Pi_t^E = \frac{p_t - \Delta p}{p_a}, \quad \forall t$$
$$\Pi_t^C = \frac{p_t + \Delta p}{p_a}, \quad \forall t$$

Having the respective pressure ratios as well as the processed amounts of air, the specific energies can be obtained from the efficiency maps of the turbomachines. We use ANNs to

represent the efficiency correlations. We use one efficiency map for the expansion and one for the compression train, each at optimized operating points. More precisely, the efficiency maps are obtained from individual operational optimizations of the turbine/compressor configuration by optimally setting pressure ratios between the single turbomachines that lead to the lowest energy consumption/highest energy production for given total flows and pressure ratios. This procedure is explained in detail in the work of Schweidtmann and Mitsos [134]. The ANN functions are denoted by $f^E_{\text{ANN}}(\cdot)$ and $f^C_{\text{ANN}}(\cdot)$. In both cases, feed-forward ANNs with hyperbolic tangent activation functions comprising one hidden layer and 25 neurons are used. Training of the ANNs is conducted using the Levenberg-Marquardt algorithm from the MATLAB R2019a Deep Learning Toolbox (Mathworks, Inc.). Overall, this results in adding two nonlinear functions for calculating the specific energies w^E_t and w^C_t:

$$\begin{aligned} w^E_t &= f^E_{\text{ANN}}(m^E_t, \Pi^E_t), \quad \forall t \\ w^C_t &= f^C_{\text{ANN}}(m^C_t, \Pi^C_t), \quad \forall t. \end{aligned}$$

Furthermore, we also represent the operating bounds of the turbomachines, such as surge limits, via ANNs. In total, there are four operating bounds for both the expansion and the compression train. In each case, two functions define maximum pressures (left and upper boundaries of the efficiency maps represented by functions $g^E_{\text{ANN},1}(\cdot)$, $g^E_{\text{ANN},2}(\cdot)$, $g^C_{\text{ANN},1}(\cdot)$ and $g^C_{\text{ANN},2}(\cdot)$) and two functions define minimum pressures (right and lower boundaries of the efficiency maps represented by functions $g^E_{\text{ANN},3}(\cdot)$, $g^E_{\text{ANN},4}(\cdot)$, $g^C_{\text{ANN},3}(\cdot)$ and $g^C_{\text{ANN},4}(\cdot)$). ANN-functions defining the operating bounds comprise one hidden layer with 12 neurons. We thus add the following set of constraints on the cavern pressure:

$$\begin{aligned} p_t &\leq g^E_{\text{ANN},1}(m^E_t) + (1 - y^E_t) \cdot M, \quad \forall t \\ p_t &\leq g^E_{\text{ANN},2}(m^E_t) + (1 - y^E_t) \cdot M, \quad \forall t \\ p_t &\geq g^E_{\text{ANN},3}(m^E_t) - (1 - y^E_t) \cdot M, \quad \forall t \\ p_t &\geq g^E_{\text{ANN},4}(m^E_t) - (1 - y^E_t) \cdot M, \quad \forall t \\ p_t &\leq g^C_{\text{ANN},1}(m^C_t) + (1 - y^C_t) \cdot M, \quad \forall t \\ p_t &\leq g^C_{\text{ANN},2}(m^C_t) + (1 - y^C_t) \cdot M, \quad \forall t \\ p_t &\geq g^C_{\text{ANN},3}(m^C_t) - (1 - y^C_t) \cdot M, \quad \forall t \\ p_t &\geq g^C_{\text{ANN},4}(m^C_t) - (1 - y^C_t) \cdot M, \quad \forall t. \end{aligned}$$

Choosing a sufficiently large constant M, which needs to be as high as the maximum cavern pressure that can be reached, ensures that the equations are always fulfilled in time intervals, where the operating mode is not active. Note that we thereby use big-M formulations for the disjunctions, although convex hull formulations generally lead to tighter relaxations [149]. Finally, we prevent emptying of the cavern over the horizon by demanding:

$$0 = \sum_{t=1}^{T} (m^C_t - m^E_t).$$

3.3.2. Degrees of freedom and reduced-space formulation

Revisiting the model equations in the previous subsection, one recognizes that if values are given for the process in- and outflow and the operating modes for all t, all other model equations can be evaluated in an explicit manner. The scheduling problem for the CAES plant can thus be written as a reduced-space formulation that is equivalent to (3.1a)-(3.1e) with $d_{1,t} = m_t^E$, $d_{2,t} = m_t^C$, $y_{1,t} = y_t^E$, $y_{2,t} = y_t^C$ corresponding to the process inputs, i.e., the DoFs, and expressing the objective and concatenating all inequality constraints as explicit functions thereof. Note however, that the objective function cannot be written as a sum over interval-specific objective functions like in (3.1a). We however emphasize that this does not affect the performance of the algorithm. A trivial solution to the reduced-space problem is selecting all DoFs to be zero ($m_t^E = m_t^C = y_t^E = y_t^C = 0$), resulting in an initial upper bound for the objective function $\Phi_0 = 0$.

For the application of the grid adaptation algorithm, we select the same matrix $\boldsymbol{A}^s$ for all process inputs within one iteration. In fact, this means that if using a grid with $\hat{T}$ grid points, there are always $4\hat{T}$ optimization variables. Due to the special structure of the problem, it appears to be appropriate to decide on grid point insertion and deletion (i.e., on increasing or decreasing the number of optimization variables) based on the analysis of the net flow $m_t^N = m_t^C - m_t^E$. This selection originates from the fact that in each interval either m_t^E or m_t^C is allowed to be nonzero. Furthermore, choosing more decision variables for optimization for y_t^E and y_t^C than for m_t^E and m_t^C is pointless. Although a reduction of the number of binary variables might be computationally highly beneficial, we consciously disregard the possibility to represent y_t^E and y_t^C by substantially fewer decision variables than m_t^E and m_t^C, as the refinement procedure originally proposed by Schlegel et al. [142] solely focused on continuous optimization problems. Refinement strategies explicitly targeting the efficient handling of binary variables by accounting for their characteristics (commonly, substantially less frequent "on/off" fluctuations) is thus a promising future work.

3.4. Performance assessment for the original algorithm

The reduced-space scheduling problem for the first case study (cf. Section 3.3) is implemented in our in-house open-source software for deterministic global optimization MAiNGO [150], which utilizes the concept of McCormick relaxations [135, 151, 152]. `C++` header files containing the model formulation used in MAiNGO can be downloaded from `https://git.rwth-aachen.de/avt.svt/public/glopse`. The reduced-space implementation substantially reduces the number of optimization variables when using ANNs, leading to significant improvements in the computational performance compared to commercial state-of-the-art solvers [134]. Note that we indeed also tested full-space formulations of the process model in GAMS (GAMS Development Corp.) and tried a solution using BARON [153]. However, the computational performance of the reduced-space formulation applied in MAiNGO was found superior, which is in good agreement with previous work [134]. Default settings are used within MAiNGO for the lower and the upper bounding procedure. Optimizations are stopped either if a relative optimality gap of $optcr = 1\%$ is undercut or if a CPU time limit of $5 \cdot 10^5$ s ($\sim$6 d) is exceeded. All calculations are conducted on an Intel® Xeon® CPU E5-2630 v3 @ 2.40GHz with 192 GB RAM. Calculations are performed using a single CPU of the server only and thus do not make use of

MAiNGO's parallel computing capabilities. We test three solvers in the upper bounding procedure (SLSQP [154], IPOPT [155], and KNITRO [156]). CPLEX (IBM Corporation) is always used as solver in the lower bounding.

The grid adaptation algorithm is always initialized with four equally-distributed grid points, giving four optimization variables per process input, i.e., the same DoFs are used for the 25% of intervals with the highest electricity prices, etc. The grid adaptation is stopped either if the relative change in the objective function between subsequent iterations is below the relative optimality gap, i.e., $\Delta\Phi^{s*} < optcr = 1\%$ or if the finest resolution is reached and no further grid points can be inserted. The applied threshold values $\epsilon^d = 10^{-4}$ and $\epsilon^i = 0.7$ have been found by trial and error and provide a good balance between improving the objective function and maintaining a low number of optimization variables.

3.4.1. Convergence behavior for increasing scheduling horizons

Figure 3.4 depicts the progress of the grid adaptation algorithm for varied lengths of the scheduling horizon with $N = 4, 5, 6, 7$, which corresponds to $T = 16, 32, 64, 128$ time intervals. Confining to an hourly discretization in accordance with the electricity market, we thus consider scheduling horizons between 16 h and ∼5 d. For $N = 4, 5$, we also show the results of a deterministic global search considering individual DoFs for each interval, i.e., considering the full time grid. As can be seen, a global search considering the full time grid becomes computationally prohibitive. In fact, no feasible point with objective value better than Φ_0 can be found within the time limit of $5 \cdot 10^5$ s for $N \geq 5$. Results depicted are produced using SLSQP as solver in the upper bounding. Using KNITRO as solver with mixed-integer capabilities gives no improvements. We thus conclude that even for moderately long scheduling horizons, the generation of favorable feasible points is not a simple task. Moreover, we expect that for $N = 5$ the current lower bound after exceeding the time limit still substantially underestimates the global minimum. Consequently, an application of deterministic global solution algorithms to scheduling problems of comparable complexity considering the full time grid will not lead to desired results, emphasizing the need for advanced solution algorithms.

As can be seen from Figure 3.4, the grid adaptation algorithm reaches one of the stopping criteria for all scheduling horizons within at most four iterations. Note that only for $N = 5$, the fourth iteration of the algorithm also results in an improvement of $\Delta\Phi^{3*} > 1\%$; however, no further iterations are performed as a further grid point insertion is only possible for insignificant parts of the time grid. The reason for this anomaly lies in a very distinct price peak in the underlying time series of electricity prices. Here, an exceptionally high electricity price occurs for only one hour of the horizon. As in the last iteration of the algorithm, individual DoFs are used exclusively in this singular hour, the improvement in the objective function is comparably high between the third and the fourth iteration.

Final solutions depicted in Figure 3.4 use 12, 14, 16, and 17 grid points, i.e., optimization variables per process input, to represent the full time grid. We emphasize that for the longest scheduling horizon of 128 hourly intervals, this corresponds to a reduction in the number of optimization variables by almost 90% compared to a solution approach assigning individual DoFs to each interval of the horizon. In the case of $N = 4$, i.e., $T = 16$, the objective value after the last iteration of the grid adaptation can be compared to the global optimum obtained by a deterministic search under consideration of the full time grid. The error between the objective values is substantially less than 1% and thus lower than the

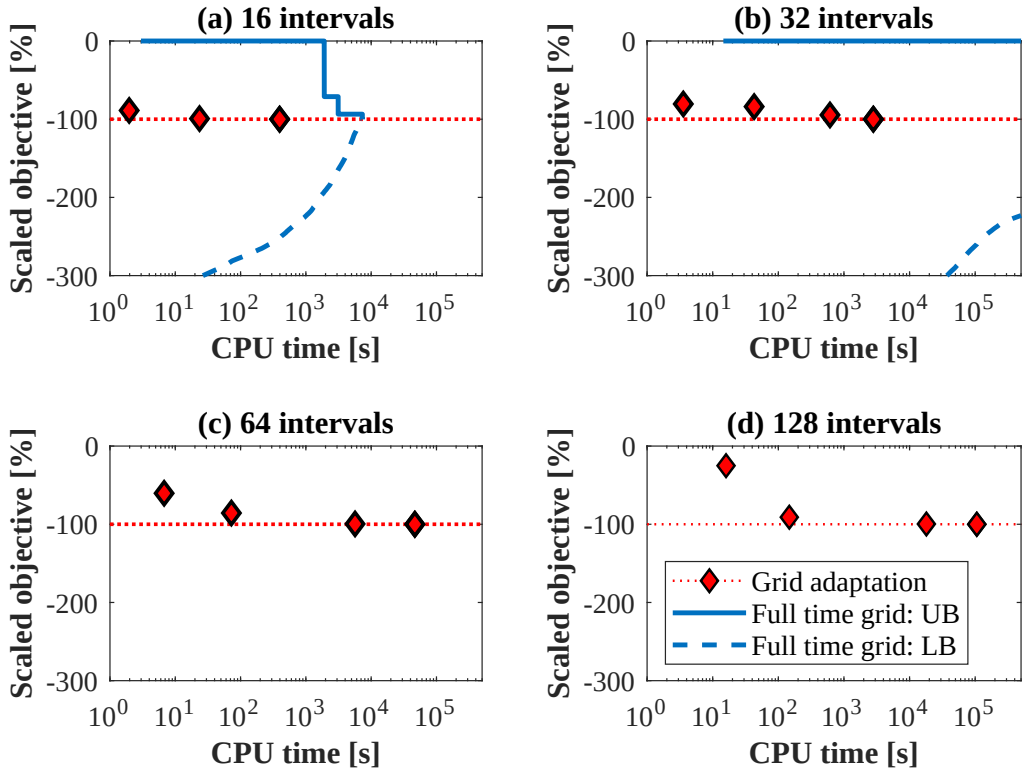

Figure 3.4.: Progress of the grid adaptation and comparison to a global optimization considering the full time grid. Results are depicted for scheduling horizons of $T = 16, 32, 64, 128$ intervals ((a) - (d)). Objective values are relative to the best feasible point furnished by the algorithm (dotted red line). Red diamonds ($\Diamond$) depict the outcome of the algorithm after each iteration. The blue solid lines give current upper bounds (UB) and dashed lines current lower bounds (LB) when solving the problem considering the full time grid.

remaining optimality gap of the global search. The comparison to the global optimum is not possible for $N \geq 5$, as the global solver does not converge within the time limit. However, also in these cases, our results indicate that solution points close to the global optima have been furnished. In particular, the sequences for $N = 6$ and $N = 7$ seem to converge with the improvements between subsequent iterations consistently decreasing, e.g., for $N = 7$, the improvement between the zeroth (four grid points) and the first iteration (eight grid points) is $\Delta\Phi^{1*} = 72\%$. Further increasing the number of grid points to 14 in the second iteration gives $\Delta\Phi^{2*} = 9\%$ and finally $\Delta\Phi^{3*} < 0.1\%$ using 16 grid points in the third one. Moreover, comparing the absolute values of the objective functions for different N allows for a rough plausibility check: if the fluctuations show repeating patterns with similar amplitudes, savings when doubling the horizon are expected twice as high. As can be seen in Table 3.1, our results are in good agreement with this rule of thumb. In fact, when comparing the objective values for $N = 5, 6$ with that for $N - 1$, the improvements are larger than a factor of two, which can be explained by larger spreads of the electricity prices. For $N = 7$, the factor is slightly lower than two; however, in this case, storage capacities become a relevant factor (cf. Figure 3.5), limiting the achievable savings.

Table 3.1.: Comparison of absolute savings for different lengths of the scheduling horizon and the fluctuations of the corresponding electricity price time series.

#Intervals [-]	Objective value [EUR]	Standard deviation of price time series [EUR/MWh]
$T = 16$ ($N = 4$)	-72.5	22.13
$T = 32$ ($N = 5$)	-156.5	25.14
$T = 64$ ($N = 6$)	-354.3	32.60
$T = 128$ ($N = 7$)	-684.5	32.57

We further tested the two optional a posteriori search strategies outlined in Subsection 3.2.3 on the original problem considering the full time grid for all scheduling horizons. The local search strategies ((i) local searches considering all optimization variables using mixed-integer solvers and (ii) local searches considering continuous optimization variables only using NLP solvers) both yielded slightly improved solutions within a short time. However, improvements are considered insignificant (less than 1% and thereby less than the final optimality gap of the global solution from the last iteration). Global solution strategies starting from the outcome of the grid adaptation algorithm did not result in any further improved feasible points within a CPU time limit of $5 \cdot 10^4$ s. We thus assume the feasible points that are obtained after stopping of the grid adaptation to be a close approximation of the global solution of the optimization problem considering the full time grid.

Looking at the solution times given in Figure 3.4 allows for estimating the savings in computational times when applying the grid adaptation. For $N = 4$, the algorithm furnishes a feasible point with an objective value extremely close to the global minimizer as stated above within less than a CPU time of 400 s by using 12 grid points, i.e. 12 optimization variables per process input, to represent the full time grid. The computational time for a deterministic global search considering the full time grid, i.e., 16 optimization variables per process input, is however ∼8,000 s and thus higher by a factor of ∼20. In the case of $N = 5$, the grid adaptation algorithm terminates after a CPU time of less than 3,000 s finally using 14 optimization variables per process input, compared to 32 optimization variables per process input when considering the full time grid. A deterministic global search on the original problem considering the full time grid is stopped after exceeding the CPU time limit of $5 \cdot 10^5$ s with a substantial remaining optimality gap. Time savings enabled by the grid adaptation thus already correspond to a factor of $\gg$100, i.e., to multiple orders of magnitude. Extrapolating the results presented in Figure 3.7 and discussed in Subsection 3.4.2 allows for approximating solution times when considering the full time grid. For instance, for $N = 5$, solution times in the order of $10^7 \ldots 10^8$ s are expected. Solution times for $N \geq 5$ thus appear prohibitive even when using multiple cores for parallel computing.

In Figure 3.5, the hourly operation schedule of the CAES plant after each iteration of the grid adaptation is given for a horizon of $T = 128$ intervals, i.e., $N = 7$. Note that this corresponds to a planning horizon of more than 5 days, which is a scheduling horizon with relevant length considering the limited predictability of future electricity prices. Recap that the depicted storage level is not a DoF of the optimization problem. In contrast, it

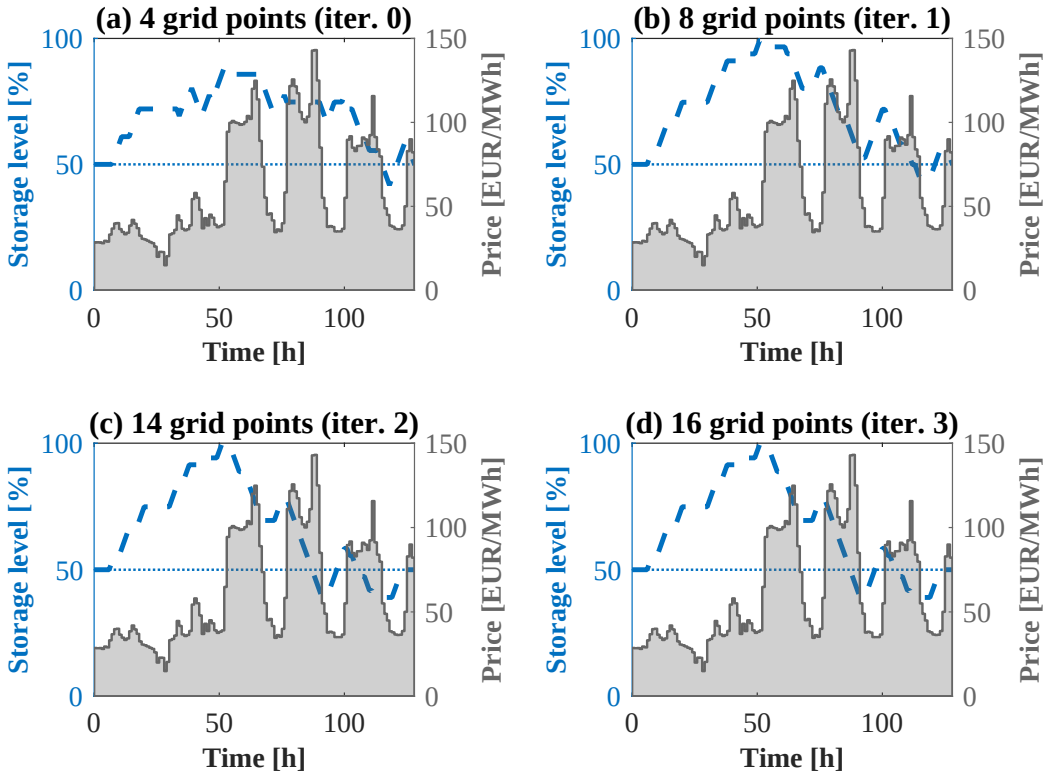

Figure 3.5.: Optimized operation schedule of the CAES plant after each of the four iterations of the grid adaptation ((a) - (d)). A scheduling horizon of $T = 128$ intervals is considered. Dashed thick blue lines depict the storage level (i.e., the cavern pressure), the dotted blue line depicts the production target. The electricity price profile for the case study is shaded in gray.

integrates the net mass flow and thereby the two continuous DoFs (cf. Subsection 3.3.1). Thus, the slope of the storage level in one interval corresponds to the instantaneous value of the DoFs. Besides, a great advantage of depicting the storage level lies in the fact that it is highly relevant for the limiting constraints in the operational optimization and thereby facilitates discussions of the schedule as will be explained in the following. Furthermore, a schematic depiction of the wavelet transform of the corresponding optimal solutions (cf. (3.5)) and thus of the numbers of grid points used as well as their distributions is given in Figure 3.6. Iteration 0 starts with using only four grid points, i.e., four optimization variables per process input, to represent the 128 hourly intervals of the scheduling horizon, i.e., the same DoFs are assigned to the 32 hourly intervals with the highest electricity prices, etc. The characteristics of the price profile are however such that the most intuitive solution, i.e., filling the cavern in the 32 intervals with lowest electricity prices and emptying the cavern in the 32 intervals with highest prices, is not feasible. This explains the very low profit in iteration 0 (cf. Figure 3.4). Consequently, additional grid points are inserted, causing an increased number of optimization variables in the subsequent iteration. Next, eight equally-distributed grid points are used that suffice to reproduce an intuitive behavior, i.e., a completely filled cavern after $\sim$50 h when electricity prices rise steeply. From then on, only minor adjustments of the charging/discharging strategy occur that slightly tend

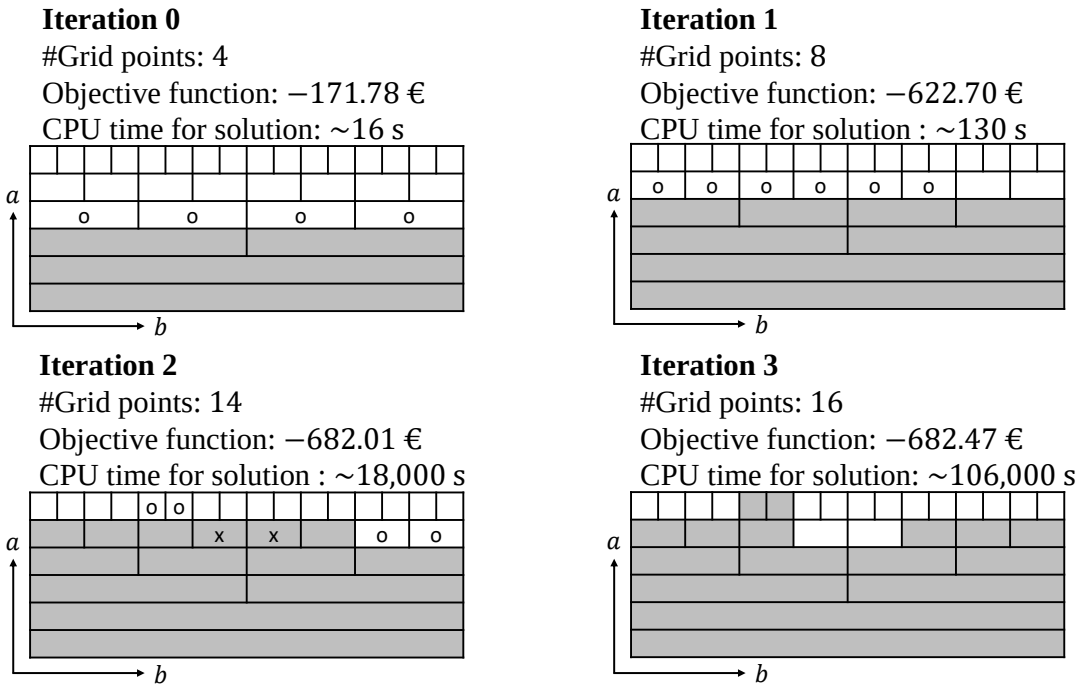

Figure 3.6.: Schematic depiction of the wavelet transform of the optimal solution (cf. (3.5)) during the iterations for a scheduling horizon of $T = 128$ intervals. Used (nonzero) wavelet coefficients are in gray, giving the number of grid points in the iterations. Wavelet coefficients selected for grid point insertion (increasing the number of optimization variables) are marked with (o). Wavelet coefficients selected for grid point deletion are marked with (x) (decreasing the number of optimization variables).

to a finer resolution when representing the parts of the time grid with higher prices. After four iterations, the finest resolution used in the linear mapping is an assignment of one DoF to four intervals with similar prices. The coarsest resolution is an assignment of one DoF to 16 intervals. It is assumed unlikely that the introduction of more than the final 16 grid points would lead to substantial improvements in the objective value. More generally speaking, we assume that 16 optimization variables per process input suffice for the exploitation of all relevant fluctuations of the electricity price in spite of the large range of prices within the considered horizon.

3.4.2. Computational performance and scaling of the algorithm

As computational times represent the main bottleneck for the solution of nonlinear scheduling problems with time-variable electricity prices on relevant planning horizons, we briefly discuss the scaling of the presented algorithm. Here, one has to differentiate between the consideration of larger planning horizons, i.e., more intervals, and the increase in the number of optimization variables per process input from adding additional grid points. Note that in solution approaches considering the full time grid, the latter directly follows from the former. However, in the case of the presented grid adaptation algorithm, the consideration of larger time horizons does not necessarily mean a proportional increase in the number of optimization variables. Solely the number of constraints that need to be satisfied, i.e., the dimensionality of (3.1b) in the general problem formulation from Section

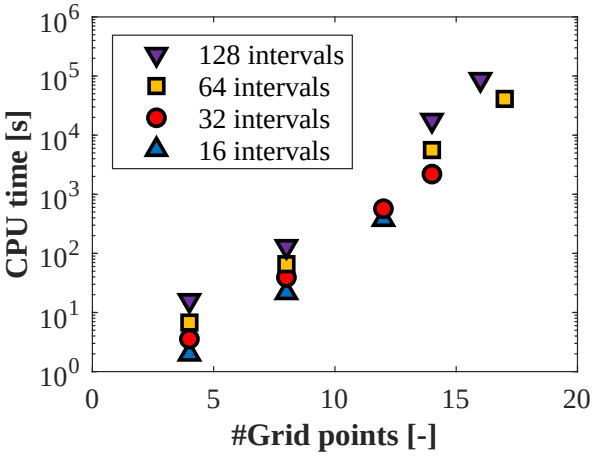

Figure 3.7.: CPU time per iteration of the grid adaptation algorithm as a function of the number of grid points, i.e., the number of optimization variables for each of the process inputs, in the corresponding iteration. Note that there are four process inputs. Symbols depict the varying scheduling horizons. Blue upwards-pointing triangles (△): 16 intervals, red circles (○): 32 intervals, yellow squares (□): 64 intervals, purple downwards-pointing triangles (▽): 128 intervals.

3.1, scales linearly with the length of the horizon.

In Figure 3.7, solution times for each individual iteration from Figure 3.4 are depicted. The plot uses a logarithmic scale for the CPU times. If the number of intervals considered remains unchanged, the solution times approximately lie on a straight line, corresponding to an exponential scaling of the solution time when increasing the number of optimization variables, i.e., the number of grid points, which matches the worst-case scaling of branch-and-bound solution approaches to deterministic global optimization. In fact, the introduction of four additional grid points results in an increase in the CPU time by a factor of ten. This finding is valid for all considered scheduling horizons, thus giving parallel lines for the respective horizons. Importantly, the distance between the lines for N and $N + 1$ is approximately constant for all N. This corresponds to a linear relation. More precisely, for a fixed number of grid points, the consideration of twice as many intervals ($N' = N + 1$) leads to a doubled solution time. This scaling behavior is considered an important feature of the grid adaptation, as it allows for the consideration of relevant planning horizons of several days to weeks even for complex nonlinear scheduling problems if an appropriate assignment of DoFs to time intervals is chosen and thus sufficiently few decision variables are exposed to the optimization algorithm.

3.5. Further development of the algorithm: Truncated wavelet transforms and sensitivity-based refinements

Despite the successful demonstration in the previous section, showing that near-optimal schedules can be furnished by applying tailored nonuniform time grids using only a limited number of optimization variables, the following issues remain:

- The refinement procedure presented above needs restrictions, limiting the possibilities

for inserting additional grid points in one iteration in order to guarantee that the derivation of the assignment of DoFs to intervals is unambiguous (cf. Appendices A.1 and A.2).

- Furthermore, the grid adaptation algorithm makes use of a systematic yet purely heuristic procedure for introducing additional optimization variables. Consequently, there is no quantitative measure for the expected improvement in the objective function through the refinement step providing a sound basis for the decision-making.

- As a consequence of the previous points, the algorithm's performance can be poor in cases where the optimal schedule is strongly governed by active ramping constraints, i.e., too many iterations are required to identify promising schedules, which involves the introduction of actually negligible optimization variables and thus unnecessarily prolong computational times.

Overcoming the shortcomings of the algorithm presented in Section 3.2 relies on two ideas that will be explained in the following: (i) we perform optimizations directly in the space of coefficients of a truncated wavelet transform of the time series of DoFs, making the derivation of a (tailored) time grid superfluous, and (ii) we decide on refinements based on the marginal effect of relaxing the constraints truncating the wavelet transform, i.e., the Lagrangian multipliers, which provides a quantitative measure for the improvements in the objective function from the introduction of additional optimization variables and thereby enables decisions based on sensitivity information instead of pure heuristics. We remark that the proposed refinement strategy using Lagrangian multipliers shows conceptual similarities with approaches for column generation that have recently also gained attention in the context of long-term operational decisions [157–159]; these approaches rely on initially exposing only a restricted number of variables to the optimizer and then iteratively increase this number based on Lagrangian multipliers as well.

3.5.1. Preliminaries

In this section, we address the special case of the general problem formulation (3.1a)-(3.1e) without binary variables, thus reading:

$$\min_{d_{i,t}} \ \Phi(d_{i,1},\dots,d_{i,T}) = \sum_{t=1}^{T} \phi_t(d_{i,t}) \tag{3.6a}$$

$$\text{s.t.} \ \ 0 \leq g_{l,t}(d_{i,1},\dots,d_{i,t}), \ \ \forall l,t \tag{3.6b}$$

$$0 = h_m(d_{i,1},\dots,d_{i,T}), \ \ \forall m \tag{3.6c}$$

$$d_i^l \leq d_{i,t} \leq d_i^u, \ \ \forall i,t. \tag{3.6d}$$

Note that the limitation to purely continuous problems is not as restrictive as it appears: in fact, binary variables often solely stem from piecewise linearizations of nonlinear functions (cf. [14]). Again, we confine to cases with horizons lengths of $T = 2^N,\ N \in \{1,2,3\dots\}$ for the ease of presentation. Moreover, we continue to use the permutation procedure using

matrix $\boldsymbol{P} \in \mathbb{R}^{T \times T}$ that orders the time series of DoFs according to the electricity prices, i.e.,

$$(d_{i,1}, \ldots, d_{i,T})^T = \boldsymbol{P}^{-1} \cdot (\tilde{d}_{i,1}, \ldots, \tilde{d}_{i,T})^T \tag{3.7}$$

with $\tilde{d}_{i,1}$ denoting the value of the process variable d_i in the interval with highest electricity price. Furthermore, we introduce the inverse Haar-wavelet transform $\boldsymbol{\mathcal{W}}^{-1}_{\psi_{Haar}}(\cdot)$. In particular, we use this to recover the time series of permuted DoFs $(\tilde{d}_{i,1}, \ldots, \tilde{d}_{i,T})^T$ from the coefficients of its wavelet transform, denoted by $\lambda_{i,a,b}$ and summarized in vector $\boldsymbol{\lambda}_i$, i.e.,

$$(\tilde{d}_{i,1}, \ldots, \tilde{d}_{i,T})^T = \boldsymbol{\mathcal{W}}^{-1}_{\psi_{Haar}}(\boldsymbol{\lambda}_i). \tag{3.8}$$

3.5.2. Dimensionality reduction through truncated wavelet transforms

In Section 3.2, we use the information contained in the coefficients of the wavelet transform to iteratively derive a tailored time grid for the scheduling problem, assigning one DoF to multiple generally nonconsecutive intervals. In contrast, we herein propose an alternative procedure making the derivation of a time grid superfluous in the case of scheduling problems with only continuous decision variables. For this purpose, we insert both the permutation (3.7) and the inverse Haar-wavelet transform (3.8) into the optimization problem (3.6a)-(3.6d), leading to an equivalent reduced-space reformulation exposing the wavelet coefficients $\boldsymbol{\lambda}_i$ as only variables to the optimizer:

$$\min_{\boldsymbol{\lambda}_i} \quad \Phi^{\psi}(\boldsymbol{\lambda}_i) \tag{3.9a}$$

$$\text{s.t.} \quad 0 \leq g^{\psi}_{l,t}(\boldsymbol{\lambda}_i), \quad \forall l, t \tag{3.9b}$$

$$0 = h^{\psi}_{m}(\boldsymbol{\lambda}_i), \quad \forall m \tag{3.9c}$$

$$\begin{pmatrix} d_i^l \\ \vdots \\ d_i^l \end{pmatrix} \leq \boldsymbol{\mathcal{W}}^{-1}_{\psi_{Haar}}(\boldsymbol{\lambda}_i) \leq \begin{pmatrix} d_i^u \\ \vdots \\ d_i^u \end{pmatrix}, \quad \forall i. \tag{3.9d}$$

Here, we formulate the objective as well as all constraints as explicit functions $\Phi^{\psi}(\cdot)$, $g^{\psi}_{l,t}(\cdot)$, and $h^{\psi}_{m}(\cdot)$ of the coefficient vector $\boldsymbol{\lambda}_i$.

As the inverse Haar-wavelet transform $\boldsymbol{\mathcal{W}}^{-1}_{\psi_{Haar}}(\cdot)$ corresponds to a matrix multiplication by definition, i.e., each value of the series $(\tilde{d}_{i,1}, \ldots, \tilde{d}_{i,T})^T$ is a linear combination of $2^N + 1$ coefficients, we do not expect a substantial deterioration of the computational performance of optimization algorithms if solving the formulation (3.9a)-(3.9d) instead of (3.6a)-(3.6d). Moreover, optimizations in the space of wavelet coefficients enable an exploitation of their physical meaning. In particular, as the coefficient $\lambda_{i,-1,0}$ corresponds to the scaled mean of the transformed time series, summation constraints on $(d_{i,1}, \ldots, d_{i,T})^T$ that are commonly used to guarantee reaching desired production targets in scheduling problems can be equivalently reformulated by fixing the coefficient $\lambda_{i,-1,0}$:

$$\sum_{t=1}^{T} d_{i,t} = \gamma \Leftrightarrow \lambda_{i,-1,0} = \frac{\sqrt{2}^{\log_2 T} \cdot \gamma}{T}.$$

Using the scheduling formulation that considers the wavelet coefficients as optimization variables, i.e., (3.9a)-(3.9d), dimensionality reductions correspond to truncations of the wavelet transform, essentially taking up a key idea from image compression using wavelet transforms [160]. Consequently, we only allow a subset of all coefficients denoted by U_i to be nonzero. Coefficients $\lambda_{i,a,b},\ (a,b) \notin U_i$ are set to zero and thereby no longer treated as optimization variables, adding (3.10) to the scheduling problem (3.9a)-(3.9d).

$$\lambda_{i,a,b} = 0, \quad \forall i, (a,b) \notin U_i \tag{3.10}$$

We finally highlight that for any U_i, any feasible point of the scheduling problem (3.9a)-(3.10) becomes a feasible point of the original problem (3.6a)-(3.6d) if transformed back using (3.8) and (3.7).

3.5.3. Sensitivity-based refinement strategy

We target the development of an algorithm for the efficient iterative refinement of the set of used, i.e., not-truncated, wavelet coefficients U_i, making the solution of (3.9a)-(3.10) converge to the solution of (3.6a)-(3.6d) with a substantially smaller number of optimization variables. To this end, we herein propose to decide on adding a specific truncated pair (a,b) to the set U_i^{s+1} for the next iteration $s+1$ based on the marginal effect of relaxing the constraints (3.10) in the current iteration s (Figure 3.8), thus exploiting the relation:

$$\left.\frac{\partial \Phi^{\psi,s*}}{\partial \lambda_{i,a,b}}\right|_{\lambda_{i,a,b}=0} = \mu_{i,a,b}^{s*}, \quad \forall i, (a,b) \notin U_i \tag{3.11}$$

with $\Phi^{\psi,s*}$ being at least a local and preferably the global minimizer of (3.9a)-(3.10) and $\mu_{i,a,b}^{s*}$ denoting the Lagrangian multipliers of (3.10). Following this interpretation, we add those truncated pairs (a,b) to U_i^{s+1} where the (local) sensitivities $|\mu_{i,a,b}^{s*}|$ of the objective function with respect to the value of the coefficient are largest. Thereby, (3.11) provides a quantitative measure for comparing the benefits in terms of objective value from adding different coefficients to the wavelet representation, thus overcoming an important weakness of the original algorithm as discussed above.

As a potential further development of the algorithm in the future, it might be promising to investigate benefits from using higher-order sensitivity information, such as a quadratic approximation of the improvements in the objective function. Along these lines, there would also arise opportunities for an optimization-based selection of truncated pairs (a,b) to add to U_i^{s+1}. That is, having an accurate approximation would allow for constructing subproblems for determining the truncated pairs (a,b), whose utilization would lead to the maximum improvement in the objective function, which could improve the convergence behavior.

Considering decisions on deletion of insignificant optimization variables, i.e., determination of pairs (a,b) to be truncated, we continue to apply the criterion proposed by Schlegel et al. [142] and applied in Section 3.2. That is, we continue to decide on deletions by comparing the absolute values of used coefficients to a relative threshold ϵ_i^d, with index i indicating that individual thresholds can be used for each DoF. Thus, if $\left|\lambda_{i,a,b}^{s*}\right| < \epsilon_i^d \cdot \|\boldsymbol{\lambda}_i^{s*}\|_2$, the pair (a,b) is removed from U_i^{s+1}. We emphasize that this criterion is already based on sound considerations. In particular, using a threshold value for determining pairs (a,b)

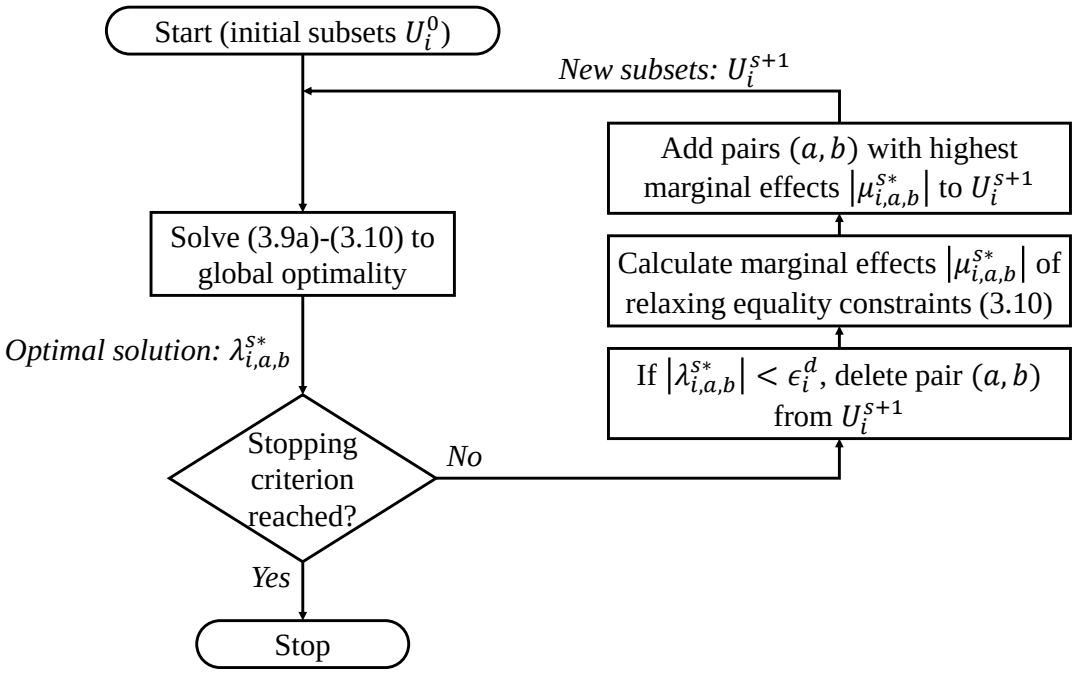

Figure 3.8.: Flowchart of the algorithm for sensitivity-based refinements of truncated wavelet transforms.

to be truncated is reasoned by the scaling of wavelet basis functions that ensures a norm equality [144]:

$$\left\| (\tilde{d}_{i,1}^{s*}, \ldots, \tilde{d}_{i,T}^{s*})^T \right\|_2 = \|\boldsymbol{\lambda}_i^{s*}\|_2 ,$$

which implies that truncating small absolute wavelet coefficients has only minor effects on the time series of process variables and thereby will also only insignificantly impair objective values.

We emphasize that in contrast to the heuristic procedure for grid point insertion (cf. Subsection 3.2.2), the sensitivity-based refinement procedure described in this section does not need to obey certain rules; instead, any refinement step is possible at any time. In particular, this enables the use of wavelet coefficients $\lambda_{i,a,b}^s$ of a high level of detail, i.e., large a, without previously exploring all levels below. Thereby, we are able to circumvent additional iterations of the algorithm and avoid the introduction of insignificant optimization variables, which otherwise unnecessarily prolongs the computational times for identification of a suitable approximation. Anticipating the presentation of the results below, this allows for an efficient treatment of cases, in which optimal schedules are strongly governed by active ramping constraints.

Finally, the further developed algorithm still relies on heuristic stopping criteria for terminating the refinement procedure, such as a threshold on the relative improvement in objective values between consecutive iterations (cf. Subsection 3.2.3) or simply a maximum number of used optimization variables. We again emphasize that these criteria are required in general to indicate that sufficiently near-optimal solutions have been furnished. Note that this is reasoned by the absence of global lower bounds, i.e., from solving relaxations of the problem considering an untruncated wavelet transform, in the general case, and thus

does not involve a certificate of optimality. Moreover, we also highlight that the additional discussions from Subsection 3.2.3 concerning the convergence of the algorithm, its tuning, and the opportunities for a posteriori searches remain valid. In other words, the extended algorithm presented herein does not change the fundamentals of the algorithm proposed in Section 3.2; it is rather meant to enhance the performance for a special case of optimization problems.

3.6. Case study 2: Scheduling of a flexible air separation process

Due to their high electricity consumption as well as inherent possibilities for liquid product storage, scheduling of cryogenic ASUs is one of the most vibrant research fields in the context of DSM [4]. We herein focus on an industrial-scale ASU producing both high-purity liquid nitrogen (LIN) and oxygen (LOX), which is stored in buffer tanks and distributed to off-site customers. The considered process configuration has been tailored to high load flexibility and is described in detail in literature [16, 101]. Most importantly, the process configuration comprises an integrated liquefaction cycle enabling a high ratio of storable liquid products. Moreover, the configuration allows for liquid-assisted operation, where stored cryogenic LIN is fed back to the distillation section. Thereby, the internal column flows and the separation performance can be maintained in the desired operating range, although substantially reducing the power consumption of the process and hence the liquefaction capacity.

The resulting wide operating range makes the considered process configuration on the one hand promising for performing DSM with profitable capabilities for the exploitation of spot electricity market price spreads on short to medium time-scales (e.g., between day and night or between working days and weekends). However, on the other hand, the turbomachines applied in ASUs are commonly characterized by narrow low-loss operating ranges and significant off-design efficiency losses if leaving these ranges, resulting in nonlinear process characteristics (cf. Figure 3.9). Consequently, scheduling of the considered process has to be conducted carefully by accounting for its nonlinearities as precisely as possible.

3.6.1. Nonlinear scheduling model

We herein consider the scheduling of the described ASU subject to time-variable hourly day-ahead spot market prices. We assume that all electricity is purchased at this market and that prices for the entire planning horizon are known. Price data is retrieved from EPEX SPOT SE (`www.epexspot.com`) and corresponds to historic German day-ahead prices from end of September/beginning of October 2018. Having fixed cumulative production targets, the objective function to be minimized $\Phi(\cdot)$ corresponds to the costs for electricity purchase, in accordance with the formulation in (3.6a), and is given by:

$$\Phi(LIN_t, LOX_t) = \sum_{t=1}^{T} \phi_t(LIN_t, LOX_t) \quad \text{with} \quad \phi_t(LIN_t, LOX_t) = c_t \cdot P(LIN_t, LOX_t)$$

where c_t denotes the day-ahead price for hour t and $P(\cdot)$ the instantaneous power consumption given as a function of the production rates LIN_t and LOX_t. Note that we again

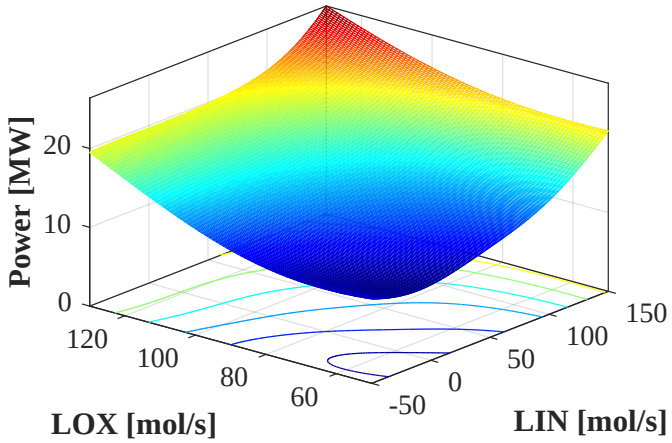

Figure 3.9.: Operating range and power consumption characteristic of the considered ASU configuration comprising an integrated liquefaction cycle and allowing for liquid-assisted operation. The power consumption is calculated using an artificial neural network fitted to optimized steady-state operating points using a rigorous process model from literature [101].

omit the multiplication with the interval length $\Delta t = 1$ h for readability. Furthermore, we impose operating limits on the two production rates:

$$\begin{aligned} -50 \text{ mol/s} \leq LIN_t \leq 150 \text{ mol/s}, \quad \forall t \\ 50 \text{ mol/s} \leq LOX_t \leq 130 \text{ mol/s}, \quad \forall t. \end{aligned}$$

Note that in the inequalities above, the negative lower bound for the LIN production corresponds to the aforementioned possibility to feed back stored LIN from the tank. In order to ensure reaching the production targets, we require the average production rates to equal the nominal production of $LIN_{\text{nom}} = LOX_{\text{nom}} = 120$ mol/s in each case, thus demanding:

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^{T} LIN_t = LIN_{\text{nom}} \\ \frac{1}{T} \sum_{t=1}^{T} LOX_t = LOX_{\text{nom}}. \end{aligned}$$

The power consumption of the process $P_t(LIN_t, LOX_t)$ is calculated using an ANN as nonlinear surrogate model, resulting in the surface given in Figure 3.9. Note that this surface is nonconvex, motivating the use of deterministic global optimization methods for solving the scheduling problems in the following. In particular, one finds univariate functions (fixed LIN and varying LOX and vice versa) with flat plateaus comprising inflection points and partially even with local minima. For generating the training data, we herein used steady-state optimizations of the rigorous dynamic process model from [101],

extended by the consideration of off-design efficiency losses in the turbomachines. Using the MATLAB R2019a Deep Learning Toolbox (MathWorks, Inc.), the ANN is fitted to 184 optimized stationary operating points, which are equally distributed across the entire operating range, and comprises one hidden layer with 6 neurons and hyperbolic tangent activation functions. The ANN fits the training data with a maximum relative error below 2% and a mean percentage error of ∼0.5%. The use of a nonlinear surrogate model herein corresponds to an advancement compared to the vast majority of previous literature that either use simplified linear characteristics [5–7, 30], which potentially introduces significant approximation errors, or rely on the identification of piecewise-linearized surrogate models [8, 17, 64–67], which potentially requires large numbers of disjunctions for being accurate. We further assume that the process can be operated in a continuous way within its operating range. In this case, we do not need additional binary variables indicating active modes. In particular, we thereby also omit possibilities for temporary shut-downs of the entire plant, which have been considered in most of the referenced works. However, in the case of the herein considered process configuration, opportunities for temporary shut-downs are of limited additional economic potential considering the existing possibilities for reductions of the power consumption by >50% (cf. Figure 3.9) on the one hand and large durations with off-spec production during shut-down or start-up periods on the other hand [161].

As ASUs are characterized by large liquid volumes [86, 87] that lead to time constants in a similar order as the time scales of spot electricity markets [10], process dynamics further need to be accounted for in the scheduling [78], as it is also discussed in more detail in Subsection 2.1.2. For this purpose, we impose constraints on the change in production rates, as widely done in scheduling practice to ensure that load changes can be tracked without violating product requirements (cf. [8, 17, 30]). Revisiting the results presented in [16] concerning the process dynamics, the following constraints appear suitable, so that common industrially applied control systems would be able to track all scheduled set-point changes.

$$
\begin{aligned}
&-15\ \text{mol/s} \leq LIN_t - LIN_{t-1} \leq 15\ \text{mol/s}, \quad \forall t \\
&-15\ \text{mol/s} \leq LOX_t - LOX_{t-1} \leq 15\ \text{mol/s}, \quad \forall t.
\end{aligned}
$$

Note that changing the control system to a more advanced one, such as NMPC, could involve substantially loosened ramping limits [89, 117] and thereby increase potential savings, which should be thoroughly investigated in the future. Moreover, we acknowledge that numerous works argue to construct dynamic surrogate models [10, 80–82] for explicitly capturing the process dynamics and demonstrate their successful application to the scheduling of ASUs [14, 18, 88]. This discussion is however out of the scope of this chapter. Nevertheless, we emphasize that the presented approach is conceptually also applicable to scheduling problems involving dynamic process models as will be briefly discussed in Section 3.8, but leave this for future work.

Finally, we consider that the storable amount of product is limited. In particular, we restrict the cumulative over-/underproduction, so that the nominal production can be kept up for not more than 12 h by using the available storage capacities, thus requiring:

$$-6 \cdot LIN_{\text{nom}} \leq \sum_{\tau=1}^{t} (LIN_\tau - LIN_{\text{nom}}) \leq 6 \cdot LIN_{\text{nom}}, \quad \forall t$$

$$-6 \cdot LOX_{\text{nom}} \leq \sum_{\tau=1}^{t} (LOX_\tau - LOX_{\text{nom}}) \leq 6 \cdot LOX_{\text{nom}}, \quad \forall t.$$

3.6.2. Degrees of freedom and reduced-space formulation

The entire scheduling problem presented in the previous subsection can be formulated equivalently to (3.6a)-(3.6d) by selecting $d_{1,t} = LIN_t$ and $d_{2,t} = LOX_t$, making the algorithm proposed in Section 3.5 applicable to this problem. Consequently, we expose the coefficients $\lambda_{1,a,b}$ and $\lambda_{2,a,b}$ of the wavelet transforms of the permuted time series of production rates as optimization variables to the solver. The objective and every presented constraint are formulated as explicit functions thereof by using (3.7) and (3.8). Note that according to the discussion about the physical meaning of the wavelet coefficients, the constraints that ensure reaching the production target can be reformulated fixing the first wavelet coefficients to $\lambda_{1,-1,0} = \sqrt{2}^{\log_2 T} \cdot T \cdot LIN_{\text{nom}}$ and $\lambda_{2,-1,0} = \sqrt{2}^{\log_2 T} \cdot T \cdot LOX_{\text{nom}}$.

3.7. Performance assessment for the extended algorithm

As in Section 3.4, the reduced-space scheduling problem presented in the previous section is implemented and solved using our in-house open-source software for deterministic global optimization MAiNGO [150] based on McCormick relaxations [135, 151, 152] - again exploiting the reduction in the number of optimization variables when using ANNs in reduced-space formulations, which lead to significant improvements in the computational performance [134]. `C++` header files containing the model formulation used in MAiNGO can be downloaded from `https://git.rwth-aachen.de/avt.svt/public/glopse`. We use IPOPT [155] as solver in the upper bounding procedure and CPLEX (IBM Corporation) as solver in the lower bounding. We apply a relative optimality tolerance of 2% in accordance with the maximum regression error of the ANN. Furthermore, we impose a CPU time limit of 10^5 s (∼28 h). In order to get the values of the Lagrangian multipliers of constraints (3.10), we perform an additional a posteriori local search starting from the best feasible point using IPOPT. All calculation are conducted on one core of an Intel® Xeon® CPU E5-2630 v3 @ 2.40GHz with 192 GB RAM.

Moreover, we apply threshold values $\epsilon_1^d = \epsilon_2^d = 10^{-3}$ for truncating coefficients of the wavelet transform in the optimization problem in the next iteration. Afterwards, we add those pairs (a, b) of truncated wavelet coefficient to the set U_i^{s+1} for the next iteration, where the marginal effect of relaxing (3.10), i.e., the absolute of the Lagrangian multiplier, is highest. This is done until the total number of optimization variables between two iterations is increased by four. Note that this procedure is primarily chosen for illustration

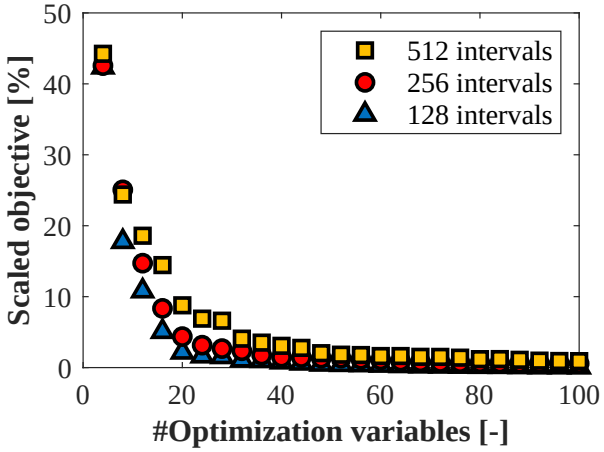

Figure 3.10.: Optimal objective value in the course of the algorithm as a function of the number of optimization variables in the corresponding iterations. Values are scaled between a constant operation (100%) and the benchmark from multi-start local optimization when using the untruncated wavelet transform (0%). Symbols depict the varying scheduling horizons. Blue upwards-pointing triangles (△): 128 intervals, red circles (○): 256 intervals, yellow squares (□): 512 intervals.

purposes, as it allows for displaying the effects of adding additional optimization variables in a high resolution. With regard to a practical application of the algorithm, other criteria are possible as well as previously discussed.

3.7.1. Convergence behavior for increasing scheduling horizons

In this section, we analyze the course of objective and optimal variable values over the iterations of the algorithm. Note that global deterministic solutions when considering the full dimensionality, i.e., using an untruncated wavelet transform, are computationally intractable for the applied scheduling horizons. Thus, a stochastic global approach relying on multi-start local searches is applied to obtain benchmark values for using the untruncated transforms. This choice is justified by the good performance of local searches with IPOPT performed during pre-processing in MAiNGO, i.e., in contrast to the results presented in Section 3.4, no improvements of the upper bounds have been observed within the CPU time limit in any case. Consequently, we herein scale all objective values for comparison purpose between the benchmark from multi-start local optimizations using the untruncated transform (0%) and the objective value of a constant production using only the fixed coefficients $\lambda_{1,-1,0}$ and $\lambda_{2,-1,0}$ (100%). Consequently, the scaled objective values indicate which percentage of the maximum achievable cost savings are discarded by truncating the wavelet transform.

In Figure 3.10, we show the course of the scaled objective value over the iterations of the algorithm for the different horizon lengths of $T = 128, 256, 512$ hourly intervals. The algorithm furnishes a series of feasible solutions with objective values that decrease monotonically and converge towards the benchmark values from multi-start local optimizations, i.e., the presumable global optima using the untruncated wavelet transform, for all considered horizon lengths. Moreover, it can be seen that high percentages of the maximum

Table 3.2.: Comparison of the minimum numbers of required optimization variables for reaching a desired scaled objective value in case of varying horizon lengths.

#Intervals [-]	**#Opt. variables for desired objective value [-]**			
	10%	**5%**	**2%**	**1%**
$T = 128$	13	17	22	35
$T = 256$	15	20	35	62
$T = 512$	20	31	49	92

achievable cost savings can be reached even when using substantially truncated wavelet transforms. Table 3.2 gives a comparison of the minimum numbers of optimization variables that are required for reaching specific desired percentages of the maximum savings. Thereby, we illustrate how the algorithm circumvents a linear scaling in the number of optimization variables with the number of considered scheduling intervals by confining to low-dimensional yet accurate approximations that suffice for exploiting large proportions of the saving potentials. For instance, if comparing the 128 h scheduling horizon to the 512 h, i.e., four times as many intervals, we find that only ∼50% more optimization variables are required if confining to 90% of the maximum savings. Likewise, ∼80% more optimization variables are required for 95% of the maximum savings and 2.2 times as many for 98%. Even if demanding the exploitation 99% of the saving potentials, 2.6 and thus still substantially less than four times as many optimization variables are required.

In addition to the discussion of objective values, the courses of the optimal variable values over the iterations of the algorithm, i.e., the optimized production schedules, are given in Figure 3.11. It can be seen that the algorithm furnishes feasible solutions that converge to the benchmark time series considering the full dimensionality, i.e., an optimal untruncated wavelet transform, even in cases where the optimum is strongly governed by active ramping constraints. Furthermore, we notice that only a very limited number of nonzero wavelet coefficients, i.e., optimization variables, suffices for detailed reproduction of the most important fluctuations of the optimum when considering the full dimensionality. In particular, the iterations in Figure 3.11 correspond to reductions in the number of optimization variables compared to an untruncated wavelet transform (d) of 93.75% (a), 87.5% (b), and 81.25% (c) respectively, demonstrating the ability of accurately reproducing large and strongly fluctuating time series from only a few important wavelet coefficients.

Finally, in Figure 3.12, we compare the convergence behavior of the sensitivity-based refinement strategy proposed in Subsection 3.5.3 to the systematic yet heuristic procedure from Subsection 3.2.2 for a horizon length of $T = 128$ hourly intervals. As can be seen, the ability to identify suitable low-dimensional approximations that take the characteristics of the original problem, such as active ramping constraints, into account is substantially enhanced by the sensitivity-based refinement procedure. Thus, if we instead apply the algorithm from Subsection 3.2.2, the performance is poorer, so that similarly close approximations of the optima when considering the full dimensionality in terms of objective and variable values require more optimization variables. Note that due to a worst-case exponential scaling of computational times when performing deterministic global optimizations in both cases (cf. Subsections 3.4.2 and 3.7.2), these achieved reductions in the number of optimization variables through the sensitivity-based refinement strategy correspond to

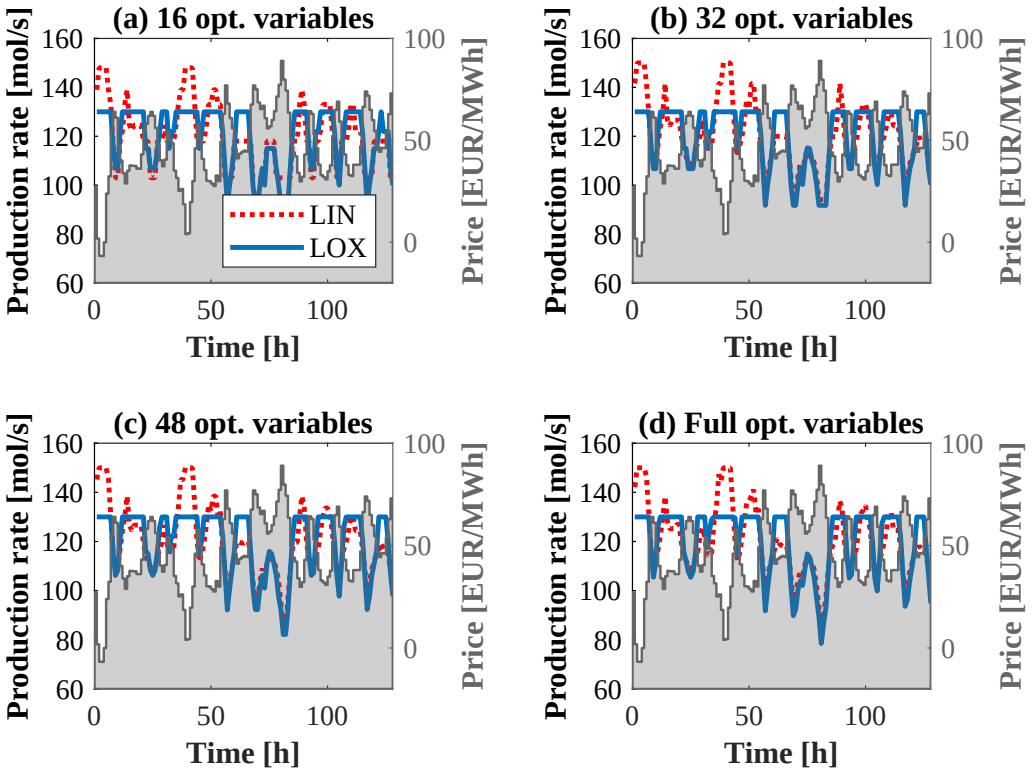

Figure 3.11.: Optimized production schedules after selected iterations of the algorithm considering a scheduling horizon of 128 intervals. (a): using 16 optimization variables, (b): using 32 optimization variables, (c): using 48 optimization variables, and (d): from multi-start local optimizations considering the full dimensionality, i.e., using an untruncated wavelet transform with 256 optimization variables. Dotted red lines depict the LIN production, solid blue lines correspond to LOX. The electricity price profile for the case study is shaded in gray.

highly promising times savings. Likewise, iterations using a similar number of optimization variables and thus resulting in similar computational times furnish improved approximations in case of the sensitivity-based refinement strategy.

3.7.2. Computational performance and scaling of the algorithm

In this section, we discuss the savings in computational times enabled by the use of the proposed algorithm in comparison to solution approaches considering the full dimensionality, i.e., using an untruncated wavelet transform. In particular, as in Subsection 3.4.2, we observe beneficial effects of reducing the number of optimization variables, i.e., truncating the wavelet transform in this case, for converging the lower bound in branch-and-bound algorithms for deterministic global optimization, that allow for obtaining converged solutions within the time limit. For illustration, we summarize the CPU times spent in MAiNGO's branch-and-bound procedure in Figure 3.13 for all problems, where a converged solution could have been obtained within the time limit. Note that the left part of this figure essentially leads to similar observations as Figure 3.7. That is, using the sensitivity-based

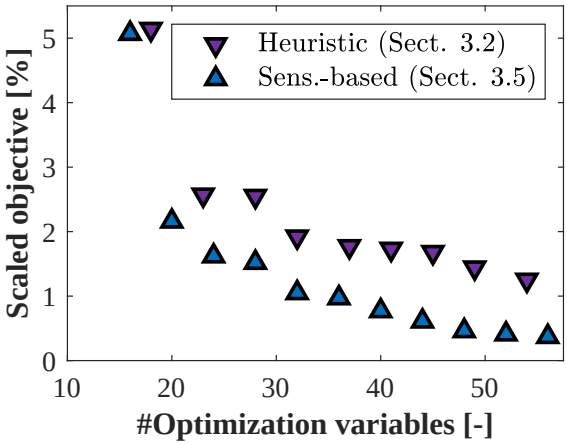

Figure 3.12.: Comparison of the course of the optimal objective value for a horizon length of 128 intervals. We compare the sensitivity-based refinement strategy (blue upwards-pointing triangles ($\triangle$)) presented in Section 3.5 to the wavelet-based grid adaptation from Section 3.2 using a heuristic criterion for refinements (purple downwards-pointing triangles ($\triangledown$).

refinement strategy again successfully circumvents an unfavorable, worst-case exponential, scaling of computational times with the horizon length, that would be observed in solution approaches considering the full dimensionality. One rather finds that the CPU times only scale exponentially with the number of optimization variables (given a fixed horizon length), whereas they scale linearly with the horizon length (given a fixed number of optimization variables). Thus, by sensibly truncating the number of optimization variables, solutions are enabled in a reasonable time that leave only small percentages of the maximum achievable savings unexploited, which are in a similar order of magnitude as the relative optimality tolerance of the solver. More precisely, the 2% gap in savings is undercut after a CPU time in the order of 10^1 s in case of $T = 128$ hourly scheduling intervals. In the case of $T = 256$, this can still be achieved within $10^3 \ldots 10^4$ s. Even for $T = 512$, we expect this to be computationally tractable, considering ready-to-use opportunities for speed-ups by using parallel computing capabilities on local machines ($\sim$10 cores). Besides these opportunities for reducing the computational times for obtaining converged global optima, we observe a nearly linear scaling of computational times in the pre-processing with the number of optimization variables. Consequently, the algorithm appears also promising for reducing the computational burden for obtaining stationary feasible points.

3.8. Summary and recommendations for future research

We present an algorithm for identifying low-dimensional yet highly accurate approximations of discrete-time nonlinear scheduling problems subject to time-variable electricity prices. Both the original and the extended version of the algorithm rely on a reduced-space formulation of the optimization problem. Therein, we explicitly allow for time-coupling constraints, such as storage or ramping limits. The key idea of the proposed algorithm is to substitute the original DoFs of the scheduling problem, which represent each time

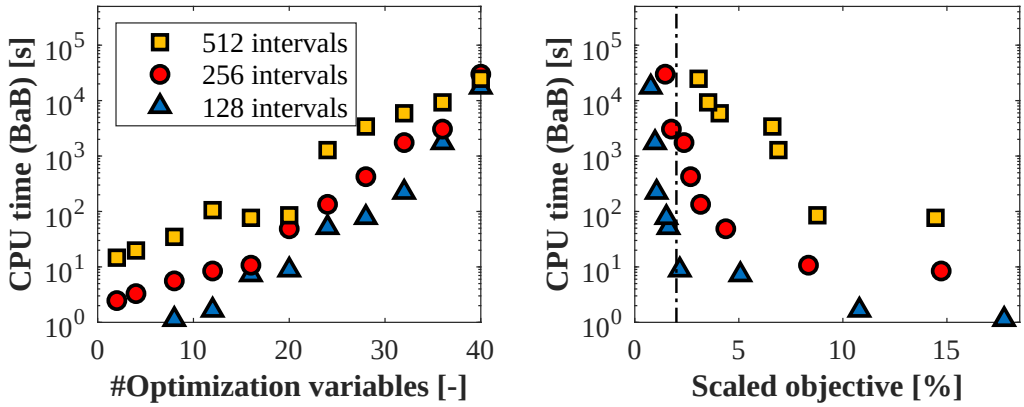

Figure 3.13.: CPU time spent in MAiNGO's branch and bound (BaB) for converging the objective value to within the optimality tolerance. (a): given as a function of the number of optimization variables, (b): plotted over the corresponding objective value. Symbols depict the varying scheduling horizons. Blue upwards-pointing triangles (△): 128 intervals, red circles (○): 256 intervals, yellow squares (□): 512 intervals. The vertical dashed-dotted line further gives the applied relative optimality tolerance for MAiNGO for comparison.

interval of the scheduling horizon individually, by a reduced number of new decision variables. More precisely, in the original version of the algorithm, one DoF is now assigned to multiple intervals of the scheduling problem that comprise similar electricity prices, giving a tailored nonuniform time grid. A systematic approach using a wavelet-based analysis proposed in literature iteratively refines the time grid. We apply the grid adaptation algorithm to the operational optimization of a CAES plant. There, substantial nonlinearities in the optimization problem arise from the efficiency characteristics of the turbomachines. These are accounted for by using ANNs as surrogate models. We show that the algorithm is able to identify feasible schedules that give similar objective values as deterministic global searches with consideration of the full time grid but require substantially less computational time. Moreover, we find that the solution times scale approximately linearly with the horizon length when maintaining the same number of grid points in the optimization problem, circumventing the worst-case exponential scaling of computational times with the horizon length in solution approaches considering the full time grid. In fact, this enables the consideration of scheduling horizons, which are prohibitively long without the grid adaptation.

Furthermore, we present a further development of the proposed algorithm targeting the special case of having only continuous variables in the optimization problem. Therein, we replace the systematic yet purely heuristic refinement procedure by much sounder considerations based on sensitivity analysis. For this purpose, we make use of an inverse wavelet transform that allows for formulating a reduced-space optimization problem exposing only the coefficients of the wavelet transform to the optimizer. Thereby, we enable reductions in the dimensionality of the optimization problem by truncating the wavelet transform. Decisions concerning the use of specific wavelet coefficients are iteratively made based on sensitivity information from the Lagrangian multipliers, giving the marginal effect of relaxing the constraint truncating a coefficient. The extended algorithm is applied to the scheduling of an industrial-scale ASU, where we again use an ANN as a surrogate model

representing the nonlinear process characteristics. To this end, we fit the ANN to the power consumption characteristic of the plant, which is obtained from optimizations using a rigorous process model. We again demonstrate that using a substantially reduced number of optimization variables compared to solution approaches considering the full dimensionality is suitable for identifying feasible near-optimal schedules. In particular, we show that substantially truncated wavelet transforms are still able to reproduce time series that allow for the exploitation of high percentages of the maximum achievable savings. Moreover, we demonstrate that the sensitivity-based refinement procedure can yield similarly close approximations as the heuristic one while using even fewer optimization variables. The reductions in the dimensionality of the optimization problems again lead to significant savings in computational times when performing deterministic global optimizations, circumventing the worst-case exponential scaling of solution times with the horizon length if considering the full dimensionality.

Due to the ability to furnish feasible near-optimal schedules in a reasonable time, the primary field of application for the proposed algorithms will lie in solving strongly nonconvex optimization problems involving multiple local solutions, where even the generation of favorable feasible points is a highly challenging task if considering the full dimensionality. In addition to that, the identified savings in computational times for local searches make the algorithm also promising for applications, where solution times are of crucial importance, e.g., to ensure real-time capability. Here, we highlight in particular that single-shooting approaches to dynamic optimization are formulated in a similar way as the reduced-space quasi-stationary scheduling addressed in this chapter - with both the objective as well as the path-constrained variables being explicit functions of the parametrized, commonly piecewise-constant input signals while hiding the actual model, that is, the state integration in case of dynamic optimization, from the optimizer (cf. [162]). Consequently, such a formulation would also allow for the usage of the presented algorithms, which should not be surprising, as the heuristic refinement procedure of the wavelet-based grid adaptation was originally proposed for dynamic optimizations. Thus, promising future fields of application will certainly also include moving horizon applications involving nonlinear dynamic optimizations, which aim at integrating both the scheduling and the control perspective. The application of the presented algorithm for scheduling with low-dimensional dynamic models (e.g., [18, 22, 88]) and economic optimizations directly in the control layer using a detailed dynamic model (cf. [16] as well as Chapter 4) is thus left for future work. In turn, utilizing concepts from previous works in the field of dynamic optimization concerning the detection of the switching structure of the optimization problem [163–165] might become promising for solving scheduling problems as well. We highlight that these two research perspectives indicate a vanishing border between scheduling, which classically relies on quasi-stationary models, and dynamic optimization problems from both a methodological and an application-related point of view.

A further promising direction for future research can be identified, lying in the extension of the further developed algorithm towards the handling of scheduling problems that comprise logic disjunctions and conditional statements, i.e., towards those problems where binary variables in state-of-the-art formulations do not solely stem from piecewise linearizations and could thus not be eliminated by using the original (smooth) nonlinear function. Here, it should be thoroughly investigated if alternative formulations exist that describe the relationships but forgo binary variables. In particular, progress made - again mostly in the area of dynamic optimization - regarding nonsmooth optimization [166–169] as well as

the handling of complementarity constraints [170–173] already bear a potential for eliminating the binary variables in many conventional problem formulations, which would in these cases allow for the direct application of the proposed sensitivity-based refinement strategy using Lagrangian multipliers. As an alternative to nonsmooth formulations, we are also interested in a tailored treatment for binary DoFs similar to that of continuous ones, i.e., using truncatable transforms and providing quantitative measures for decisions on refinements. Along these lines, future research should also investigate if the underlying concepts of our approaches (in particular, the reduced-space optimization, but also the truncated wavelet transforms) are computationally beneficial as well in the case of MILP scheduling problems, which have been extensively studied in the literature.

Finally, it is worth exploring further fields of application for the proposed algorithms on a higher-level perspective. For instance, we see great potential in applying the algorithms to two-stage stochastic programs, where the consideration of individual recourse actions for each scenario easily involves a large computational burden that could be reduced by using a reduced-space formulation either assigning the same recourse action to multiple similar scenarios in the mixed-integer case or exposing a truncated transform of the recourse decisions to the optimizer and using sensitivity-based refinements. In the context of the present thesis, we are particularly interested in extending the approaches towards a two-dimensional analysis considering scheduling problems with electricity price fluctuations over multiple time intervals and in multiple scenarios due to the uncertainties.

4. Hybrid mechanistic/data-driven models for real-time capable electricity market participation at the control layer

In this chapter, we address the development of a reduced hybrid mechanistic/data-driven dynamic model for distillation columns with a particular focus on its application to the control of ASUs participating in continuous electricity markets in real-time. First, we give a brief overview of existing approaches for reduced dynamic modeling of distillation columns in Section 4.1. In particular, we therein discuss their individual strengths and weaknesses and present a characterization of desired novel reduction approaches. In Section 4.2, we show how the integration of machine learning can overcome the shortcomings of classical compartmentalization approaches. To this end, we first repeat the classical compartmentalization approach and afterwards propose its enhancement by replacing complex nonlinear thermodynamic stage-to-stage calculations with ANNs as surrogate models for the intra-compartment input/output relations. Moreover, we briefly cover the training procedure for the ANNs due to its crucial influence on model performance. In Section 4.3, we introduce the case study for the performance evaluation of the proposed ANN-based compartment model, namely a well-investigated single-product nitrogen plant. Therein, we first provide a brief description of the process giving all manipulated and controlled variables. Second, we present the formulation of the dynamic optimization problem that we solve in open-loop. Third, we introduce the control structure and characterize the operating scenarios considered in the closed-loop simulations. In Section 4.4, the results of the open-loop optimizations are discussed focusing on both the prediction accuracy and the solutions times when applying the ANN-based compartment model in comparison to a full-order stage-to-stage model. The achieved trade-off between the loss in accuracy and the reduction of computational times is also put into the context of the relevant literature. In Section 4.5, we present the results of the closed-loop simulations. Therein, a distinct focus is set on the in silico demonstration of the real-time capability of the approach. Moreover, we discuss the controller performance both in the absence and presence of unmeasured disturbances. Finally, we conclude the chapter in Section 4.6. Here, we present perspectives for further improvements in both the model formulation and the control architecture.

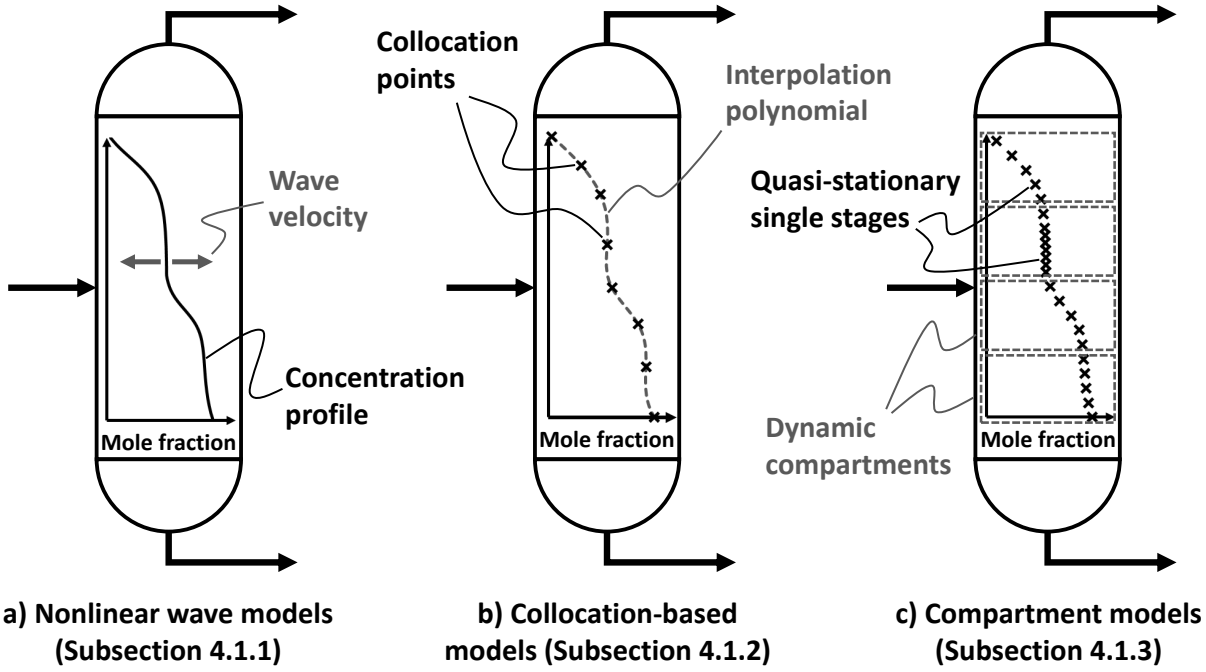

Figure 4.1.: Illustration of the key concepts of classical dynamic model reduction approaches for distillation columns.

4.1. Classical reduced dynamic modeling approaches for distillation columns

As stated in Subsection 2.1.2, existing approaches for reduced dynamic modeling of distillation columns can be classified into three categories (Figure 4.1): (i) nonlinear wave models, (ii) collocation-based approaches, and (iii) compartmentalization/stage aggregation. In the following, we briefly review their concepts and limitations. Based on that, we derive desired attributes for the novel reduction approach that we propose in the remainder of this chapter.

4.1.1. Nonlinear wave models

The oldest of the three classes of dynamic reduction approaches for distillation columns interprets the column profiles as traveling waves and describes their movement in space and time by partial differential equations [174]. Nonlinear wave models have originally been proposed considering packed columns. Nevertheless, they can also be used to approximate the profiles of staged columns described by stage-to-stage models. The early work of Gilles and Retzbach [102] proposes a first wave model for binary columns assuming sharp, i.e., discontinuous, temperature and concentration profiles. Later on, the wave model presented by Marquardt [103] overcomes the limitation to sharp profiles. The work by Kienle [104] finally generalizes the concept to multi-component mixtures. Under certain thermodynamic assumptions (in particular, constant molar overflow and relative volatilities), the wave equations can be solved analytically and embedded into overall column balances. For this purpose, Caspari et al. [105] recently extended the nonlinear wave model by an overall

mass and energy balance that account for varying column hold-ups occurring in case of load changes. In fact, the authors thereby embed the wave equations into a simplified compartmentalization approach (cf. Subsection 4.1.3).

Nonlinear wave models correspond to low-dimensional representations of distillation columns by definition, enabling promising opportunities for savings in computational times. As an important advantage compared to collocation-based or compartmentalized models, the size of the DAE system describing the nonlinear wave model is independent of the size of the original full-order stage-to-stage model. Moreover, the approach involves only a few constitutive parameters that can be estimated from stationary profiles and does not rely on a priori definitions of collocation points, compartment boundaries, aggregated stages, etc., which mitigates the risks for an impaired model performance from a suboptimal model set-up.

As a limitation, nonlinear wave models inherently rely on simplifying assumptions on the thermodynamic properties. Consequently, there is a significant dependence of the predicting capabilities of the models on the considered mixture. For instance, when applying the extended multi-component nonlinear wave model to the high-pressure nitrogen column of an ASU, Caspari et al. [105] show good prediction capabilities in case of rarely nonideal behavior (a liquid air mixture containing mostly nitrogen and oxygen). At the same time, large deviations from full-order stage-to-stage models in case of strongly nonideal mixtures are likely and restrict the applicability of nonlinear wave models. Note that this already holds for other compositions of a liquid air feed, particularly for higher argon contents that occur in further distillation columns of ASUs.

4.1.2. Collocation-based approaches

Collocation-based model reduction approaches approximate the column profile by interpolations between a finite number of collocation points. By selecting substantially fewer collocation points than theoretical stages of a corresponding full-order stage-to-stage model, a substantial model reduction is achieved [106, 107]. The column states above and below each collocation point, which are required for the mass and energy balances, are calculated through interpolation. For instance, Cao et al. [115] apply second-degree Lagrange polynomials. The locations of the collocation points can but do not necessarily need to coincide with the locations of actual stages.

Using a collocation-based approach, the number of MESH-equations is reduced, leading to a smaller DAE system for a given distillation column. As the structure of the DAE system however remains almost unaffected, the reduction of the size of the system directly translates into savings in computational times for integration. For instance, Cao et al. [115] find a nearly linear scaling of simulation times with respect to the number of collocation points used to approximate the column profile. This behavior constitutes an important strength of collocation-based models: the degree of reduction can a priori be chosen by selecting the number of collocation points.

The achievable degree of reduction of collocation-based models is obviously limited by accuracy requirements. In other words, too high degrees of reduction lead to substantial deviations from the predictions of full-order models for both the stationary and the transient behavior. For a detailed study on this trade-off, we refer to the work of Carta et al. [175]. Most importantly, the authors find that there is a minimum number of collocation points that is required to accurately reproduce the characteristics of the transient behavior,

which strongly depends on the considered mixture. Consequently, desirable large degrees of reduction might easily introduce unphysical behavior, such as additional steady-states.

4.1.3. Compartmentalization and stage aggregation

Compartmentalization has been proposed as an alternative to collocation-based approaches [109]. Its mathematical derivation is given by Lévine and Rouchon [111] and builds on singular perturbation theory. In addition, there is a physically meaningful interpretation (cf. [112]): a column model is split up into a number of compartments. One compartment is described by the dynamic compartment balances, stationary single-stage balances for all included stages except for one so-called sensitivity stage, and equilibrium calculations for all stages. Thus, the approach applies a separation by time scales assuming single-stage dynamics to be negligible compared to overall compartment dynamics.

Compartment models reproduce exactly the steady-state of the corresponding full-order stage-to-stage models. This finding is independent of the degree of reduction and thus represents a major advantage compared to collocation-based reduction approaches. In contrast, the transient behavior of compartment models depends on the number of compartments, their sizes, and the locations of the sensitivity stages. Horton et al. [110] present heuristics for these choices and demonstrate that dynamic responses of compartment models can be close to those of the corresponding full-order models, even if relying on substantially fewer differential equations. Likewise, Bian et al. [119] apply the compartmentalization approach to a distillation column of an ASU and find only small influences on the dynamic behavior even if reducing the number of differential equations by one order of magnitude.

Linhart and Skogestad [112] propose a different perspective on compartment models and introduce the concept of stage aggregation. More precisely, the mathematical formulation of a compartmentalized column is equivalent to interconnecting dynamic stages (with aggregated hold-ups) and stationary columns (without hold-ups), making the definition of compartments superfluous. Instead, the selection of aggregation stages and the definition of their initial hold-ups become relevant. The authors also provide a study on the computational efficiency of stage aggregation (and hence also of compartment) models. Summarizing their findings, the reduction approaches manipulate the stiffness of the DAE system instead of reducing its size. Consequently, improvements in the computational efficiency strongly depend on the integration method as well as the time constants of the input signal and are in no case guaranteed. In a second study, the same authors identify the solution of the large-scale nonlinear algebraic equation system describing the intermediate stationary columns to be the computational bottleneck of compartment and stage aggregation models [176]. To overcome this limitation, they propose the use of explicit functions interpolating between precalculated solutions, which however exhibits an unfavorable scaling with both the level of detail of the table and the number of components in the column.

4.1.4. Summary of shortcomings

As can be seen, a variety of dynamic model reduction approaches is available for distillation columns with numerous successful applications in literature for all of them, noticeably in all cases also with applications to the modeling and the dynamic optimization of ASUs (cf. the

literature referred to in Subsection 2.1.2). Moreover, one finds no general superiority of one approach. That is, each of the approaches offers certain promising advantages, which make it promising for specific use cases. Nevertheless, these advantages are in all cases accompanied by disadvantages that certainly limit the performances:

- Nonlinear wave models allow for representing an entire distillation column by only one wave equation per component. However, they make strong assumptions on the underlying thermodynamics when deriving the solution, limiting their applicability to near-ideal mixtures.
- Collocation-based models represent a convenient way to reduce the number of MESH-equations and hence the size of the DAE system. However, they bear a severe risk of introducing unphysical behavior for desired large degrees of reduction.
- Compartmentalization and stage aggregation make use of separation by time scales and do not depend on the thermodynamic properties. Thereby, they effectively reduce the number of differential equations. In standard variants, their computational performance is however not promising.

Desired novel reduction approaches should combine the benefits and avoid the drawbacks by (i) enabling substantial reductions of the size of the DAE system that directly translate into savings in computational times, (ii) reproducing physically meaningful behavior even for high degrees of reduction, and (iii) being independent of the considered thermodynamic system in terms of performance.

4.2. Reduced hybrid dynamic model based on compartmentalization and artificial neural networks

Machine learning techniques have already been applied in PSE several decades ago (cf. [177]) and are currently regaining attention after a period of low activity (cf. [178, 179]). In particular, the use of hybrid mechanistic/data-driven (also called semi-parametric/empirical or gray box) modeling approaches appears promising, as it enables a balance between first-principle mechanistic (white box) and data-driven (black box) models [180–182]. Consequently, we herein propose a reduced hybrid mechanistic/data-driven dynamic model for distillation columns enhancing the classical compartmentalization by integrating machine learning to overcome the aforementioned shortcomings (cf. Subsection 4.1.4).

4.2.1. Step 1: Classical compartmentalization

In order to facilitate both the derivation of the reduced hybrid model and its discussion in the next subsection, we first present the rigorous full-order stage-to-stage model forming the basis of the reduced model and show how the classic compartmentalization is conducted, although both steps are standard in the literature. For the DAE system describing the full-order model, we apply the following assumptions that have been widely used and validated in literature for modeling of ASUs. Thus, we also consider this as a suitable replacement of a real distillation column in an ASU. Note that neither the compartmentalization nor

the surrogate modeling actually requires one of these assumptions, i.e., dropping one of them would only involve adjustments in the compartment balances and/or the equation system to be replaced by the ANNs.

- The vapor hold-up on a column stage is negligible compared to the liquid hold-up [96, 115, 119, 183].
- Exiting liquid and vapor streams on each stage are in thermodynamic equilibrium [96, 115, 119, 183].
- The liquid on each stage is ideally mixed [96, 115, 119, 183].
- The dynamics of the temperature are much faster than the mass exchange, i.e., a pseudo steady-state energy balance can be used. Note that this also reduces the differential index of the equation system from two to one [115, 119].
- There is a constant pressure difference between neighboring stages [96, 119].
- The mass hold-up in the condenser is negligible [96, 115, 119, 183].
- There is a linear relation for stage hydraulics with constant proportional factor k_d on each stage [96, 115, 119].

Following these assumptions, the DAE system for a single column stage $i \in \{1, \dots, N\}$ with components $j = 1 \dots C$ reads:

$$\dot{M}_i = L_{i-1} + V_{i+1} - L_i - V_i \tag{4.1}$$

$$M_i \cdot \dot{x}_i^j = L_{i-1} \cdot (x_{i-1}^j - x_i^j) + V_{i+1} \cdot (y_{i+1}^j - x_i^j) - V_i \cdot (y_i^j - x_i^j), \quad j = 1 \dots C-1 \tag{4.2}$$

$$\sum_{j=1}^{C} x_i^j = 1 \tag{4.3}$$

$$0 = L_{i-1} \cdot (h_{i-1}^L - h_i^L) + V_{i+1} \cdot (h_{i+1}^V - h_i^L) - V_i \cdot (h_i^V - h_i^L) \tag{4.4}$$

$$\begin{pmatrix} \boldsymbol{y_i} \\ T_i \end{pmatrix} = \text{px-flash}(\boldsymbol{x_i}, p_i) \tag{4.5}$$

$$h_i^L = \mathrm{h^L}(\boldsymbol{x_i}, T_i, p_i) \tag{4.6}$$

$$h_i^V = \mathrm{h^V}(\boldsymbol{y_i}, T_i, p_i) \tag{4.7}$$

$$L_i = k_d \cdot M_i \tag{4.8}$$

with M_i denoting the molar hold-up, L_i and V_i the liquid and vapor streams leaving stage i, x_i^j and y_i^j as mole fractions of component j in liquid and vapor phase respectively (which

can be combined as vectors $\boldsymbol{x_i}$ and $\boldsymbol{y_i}$), T_i denoting the temperature, p_i the pressure, and h_i^L and h_i^V as liquid and vapor enthalpies. The specific functions px-flash$(\cdot)$, $\mathrm{h^L}(\cdot)$, and $\mathrm{h^V}(\cdot)$ depend on the thermodynamic models used. In case of additional feeds to the stage or product withdrawals, the corresponding streams are added to the balance equations. The entire model for a distillation arises from formulating the equations (4.1)-(4.8) for all theoretical stages $i = 1 \dots N$ and adding the equation systems for condenser and reboiler.

As stated above, one compartment is now defined to consist of multiple single stages. Without loss of generality, balance equations (4.1)-(4.4) for one single stage, the so-called sensitivity stage $N_{c,s}$, can be replaced by the overall compartment balances. Assuming that stages $N_{c,1}$ to $N_{c,2}$ form compartment c, we get:

$$\dot{M}_c = L_{N_{c,1}-1} + V_{N_{c,2}+1} - L_{N_{c,2}} - V_{N_{c,1}} \tag{4.9}$$

$$\begin{aligned} M_c \cdot \dot{x}_c^j = {} & L_{N_{c,1}-1} \cdot (x^j_{N_{c,1}-1} - x_c^j) + V_{N_{c,2}+1} \cdot (y^j_{N_{c,2}+1} - x_c^j) \\ & - L_{N_{c,2}} \cdot (x^j_{N_{c,2}} - x_c^j) - V_{N_{c,1}} \cdot (y^j_{N_{c,1}} - x_c^j), \quad j = 1 \dots C-1 \end{aligned} \tag{4.10}$$

$$\sum_{j=1}^{C} x_c^j = 1 \tag{4.11}$$

$$\begin{aligned} 0 = {} & L_{N_{c,1}-1} \cdot (h^L_{N_{c,1}-1} - h_c^L) + V_{N_{c,2}+1} \cdot (h^V_{N_{c,2}+1} - h_c^L) \\ & - L_{N_{c,2}} \cdot (h^L_{N_{c,2}} - h_c^L) - V_{N_{c,1}} \cdot (h^V_{N_{c,1}} - h_c^L). \end{aligned} \tag{4.12}$$

In equations (4.9)-(4.12), the compartment states M_c, x_c^j and h_c^L are introduced and defined as follows:

$$M_c = \sum_{i=N_{c,1}}^{N_{c,2}} M_i$$

$$x_c^j = \frac{1}{M_c} \sum_{i=N_{c,1}}^{N_{c,2}} M_i \cdot x_i^j, \quad j = 1 \dots C$$

$$h_c^L = \frac{1}{M_c} \sum_{i=N_{c,1}}^{N_{c,2}} M_i \cdot h_c^L.$$

The fundamental idea of compartment models lies in assuming that the compartments are sufficiently large, so that single-stage dynamics can be neglected compared to overall compartment dynamics. Consequently, single-stage balance equations are assumed stationary. Thus, the entire DAE system for compartment c consists of equations (4.9)-(4.12) plus equations (4.5)-(4.8) for stages $i \in \{N_{c,1}, \dots, N_{c,2}\}$ and the steady-state versions of equations (4.1)-(4.4) for stages $i \in \{N_{c,1}, \dots, N_{c,2}\} \setminus \{N_{c,s}\}$, i.e., all stages except for the sensitivity stage. An overall compartmentalized model for a distillation column arises from conducting these steps for all compartments $c \in \{1, \dots, N^C\}$, thus reducing the number of differential equations compared to a full-order stage-to-stage model by a factor of N/N^C with the the number of algebraic equations remaining (nearly) constant. Note that if the DAE system for single stages is of differential index one, so is the DAE system for an entire compartment and thus also the overall compartment model.

4.2.2. Step 2: Surrogate modeling with artificial neural networks

In order to derive the proposed model, we formulate the equation system describing a classical compartment as an index-1 DAE system:

$$\dot{\hat{\boldsymbol{x}}}(t) = \hat{\boldsymbol{f}}(\hat{\boldsymbol{x}}(t), \hat{\boldsymbol{u}}(t), \hat{\boldsymbol{y}}(t)) \tag{4.13}$$

$$\mathbf{0} = \hat{\boldsymbol{g}}(\hat{\boldsymbol{x}}(t), \hat{\boldsymbol{u}}(t), \hat{\boldsymbol{y}}(t), \hat{\boldsymbol{z}}(t)). \tag{4.14}$$

In equations (4.13) and (4.14), we introduce the following notation: differential compartment states are denoted by $\hat{\boldsymbol{x}}(t)$. Compartment inputs, that are handed over by the neighboring compartments, are denoted by $\hat{\boldsymbol{u}}(t)$, which corresponds to states of column stages $N_{c,1} - 1$ and $N_{c,2} + 1$. Compartment outputs, that are required in the equation systems of neighboring compartments, are denoted $\hat{\boldsymbol{y}}(t)$, which corresponds to the states of column stages $N_{c,1}$ and $N_{c,2}$. The remaining (algebraic) compartment states are denoted by $\hat{\boldsymbol{z}}(t)$.

As analyzed by Linhart and Skogestad [176] (see, above), when solving equations (4.13) and (4.14), the main computational effort is spent in the solution of the highly nonlinear algebraic equation system (4.14) that mainly originates from the thermodynamic relations (4.5)-(4.7), such as a G^E-model (Margules, NRTL, etc.) or an equation of state (Peng-Robinson, PC-SAFT, etc.). To reduce the computational effort, Linhart and Skogestad [112] propose interpolation between tabulated precalculated solutions of (4.14), giving the compartment outputs as an explicit function of the compartment inputs and the differential states:

$$\begin{pmatrix} \hat{\boldsymbol{y}}(t) \\ \hat{\boldsymbol{z}}(t) \end{pmatrix} = \hat{\boldsymbol{g}}^{-1}(\hat{\boldsymbol{x}}(t), \hat{\boldsymbol{u}}(t)). \tag{4.15}$$

As an alternative avoiding the unfavorable scaling of the interpolation approach with the level of detail in the table, we herein propose to use a continuous surrogate model for the solution of the algebraic equations, i.e., for (4.15), to further improve the computational efficiency of the model and make it readily applicable in state-of-the-art frameworks for dynamic optimization. Due to their recent successful application to numerous chemical engineering problems involving complex thermodynamics - with Schweidtmann et al. [138] replacing implicit thermodynamic functions for fluid properties, Nentwich et al. [140] and Huster et al. [184] replacing the iterative computation of mixture properties, as well as Schweidtmann et al. [139] replacing entire flash calculations for vapor-liquid equilibria - we herein use ANNs as surrogate model replacing (4.15). Note that ANNs can be evaluated efficiently in a forward manner. Likewise, the use of backpropagation algorithms (as a special case of automatic differentiation for ANNs) allows for efficient gradient calculations even for very complex network structures [185]. However, the latter opportunity is currently omitted in our computational framework. Instead, analytical Jacobians are used for the ANN-functions, which at some point will presumably cause limitations in problem complexities. Furthermore, sophisticated computer codes are readily available for efficient training of the ANNs. Thereby, ANNs can be fitted very efficiently to large data sets, that arise from the sampling of a high-dimensional input space [186].

The DAE system representing the ANN-based compartment model then reads:

$$\dot{\hat{\boldsymbol{x}}}(t) = \hat{\boldsymbol{f}}(\hat{\boldsymbol{x}}(t), \hat{\boldsymbol{u}}(t), \hat{\boldsymbol{y}}(t)) \tag{4.16}$$

$$\hat{\boldsymbol{y}}(t) = \hat{\boldsymbol{g}}_{\mathrm{ANN}}(\hat{\boldsymbol{x}}(t), \hat{\boldsymbol{u}}(t)). \tag{4.17}$$

We highlight that the model formulation (4.16)-(4.17) is only one possibility of an ANN-based compartmentalization approach. The choice of this system however seems appealing, as it shows an analogy to the dynamic modeling of simple flash units. That is, the model outputs can be calculated as an explicit function of the model inputs (cf. a Tx-flash). An ANN-based compartmentalized distillation column model can thus still be considered as interconnected MESH-equations. Other possibilities for using ANNs within hybrid compartmentalization approaches exist as well. For instance, using a surrogate model for the ordinary differential equation (ODE) form of the index-1 DAE system (4.13)-(4.14) represents an alternative. Likewise, ANNs could also be used to replace only specific parts of the algebraic equation system (4.14), such as single-stage flash calculations (4.5). Further note that the stage aggregation analogy from Linhart and Skogestad [112] again allows for an alternative formulation of a reduced hybrid model with the ANNs replacing a series of interconnected stationary stages, which corresponds to the learning of the input/output behavior of stationary distillation columns.

The ANN-based compartment model inherits the ability of the classical compartment model to substantially reduce the number of differential equations, which is a desired attribute with regard to computational efficiency. Furthermore, in the ANN-based compartment, we facilitate the solution of the algebraic equation system describing the input/output relations inside the compartment. More precisely, we confine to solving the system only for those variables that are relevant for the compartment balances and introduce explicit functions for this purpose. As stated above, an ANN-based compartmentalized column still considers physical insights in the form of integral balance relations.

The use of surrogate models represents a second source of errors in addition to potential errors from manipulating the stiffness of the DAE system through compartmentalization. Whereas the latter only affects the transient model behavior, prediction errors of the ANNs are also present in steady-state, so that in contrast to the classical compartmentalization, the steady-state of the ANN-based compartment model does not necessarily coincide with the steady-state of the corresponding full-order stage-to-stage model. However, assuming a sufficiently good estimation accuracy of the function $\hat{\boldsymbol{g}}_{\mathrm{ANN}}(\cdot)$, deviations are expected to be small. One disadvantage of the ANN-based compartmentalization is that it requires a significant preparation effort for obtaining $\hat{\boldsymbol{g}}_{\mathrm{ANN}}(\cdot)$. For instance, the wall time for the training of the most complex single ANNs used in the case study was 10-20 h by applying the procedure presented below. This preparation effort represents a shortcoming compared to collocation-based and nonlinear wave models that can be set up much more easily. However, the preparation is only done once offline and in an automated way. The CPU time savings when applying the reduced hybrid model for control purposes are in contrast online and repeatedly seen.

Extrapolability of the ANN functions cannot be expected in general. It is thus crucial to cover the required operating range well. Further note that using the hyperbolic tangent activation function of the ANN inherently introduces nonconvexities that need to be handled within the optimization problem [134]. Consequently, in the case of overfitting the relation (4.15), these nonconvexities might also introduce multiple unphysical local minima to the

optimization problem. However, this risk can be mitigated by applying regularization approaches during the training, such as the commonly used early stopping criterion (cf. [185]) as done in this work or more sophisticated approaches such as dropout (cf. [187]). The results of the case study conducted in this work further indicate that the discussed issues might be of minor relevance in practice.

4.2.3. Comments on the training procedure

Obtaining the required function $\hat{\boldsymbol{g}}_{\mathrm{ANN}}(\cdot)$ in (4.17) consists of the three following major steps: (i) sampling the input space, (ii) solving the nonlinear equation system, and (iii) fitting the ANNs. In the following, we give a brief description of how we have conducted them for the case study and discuss potential further developments for the future.

As mentioned above, it is crucial that the relation $\hat{\boldsymbol{g}}_{\mathrm{ANN}}(\cdot)$ produces accurate predictions across the entire operating range of the compartment section. We therefore aim to provide input samples (i.e., samples of $\hat{\boldsymbol{x}}(t)$ and $\hat{\boldsymbol{u}}(t)$) that cover this range well. Note that if the range was known exactly, problems with extrapolation would not arise. In practice, the identification of the actual operating range is not straight-forward. Therefore, the following approach is proposed: first, characteristic load changes across the entire operating range of the distillation column are performed by simulating the full-order stage-to-stage model. Thereby, lower and upper bounds for the differential compartment states $\hat{\boldsymbol{x}}(t)$ and input variables $\hat{\boldsymbol{u}}(t)$ can be obtained for each compartment. Afterwards, the space spanned by the bounds can be sampled by means of appropriate techniques, e.g., the well-known Latin hypercube method [188]. Here, we recommend to add adequate safety factors to the bounds of $\hat{\boldsymbol{x}}(t)$ and $\hat{\boldsymbol{u}}(t)$.

In a second step, equation (4.14) is solved for all samples of $\hat{\boldsymbol{x}}(t)$ and $\hat{\boldsymbol{u}}(t)$. As this corresponds to the solution of a nonlinear algebraic equation system, various computer codes can be used for the solution. As all models in the case study are implemented in the modeling language Modelica and made accessible to our in-house software package for dynamic simulation and optimization (cf. Subsection 4.3.2), we herein use the internal capabilities of this software package for solving nonlinear algebraic equation systems.

As by solving the nonlinear equation system, the required input/output samples for equation (4.15) are obtained, the function $\hat{\boldsymbol{g}}_{\mathrm{ANN}}(\cdot)$ can be fitted using state-of-the-art computer codes for training of ANNs. For instance, in this work, we use the MATLAB R2018a Neural Network Toolbox (Mathworks, Inc.). We investigated different network architectures and obtained the best results by representing each compartment output separately with a multiple-input-single-output ANN with one hidden layer. For good performance of the ANNs, we further recommend using the same scaling procedure for composition variables as applied by Cao et al. [115] when applying the ANN-based compartment model to case studies focusing on the high-pressure columns of ASUs. More specifically, we use the logarithm of molar fractions (for $j = \mathrm{O_2}$ and $j = \mathrm{Ar}$) or the logarithm of one minus the molar fraction (for $j = \mathrm{N_2}$) within the ANN instead of the molar fractions directly. For determining the number of neurons per ANN, an iterative procedure is applied that increases the number of neurons within the hidden layer of the ANNs until a minimum performance criterion is met.

We highlight that potential improvements in the training procedure exist in each of the three steps. In particular, we see a great perspective in using both more sophisticated (adaptive) sampling methods (e.g., [140, 189, 190]) and more complex net-

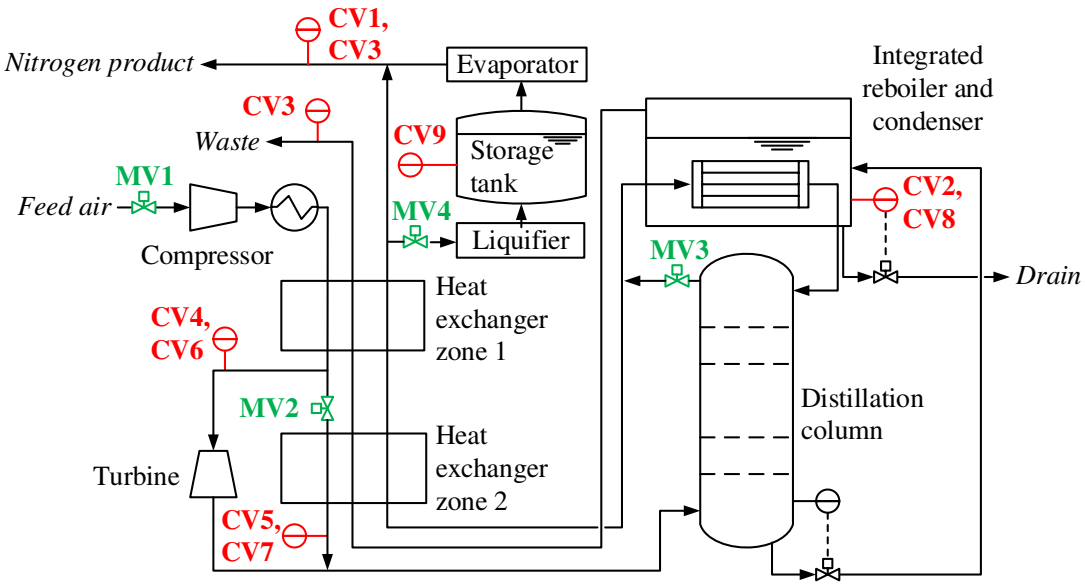

Figure 4.2.: Flowsheet of the process considered in this chapter. Manipulated (MVs) and controlled variables (CVs) are highlighted in the flowsheet and listed in Tables 4.1 and 4.2 respectively.

work structures using multiple hidden layers, potentially enhanced by ensemble learning (e.g., [191, 192]). Assuming maximum internal column flows to be known for a specific application and considering that compositions are inherently bounded between zero and one, these further developments could thereby potentially allow for ANNs having accurate prediction capabilities across the entire input range and thus entirely circumventing the risk for extrapolation.

4.3. Case study: Optimal operation of a nitrogen plant

We focus on a process for producing highly pure gaseous nitrogen for an on-site customer, such as a chemical plant (Figure 4.2). Similar process designs are well-investigated in the literature (e.g., [10, 14, 83, 87, 88, 98, 99, 105, 116, 193]). In the process, ambient air is pressurized, cooled down to the two-phase region, and fed to a distillation column. On top of the column, the nitrogen product stream is drawn off. The bottom stream of the column is expanded and used for condensing the column reflux. Both the gaseous product and the waste stream are used to cool down the incoming feed air in a counter-current heat exchanger. The thermal couplings of the column's in- and outputs make this ASU configuration an interesting case study for control purposes.

We herein aim at identifying operating strategies that reduce the operational, i.e., electricity, costs for the plant while satisfying the customer's product demand. We highlight the external liquefaction cycle of the plant that can act as a buffer for storable liquified products and hence allows a flexibilization of the operation of the ASU, which is - to some degree - independent from the customer's operating schedule. In times of low electricity prices and low utilization rates, nitrogen can be overproduced, liquified, and stored. In

Table 4.1.: Manipulated variables (MVs, i.e., process inputs) in the case study with their respective bounds.

	Description [Unit]	Bounds
MV1	Inlet flow main air compressor $[\frac{\text{mol}}{\text{s}}]$	[301.5, 368.5]
MV2	Inlet flow zone 2 of heat exchanger $[\frac{\text{mol}}{\text{s}}]$	[36.75, 44.90]
MV3	Distillation column product withdrawal $[\frac{\text{mol}}{\text{s}}]$	$[0, \infty)$
MV4	Inlet flow product storage tank $[\frac{\text{mol}}{\text{s}}]$	$[0, \infty)$

Table 4.2.: Controlled variables (CVs) in the case study with their respective bounds.

	Description [Unit]	Bounds
CV1	Product impurity [ppm]	[0, 20]
CV2	Temperature difference (integrated reboiler and condenser) [K]	$[0, \infty)$
CV3	Temperature difference (inlet heat exchanger zone 1) [K]	$[0, \infty)$
CV4	Temperature difference (outlet heat exchanger zone 1) [K]	$[0, \infty)$
CV5	Temperature difference (outlet heat exchanger zone 2) [K]	$[0, \infty)$
CV6	Temperature at turbine inlet [K]	$[115, \infty)$
CV7	Temperature behind heat exchanger [K]	[0, 95]
CV8	Relative hold-up (reboiler) [%]	[95, 105]
CV9	Relative hold-up (storage tank) [%]	[0, 200]

turn, this allows for load reductions in case of high electricity prices and high utilization rates. The plant's operation can be defined using four manipulated variables (MVs, cf. Figure 4.2), whose operating ranges can be found in Table 4.1. Additionally, Table 4.2 lists the considered controlled variables (CVs, cf. Figure 4.2). Note that in the following, the controller is required to keep the CVs within their respective bounds and not at a given set-point as is done in classic regulatory control.

In the following, we first briefly describe the applied models for the unit operations. Afterwards, we present the formulation of the dynamic optimization problem that we solve in open-loop in Section 4.4 to assess both the prediction accuracy and the computational efficiency of the proposed reduced hybrid mechanistic/data-driven dynamic model. Afterwards, we apply this model within a closed-loop eNMPC framework, which we introduce at the end of the section. Furthermore, we describe the operating scenario for assessing the performance of the proposed eNMPC framework, i.e., the applied market, demand, and disturbance models, as well as the controller parameters.

4.3.1. Process modeling

We herein differentiate between the model used when solving the dynamic optimization problems (cf. Subsection 4.3.2) and the model used as the replacement for the actual plant. The only difference between the two lies in the way that the distillation column is modeled. For all other unit operations, we use the same models as in Caspari et al. [98]

both when solving the dynamic optimization problems and when simulating the dynamic response of the plant replacement. Details on those units can be found in Appendix B. In the plant replacement, we use a full-order stage-to-stage model making use of the assumptions in Subsection 4.2.1 for the distillation column. The model comprises 40 theoretical stages, a steady-state condenser, and a single flash unit for the reboiler. When solving the dynamic optimization problems, we apply the presented ANN-based compartment model. For comparison purposes, we however also consider benchmark cases using either a full-order model (no model/plant mismatch) or the corresponding classical compartment model (with errors only due to compartmentalization).

We investigated different compartmentalization schemes for the distillation column and gained valuable insights into the influence of the number of compartments on model accuracy, which are used for setting up the ANN-based compartment model. That is, confining to compartments of almost equal size, we found the use of four compartments (i.e., approximately ten stages per compartment) to be best suited for the application, which is in good agreement with the results of Bian et al. [119]. This corresponds to a reduction of the size of the differential equation system describing the distillation column by 90%. Increasing the number of compartments to five, i.e., a lower degree of reduction, did not show significant increases in the accuracy, but substantially prolonged computational times, whereas decreasing the number to three, i.e., a higher degree of reduction, led to an undesired inverse-response behavior when simulating step responses. In contrast to the number of compartments, the complexity of the ANNs did not show a substantial effect on computational times, so that adequately complex ANN structures can be applied without a need to balance between accuracy and computational performance. Details on the number of neurons used for fitting all required input/output functions inside the compartments are also given in Appendix B. As a further adjustment to the classical compartment, we include the (stationary) condenser into the upper compartment of the ANN-based model. In contrast, the feed stage and reboiler are still modeled separately.

4.3.2. Formulation of the optimal control problem

Optimal values for the process inputs $\boldsymbol{u}(t)$, i.e., the MVs, are calculated by solving a dynamic optimization problem (4.18a)-(4.18g) on the finite prediction horizon $\mathcal{T} = [t_0, t_f]$ minimizing the electricity costs at final time $\phi\left(\boldsymbol{x}\left(t_f\right)\right)$, which is referred to as the optimal control problem (OCP).

$$
\begin{aligned}
\min_{\boldsymbol{u}(t),\boldsymbol{x}(t),\boldsymbol{y}(t)} \quad & \phi\left(\boldsymbol{x}\left(t_f\right)\right) && && \text{(4.18a)}\\
\text{s.t.} \quad & \dot{\boldsymbol{x}}(t) = \boldsymbol{f}\left(\boldsymbol{x}(t), \boldsymbol{y}(t), \boldsymbol{u}(t)\right), && \forall t \in \mathcal{T} && \text{(4.18b)}\\
& \boldsymbol{0} = \boldsymbol{g}\left(\boldsymbol{x}(t), \boldsymbol{y}(t), \boldsymbol{u}(t)\right), && \forall t \in \mathcal{T} && \text{(4.18c)}\\
& \boldsymbol{0} = \boldsymbol{h}\left(\boldsymbol{x}(t_0), \boldsymbol{y}(t_0), \boldsymbol{u}(t_0)\right) && && \text{(4.18d)}\\
& \boldsymbol{u}_{lb} \leq \boldsymbol{u}(t) \leq \boldsymbol{u}_{ub}, && \forall t \in \mathcal{T} && \text{(4.18e)}\\
& \boldsymbol{y}_{lb} \leq \boldsymbol{y}(t) \leq \boldsymbol{y}_{ub}, && \forall t \in \mathcal{T} && \text{(4.18f)}\\
& \boldsymbol{0} = \boldsymbol{c}^{\boldsymbol{f}}\left(\boldsymbol{x}(t_f)\right) && && \text{(4.18g)}
\end{aligned}
$$

The semi-explicit index-1 DAE system (4.18b)-(4.18c) corresponds to the process model considered within the OCP and incorporates - besides the model equations for the other

process units (cf. Appendix B) - either the DAE system describing the ANN-based compartment model, the classical compartment, or the full-order stage-to-stage model. Recall that by using the ANN-based model for the distillation column instead of the full-order one, the number of differential states $\boldsymbol{x}(t)$ as well as the number of algebraic states $\boldsymbol{y}(t)$ are substantially reduced, which is beneficial with regard to solution times. Equation (4.18d) further defines the initial state of the DAE system.

Calculation of the electricity costs as objective to be minimized enters the DAE system by using the following differential equation:

$$\dot{\phi}(t) = \hat{c}_E(t) \cdot P_t(t)$$

with $\hat{c}_E(t)$ denoting the (predicted) instantaneous electricity price and $P_t(t)$ the net electricity consumption of the entire plant (consumption of the compressor and liquifier minus production of the turbine).

Hard path constraints (4.18e)-(4.18f) on process inputs, i.e., the MVs, as well as selected algebraic states, i.e., the CVs, are used to ensure that the plant remains within its operating range (Tables 4.1 and 4.2). In particular, compressor and turbine flows (MV1 and MV2) can be varied within ±10% compared to their nominal stationary operating point and a maximum product impurity (CV1) of 20 ppm is demanded. In addition to that, we force driving temperature differences at various points of the plant to be positive. Furthermore, we require the temperature of the air flow at the turbine inlet to remain above and that behind the second heat exchanger zone below the boiling point. Due to delivery commitments, the on-site gaseous product demand needs to be fulfilled at every point in time either via the column withdrawal or re-vaporization of the tank content. In order to prevent emptying the tank, a terminal equality constraint (4.18g) is used for the final state of the storage tank, i.e., we constrain the tank holdup at the end of the horizon to its initial value. Note that a wide range of additional operational constraints has been used in literature, such as flooding limits (cf. [10, 116]). For the sake of simplicity, these are discarded here. However, their integration would be straight-forward. Assuming that the constrained variable can be calculated as a function of column states of the full-order stage-to-stage model (as is the case for, e.g., the flooding velocity), we can also use a surrogate model for its calculation inside a compartment, which essentially corresponds to adding one ANN per additional CV.

Solution of the OCP (4.18a)-(4.18g) is realized using the in-house software package DyOS [194]. The models themselves are set up in the object-oriented modeling language Modelica and made accessible for DyOS via exportation as a functional mock-up unit (FMU) using Dymola (Dassault Systèmes®). DyOS solves the OCP in a sequential approach using a single-shooting algorithm [195, 196]. Therein, the OCP is converted into an NLP by control vector parametrization and then solved via sequential quadratic programming (SQP) using SNOPT [197] version 7.2-4. As integrator, we apply NIXE [198], but also compare to using LIMEX [199] in Subsection 4.4.2. When solving the OCP once offline in open-loop in Section 4.4, we consider a prediction horizon of 24 h and use a piecewise-constant parametrization for each control vector with twelve equally-sized intervals, giving 48 DoFs for the optimizer. We apply historic German day-ahead spot electricity prices retrieved from EPEX SPOT SE (`www.epexspot.com`) for January 4$^{\text{th}}$, 2018 for illustration. Furthermore, we consider a constant withdrawal of 155 $\frac{\text{mol}}{\text{s}}$ of gaseous nitrogen by an on-site customer. In Section 4.5, the OCP is solved repeatedly on a moving horizon, as

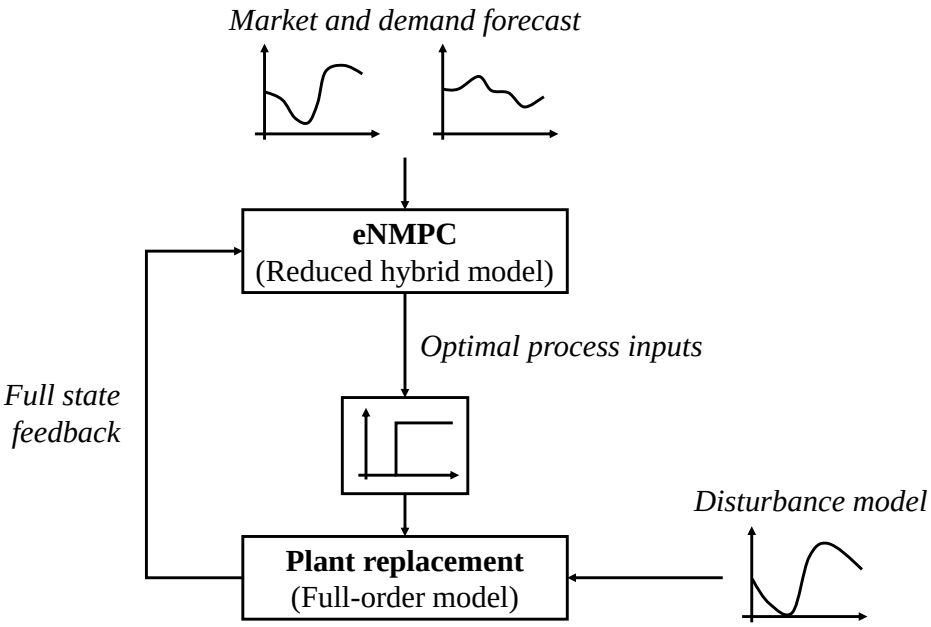

Figure 4.3.: Single-layer eNMPC framework considered in the closed-loop case study in Section 4.5.

is presented in the following subsection.

4.3.3. Control structure and closed-loop scenarios

We apply the single-layer eNMPC framework given in Figure 4.3 to allow for the participation of the ASU in a continuous intra-day electricity market directly on the control layer. In particular, we demonstrate in silico that real-time capable decision-making concerning energy trading is enabled by a combination of the reduced hybrid mechanistic/data-driven dynamic model presented above with a suboptimal fast update scheme from literature. We herein focus on the German continuous intra-day market with hourly electricity products. Note that this market is indeed an auction market following the pay-as-bid principle, so that the ask prices submitted to the market would strictly speaking define a probability of acceptance (cf. Chapter 5). However, this perspective is out of the scope of this chapter, such that a simplified market model is applied with the following constituting rules:

- During one hour, the price is known and remains constant.
- Electricity is both bought and sold at the same price.
- There is no delay between the purchase and consumption of energy.
- Energy can only be bought and sold for the current hour.

As the development of the electricity price in the course of the day is inherently uncertain, we construct price forecasts that are used in the eNMPC by assuming a geometric Brownian motion (GBM) process [200]. Thereby, we assume that the long-term drift of the price

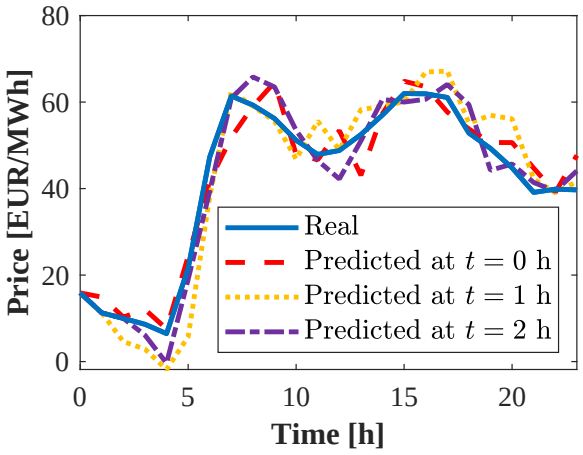

Figure 4.4.: Electricity prices for the computational study. The solid blue line corresponds to the actual prices at the German intra-day continuous market, the other lines depict forecasts used in the eNMPC for the first three hours of the horizon. Dashed red line: forecast at $t = 0$ h, dotted yellow line: forecast at $t = 1$ h, dashed-dotted purple line: forecast at $t = 2$ h.

development is known but superimposed with a random process [201]. Tuning parameters can be used for balancing between these two effects. The GBM assumes that in the t'th hour, the k-hours-ahead forecast of the electricity price $\hat{c}_E^{t+k|t}$ can be calculated using the equation below with c_E^{t+k} denoting the real price development:

$$\hat{c}_E^{t+k|t} = \hat{c}_E^{t+k-1|t} + \theta \cdot (c_E^{t+k} - \hat{c}_E^{t+k-1|t}) + \xi^{t+k|t} \quad \text{with} \quad \xi^{t+k|t} \sim \mathcal{N}(0, \sigma).$$

Here, $\xi^{t+k|t}$ is a random variable and corresponds to a zero-mean white Gaussian noise added to the signal. Parameter values $\theta = 1$ and $\sigma = 0$ correspond to perfect forecasts. For $\theta = 0$ we find a random walk with variance σ^2. Parameter values of $\theta = 0.8$ and $\sigma = 5$ are used to construct forecasts for the case study. Figure 4.4 compares the real course of the electricity price for the case study, which corresponds to German intra-day continuous market data from December 18$^{\text{th}}$, 2018 retrieved from EPEX SPOT SE (`http://www.epexspot.com`), to the predicted one from using the GBM in the first three hours of the considered horizon.

Although exploitation of temporal price spreads at electricity markets appears highly promising, the primary target of the ASU remains to safely satisfy its on-site customer demands at any time. We thus assume the following operating scenario: the customer submits its schedule beforehand, i.e., the schedule is known in the eNMPC layer. However, due to a disruption in the customer's operation, the schedule is changed. We assume that this change of schedule is communicated to the ASU operator 1 h before its occurrence (Figure 4.5). From then on, the eNMPC can react to this upcoming change in the customer's demand.

We further consider a heat input per tray due to imperfect isolation as a disturbance as done by Caspari et al. [100]. We assume that this disturbance follows a sine curve. The curve is shifted, such that it matches the typical course of temperature over one day (i.e., the minimum is around 5:00 am, cf. Figure 4.6). As will be shown later, this additional heat stream is a severe disturbance to the energy balance of the column, such that an

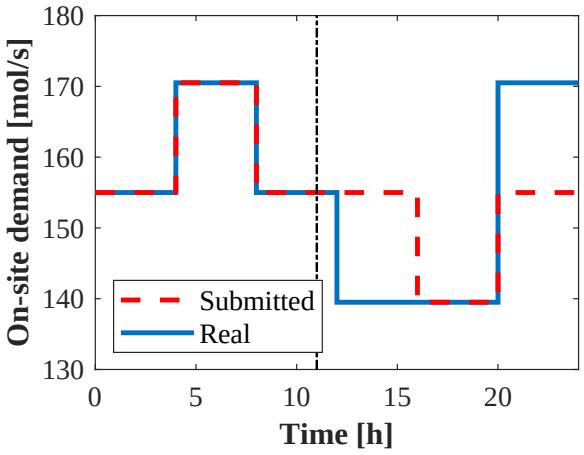

Figure 4.5.: On-site gaseous product demand for the computational study. The dashed red line depicts the originally submitted schedule of the customer, the solid blue line corresponds to the actual schedule. The vertical line gives the point in time when the change in the schedule is submitted and can be used in the eNMPC.

increased amount of heavy boiling oxygen and argon would contaminate the product without countermeasures. We assume that the disturbance is neither measured nor estimated. Thus, the controller rejects the disturbance only by using process feedback. Note that in practice, measurements and forecasts of the course of the temperature would be possible and might allow for a feed-forward component in the control structure using a suitable model for its relation to the heat stream.

In order to reject the disturbance as well as to react on both changes in the on-site demand and newly available price forecasts, the OCP (4.18a)-(4.18g) is solved repeatedly on a moving horizon using the ANN-based compartmentalization for the modeling of the distillation column. We choose a sampling time of 15 min for the eNMPC that we consider adequate compared to the time constants of the plant. Note that this is much longer than what is common in industrial practice with LMPCs. However, the mismatch between the applied reduced hybrid nonlinear model and the actual plant is also substantially smaller than in the case of linearized control models. Due to the nonetheless existing model/plant mismatch, we add a safety factor to the bounds of the limiting CV, i.e. we use a constraint back-off. More precisely, a product impurity bound of 19 ppm is used instead of 20 ppm that would actually correspond to the product requirement given in Table 4.2. The constraint back-off is implemented by adjusting (4.18f). We use no special treatment for infeasible optimization problems, such as a reformulation of the hard path constraints as soft ones. In particular, this also holds in case of initial constraint violations due to model/plant mismatch, i.e., normal iterations of the solver are conducted for driving the CVs back within their bounds in the course of the horizon. We assume full state feedback and calculate the initial state of the ANN-based compartment model for optimization from the states of the full-order stage-to-stage model (cf. equation (4.18d)), which we use as plant replacement to simulate dynamic responses. Both the prediction and the control horizon are chosen to be 12 h, which enables the exploitation of the most distinct daily price spreads (cf. Figure 4.4) by the eNMPC. Using an hourly piecewise-constant control vector parametrization thus results in 48 DoFs for the optimizer. In the

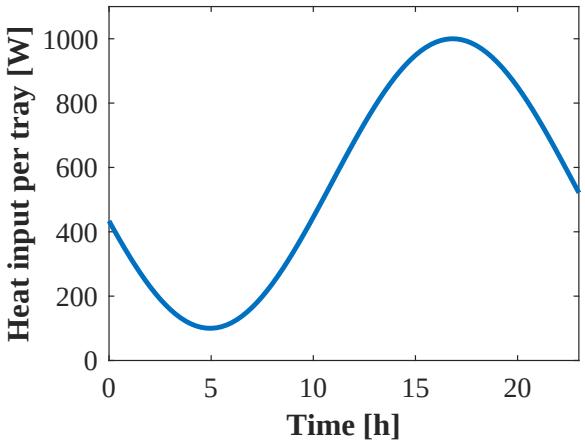

Figure 4.6.: Disturbance model for the computational study. An unmeasured heat input per tray is assumed following a sine curve.

first cycle, the optimization is initialized at the optimal steady-state in case of nominal production. All subsequent optimizations are initialized using the optimal solution from the previous iteration.

In order to limit the computational burden for solving the OCP (4.18a)-(4.18g), we rely on the suboptimal eNMPC scheme proposed by Caspari et al. [98], essentially utilizing ideas originally proposed by Diehl et al. [202]. More precisely, we allow for at most five major SQP iterations without performing an interim linesearch, which in fact means that we confine to a limited number of five DAE and sensitivity integrations, that are responsible for the vast majority of the computational time. We account for the solution times by assuming that the optimal MVs reach the process after a fixed dead-time of 30 s. Within that time, the optimal MVs from the last optimization are used further on. Comparing this to the results presented in Subsection 4.5.1, we find the assumed time delay to be even rather conservative. No further measures to compensate for the delay, such as delay compensation by prediction, etc. (cf. [203]), are taken. Note that selecting the number of DAE integrations corresponds to finding a balance between short time delays and better (i.e., more converged) solutions. The comparison of the performance of the presented framework to an ideal controller that uses the exact full-order process model and has perfect knowledge (cf. Subsections 4.5.2 and 4.5.3) indicates that the chosen parameters balance this trade-off in an appropriate way for the considered case study.

The economic performance of the presented eNMPC framework considering the introduced market, demand, and disturbance models is evaluated in comparison to an optimized nonflexible operation and the optimal flexible production schedule from solving an open-loop OCP offline assuming perfect knowledge about markets, demands, and disturbances and applying a full-order stage-to-stage model. In the following, we distinguish between operating scenarios without (SI.x) and with (SII.x) process disturbances. All scenarios are compared on the basis of equal cumulative production in the course of one day. This means that the net production that is achieved by the eNMPC framework defines the net production for the comparative cases. Note that this allows for a simple comparison of the electricity costs of each operating strategy.

Scenario SI.1 corresponds to the online application of the presented eNMPC framework

Table 4.3.: Characterization of operating scenarios without consideration of process disturbances (SI.1-SI.3). ANNCM abbreviates the ANN-based compartment model, FSM the full-order stage-to-stage model.

	SI.1 eNMPC	**SI.2 OCP**	**SI.3 Opt. steady-state**
Optimization	Online	Offline	Offline
Sampling time	15 min	-	-
Time delay	Yes	No	No
Control and prediction horizon	12 h	24 h	24 h
Control vector parametrization	1 h	15 min	24 h
Distillation model (eNMPC)	ANNCM	FSM	FSM
Distillation model (plant)	FSM	FSM	FSM
Electricity prices	Predicted	Known	Known
Customer demand	Predicted	Known	Known

under consideration of time delays due to solution times as well as imperfect knowledge about the market. This scenario is compared to two benchmark operation strategies. Scenario SI.2 is the solution of an offline OCP assuming perfect knowledge (no model/plant mismatch and perfect forecasts). It thus represents the economically optimal operation for a given sampling time that cannot be outperformed by the eNMPC. In contrast, scenario SI.3 corresponds to a nonflexible operation mode, i.e., the process is operated at a constant production rate yielding the same net production as the eNMPC framework in scenario SI.1. Scenario SII.1 essentially repeats the computations from scenario SI.1. However, the unmeasured disturbances (cf. Figure 4.6) are used as additional input for simulating the plant replacement's response. From a control perspective, this means that the controller now needs to reject disturbances in addition to the handling of model/plant mismatch. Scenario SII.2 again represents the ideal solution for benchmarking the performance of the eNMPC framework. Note that we herein also assume the disturbance profile to be known in the OCP. Finally, in scenario SIII.3, the optimal trajectories from scenario I.2 are used for simulating the process under disturbances in a feed-forward manner. This scenario is not relevant for the economic comparison, however, it allows for assessing the severeness of the considered disturbances. Short characterizations for all considered scenarios are given in Tables 4.3 and 4.4.

4.4. Performance in open-loop control

The OCP is implemented and solved using DyOS as described in Subsection 4.3.2. For SNOPT, we apply an optimality tolerance of 10^{-2} and a feasibility tolerance of 10^{-6}. For both integrators, NIXE and LIMEX, we apply absolute and relative tolerances of 10^{-6}. Dynamic responses of the plant replacement are simulated using the same integration tolerances. All calculations are conducted on an Intel® Xeon® CPU E5-2630 v2 with 2.60 GHz and 128 GB RAM.

Table 4.4.: Characterization of operating scenarios with consideration of process disturbances (SII.1-SII.3). ANNCM abbreviates the ANN-based compartment model, FSM the full-order stage-to-stage model.

	SII.1 eNMPC	**SII.2 OCP**	**SII.3 OCP**
Optimization	Online	Offline	Offline
Sampling time	15 min	-	-
Time delay	Yes	No	No
Control and prediction horizon	12 h	24 h	24 h
Control vector parametrization	1 h	15 min	15 min
Distillation model (eNMPC)	ANNCM	FSM	FSM
Distillation model (plant)	FSM	FSM	FSM
Electricity prices	Predicted	Known	Known
Customer demand	Predicted	Known	Known
Disturbances	Unknown	Known	Unknown

4.4.1. Prediction capabilities of the hybrid mechanistic/data-driven model

In Figure 4.7, we present the optimal profiles for selected MVs, namely the gaseous feed air (MV1) as well as the liquified and stored product (MV4), obtained when solving the OCP embedding the different models for the distillation column. As can be seen, there is an almost perfect agreement between the optimal MV profiles obtained when using the ANN-based compartment model, the classical compartment model, and the full-order stage-to-stage model within the OCP. In fact, the highest relative error compared to the optimal MV profile from using the full-order model is 0.07% for the ANN-based compartment model and 0.04% for the classical compartment model.

In Figure 4.8, we further show the actual responses of selected CVs, namely the nitrogen product purity (CV1) and the temperature at the turbine inlet (CV6), when simulating the plant replacement using the full-order stage-to-stage model and applying the optimal MV profiles and compare them to the corresponding predictions made in the OCP. We thereby assess the errors introduced through the reduction approaches, namely through both the compartmentalization and the surrogate modeling using ANNs. However, one should consider the errors in relation to the agreement between the full-order model and the actual responses of a physical plant. With regard to control applications, the model/plant mismatch is commonly handled efficiently in a closed-loop manner by using feedback of the process. Simply speaking, if errors that are introduced during the (hybrid) model reduction are below the inherent model/plant mismatch, the accuracy of the reduced (hybrid) model is not critical. For instance, in the case of the considered ASU, we assume a mismatch between physical plants and full-order stage-to-stage model according to Subsection 4.2.1 in the order of few ppm for the most important (and also the limiting) process variable, the nitrogen product purity (CV1), to be realistic.

The results given in Figure 4.8 validate the good prediction capabilities of the classical compartment model. In particular, our results are in good agreement with the findings

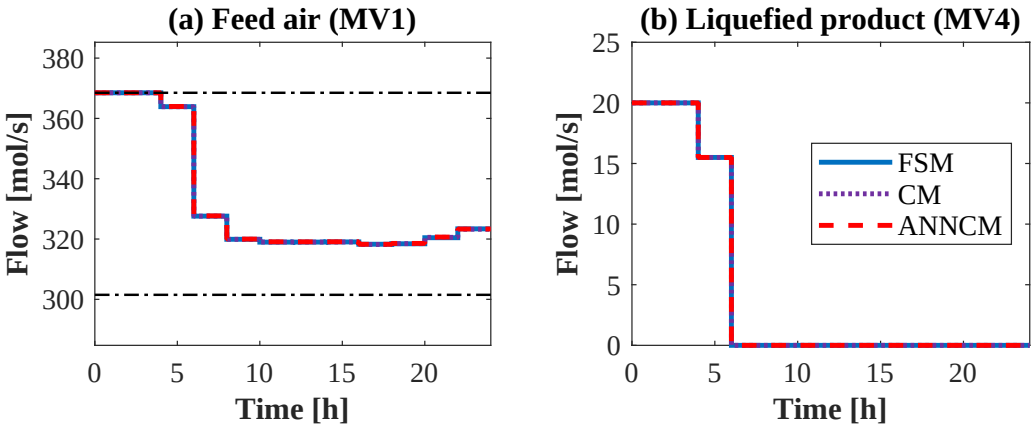

Figure 4.7.: Selected optimal MV profiles obtained when embedding the different distillation column models in the OCP. Solid blue trajectories are optimized using the full-order stage-to-stage model (FSM), dotted purple lines using the classical compartment model (CM), and dashed red lines using the proposed ANN-based compartment model (ANNCM). Thin dashed-dotted black lines give the respective bounds for the MVs. (a) MV1: feed air stream entering the main air compressor, (b) MV4: liquified product stream entering the storage tank.

from Bian et al. [119], indicating that for columns comprising around 40 or more stages, ten single stages can easily be combined into one compartment without significant influences on the dynamic responses. For instance, we find that for the impurity in the product stream (CV1), the root mean square error (RMSE) between the prediction using the classical compartment model and the actual response of the full-order stage-to-stage model is $\sim$0.5 ppm, which corresponds to an error of $\sim$2.5% relative to the impurity limit of 20 ppm. Following the discussion above, this accuracy appears adequate for control purposes, as we assume that a sufficiently large safety factor $\gg$2.5% would be added in practice to bounds on crucial CVs. Due to the derivation of the ANN-based compartment model, the accuracy achieved by classical compartmentalization represents the actual benchmark for the hybrid approach, i.e., a better agreement between the ANN-based compartment model and the full-order model than between the classical compartment model and the full-order model could only occur by chance through the vanishing of two error sources.

We further see from Figure 4.8 that - as expected - the ANN-based compartment model inherits the capabilities of the classical compartment model to reproduce the transient behavior of full-order models well while reducing the number of differential equations by 90%. More precisely, for the predictions of the ANN-based compartment model, an RMSE of 0.67 ppm in the product impurity (CV1) is found, which is only slightly worse than the RMSE when applying the classical compartmentalization. The small deviations between the two performances can be explained by the remaining prediction errors of the embedded ANNs. These errors are present both in transient and in (almost) stationary operation, where the classical compartment model achieves a perfect agreement between predictions and actual responses. However, since ANNs are universal approximators (cf. [185]), the prediction errors of the ANNs (i.e., the deviations between the hybrid and the classical compartmentalization) can be decreased to an arbitrary size if sufficiently many hidden neurons are applied and data points for training are used. In practice, the prediction

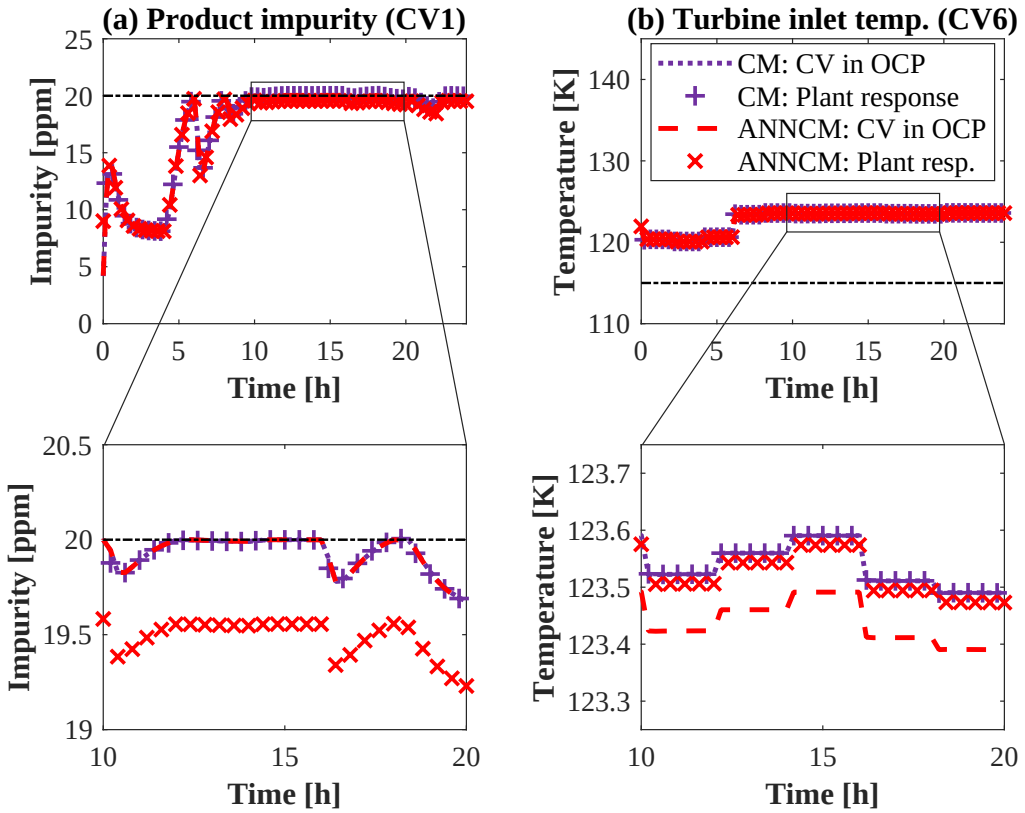

Figure 4.8.: Comparison of predicted trajectories of selected CV when solving the OCP embedding the reduced model to the actual simulated dynamic response of the plant replacement using a full-order stage-to-stage model. Dotted purple lines show predictions using the classical compartment (CM) in the OCP, dashed red lines the predictions using the ANN-based compartment (ANNCM). Symbols depict the dynamic responses of the full-order model when applying the obtained optimal MV profiles. Thin dashed-dotted black lines are the respective bounds for the variables. Lower figures show detailed views marked in the upper ones. (a) CV1: impurity in the nitrogen product stream, (b) CV6: absolute temperature of the feed air at the turbine inlet.

accuracy of the ANNs is nevertheless limited by the actual number of neurons applied and the size of the data set used for training. Using the obtained ANNs (cf. Appendix B), this additionally introduced error is always below 0.5 ppm for the product purity and thus already in the same order as the inherent limitations in the prediction accuracy of the compartmentalization approach applied. We highlight that the prediction errors of the ANN-based compartment model are thus also smaller than assumed safety factors and that the actual model/plant mismatch between real plants and full-order stage-to-stage models will likely be at least in a comparable order of magnitude. Qualitatively similar results are obtained for different process variables as well. For instance, the RMSE for predicting the temperature at the turbine inlet (CV6), which depends on the molar streams and enthalpies of the product and waste streams, is even below 0.1 K when applying the ANN-based compartment model.

Table 4.5.: Average CPU times over 10 runs for solving the OCP when using the different models for the distillation column.

Distillation model in OCP	CPU time (NIXE) [s]	CPU time (LIMEX) [s]
Full-order stage-to-stage model	2,490	4,909
Classical compartment model	14,928	12,099
ANN-based compartment model	162	123

4.4.2. Computational performance and savings in solution times

Table 4.5 shows CPU times for solving the OCP (4.18a)-(4.18g) when using both the ANN-based and the classical compartment model as well as when using the full-order stage-to-stage model. Results are presented separately for applying the integrators NIXE and LIMEX. We find that the ANN-based compartment model enables high-quality prediction capabilities (Figures 4.7 and 4.8) while substantially reducing the computational effort for solving the OCP by more than one order of magnitude compared to using a full-order model. This holds for both integrators tested. However, if LIMEX is applied, the benefits appear even stronger (factor of ∼40 when using LIMEX vs. ∼15 when using NIXE). Similar behavior can be observed for the comparison of the hybrid compartmentalization to the classical version. Again, the relative improvement is higher when using LIMEX. Noticeably, for both integrators tested, the use of the classical compartment model in our numerical set-up leads to a significant increase in computational times, that is, solution times when using the classical compartment model are higher by factors of 2.5 (LIMEX) and 6 (NIXE) compared to the full-order stage-to-stage model respectively. As discussed in Subsection 4.1.3, this behavior has also been observed in the literature [176].

An extensive study might become necessary in the future for verification and generalization of the exceptionally good computational efficiency of our reduction approach to other solution methods for dynamic optimization. In particular, we recommend the application of such a reduced hybrid mechanistic/data-driven model within simultaneous solution methods (e.g., [204]) as well for future research.

4.4.3. Comparison with literature results

The works of Khowinij et al. [118] and Bian et al. [119] for the application of classical compartment models, the works of Cao et al. [115, 116] for the application of collocation-based models, and the recent work by Caspari et al. [105] applying an extended nonlinear wave model appear suitable for comparing our results to findings from the literature, as similar processes are considered. Furthermore, these works apply full-order stage-to-stage models using the same assumptions as we did for benchmarking the respective reduction approaches. Khowinij et al. [118] report a fivefold reduction of computational times using MATLAB's ode15s (Mathworks, Inc.) as integrator for a single nitrogen column. For another column producing both nitrogen and oxygen tested by Bian et al. [119] using the same methods, only insignificant reductions were observed. In comparison, our results even show an increase in computational times by applying a classical compartment model instead of a full-order stage-to-stage model when using state-of-the-art integrators. However, we

do not exclude the possibility that highly tailored integration methods might be able to efficiently exploit the structure of the classical compartment model. We however validate that compartmentalization approaches (as long as not too many stages are combined within one compartment) are exceptionally suited for reproducing dynamic responses from full-order models and thus efficiently reduce the number of differential equations. We highlight that this fact is believed to be the key to the computational performance of the presented ANN-based compartmentalization approach.

Concerning the use of collocation-based models, Cao et al. [116] report a reduction in simulation times in gPROMS (PSE, Ltd.) of ∼70% when reducing the model size by ∼60% for a similar case study (optimal operation of a nitrogen plant with a liquid buffer subject to fluctuating electricity prices). The authors further report absolute errors in relevant product purities of below 1 ppm for that model [115], which appears comparable to the accuracy of the presented ANN-based compartment model. Unfortunately, the authors do only report time reductions for simulating the process responses and not for solving the OCP. However, as a first approximation, we estimate savings in solving the OCP to be of a similar order of magnitude. We thus expect the proposed hybrid compartmentalization approach to enable a substantially better balancing between model accuracy and computational demand than what could be achieved by using collocation-based approaches instead.

Finally, Caspari et al. [105] propose an extension of the nonlinear wave model from Kienle [104] accounting for varying column hold-ups by using integral dynamic balance equations and apply this model to both open-loop and closed-loop case studies considering the same process configuration as we do. Comparing their reported open-loop results to our work, we find slightly larger deviations from the full-order stage-to-stage model with regard to both the dynamic and the stationary behavior when applying the nonlinear wave model. Note that these deviations likely stem from (i) a higher degree of reduction of differential states and (ii) simplifying thermodynamic assumptions when deriving the solution of the wave equations. However, the authors demonstrate that the deviations can be handled effectively in closed-loop. Thus, the reported savings in computational times for solving OCPs, which are in a similar order of magnitude as we find for the ANN-based compartment model, make the nonlinear wave model a promising candidate for application in an eNMPC framework targeting the participation of the considered ASU in a continuous intra-day electricity market in real-time directly on the control layer. However, we emphasize that the prediction capabilities of nonlinear wave models - in contrast to those of the ANN-based compartment model - strongly depend on the nonideality of the considered thermodynamic system, so that their applicability to other columns in ASUs, particularly those with higher argon fractions, will certainly be limited.

4.5. Performance in closed-loop control

All computations are again conducted in the in-house software package DyOS. MATLAB R2018a (Mathworks, Inc.) is used for setting up the closed-loop framework and thus for sequencing the tasks. Within DyOS, integrator tolerances (both absolute and relative) are set to 10^{-2} when solving the OCP in the eNMPC layer and 10^{-4} for simulating plant responses. For solution of the benchmark calculations (cf. Tables 4.3 and 4.4), all tolerances (including optimality and feasibility tolerances of SNOPT) are set to 10^{-4}. All calculations

Table 4.6.: Solution times for the OCP in the eNMPC layer, i.e., the time for five major SQP iterations without interim linesearch. The first row corresponds to the eNMPC framework using the reduced hybrid model (scenarios I.1 and II.1). The second row corresponds to the same operating scenario when using the full-order model for comparison.

Distillation model in OCP	**CPU time [s]**		
	Mean	**Std. dev.**	**Max.**
ANN-based compartment model	19.00	1.90	26.75
Full-order stage-to-stage model	314.15	124.73	563.25

are conducted on an Intel® Xeon® CPU E5-2630 v2 with 2.60 GHz and 128 GB RAM.

4.5.1. Discussion of real-time capability

Average solution times spent in the eNMPC layer in closed-loop simulations of SI.1 and SII.1 are listed in Table 4.6. We further compare the solution times to a scenario without model/plant mismatch that is not real-time capable, i.e., using a detailed full-order stage-to-stage model in the OCP formulation (4.18a)-(4.18g) of the eNMPC . All other closed-loop settings, in particular the time delay, are kept constant. CPU time savings achieved by using the ANN-based compartment model instead of a full-order stage-to-stage model are ∼95% and therefore in good agreement with the open-loop results shown in Subsection 4.4.2. Furthermore, Table 4.6 gives a ratio of the standard deviations to the means of the solution times that is substantially smaller when using the reduced hybrid model (10% vs. ∼40%). This finding is of practical importance for suboptimal eNMPC schemes, as solution times and hence time delays for updating the MVs can be estimated very precisely.

We highlight that by using the presented eNMPC framework applying the ANN-based compartment model, the solution time spent in the eNMPC layer never exceeded the assumed time delay of 30 s. In fact, this means that the presented results for the eNMPC framework represent a conservative estimate. In practice, either a fixed solution time for the optimizer would allow for more SQP iterations and thus for more converged solutions or the updated inputs after a fixed number of SQP iterations would reach the process with less time delay in most of the cycles improving the ability of the control system to react on changes/disturbances.

4.5.2. Controller performance without disturbances

Optimal MV profiles for the gaseous feed air (MV1) and the liquified product (MV4) in SI.1, SI.2 and SI.3 are given in Figure 4.9. We demonstrate that a near-optimal operation can be achieved using the presented eNMPC framework. More precisely, one can see that only small deviations occur between the MV profiles computed in the eNMPC layer (SI.1) and the ideal MV profiles obtained by solving an OCP without model/plant mismatch and with perfect forecasts about electricity prices and customer demands offline (SI.2). Furthermore, we see that fluctuations around the optimal steady-state operation (SI.3) are substantial and extend nearly over the entire operating range of the process. These

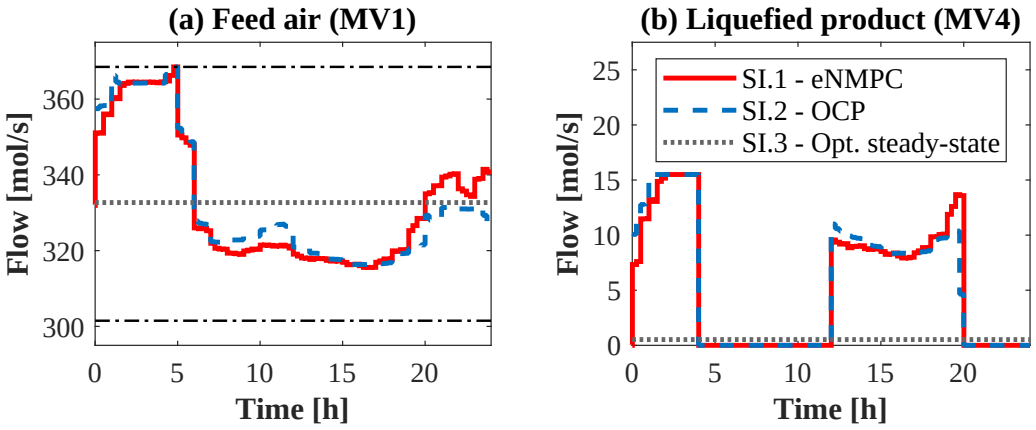

Figure 4.9.: Profiles for selected MVs in scenarios SI.1-SI.3. The solid red line corresponds to using the described eNMPC framework. The dashed blue line is the offline solution of an OCP with perfect knowledge. The dotted gray line gives the optimal stationary operating point for the process. Thin dashed-dotted black lines give the respective bounds for the MVs. (a) MV1: feed air stream entering the main air compressor, (b) MV4: liquified product stream entering the storage tank.

fluctuations (overproduction in times of low prices and reduced production in times of high prices) form the basis of economic improvements (cf. Table 4.7). We emphasize that even the offline solution of the OCP as the limiting case for the eNMPC does not yield control signals that are constant during one hour in accordance with the varying electricity prices, but rather show a ramping behavior. Note that this finding represents in fact an economic motivation for sampling times of the eNMPC below one hour irrespective of the compensation of model/plant mismatch and disturbance rejection.

Selected trajectories of CVs in SI.1-SI.3 are given in Figure 4.10. We find the product purity (CV1) to be again the limiting constraint. No further constraint is active at any time. These results highlight the crucial influence of product purities on process economics. Producing product in higher purity as required would decrease the economic performance. At the same time, this also causes the need for highly accurate process models for the distillation column. Most importantly, the eNMPC framework satisfies the product requirements at any time. However, the bound used within the eNMPC layer including the constraint back-off for the product purity is exceeded regularly due to a combination of model/plant mismatch and time delays in updating the MVs. Using hard path constraints for the CVs in the OCP thus regularly causes infeasibility at the initial conditions and consequently infeasible optimization problems. However, the conducted SQP iterations are found suitable for identifying MV trajectories that drive the impurity below its bound again within one sampling interval. The highest constraint violation occurs after ∼6 h, when the feed air drops by >10% within short time (cf. Figure 4.9 (a) and 4.10 (a)), causing a sharp peak in the impurity trajectory. This finding however is in good agreement with previous results that the highest deviations between full-order stage-to-stage models and compartmentalized models occur for fast input dynamics. Nevertheless, even a relatively low safety factor (19 ppm as bound instead of the actually required 20 ppm) appears already sufficient - in particular, when considering that the assumed time delay is estimated conservatively. Trajectories from the offline solution of the OCP and the

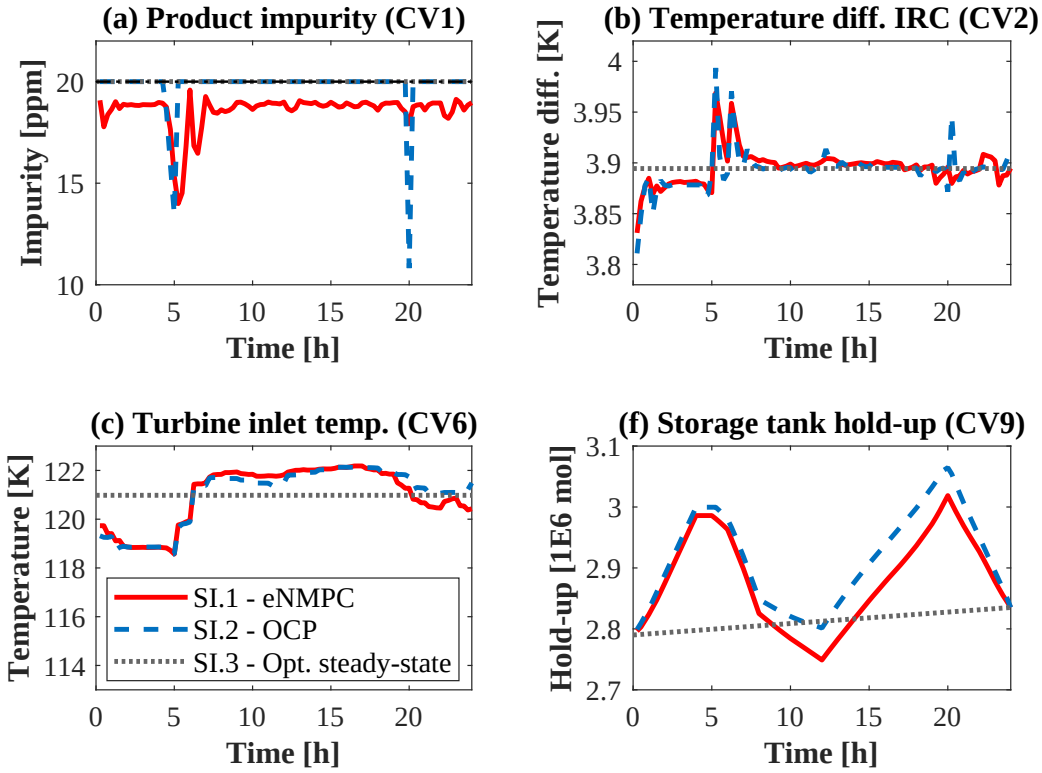

Figure 4.10.: Simulated closed-loop trajectories for selected CVs in scenarios SI.1-SI.3. The solid red line corresponds to using the described eNMPC framework. The dashed blue line is the offline solution of an OCP with perfect knowledge. The dotted gray line gives the optimal stationary operating point for the process. Thin dashed-dotted black lines are the respective bounds for the variables. Note that in SI.1, a safety factor is added to the bound, keeping the product impurity below its actual specification. (a) CV1: impurity in the nitrogen product stream, (b) CV2: driving temperature difference in the integrated reboiler and condenser, (c) CV6: absolute temperature of the feed air at the turbine inlet, (d) CV9: liquid hold-up of the storage tank.

steady-state process optimization are at the impurity bounds (nearly) all the time. Only if very steep changes in the MV profiles occur, the product impurity also decreases such that the constraint is temporarily not active. However, the trajectory reaches the bound again after less than two sampling times.

Comparing the economic performance for a given production target, which is defined by the net production when applying the eNMPC framework, in Table 4.7, we can quantify the near-optimal operation of the eNMPC . That is, given a quarter-hourly control vector parametrization and respecting the operating bounds from Tables 4.1 and 4.2, a maximum economic improvement of 0.63% versus an optimal stationary operation can be achieved for the electricity market fluctuations depicted in Figure 4.4. Approximately 75% of these savings can also be achieved by using the eNMPC framework presented, although we therein consider the inherent uncertainties at real-time markets as well as in customer demands. Note that these savings in the order of ∼0.5% in electricity costs are already promising regarding the production of commodities while at the same time safely satisfying all customer demands. Moreover, the considered price fluctuations are rather moderate,

Table 4.7.: Operational (electricity) costs in the considered scenarios. Costs are normalized to the offline OCP benchmark calculations (scenario SI.2 and SII.2) respectively.

	Scenario	Costs
Scenarios without disturbances	SI.1 - eNMPC	$100 + 0.16\%$
	SI.2 - OCP (offline)	100%
	SI.3 - Optimal steady-state	$100 + 0.63\%$
Scenarios with disturbances	SII.1 - eNMPC	$100 + 0.19\%$
	SII.2 - OCP (offline)	100%

such that even nowadays, one could find periods with potentially higher profit margins.

Finally, we briefly discuss the economic gap between the offline OCP (SI.2 and SII.2) and the eNMPC strategy (SI.1 and SII.1). For this purpose, we have conducted additional in silico studies varying the constitutive parameters of scenario SI.1 and (partially) aligning them with the ones of scenario SI.2 (cf. Table 4.3). In particular, we thereby assess the effect of (i) computational delays in updating the MVs, (ii) the use of imperfect market forecasts, and (iii) a limited prediction horizon below one entire day:

(i) The influence of the computational delay is negligible, i.e., when disregarding it, relative changes in performance-relevant process variables, such as cumulative production or total electricity costs, are $\ll 0.01\%$.

(ii) The effect of the imperfect electricity price forecasts is also comparably minor. This can be explained by the fact that the relevant price spread between the hours where the load is increased (1-6 h) and where it is reduced (7-21 h) is substantially larger than the noise term. Thus, price forecasts do not need to be perfect but only reproduce the relevant price spreads.

(iii) Among the three influences, increasing the horizon of the eNMPC shows the biggest impact. Nevertheless, by increasing the horizon from 12 to 24 h, the gap between the economic performance when solving the OCP offline as limiting case and the eNMPC framework is only reduced by one fourth coming at the price of a higher computational burden.

Summarizing, a small potential for optimizing the eNMPC parametrization by increasing the prediction horizon is identified here, as solution times and thus computational delays do not appear restrictive. However, the results also indicate that the most important fact causing the loss of economic performance stems from the model/plant mismatch. More precisely, by using the constraint back-off, one actively prevents the eNMPC from operating the plant at its impurity bound and thus from the most efficient operation (i.e., highest production for given feed and thus given electricity consumption).

4.5.3. Controller performance with disturbances

Optimal MV profiles for the gaseous feed air (MV1) and the liquified product (MV4) in SII.1, SII.2, and SII.3 are given in Figure 4.11. Note that the MVs in SII.3 are equal

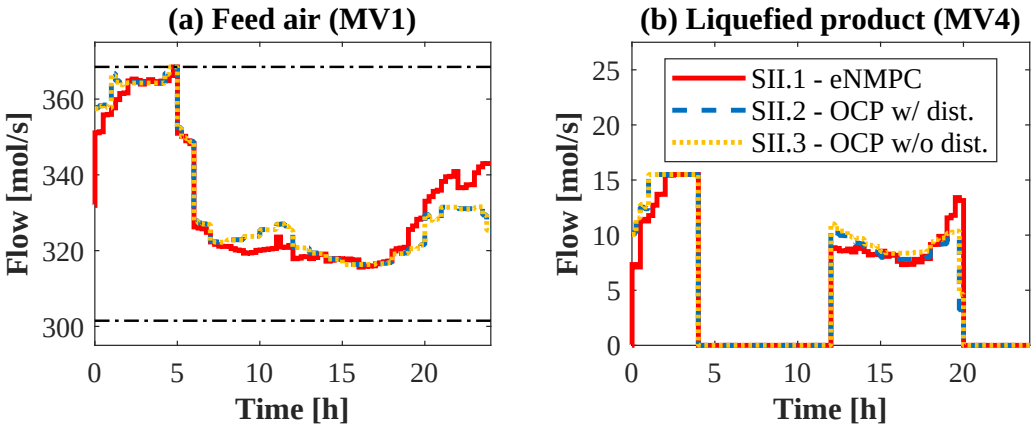

Figure 4.11.: Profiles for selected MVs in scenarios SII.1-SII.3. The solid red line corresponds to using the described eNMPC framework. The dashed blue line is the offline solution of an OCP with perfect knowledge. The dotted yellow line repeats the MV profiles from SI.2. Thin dashed-dotted black lines give the respective bounds for the MVs. (a) MV1: feed air stream entering the main air compressor, (b) MV4: liquified product stream entering the storage tank.

to those in SI.2 and only repeated for comparison purposes. As can be seen, the total inflow to the process (MV1) remains nearly unchanged, i.e., optimal offline solutions of the OCP with and without consideration of disturbances are very similar. Small differences occur in the amount of stored product (MV4). These differences however already explain how a constraint violation is circumvented by the control system: a temporarily lowered distillate-to-feed ratio is used for the column to enhance the separation effect, counteracting the additional heat input, i.e., the disturbances. Again, optimal MV profiles when using the eNMPC framework are close to those from the offline solution of the OCP and follow the same trends.

Selected trajectories of CVs in SII.1-SII.3 are given in Figure 4.12. This illustrates the severity of the assumed disturbance model. If no countermeasures are taken and the MV profiles from solving the OCP without consideration of the disturbances offline (SII.3) are applied, the additional heat input to the distillation column causes the product impurity (CV1) to substantially exceed its bounds by more than 10 ppm, which in fact corresponds to an impurity that is >50% higher than allowed. No other constraint violation occurs. However, the depicted influence on the process is already remarkable. Similarly to SI.2, using the MVs obtained by solving the OCP with perfect knowledge offline (SII.2) allows for operating the process at the limiting impurity bound nearly constantly. The process operation applying the eNMPC framework again results in low-amplitude fluctuations around the bound used in the eNMPC layer including the constraint back-off. Note that these fluctuations are neither more frequent nor of a substantially increased amplitude compared to the results in Figure 4.10 (a). We highlight that again no violation of the actual product requirement is observed, even when applying the same safety factor as in the nondisturbed case. Generalizing the results, we conclude that the presented eNMPC framework enables an effective rejection of disturbances with slow dynamics (e.g., disturbances caused by differences between day and night, etc.) by utilizing feedback from the process on a quarter-hourly scale.

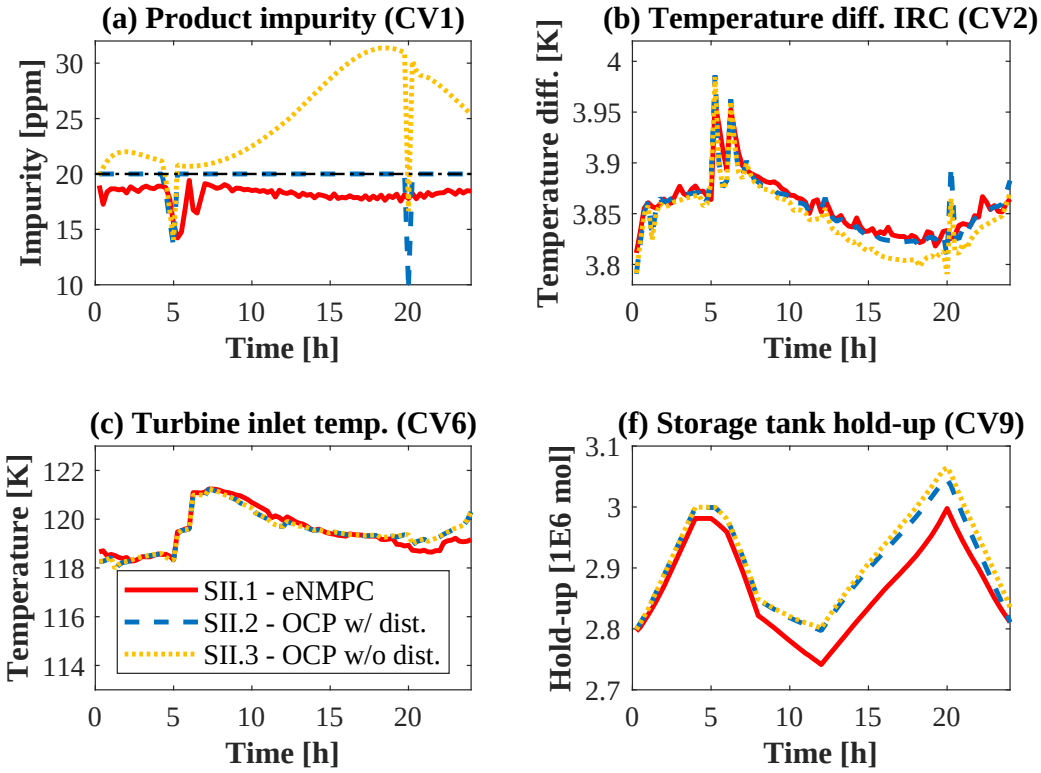

Figure 4.12.: Simulated closed-loop trajectories for selected CVs in scenarios SII.1-SII.3. The solid red line corresponds to using the described eNMPC framework. The dashed blue line is the offline solution of an OCP with perfect knowledge. The dotted yellow line gives the response of the plant when using the MV profiles from SI.2. Thin dashed-dotted black lines are the respective bounds for the variables. Note that in SII.1, a safety factor is added to the bound, keeping the product impurity below its actual specification. (a) CV1: impurity in the nitrogen product stream, (b) CV2: driving temperature difference in the integrated reboiler and condenser, (c) CV6: absolute temperature of the feed air at the turbine inlet, (d) CV9: liquid hold-up of the storage tank.

A second economic evaluation is conducted for assessing the performance of the eNMPC framework in operating scenarios with disturbances. The lower rows of Table 4.7 give a comparison of the actually achieved operation when applying the eNMPC framework with the results from the offline solution of the OCP - again on the basis of equal net production. Note that the comparison to an operating schedule with a constant operation is not informative, as the operating point would have to be chosen in a way that even the strongest disturbance intensity would not cause a constraint violation. Note that in reality, one would instead use a subordinated regulatory control maintaining the desired set-points for the CVs. Besides, a comparison of SII.1 and SII.2 to SII.3 is pointless, as in II.3, a significant violation of the product requirements occurs. Furthermore, we highlight that the results for SII.1 and SII.2 are based on a different production target than the results for SI.1-SI.3. Comparing the costs per produced amount of nitrogen allows for estimating the impact of the chosen disturbance model on process economics. That is, comparing both SI.1 with SII.1 and SI.2 with SII.2 gives an increased cost per mole of nitrogen by ∼1%. This loss in economic performance can be explained by the countermeasures for

circumventing the constraint violation, which increases the amount of feed air processed and thus the electricity consumption per amount of product. Nevertheless, the results for SII.1 and SII.2 are in good agreement with the results for SI.1 and SI.2. That is, even in the presence of unmeasured disturbances to the process, the economic performance of the eNMPC framework is close to solving an OCP with perfect knowledge offline.

4.6. Summary and recommendations for future research

We propose a reduced hybrid mechanistic/data-driven dynamic model for distillation columns that is based on compartmentalization and combines it with ANNs. The proposed ANN-based compartment model benefits from the principle of a separation by time scales to efficiently reduce the number of differential equations. Time-consuming stage-to-stage calculations for the intra-compartment input/output relations are avoided by using ANNs as explicit surrogate models. We successfully apply the reduced hybrid model in a case study addressing the optimal flexible operation of a nitrogen plant. The proposed ANN-based compartmentalization reduces the number of differential equations for modeling the distillation column by 90% compared to a full-order stage-to-stage model. The prediction error of the reduced hybrid model compared to the open-loop response of the full-order model is below 1 ppm in the limiting product purity. This error is partially introduced by the compartmentalization itself, additional errors introduced by using the ANNs are in a comparable order of magnitude. Moreover, we demonstrate that by embedding the ANN-based compartment model into an OCP, solution times can be substantially reduced compared to using the full-order stage-to-stage model or the classical compartment model by more than one order of magnitude.

We further present a real-time capable single-layer eNMPC framework combining the proposed reduced hybrid dynamic model with a suboptimal fast update method. That is, when solving the OCP in closed-loop on a moving horizon, we confine to a limited number of major SQP iterations without interim linesearch. The eNMPC framework is applied to the control of the nitrogen plant and targets the participation in a continuous intra-day electricity market. In the in silico study, we explicitly consider the mismatch between the reduced hybrid model used in the OCP formulation and the detailed full-order model used as plant replacement. We demonstrate that the single-layer eNMPC framework enables adjustments of the production schedule in real-time, which are based on newly available information concerning the uncertain electricity market, i.e., updated price forecasts, and allows for reacting on upcoming changes in on-site customer demands. Applying the eNMPC framework further allows for economic savings compared to stationary process operation strategies that are close to the maximum achievable savings when assuming perfect knowledge, i.e., no model/plant mismatch, perfect price forecasts, and known on-site demands. Additionally performed investigations considering the presence of an unmeasured heat input to the column as disturbance show that the proposed eNMPC framework can also effectively reject disturbances with slow dynamics (e.g., fluctuations between day and night).

Important research potentials can be identified that should be addressed in the future. From the modeling perspective, the derivation of the ANN-based compartment model should be thoroughly analyzed. In particular, we recommend detailed studies on both the compartmentalization itself, i.e., on which and how many stages to aggregate, and

the surrogate modeling, i.e., on which surrogate model structures to apply for modeling which input/output relations. These studies are intended to validate the insights gained so far. That is, a minimum number of compartments is required to adequately reproduce dynamic responses for given input signals and any introduction of further compartments unnecessarily prolongs solution times. Moreover, our results suggest that the surrogate model complexity does not substantially influence computational times and should thus not be limiting. Detailed empirical knowledge would thereby enable further improvements in the model performance with regard to both the prediction accuracy and computational efficiency. Along these lines, we also recommend considering the previously described opportunities to enhance the performance of the surrogate models by using more sophisticated sampling and training methods. Finally, due to the good performance of the extended nonlinear wave model presented by Caspari et al. [105] in case of rarely nonideal thermodynamic behavior, we consider a combination of both approaches to be a promising perspective for modeling of complex distillation systems that occur, e.g., in larger ASU configurations. More precisely, one can think of using compartment balances embedding wave equations for column sections that can be represented well with this approach, i.e., sections with nearly constant relative volatilities and molar overflow, and use ANN-based compartments for the others.

Future work should thus also address case studies with larger ASU configurations comprising multiple distillation columns producing oxygen and argon as well (e.g., [16, 18, 86, 100]). To this end, replacing the suboptimal update method through more sophisticated sensitivity-based update methods (cf. [205–207]) appears promising to retain low sampling times. If necessary, the application of two-layer eNMPC structures (cf. [208, 209]) represents a further alternative that could allow for the separation in time scales between economic optimization (acceptable if performed on slower scales) and disturbance rejection (faster time scale than disturbance is required). Along these lines, the aforementioned opportunity to apply the methods proposed in Chapter 3 to dynamic optimization problems subject to time-variable electricity prices bears great potential for reducing the dimensionality of the OCP and thereby further reduce the computational burden. With regard to the practical implementation, the eNMPC framework also needs to be extended to cope with a limited number of available measurements and thus requires the integration of state-of-the-art methods for state estimation (cf. [210, 211]).

Finally, as analogies between mathematical models for distillation columns and those for other spatially distributed process units have already motivated the adaptation of classical dynamic model reduction approaches for distillation columns to further spatially distributed process units, including fixed bed reactors [174, 212] and heat exchangers [115, 174], using the reduced hybrid mechanistic/data-driven modeling approach for these purposes appears highly promising. In particular, we consider the operation of flexible energy processes that are designed to cope with fluctuating sources, e.g., organic Rankine cycles for waste heat recovery from transient heat sources (e.g., [213]), to be a promising field of application. Similarly to the distillation columns considered in this chapter, these processes are governed by strong nonlinearities and transient behavior - this time in the heat exchangers - motivating the development of advanced control methods that require low-dimensional yet accurate nonlinear models. Here, we emphasize that many relevant applications involve the appearance and disappearance of phases, e.g., condensation/evaporation during start-up/shut-down of flexibly operated distillation columns (e.g., [161]) or moving phase boundaries in transiently operated evaporators (e.g., [214]. Consequently,

future work should also revisit the proposed hybrid mechanistic/data-driven compartmentalization in the light of optimization-oriented model formulations accounting for discrete events, such as appearing/disappearing phases (cf. [171, 173]).

5. Model-based bidding strategies for optimal decision-making at pay-as-bid markets

In this chapter, we formulate the optimal decision-making at a pay-as-bid market for ancillary services, namely the German market for PRL, as a two-stage stochastic program. Therefore, in Section 5.1 we first provide a brief characterization of the PRL market and compare it to the day-ahead spot market, whose price spreads are exploited by a subordinated scheduling optimization. Afterwards, we present the proposed two-stage stochastic problem formulation in Section 5.2. Therein, we first introduce the first and second stage problem formulations, i.e., the optimization of the actual bidding strategy and the subordinated scheduling problem. Then, we discuss a tailored solution approach based on enumeration, decomposing the MINLP formulation of the deterministic equivalent into a finite number of NLPs and MILPs. In Section 5.3, the case study for assessing the economic benefits of the optimized bidding strategies is presented. To this end, we first introduce an autoregressive time-series model for forecasting the upcoming maximum accepted ask price at the PRL market and give the corresponding formulation of the first-stage. Second, we present an MILP scheduling model for the operational optimization of a flexibilized aluminum electrolysis cell. The results of the case study are shown in Section 5.4. Therein, we first discuss solution times as well as the scaling of the algorithm with respect to the number of admissible bids. Then, we present the results of the economic performance evaluation for the optimized bidding strategies. In particular, we also compare the savings in operational costs to the pure exploitation of spot market price spreads. Moreover, the influence of constitutive parameters of the subordinated scheduling model on the optimal bidding strategy is investigated within a parameter study and the goodness of price forecasts is discussed. Finally, the chapter is concluded and further research perspectives are sketched in Section 5.5.

5.1. Characterization of the considered electricity markets

The previous chapters focus on the participation of flexible processes in energy-only markets, such as the day-ahead or the continuous intra-day market, where physical amounts of electricity are traded for a fixed time of delivery following the principle of balancing supply and demand. In contrast to that, in ancillary service markets, participants offer balancing capacities that can be retrieved to maintain the stability of the electricity grid. In turn, participants that reserve capacities for this purpose receive compensation payments from the TSO. Among the German markets for ancillary services, the market for PRL is of particular importance, as provided PRL capacities are retrieved automatically within a few seconds if the grid frequency exceeds/falls below a threshold value. Consequently, high

hurdles exist for entering the PRL market. At the same time, offering PRL is economically promising due to valuable compensation payments. As this chapter focuses on the participation in the PRL market, its main market rules and mechanisms are explained in the following.

The PRL market is a weekly auction, where the participants are allowed to place single or multiple bids. The number of bids allowed to be submitted by one participant is not restricted by market regulations. Each individual bid consists of two parts: (i) the amount offered, i.e., the capacity reserved for PRL that the participant is willing to provide and (ii) the corresponding ask price, i.e., the compensation payment that the participant is willing to accept for providing the offered PRL capacity. Once the auction is closed, those bids with the lowest ask prices are accepted until the total demand for PRL that the TSO has communicated is satisfied. In contrast to the day-ahead market as well as several of the US-American ancillary service markets, there is no uniform pricing. Instead, if a bid is accepted, the corresponding individual ask price is paid (pay-as-bid). Also if multiple bids have been placed by the same bidder, each one is compensated individually at the corresponding ask price. The smallest amount that can be offered is 1 MW. The increment is also 1 MW. Note that the case study presented below considers the circumstances from 2018 as described above and that the regulations for ancillary service markets are steadily adjusted. We however emphasize that this does not affect the applicability of the proposed methods in principle in case of pay-as-bid auctions.

5.2. Two-stage stochastic problem formulation

The key idea of the presented approach is to identify a bidding strategy that minimizes expected weekly production costs. For this purpose, mathematical formulations are introduced for both the bidding at the PRL market and an integrated scheduling model, which considers alternative marketing opportunities at the day-ahead spot market. Thereby, a two-stage stochastic problem formulation arises, where optimizing the bids to be placed at the PRL market is the first stage and marketing the remaining flexibility at the spot market is the second stage. Afterwards, we present a solution approach based on enumeration.

5.2.1. Mixed-integer nonlinear bidding problem

The optimal bidding strategy is defined as the set of bids, which consists of offered PRL capacities and the corresponding ask prices, that lead to the lowest production costs. Note that revenues from the PRL market, i.e., the compensation payments from the TSO, are herein considered as negative costs. As the maximum accepted ask price is not known beforehand, selecting the optimal bidding strategy is not straight-forward. Instead, one has to balance between higher potential revenues and a lower probability of acceptance of the bid if high prices are asked. If a distribution is available for the maximum accepted ask price, the bidding problem can be formulated as a stochastic program. Therein, the objective EC corresponds to the expected weekly production costs, which can be written as a probability-weighted sum over different scenarios. In the following, we present the formula for the general setting where up to N bids can be made (cf. (5.2)). Here, N pairs of capacities P_j^{PRL} [MW] and ask prices A_j^{PRL} [EUR/MW] are optimized. These are expressed as vectors $\boldsymbol{P^{PRL}}$ and $\boldsymbol{A^{PRL}}$. The N ask prices $\boldsymbol{A^{PRL}}$ are further assumed to

be ordered, i.e.,

$$A_j^{PRL} \leq A_{j+1}^{PRL},\ j = 1 \dots N \tag{5.1}$$

Note that P_j^{PRL} is defined as the additionally offered capacity and does not include $P_1^{PRL} \dots P_{j-1}^{PRL}$. For calculating the expected costs EC, $N+1$ mutually exclusive scenarios have to be considered that are defined as follows: in scenario i, the i bids with the lowest ask prices, i.e., the bids $1 \dots i$ are accepted and the respective compensation payments are received. All other bids $i+1 \dots N$ are refused. We also introduce scenario 0, in which no bid is accepted. $p_0 \dots p_N$ denote the probability for the respective scenarios, such that $\sum_{i=0}^{N} p_i = 1$ holds. The revenues from the PRL market that need to be considered in scenario i can be calculated as $\sum_{j=1}^{i} P_j^{PRL} \cdot A_j^{PRL}$. C_i gives the corresponding weekly production costs in scenario i without consideration of the revenues from the PRL market. Herein, the C_i's are calculated via an integrated scheduling model, which further allows for exploiting price spreads on the day-ahead spot electricity market using the remaining operational flexibility. We thereby follow similar considerations as Dowling et al. [122] and Otashu and Baldea [123] (cf. Subsection 2.1.3) and market capacities that have remained unused in a longer-term market optimization on a shorter-term perspective. Note however that herein, the definitions of long- and short-term markets are adjusted. The corresponding scheduling model is presented and discussed in detail in Subsection 5.2.2 and corresponds to a subordinated optimization problem, such that a two-stage stochastic formulation arises.

Finally and following the discussion above, the expected costs can be calculated as:

$$EC = p_0 \cdot C_0 + \sum_{i=1}^{N} p_i \cdot \left(C_i - \sum_{j=1}^{i} P_j^{PRL} \cdot A_j^{PRL} \right). \tag{5.2}$$

Therein, we assume that if a bid is accepted, i.e., if the ask price is lower than the maximum accepted ask price, the complete offered capacity is compensated irrespective of the actual demand of the TSO. Since the flexibility potential of most real processes is much lower than the demand of the TSO (order of 1 GW), this assumption appears to be valid. If the distribution of the maximum accepted PRL ask price is further known, the probabilities for all scenarios $p_0 \dots p_N$ can be calculated by using the cumulative distribution function $F(\cdot)$. The probability for scenario 0, in which all bids are refused, is equal to the probability that the maximum accepted PRL ask price is below the lowest ask price, i.e., lower than A_1^{PRL}, and thus reads:

$$p_0 = F\left(A_1^{PRL}\right). \tag{5.3}$$

The probability for all intermediate scenarios $i = 1 \dots N-1$, i.e., the probability that all bids up to i'th one are accepted, equals the probability that the maximum accepted PRL ask price is between A_i^{PRL} and A_{i+1}^{PRL}, giving:

$$p_i = F\left(A_{i+1}^{PRL}\right) - F\left(A_i^{PRL}\right),\ i = 1 \dots N-1. \tag{5.4}$$

Finally, the probability of scenario N, in which all bids are accepted, is the probability that the maximum accepted PRL ask price lies above the highest ask price (A_N^{PRL}) and

can be calculated by using the summation relation:

$$p_N = 1 - F\left(A_N^{PRL}\right). \tag{5.5}$$

As discussed in Section 5.1, the market imposes increments of 1 MW and thus the N capacities to offer $\boldsymbol{P^{PRL}}$ are integer variables with values between 1 MW (smallest amount to offer) and P_{max}^{PRL}, i.e.,

$$P_i^{PRL} \in \{1, 2, \ldots, P_{max}^{PRL}\},\ i = 1 \ldots N. \tag{5.6}$$

Note that in order to lower the computational demand of the optimization, a continuous approximation of $\boldsymbol{P^{PRL}}$ might be possible. However, this approach will not be considered in the following. Further note that P_{max}^{PRL} does not stem from market regulations, but from flexibility limits of the process and thus requires physical insights or expert process knowledge. Consequently, the sum of all offered capacities P_i^{PRL} is also not allowed to exceed P_{max}^{PRL}, such that the following inequality holds:

$$\sum_{i=1}^{N} P_i^{PRL} \leq P_{max}^{PRL}. \tag{5.7}$$

5.2.2. Integrated mixed-integer linear scheduling model

As stated above, for a specific scenario i, the weekly production costs C_i without revenues from the PRL market are calculated via an integrated scheduling optimization. Therein, the remaining operational flexibility, i.e., the entire flexibility range of the process minus the capacity reserved for the provision of PRL in scenario i, is used to optimally exploit spot market price spreads and thus to minimize C_i (cf. (5.8)). Following the discussion in Subsection 2.1.1, we herein confine to discrete-time scheduling models, as the overwhelming majority of DSM-related scheduling formulations relies on this time representation. Note that this assumption is however only made for illustration purposes and not necessary for the applicability of the presented approach.

When evaluating the scheduling objective, we consider - in alignment with the case study presented in Section 5.3 - both electricity costs, which are the sum of the products of the (hourly) day-ahead electricity spot market prices c_t^E and the instantaneous electricity consumptions $P_{i,t}$ for each time interval t, and opportunity costs due to production losses Δm_i^P multiplied with the product market price c^P. Beside these components, other subordinated scheduling problems might also require the consideration of additional raw material usage (see, e.g., [21]). T denotes the horizon length of the scheduling problem, that is $T = 168$ for an hourly discretization in case of a weekly auction. The objective for the subordinated scheduling problem finally reads:

$$C_i = \sum_{t=1}^{T} c_t^E \cdot P_{i,t} + c^P \cdot \Delta m_i^P. \tag{5.8}$$

As done in the case studies of Chapter 3, we omit the multiplication with the interval length $\Delta t = 1$ h for readability. We further assume that when the bidding takes place, day-ahead prices for the entire scheduling horizon are known or rather that highly accurate forecasts are available.

In this chapter, we focus on processes that can be operated continuously within physical limits, i.e., the energy consumption is limited between a minimum and a maximum value P_{min} and P_{max}. Once an offered PRL capacity is accepted, part of this flexibility range is given to the TSO, who can use it to maintain the grid frequency within desired bounds. Recap that in scenario i, the i bids with the lowest ask prices are accepted. Consequently, the total capacity that is reserved for PRL in scenario i, denoted by $P_{Cap,i}^{PRL}$, reads:

$$P_{Cap,i}^{PRL} = \sum_{j=1}^{i} P_j^{PRL}. \tag{5.9}$$

A 100% availability of the accepted PRL capacity is strictly demanded. Therefore, the original operational flexibility range of the process in scenario i is reduced by $P_{Cap,i}^{PRL}$ (cf. (5.10)). This actually corresponds to a very conservative scheduling, as the total PRL capacity of the process will likely be retrieved in almost none of the intervals. Nevertheless, the strict formulation is justified, as the risk of penalties ranging up to losing access to the PRL market for not being able to provide the required capacity is prevented. Note that PRL capacities have to be offered symmetrically, i.e. both positive and negative powers can be retrieved by the TSO.

$$P_{min} + P_{Cap,i}^{PRL} \leq P_{i,t} \leq P_{max} - P_{Cap,i}^{PRL}, \quad \forall t. \tag{5.10}$$

The equation above represents the link between the scheduling and the bidding model from Subsection 5.2.1. It is worth emphasizing that only the offered PRL capacity has a parametric influence on the subordinated scheduling model and not the ask price. This fact will be used for the solution approach presented in the next subsection. For setting up the actual process model and constraints, we refer to the extensive DSM-related literature cited in Subsection 2.1.1. In particular, we refer to the numerous approaches to formulate scheduling problems subject to time-variable electricity prices as MILPs. A consideration of subordinated (MI)NLP scheduling problems (cf. Chapter 3) would not prohibit the application of the solution approach presented below in principle. Nevertheless, the computational performance of the solution approach relies on the efficient solvability of the subordinated scheduling problems for fixed $P_{Cap,i}^{PRL}$ as discussed in the following.

5.2.3. Solution approach based on enumeration

Concatenating the equations from Subsection 5.2.1 and those from Subsection 5.2.2 for all scenarios i, a comprehensive MINLP formulation arises determining the optimal bidding strategy by minimizing the expected costs EC. In fact, this corresponds to the deterministic equivalent of a two-stage stochastic program (cf. [215]). Due to a very large number of additional variables introduced by the subordinated scheduling model, a straight-forward solution of the deterministic equivalent with standard solvers appears prohibitive. In fact, for the case study presented in the next section, the direct optimization of the deterministic equivalent has been tried in pre-studies using state-of-the-art commercial MINLP solvers. However, no converged solution could have been obtained within an acceptable time. Thus, a different solution strategy is proposed.

Optimization of the bidding strategy can be equivalently reformulated into the two-stage stochastic program (5.11a)-(5.11e). Therein, the bids to be placed at the PRL

market (the amounts to offer and the prices to ask) are the first-stage decision variables. The weekly production costs in scenario i without consideration of revenues from the PRL market, denoted by C_i, is the optimal value of the second-stage problem, in which the scheduling problem for a specific PRL capacity to be provided is solved, i.e., the hourly energy consumption $P_{i,t}$ is optimized, depending on the outcome of the PRL market auction.

$$\min_{\boldsymbol{P^{PRL}},\boldsymbol{A^{PRL}}} \quad EC \tag{5.11a}$$

$$\text{s.t.} \quad (5.1)\text{-}(5.7) \tag{5.11b}$$

$$C_i = \min_{P_{i,t}} (5.8),\ i = 1 \dots N \tag{5.11c}$$

$$\text{s. t.} \quad (5.9), (5.10) \tag{5.11d}$$

$$\text{Process model equations} \tag{5.11e}$$

The second-stage optimal values C_i only depend on the integer upper-level variables $\boldsymbol{P^{PRL}}$. Consequently, we herein solve (5.11a)-(5.11e) by enumeration. More precisely, the second-stage problem (MILP) is parametrically solved for all possible realizations of $\boldsymbol{P^{PRL}}$, which in fact means for all possible $P_{Cap,i}^{PRL} \in \{1, 2, \dots, P_{max}^{PRL}\}$. Then, the optimal objective values of the second-stage are inserted in the first-stage problem (NLP), which is solved for the optimal value of $\boldsymbol{A^{PRL}}$ given a fixed realization of $\boldsymbol{P^{PRL}}$. This approach appears to be suitable under the following assumptions:

- There is no combinatorial explosion in the number of subproblems to solve. This appears to be the case in practice. As shown in Subsection 5.4.3, a small number of bids N to be placed already suffices. Thus, while scaling in the number of possible realizations of $\boldsymbol{P^{PRL}}$ and hence the number of NLP subproblems to be solved is exponential, i.e., considering a process with $P_{max}^{PRL} = 10$ MW as done below, there are 10 possible realizations if $N = 1$, 45 if $N = 2$, 120 if $N = 3$, and so on, this is not of practical importance.
- The single subproblems can be solved within a short time. This assumption appears to be valid for the majority of MILP scheduling problems considered in literature so far if state-of-the-art MILP solvers are used. If N is furthermore restricted to a small number (see, above), global solution of the nonconvex NLPs also appears unproblematic due to the low number of optimization variables $[\boldsymbol{P^{PRL}}^T, \boldsymbol{p}^T]^T \in \mathbb{R}^{2N+1}$.

Note here that the number of NLPs as well as their complexity are completely independent of the considered process. Consequently, the performance of the algorithm for a fixed number of bids exclusively depends on the computational tractability of the subordinated scheduling optimizations. Consequently, the approach appears adequate for the vast amount of scheduling problems that have been formulated as MILPs so far in the literature. Otherwise, the use of structure-exploiting decomposition algorithms might become necessary for the solution of the two-stage stochastic program (5.11a)-(5.11e). Along these lines, we emphasize that the presence of mixed-integer second-stage variables due to the aforementioned linearization of nonlinear functions and/or the reformulation of logical constraints substantially complicates the decomposition as it prevents the application of classical Benders cuts (cf. [216]).

5.3. Case study: Flexible aluminum production at the market for primary balancing power

Due to its high energy demand of around 14 MWh per tonne, the aluminum electrolysis is promising for performing DSM. Furthermore, efficiency losses in periods of off-design operation are rather small compared to other processes. The thermal energy balance of the electrolysis cell is however crucial for the process. That is, the process can be operated in over- or underload only for a limited period. Otherwise, in case of a too high power uptake, the cell gets damaged. Likewise, if the power uptake is too small, the molten electrolyte freezes. The electrolytic process considered in this chapter (a series connection of single electrolysis cells), which is located in Essen, Germany, and operated by TRIMET Aluminium SE, has a nominal power uptake of 90 MW. For flexible operation, the power uptake can be varied by ± 22.5 MW (i.e., ±25%). The additionally accumulated thermal energy is limited between 48 operating hours with maximum power uptake and 48 hours with minimum power uptake. As the provision of PRL capacities is accompanied by highly frequent fluctuations in the power uptake, which might strongly influence the process if occurring with large amplitudes, the maximum PRL capacity is restricted to below the maximum flexibility range of 22.5 MW and set to 10 MW.

In the following, we first briefly introduce the model for the PRL market regression that allows for computing the (expected) probabilities of acceptance for bids with given ask prices. Afterwards, we present an MILP formulation for the scheduling of the described aluminum electrolysis.

5.3.1. Market regression and formulation of the bidding problem

We adapt the approach of Ierapetritou et al. [5] to forecast the development of the maximum accepted PRL ask price. In that work, the authors originally forecast the course of spot market prices by using a generalized form of autoregressive moving average (ARMA) models that is superimposed by a constant daily pattern. For a detailed treatment of ARMA models, we refer to the textbook of Box et al. [217]. Adapting the approach, we herein assume the maximum accepted PRL ask price $\widetilde{A}^{PRL}$ for the considered week to consist of three components:

$$\widetilde{A}^{PRL} = f^{PRL}_{Short-term} \cdot f^{PRL}_{Season} \cdot \widetilde{A}^{PRL}_{Base}$$

The factors $\widetilde{A}^{PRL}_{Base}$ and f^{PRL}_{Season} account for long-term developments and yearly patterns. Mathematical expressions for both of them can be found in Appendix C. The logarithm of the short-term factor $f^{PRL}_{Short-term}$ is regressed using a stationary ARMA(P_{ARMA},Q_{ARMA}) model. Different values for P_{ARMA} and Q_{ARMA} are tried. Parameter estimation for all models is conducted in the MATLAB R2017a Econometrics Toolbox (MathWorks, Inc.). The model parameters are fitted to historical data for the maximum accepted ask price for the years 2014-2017. All data concerning PRL is retrieved from the central Internet platform `www.regelleistung.net`. As the historical time series data shows a significant autocorrelation, only models with $P_{ARMA} \geq 1$ are considered. Parameters of all identified models, as well as the Akaike information criterion AIC (cf. [218]), are given in Table 5.1. Although one finds that an increased model order slightly reduces the variance of the noise term compared to the model with only one autoregressive parameter, i.e., the ARMA(1,0)

Table 5.1.: Parameters of the identified ARMA(P_{ARMA},Q_{ARMA}) models. ϕ_p and θ_q are the regression parameters, σ^2 is the model variance. AIC gives the Akaike information criterion.

Model	$\boldsymbol{\phi_1}$	$\boldsymbol{\phi_2}$	$\boldsymbol{\theta_1}$	$\boldsymbol{\theta_2}$	$\boldsymbol{\sigma^2}$	$\boldsymbol{AIC}$
ARMA(1,0)	0.8715	-	-	-	0.0062	-465.72
ARMA(1,1)	0.9276	-	-0.2369	-	0.0059	-474.66
ARMA(2,0)	0.6578	0.2461	-	-	0.0058	-476.55
ARMA(2,1)	0.5026	0.3812	0.1659	-	0.0058	-474.82
ARMA(2,2)	1.4644	-0.5170	-0.8081	0.2631	0.0058	-474.20

model, the Akaike criterion indicates the suitability of the ARMA(1,0) model. In fact, all higher-order models appear to overfit the relation, i.e., to introduce additional parameters with only an insignificant influence on the model accuracy. Consequently, we will confine to the ARMA(1,0) model, which will be referred to as AR(1) model in the following.

The use of the AR(1) model for predicting the logarithm of the short-term factor $f^{PRL}_{Short-term}$ leads to small adjustments of the bidding problem in contrast the formulation presented in Subsection 5.2.1. Further note that by following the presented solution approach (cf. Subsection 5.2.3), the PRL capacities to be offered $\boldsymbol{P^{PRL}}$ are no longer optimization variables but rather parameters of the bidding problem. Consequently, the production costs in each scenario without revenues from the PRL market, i.e., the C_i's, are also no longer optimization variables and are instead precalculated by solving the scheduling problem presented in the following subsection. We consider three bids ($N = 3$) to be placed. In Subsection 5.4.3, the influence of different numbers of bids is however investigated. The price forecasts using the presented ARMA models assume a Gaussian distribution and give an expected value $\hat{y}^{PRL}$ with variance σ^2. Thus, the cumulative distribution function $F(\cdot)$ in equations (5.3)-(5.5) can be expressed using the error function erf$(\cdot)$. Note that this probability distribution now gives the distribution of the logarithm of the short-term price factors, denoted by $\boldsymbol{y^{PRL}}$, and not of the actual ask prices $\boldsymbol{A^{PRL}}$. The link between these two variables is realized by adding the following equation to the bidding problem:

$$y_i^{PRL} = \log(A_i^{PRL}) - \log(f^{PRL}_{Season}) - \log(\widetilde{A}^{PRL}_{Base}),\ i = 1 \dots N. \tag{5.12}$$

Obviously, if $\boldsymbol{y^{PRL}}$ is ordered, $\boldsymbol{A^{PRL}}$ is ordered as well due to the monotonicity of the logarithm. Furthermore, we introduce bounds on the ask prices. Allowing a deviation from the expected value $\hat{y}^{PRL}$ of $\pm 2\sigma$ already covers more than 95% of all values and thus appears suitable for our purpose, thus we additionally require in the bidding problem that the following holds:

$$\hat{y}^{PRL} - 2\sigma \leq y_i^{PRL} \leq \hat{y}^{PRL} + 2\sigma,\ i = 1 \dots N. \tag{5.13}$$

A truncated form of the series definition of $\mathrm{erf}(\cdot)$ is used to allow for the use in common commercial solvers for global deterministic optimization like BARON [153]. More precisely, we use the Taylor series

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \cdot \sum_{n=0}^{\infty} \frac{(-1)^n \cdot x^{2n+1}}{(2n+1) \cdot n!} = \frac{2}{\sqrt{\pi}} \cdot \left(x - \frac{x^3}{3} + \frac{x^5}{10} - \frac{x^7}{42} + \dots \right)$$

which we truncate after 12 terms to limit the maximum error to below $3 \cdot 10^{-7}$ in the relevant interval $y_i^{PRL} \in [\hat{y}^{PRL} - 2\sigma; \hat{y}^{PRL} + 2\sigma]$.

5.3.2. Mixed-integer linear scheduling model

According to the process description above, the discrete-time scheduling problem presented below is set up for the electrolytic process assuming that a fixed capacity is reserved for PRL, denoted by P_{Cap}^{PRL}. Note that for the sake of readability, we further omit the scenario index i in the following. As stated in Subsection 5.2.2, an hourly time-discretization is used in accordance with hourly available day-ahead electricity prices c_t^E on German day-ahead market that are retrieved from EPEX SPOT SE (`www.epexspot.com`) and that are assumed to be known for each hour t of the whole scheduling horizon of one week, i.e. $t \in \{1, \dots, T = 168\}$.

Besides the electricity costs for one week, the objective function C in the scheduling problem below also considers opportunity costs if the production target for the week m_{target}, which assumes 168 h of nominal operation with $P_{nom} = 90$ MW, is not met. We assume a specific electricity requirement of $\zeta_{nom} = 14$ MWh per tonne of aluminum in nominal design operation. Consequently, the production target is $m_{target} = 1080$ t. The costs for not achieving this target is accounted for by multiplying the production loss Δm^P with an average aluminum stock market price for the week, denoted by c^P and retrieved from the London Metal Exchange (`www.lme.com`). The objective thus reads:

$$C = \sum_{t=1}^{T} c_t^E \cdot P_t + c^P \cdot \Delta m^P.$$

As discussed above, the flexibility range of the process is reduced by P_{Cap}^{PRL} due to the required 100% availability of the PRL capacity, resulting in:

$$P_{min} + P_{Cap}^{PRL} \leq P_t \leq P_{max} - P_{Cap}^{PRL}, \quad \forall t$$

with $P_{min} = 67.5$ MW and $P_{max} = 112.5$ MW. Further equations are introduced to match the process description above. That is, we model the accumulation of surplus energy E_t using a balance equation and impose bounds $E_{max} = -E_{min} = 1080$ MWh to prevent the damaging of the cells as well as freezing of the electrolyte. In order to prevent a net accumulation of energy in the cells over one week, the energy is forced to reach zero at the end of the horizon.

$$E_1 = 0$$

$$E_{t+1} = E_t + (P_t - P_0), \quad \forall t$$

$$E_T + (P_T - P_0) = 0$$

$$E_{min} \leq E_t \leq E_{max}, \quad \forall t$$

Likewise, production equations for the amount of aluminum produced m_t are introduced:

$$m_1 = 0$$

$$m_{t+1} = m_t + \dot{m}_t(P_t), \quad \forall t$$

$$\Delta m^P = m_{target} - (m_T + \dot{m}_t(P_T)).$$

Here, $\dot{m}_t(P_t)$ is the - in general - nonlinear relation between the power uptake P_t and the production rate $\dot{m}_t$. For this relation, we assume that the influence of the power uptake on the specific electricity requirement can be expressed by a quadratic function over the whole operating range, i.e., $(\zeta_t - \zeta_{nom}) \propto (P_t - P_{nom})^2$. By multiplying the function $\zeta_t(P_t)$ with the instantaneous power uptake P_t, a nonlinear production characteristic arises. Here, we assume an increase in the specific electricity requirement by 1% if operating the process at $P_t = P_{min}$ and $P_t = P_{max}$. For the sensitivity analysis in Subsection 5.4.3, different values are used to evaluate the influence of this parameter on the bidding strategy. The influence of the provision of PRL on the production is considered by an offset of the nonlinear function, denoted by $\Delta\dot{m}^{PRL}$, that is assumed constant for all P_t and that increases quadratically with the PRL capacity provided. In the base case, we assume a production loss $\Delta\dot{m}^{PRL} = 0.032$ t/h, corresponding to a loss of -0.5%, if the maximum capacity of 10 MW is provided for PRL at nominal operation, i.e.,

$$\Delta\dot{m}^{PRL} = 0.0005 \cdot \frac{P_0}{\zeta_{min}} \cdot P_{Cap}^{PRL^2}.$$

In order to have an MILP representation of the production characteristic, the linear segmentation method as outlined by Floudas [219] is applied. Thus, the following set of equations is added to the scheduling problem:

$$\dot{m}_t = \sum_{s=1}^{S} (z_{s,t} \cdot a_s + b_s \cdot P_{s,t}) - \Delta\dot{m}^{PRL}, \quad \forall t$$

$$\sum_{s=1}^{S} z_{s,t} = 1, \quad \forall t$$

$$\sum_{s=1}^{S} P_{s,t} = P_t, \quad \forall t$$

$$z_{s,t} \cdot P_s^{lb} \leq P_{s,t} \leq z_{s,t} \cdot P_s^{ub}, \quad \forall s, t$$

$$z_{s,t} \in \{0, 1\}, \quad \forall s, t$$

Table 5.2.: Average CPU times for solving the subproblems when using the proposed enumeration approach.

Problem	Solver	Instances	Mean CPU time per instance [s]
Scheduling	CPLEX	10	5.21
Bidding ($N = 1$)	BARON	10	0.38
Bidding ($N = 2$)	BARON	45	2.28
Bidding ($N = 3$)	BARON	120	77.00

Here, a_s and b_s are the intercept and the slope of the linear segment s. Binary decision variables $z_{s,t}$ are added for each segment, indicating if the respective segment is active or not. We further ensure that only one segment can be active in one interval. Furthermore, auxiliary variables $P_{s,t}$ are introduced, such that they are nonzero if and only if segment s is active. P_s^{lb} and P_s^{ub} are the corresponding lower and upper bound of the segment.

Applying $s \in \{1, \ldots, S = 15\}$ for the segmentation for highly accurate computation of the opportunity costs, the overall MILP scheduling model for one week and fixed P_{Cap}^{PRL} is of moderate size comprising 3,025 continuous and 2,520 binary variables and can consequently be solved efficiently by using state-of-the-art MILP solvers.

5.4. Results

Optimal bidding strategies are calculated for the first six weeks of 2018. Both the bidding problem and the scheduling problem are implemented in GAMS v. 25.0.1 (GAMS Development Corp.). Data exchange between both GAMS models for the subsequent solution is realized by using the interface between MATLAB R2017a (MathWorks, Inc.) and GAMS. CPLEX v. 12.8 (IBM Corporation) is used as a solver for MILPs (referred to as scheduling problems) and BARON v. 17.10.16 [153] for NLPs (referred to as bidding problems). MILPs are solved until guaranteed global optimality. In BARON, a relative optimality gap of 10^{-5} is applied as a termination criterion. All calculations are conducted on an Intel® Xeon® CPU E5-2630 v2 with 2.60 GHz and 128 GB RAM.

5.4.1. Comments on solution times and their scaling

Average solution times when performing the optimizations assuming $N = 1, 2, 3$ bids placed are given in Table 5.2. It is indeed found that the solution times for the MILP scheduling problems are not prohibitive. Solution times for the NLP bidding problems strongly depend on the number of bids that are allowed to be placed. Nevertheless, even if three bids are allowed to be placed (which will be later shown to be adequate for raising the economic potential of a PRL market participation), the average solution time for one of the 120 bidding problems is less than 80 s. Summarizing, Table 5.2 indicates the validity of the previously made assumptions concerning the suitability of the solution approach, demonstrating its ability to identify optimal bidding strategies.

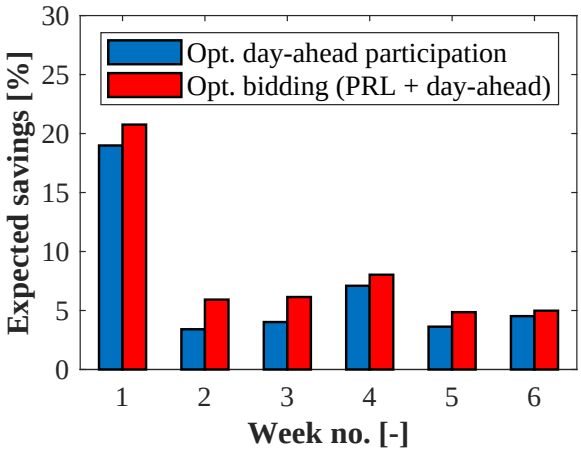

Figure 5.1.: Savings in expected weekly production costs compared to a constant production with day-ahead prices for the first six weeks of 2018. Blue bars correspond to savings by an optimized production schedule exploiting day-ahead price spreads only, red bars correspond to the optimal bidding strategy considering both revenues from the PRL market and exploitation of day-ahead price spread using remaining capacities.

Table 5.3.: Characterization of the optimal biding strategies identified for the first six weeks of 2018.

Week	Offered capacity [MW]	Lowest ask price [EUR/MW]	Price forecast [EUR/MW]
Week 1	$[4, 3, 2]^T$	3,554	3,668
Week 2	$[5, 3, 2]^T$	2,961	3,586
Week 3	$[5, 3, 2]^T$	2,613	3,368
Week 4	$[4, 2, 2]^T$	2,383	2,746
Week 5	$[4, 3, 2]^T$	2,080	2,468
Week 6	$[3, 2, 1]^T$	1,896	2,254

5.4.2. Economic evaluation of the bidding strategies

For the first six weeks of 2018, we calculate savings that can be achieved by using the presented model-based bidding strategy for a PRL market participation and compare them to a constant production at nominal operation under time-variable spot market prices in Figure 5.1. Furthermore, we compare the presented approach to the pure exploitation of spot market price spreads as done in the majority of previous studies (cf. Subsection 2.1.1). As can be seen, the active marketing of operational flexibility offers significant economic benefits compared to a constant production. In the first week of 2018, savings of more than 20% are possible that stem from both strongly fluctuating spot market prices as well as high compensation payments for the provision of PRL capacities. However, also in the following weeks, significant savings are possible in the range of 5-10%. In addition, we find that for most of the considered weeks, advantages of a PRL market participation compared to a pure spot market participation can be clearly observed. In week two, the expected

savings achieved by an optimal bidding strategy considering both revenues from PRL and spot markets are almost twice the savings from the pure exploitation of spot market price spreads (5.9% compared to a stationary operation with the optimal bidding strategy vs. 3.4% if solely participating in the spot market). Note that this high value of operational flexibility provided as PRL is found in spite of (i) a comparably high production loss due to the provision of PRL and (ii) a small maximum PRL capacity compared to the entire flexibility range.

Table 5.3 gives further details about the bids placed in each week. It can be seen that the optimal bidding strategies are very conservative, i.e., the ask prices, for which the majority of the capacities is offered, are clearly below the predicted maximum accepted ask prices. In other words, reducing the risk of a refused offer for a large capacity by lowering the ask price has a larger impact than increasing the potential revenues through higher ask prices. In particular, this holds for those weeks, where the advantage of marketing the flexibility on the PRL market rather than on spot markets is most significant, i.e., in weeks two and three. Here, half of the admissible PRL capacity is offered 600-700 EUR per MW below the expected maximum accepted ask prices, corresponding to a deviation from the expected value that is only slightly lower than 2σ, which in turn corresponds to a probability of acceptance of approximately 95%. This conservatism stems from the high monetary value of PRL in certain weeks and is lowered in weeks, where the benefits are smaller. Beside the ask prices, the capacity that is offered for PRL is also of interest. Although there is a clear relation between the capacity offered and the ask price, that is, almost half of the total capacity is offered at the lowest price in all weeks and then the additionally offered capacities are lowered with increasing ask prices, the total capacity offered however differs from 6 up to 10 MW. Again, there is the trend that in weeks, where the monetary value of PRL is high, higher capacities are offered. The optimal strategy thus aims at marketing the entire admissible capacity of 10 MW only in promising marketing situations.

Finally, it is worth mentioning that assuming the ability to optimally exploit spot market price spreads over an entire week is very optimistic. The optimal exploitation of hourly price spreads for one week would require a perfect knowledge about short- to medium-term developments of the spot market. Obviously, this knowledge is not available, such that only parts of the price spreads are disposable for exploitation. Consequently, the actually achievable savings in weekly production costs from the marketing of the operational flexibility will be lower than Figure 5.1 suggests, however, the potential of an additional PRL market participation compared to a pure spot market optimization, i.e., the differences in height of the two bars in each week in Figure 5.1, will likely increase. Furthermore, we expect that knowing about the imperfect exploitation of spot market price spreads will likely lead to even more conservative bidding strategies.

5.4.3. Parameter study

Revisiting the general problem formulation introduced in Section 5.2, one finds that for a given price forecasting model, the optimal bidding strategy exclusively depends on the parametric solution of the second-stage problem, i.e., the solution of the spot market scheduling problem as a function of the capacity reserved for PRL. In order to investigate the influence of different "costs-over-PRL-capacity" functions, the constitutive parameters of the scheduling problem are varied. For this purpose, off-design efficiency losses through an increased specific energy requirement in the range of 0-2% at the bounds of the flexibility

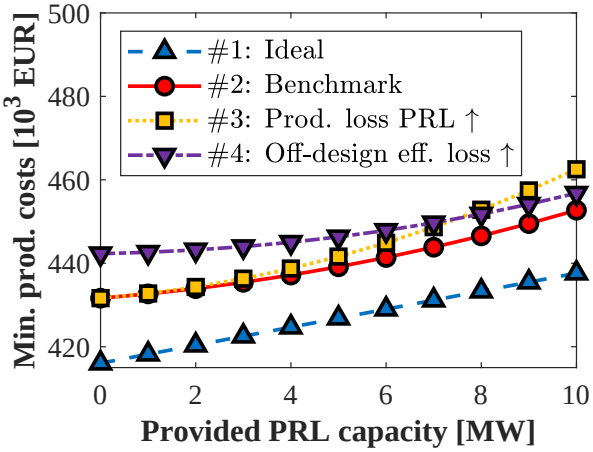

Figure 5.2.: Influence of model parameters on minimum production costs in week four when providing different PRL capacities. We consider the optimal scheduling subject to spot market prices and disregard revenues from the PRL market. Blue upwards-pointing triangles (△) with dashed line give an ideal case without losses, red circles (○) with solid line give the benchmark described in Subsection 5.3.2. Yellow squares (□) with dotted line and purple downwards-pointing triangles (▽) with dashed-dotted line assume either increased production losses due to the provision PRL or off-design efficiency losses.

Table 5.4.: Overview of the optimal bidding strategies in week four for the different minimum production cost curves given in Figure 5.2.

Curve	**Offered capacity [MW]**	**Lowest ask price [EUR/MW]**
#1	10	2,548
#2	$[4, 2, 2]^T$	2,383
#3	$[2, 1, 1]^T$	2,384
#4	$[5, 3, 2]^T$	2,334

range and production losses compared to nominal operation due to the provision of PRL capacities in the range of 0-1% at the maximum admissible PRL capacity are considered. Selected cost functions are depicted in Figure 5.2. The different shapes of these functions stem from three effects that the provision of PRL has on the process, that are: (i) the provision of PRL is associated with a production loss and therefore opportunity costs due to highly frequent fluctuations in the power uptake that lower the efficiency across the entire operating range, (ii) the provision of PRL decreases the flexibility range of the process and therefore limits the ability to exploit spot market price spreads, and (iii) as a flexible operation is usually accompanied by efficiency losses due to periods of off-design operation, the previous effect of a narrowed flexibility range might be weakened - particularly for low PRL capacities.

Different cost functions also lead to different bidding strategies. The influences of the cost functions on the optimal bids are examined for week four (Table 5.4). Starting from the benchmark parameters introduced in Subsection 5.3.2, increasing the intensity of the

production loss caused by fluctuations due to the provision of PRL (effect (i) becomes more relevant) lowers the amount offered substantially from a total of 8 MW to only 4 MW. In contrast, increasing the efficiency losses in the case of off-design operation (effect (iii) becomes more relevant) increases the amount offered to 10 MW. The influence on the actual ask price for which the majority of the capacity is offered is however low. For completeness, Table 5.4 also lists an idealized case, where only effect (ii), i.e., the decreased flexibility range at spot markets, is present. Note that in this case, neither splitting the offer nor offering less than the maximum capacity is reasonable, as the optimal ask price for each increment is the same. The higher ask price in this case can be explained by the larger slope compared to the other curves in the relevant range of PRL capacities up to 5 MW, such that higher revenues are required to compensate for the increased costs. In conclusion, it is highly crucial to know the actual influence of the provision of PRL capacities on the production process when calculating the bidding strategy. For instance, if curve #1 is assumed, but the actual relation follows curve #3, the revenues from the PRL market can no longer compensate for the production losses. Vice versa, if curve #3 is assumed and curve #1 is correct, a substantial economic potential would remain unexploited. We highlight that although the previous discussions have been conducted for the example of a flexible aluminum electrolysis cell, the general conclusions hold for different processes as well. In other words, similar effects will appear for the vast majority of processes and have to be taken into account to find a reasonable bidding strategy.

In addition to the study on the influence of process parameters, we also investigate how erroneous forecasts of the spot market price development influence the optimal bidding strategy. Therefore, we repeat calculations for week four using two perturbed time series with an equal average for the day-ahead price (cf. Figure 5.3). One of them assumes increasing volatility in the course of the week and the other one decreasing volatility. Note that deviations from the original price signal are assumed to be stronger at the end of the week, i.e., when they are farther away from the PRL auction. The corresponding bidding strategies are summarized in Table 5.5. Here, one clearly sees that in case of high volatility, i.e., of high revenue potentials at the spot market, the optimal bidding strategy leaves substantial capacities for the subordinated scheduling optimization at the day-ahead market. At the same time, the optimizer tends to higher and thus more aggressive ask prices to make revenues from the PRL market competitive to those from the spot market. In turn, in the low-volatility case, the optimal bidding strategy aims at marketing the maximum admissible range at the PRL market and prefers low-risk ask prices. Note that determining the bidding strategy using erroneous spot market time series also bears the risk of leaving economic potentials unexploited. However, in contrast to process parameters, future electricity prices are inherently uncertain. Consequently, future work should aim at incorporating the knowledge about these uncertainties into the proposed framework, e.g., by stochastic programming. Note that this would in fact lead to a significant increase in the number of scenarios that need to be considered and might thus require improvements in the solution approach. In contrast to that, adjustments of the short-term production schedule can already be addressed efficiently by means of a moving horizon recalculation of the schedule once new price information becomes available. Here, we refer the reader to the relevant literature (e.g., [88]).

Finally, we evaluate the benefits of splitting one offer into multiple single offers with individual ask prices, which is a common practice at the PRL market as can be seen from past auction results. For this purpose, Table 5.6 shows how the number of bids placed

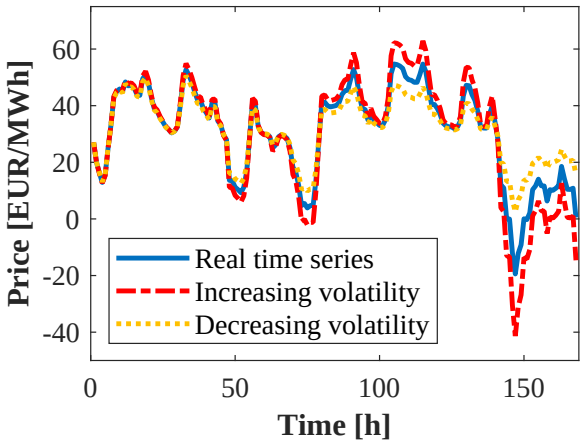

Figure 5.3.: Day-ahead price time series for calculating the optimal bidding strategy in week four. The solid blue line shows the original prices for that week, the dashed red line gives a perturbed time series with increasing volatility in the course of the week and the dotted yellow line corresponds to decreasing volatility.

Table 5.5.: Overview of the optimal bidding strategies in week four for the different day-ahead electricity price time series given in Figure 5.3. For comparison, the probability that no bid is accepted (scenario 0) is added.

Price time series	**Offered capacity [MW]**	**Lowest ask price [EUR/MW]**	**Probability of scenario 0**
Original time series	$[4, 2, 2]^T$	2,383	9.50%
Increased volatility	$[2, 2, 1]^T$	2,432	13.94%
Decreased volatility	$[5, 3, 2]^T$	2,349	6.38%

influences the expected production costs in week four. As can be seen, the improvement in relative savings through placing three instead of one bid is less than 0.05 percentage points and thus very small compared to the savings enabled by placing the first bid. Consequently, placing more than one bid does not enable additional economic benefits in the present setting. Calculations for Table 5.6 are repeated assuming two hypothetical forecasting models with equal expected values but different forecasting standard deviations σ^* (square root of the introduced variance σ^{*2}). Comparing the results for $\sigma^* = 0.5\sigma$, $\sigma^* = \sigma$, and $\sigma^* = 2\sigma$ illustrates the monetary value of improved forecasting accuracy, that is, halving the forecasting standard deviation increases the savings in weekly production costs by >15%. Furthermore, we find that improvements enabled by submitting additional bids increase with more uncertain forecasts. However, even if twice the standard deviation is assumed for the forecasts, which already corresponds to an uncertainty of several hundreds of EUR per MW, improvements appear still small. Nevertheless, splitting the total capacity to offer for PRL might still be rational from a different point of view. That is, the problem formulation in this work assumes risk-neutral market participation. However, the model could be also extended to account for other risk preferences such as in particular risk aversion.

Table 5.6.: Influence of the number of bids placed on the expected production costs in week four. Calculations are repeated assuming a doubled and a halved standard deviation for the price forecasting model.

Forecasting	**Expected savings for a varying number of bids**		
standard deviation	$\boldsymbol{N = 1}$	$\boldsymbol{N = 2}$	$\boldsymbol{N = 3}$
0.5σ	1.137%	1.157%	1.162%
σ	0.966%	0.998%	1.008%
2σ	0.792%	0.847%	0.864%

5.4.4. Goodness of price forecasts

In Figure 5.4, the identified PRL price forecasting model is validated by comparing the one-step-ahead forecasts when applying the AR(1) model with parameters from Table 5.1 to the actual PRL price data for the first 18 weeks of 2018. Therein, we find that 55.56% of the forecasting errors are in the interval $[-\sigma, \sigma]$, 72.22% lie within $[-2\sigma, 2\sigma]$ and no forecasting error is found outside $[-3\sigma, 3\sigma]$. Comparing this to the assumed normal distribution, one has to conclude that the proposed time series forecasting model performs poorer at this sample than expected from the estimated model variance (cf. Table 5.7). In particular, the model performance is poor for the second half of the validation set. The comparison of predicted and actual maximum accepted ask prices for the weeks 10-18 thereby illustrates a shortcoming of a purely autoregressive model, that is, once accepted ask prices are currently low, an autoregressive model tends to forecast low prices for the next week as well. However, at the same time, aggressive market competitors will be encouraged to increase their ask prices in order to exploit the entire price range, in which their bids are accepted. Consequently, the maximum accepted ask price increases, whereas only minor changes in the weighted average of accepted ask prices appear. In the auction for the following week, this spread will in turn encourage competitors, whose bids have not been accepted in the previous week, to underbid the prices, thus closing the gap between the maximum and the average accepted ask prices again. The fluctuating maximum accepted ask prices accompanied by an almost constant weighted average of accepted ask prices in the weeks 10-18 might indicate such behavior. Obviously, the proposed AR(1) model is not able to predict these fluctuations. Instead, the AR(1) model appears to lag behind one week.

For the six weeks considered in Subsection 5.4.2, Table 5.8 additionally lists, which of the optimized ask prices would have been accepted in the corresponding weeks. It can be seen that although the forecasts of the maximum accepted ask prices are above the actual value for almost the complete period, the vast majority of the offered capacities would still have been accepted. This is indeed a consequence of the previously discussed conservatism of the approach due to large uncertainty in the forecasts. However, this also implies that particularly if prices are substantially underestimated, economic potentials remain unexploited, as far too low prices are asked, which can only be overcome by an improved forecasting accuracy. Thus, although the proposed AR(1) model shows an acceptable performance on the data set, further research is recommended concerning the improvement of the forecasting accuracy in order to exploit the full economic potential of the approach. This might

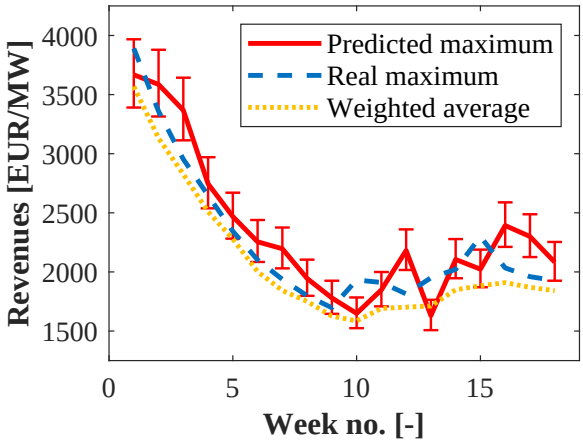

Figure 5.4.: Validation of the proposed AR(1) time series model for forecasting the maximum accepted ask price for the first 18 weeks of 2018. The solid red line shows the AR(1) forecasts. Error bars indicate the interval $\pm\sigma$. The dashed blue line gives the actual maximum accepted ask price in these weeks. The dotted yellow line corresponds to the weighted average of all accepted ask prices and is only depicted for comparison.

Table 5.7.: Comparison of the distribution of actual model forecasting errors for the first 18 weeks of 2018 with the assumed normal distribution.

Interval	**Percentage of values within interval**	
	Forecasting errors	**Normal distribution**
$[-\sigma, \sigma]$	55.56%	68.27%
$[-2\sigma, 2\sigma]$	72.22%	95.45%
$[-3\sigma, 3\sigma]$	100%	99.79%

include the transition towards more sophisticated model structures for forecasting complex market dynamics.

5.5. Summary and recommendations for future research

The optimal participation in a pay-as-bid market for ancillary services, namely the German market for PRL, is formulated as a two-stage stochastic MINLP minimizing the expected weekly production costs. Both revenues from the PRL market as well as from the day-ahead spot market are considered. More precisely, the bids to be placed at the PRL market (the amounts to offer and the prices to ask) are the first-stage decision variables, whereas the second stage corresponds to a scheduling optimization, which markets the remaining flexibility at the spot market depending on the outcome of the PRL auction. Furthermore, we herein apply an autoregressive time series model for forecasting the upcoming maximum accepted ask prices on the PRL market, which allows for calculating the probabilities that a bid is accepted. A tailored solution approach based on enumeration is proposed to decompose the problem into a finite number of NLPs and MILPs. Following this approach,

Table 5.8.: Offered capacity that would have actually been accepted in the first six weeks of 2018 when applying the optimized bidding strategy relying on the foreasted prices.

Week	**Capacity [MW/MW]**	**Maximum accepted ask price**	
	Accepted / offered	**Value [EUR/MW]**	**Forecast [EUR/MW]**
Week 1	7/9	3,891	3,668
Week 2	10/10	3,359	3,586
Week 3	10/10	2,953	3,368
Week 4	6/8	2,647	2,746
Week 5	7/9	2,344	2,468
Week 6	5/6	2,102	2,254

optimal bidding strategies are calculated for an industrial aluminum electrolysis cell. Our results indicate a high value of a flexible operation, enabling savings up to more than 20% compared to a stationary operation in specific weeks at the beginning of 2018. Moreover, we demonstrate that the optimal participation in the balancing markets can offer additional saving potentials exceeding those from the pure exploitation of spot market price spreads. Besides, we show that the optimal bidding strategy crucially depends on how the provision of PRL capacities influences the process efficiencies. If poor assumptions are made about this relation, economic potentials might remain unexploited or - in extreme cases - savings might turn into losses. Finally, the benefits from splitting the capacity to offer into several bids with individual ask prices as is common practice in the PRL market are evaluated. Our results however indicate that only insignificant improvements are enabled when submitting more than one bid.

In summary, our results demonstrate the potential of model-based strategies for marketing the operational flexibility at pay-as-bid markets for balancing power, but also show that the performance of the proposed approach, i.e., the actually realized savings, strongly depends on the quality of the forecasts of the upcoming maximum accepted ask prices. If the forecasts comprise large uncertainties, the optimizer tends to conservative bids asking low prices to ensure receiving the highly valuable compensation payments for providing PRL. Thereby, economic benefits remain unexploited, as the range of accepted ask prices is not exhausted. Here, we recommend the development of more sophisticated forecasting models than a simple autoregression. In particular, we see a high potential in applying machine learning models for that purpose, such as long short-term memory networks that have been applied to the forecasting of the development of financial markets with great success (e.g., [220]). Furthermore, there is also a possibility to include game-theoretical approaches, which aim at modeling the strategic behavior of competitors at the energy markets (cf. [221]).

Future research should additionally target a more realistic treatment of uncertainties in upcoming spot market prices. In particular, we are interested in the explicit consideration of uncertainties in the spot market price development when determining the optimal bidding strategy. Moreover, future approaches should also incorporate recalculations of the optimal short-term production schedule subject to spot market signals on a moving horizon, once new price information becomes available. Along these lines, one could also implement the actual bidding procedure within a moving horizon framework, such that production

targets need no longer to be met within one scheduling horizon and load shifts between multiple horizons get possible. Here, we emphasize that adaptations of the presented approach have to keep up with continuously occurring changes in the markets themselves. Note that for instance, a recent change towards daily auctions has partially overcome the issue of the increasing uncertainty in spot market prices in the course of the scheduling horizon. However, when focusing on the finer discretized continuous intra-day market that on the one hand enables higher savings (cf. [222]) but on the other hand comprises larger uncertainties, the described extension of the approach becomes relevant again.

Finally and as indicated above, the general multi-stage stochastic concept should be applied to other decision-making problems. Note that this requires addressing both the upper-stage bidding problems (if considering different markets for ancillary services) and the lower-stage subordinated scheduling optimizations (if considering different shorter-term spot markets). Here, a very promising research perspective lies in the consideration of the market for secondary balancing power (SRL), where compensation payments are indeed lower than at the PRL market. However, technical requirements for participation in the SRL market are also less strict, making it promising for a larger variety of processes. Note that modeling the participation in the SRL market will require an adequate consideration of the so-called energy prices that ultimately determine the amount of balancing energy that is retrieved from the individual participants. In this context, we have recently proposed in a conference paper to simplistically correlate the amount of retrieved balancing energy with the energy price using an exponential law fitted to historic data [223]. A detailed discussion of the corresponding potential model formulations is however out of the scope of the present thesis and thus foregone for the sake of brevity. Furthermore, we emphasize that the existence of more products to offer (six periods per day, each subdivided into a positive and a negative load direction, instead of one unidirectional PRL product a day), leads to a substantial increase in both the number of first-stage decision variables and the number of second-stage scenarios. Against this backdrop, we strongly recommend to revisit the proposed solution approach to cope with the increasing complexity of such a decision-making problem by either developing and/or tailoring structure-exploiting decomposition algorithms (e.g., [224, 225]) addressing a deterministic solution of the MINLP or by assessing the suitability of stochastic/semi-stochastic solution approaches.

6. Multi-objective assessment of economically and environmentally incentivized measures in industrial demand side management

In this chapter, we critically question the currently discussed measures increasing the flexibility potential of processes on the demand side from the perspective of the decarbonization of the electricity sector. In particular, we herein investigate if economically-driven measures, i.e., measures that enhance the exploitation of fluctuating electricity prices, promote the penetration of renewable electricity sources by allowing to consume electricity primarily in periods with a high renewable share. To this end, we provide a massive simulative study focusing on the German electricity sector using both historic electricity time series data as well as projected time series data for the future In Section 6.1, we first present the objective functions, which we use for quantifying the economic and environmental perspective on DSM throughout this chapter and provide rationales for these choices. Namely, we herein continue to use the total costs for electricity purchase at the day-ahead market as the economic objective and introduce the individual contribution to the integral residual load as the environmental one. In Section 6.2, we show the results from an analysis of the underlying time series, i.e., of the day-ahead electricity prices as well as of the ratio between the residual load and the network-wide electricity consumption. Thereby, we facilitate the interpretation of the results of the scheduling optimizations in the remainder of this chapter. For this purpose, we present a quantitative characterization of the time series and show the results of both a correlation and a frequency analysis. Afterwards, we present a generic process model in Section 6.3, which we use to evaluate the benefits from DSM with regard to an exploitation of the different fluctuation patterns of the two time series to minimize the corresponding objective functions. First, we present an ideal storage-type customer and discuss its constitutive parameters. Afterwards, two extensions to this basic model are given that allow for both the consideration of off-design efficiency losses and making use of shut-down opportunities. The results of the single- and bi-objective scheduling optimizations using the historic time series are given in Section 6.4. Therein, focuses are set on the effect of both average utilization rates and off-design efficiency losses on Pareto curves. Furthermore, we systematically assess the influence of storage and ramping limits on the two objectives and also evaluate the improvements when making use of opportunities for shut-downs. In addition to that, we qualitatively describe the results of the scheduling optimizations considering the projected time series but limit the discussion to the key observations. Finally, the chapter is concluded in Section 6.5. Therein, we also briefly discuss further developments in the methodology as well as extensions of the scope of the simulative study.

6.1. Objective functions for scheduling

As briefly mentioned in the introduction, the German electricity market as well as many others are characterized by a time-variable electricity supply satisfying a (mostly) inelastic electricity demand, which causes fluctuations in the electricity mix. This consequently leads to time-variables electricity prices, as the numerous generation technologies exhibit different operating costs. Whereas the previous chapters are mainly concerned with the exploitation of spot market price spreads, time-variable electricity mixes can be exploited as well from an environmental point of view (cf. Subsection 2.1.4). In this section, we thus introduce the two objective functions for the optimization that are considered in the remainder of this chapter. That is, we introduce one objective function for quantifying the economic perspective and one for the environmental perspective, and briefly discuss their interrelation.

6.1.1. Quantification of the economic perspective

From an entrepreneurial perspective, the primary objective of DSM is to minimize the costs for the electricity purchase by preferably consuming electricity in hours with low prices while respecting all requirements, e.g., obeying operating limits, satisfying demands, etc. We focus on hourly changing spot prices at the day-ahead market due to its large trading volumes. Note that we herein also give a more detailed description of its underlying price setting mechanisms that is beyond the scope of the previous chapters, as it allows for a qualitative explanation of the correlation between the economic and the environmental objective. We further remark that recent studies indeed identify larger economic benefits of participation in real-time electricity markets, such as the intra-day continuous market in Germany, as more distinct price peaks occur there [18, 222]. In the following, we however assume that the entire electricity is purchased at the day-ahead spot market for the ease of presentation. The economic objective for optimization thus equals the total electricity costs

$$C = \sum_t c_t \cdot P_t, \tag{6.1}$$

with c_t denoting the instantaneous day-ahead spot market price in hour t and P_t the power consumption.

The German day-ahead market applies uniform pricing, i.e., the market clearing price in each hour is determined by the intersection of the supply and the demand curve. Assuming a perfectly competitive market, the supply curve is the aggregated marginal cost curve of all electricity suppliers. Consequently, for known hourly demands and a given pool of suppliers with known marginal costs, the so-called marginal power plant would then solely determine the spot electricity price c_t [226]. Therefore, in periods with a high share of renewable electricity generation, which is associated with negligible marginal costs (and in Germany currently also feed-in support), spot prices are low, which is referred to as the merit-order effect of renewable energy sources [227].

6.1.2. Quantification of the environmental perspective

The described price setting procedure inherently leads to time-variable electricity mixes depending on the almost inelastic instantaneous network-wide electricity demand and the realized generation from intermittent renewable sources. Consequently, there is an analogous environmental optimization opportunity by preferably consuming electricity in those hours where the electricity generation is associated with low environmental impacts. In the previous literature referred to in Subsection 2.1.4, the environmental objective has mostly been quantified by accounting for different contributions of the individual electricity generation technologies to climate change, targeting the exploitation of time-variable carbon footprints for the electricity consumption. As an alternative, we herein measure whether electricity is primarily consumed in hours with a high share of renewable generation by introducing the hourly residual load share r_t, which is the quotient of the instantaneous residual load and the network-wide electricity consumption. Thus, r_t gives the percentage of the entire network-wide electricity consumption that needs to be satisfied by a non-renewable generation. Note that $r_t \leq 1$ thus holds and that $r_t < 0$ is also possible, indicating hours with a surplus of renewable energy, which should certainly be preferred. A suitable environmental objective then reads:

$$R = \sum_t r_t \cdot P_t \tag{6.2}$$

if assuming that the purchased electricity has the same mix as traded at the market, giving the individual contribution to the integral residual load, i.e., the part of the total individual electricity consumption that needs to be satisfied by a non-renewable generation.

We emphasize that the proposed environmental objective is not meant for a detailed life cycle analysis but should rather provide a tangible measure for the reduction in the consumption of fossil-produced electricity in favor of renewable electricity with a compact mathematical form comparable to that of the economic objective, i.e., a linear combination of the hourly electricity consumptions with fixed constants. Moreover, using this measure omits an existing optimization potential between different fossil energy sources, which we do not intend to address in this study. In particular, we thereby prevent favoring periods with a high share of generation from gas-fired plants, which on the one hand involve fewer emissions than coal-fired plants, but on the other hand commonly only contribute to the electricity mix in periods with a very low renewable generation due to their high operating costs.

6.2. Time series analysis

We consider yearly time series for both c_t and r_t when evaluating the objective functions (6.1) and (6.2), i.e., $t \in \{1, \ldots, T = 8760\}$, and investigate both the current and a projected future scenario for the German electricity sector. The first data set contains historic time series from 2017 and is provided by Agora Energiewende (`www.agora-energiewende.de/`). The 2030 data set is provided by Forschungsgesellschaft für Energiewende (`www.ffegmbh.de/`) and has been obtained in the project MONA 2030 [228] by means of system simulations assuming a continued increase in the installed renewable electricity generation capacities by ∼70% compared to 2015. Before conducting scheduling optimizations using a generic process model (cf. Sections 6.3 and 6.4), we investigate

Table 6.1.: Quantitative descriptors for the four time series. All values are normalized to the respective means.

Descriptor	Electricity price [-]		Residual load share [-]	
	2017	**2030**	**2017**	**2030**
Mean	1.00	1.00	1.00	1.00
Minimum	-2.43	0.00	0.19	-2.03
Maximum	4.78	2.68	1.46	2.37
Standard deviation	0.52	0.35	0.24	0.86

the time series, assess their correlation, and identify different characteristic fluctuation patterns. The results presented below thus allow for qualitatively anticipating the observations that we make during the scheduling optimizations described in Section 6.4 and facilitate the interpretation thereof.

6.2.1. Quantitative characterization

Table 6.1 gives quantitative descriptors for all considered time series. All values are normalized to the respective means to make the series comparable. As we confine to the discussion of relative improvements through DSM activities in Section 6.4 by comparing to a constant production, i.e., with averaged input data in (6.1) and (6.2), this normalization appears reasonable for interpreting the results. As can be seen, electricity prices in the historic 2017 time series show a very broad distribution compared to the residual load shares, as measured by both the range and the standard deviation. Consequently, higher relative optimization potentials from the economic perspective are expected compared to the environmental one. Comparing the 2017 to the 2030 time series, one finds that the distribution of the electricity prices becomes narrower, whereas the distribution of the residual load shares becomes broader with even a larger range and standard deviation than the prices. Consequently, relative economic optimization potentials are expected to decrease, whereas environmental ones are expected to increase. The latter conclusion can be explained by the increasing penetration of electricity generation from intermittent renewable sources assumed for 2030. From this perspective, the narrowing of the distribution of the electricity prices seems counter-intuitive. However, we presume that this observation stems from the assumptions behind the time series simulation for 2030. More precisely, projected prices are bounded by the minimum (zero without feed-in support) and maximum marginal costs of the power plant fleet when assuming a perfectly competitive market. Possible market behavior going beyond this assumption, which is apparently present in 2017 and causes more distinct price peaks and troughs, is not included in the 2030 price series. We however emphasize that absolute price spreads in 2030 are still higher than in 2017.

6.2.2. Correlation analysis

Considering bi-objective optimizations, i.e., trade-offs in simultaneously fulfilling both objectives, the correlation between the time series of electricity prices and residual load shares

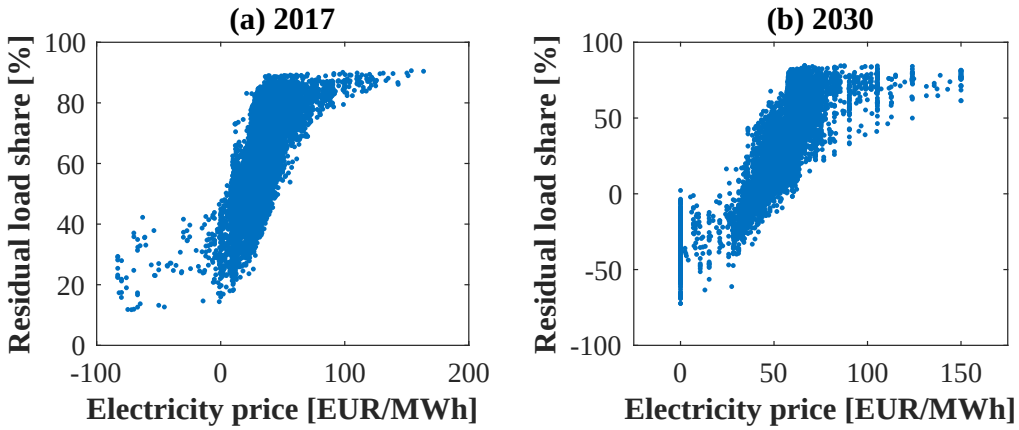

Figure 6.1.: Scatter plots for both the historic and the projected time series.

is crucial. Thus, in Figure 6.1, we show scatter plots containing the pairs (c_t, r_t) for every hour in 2017 and 2030 respectively. In both cases, an S-curve can be observed indicating a positive correlation. The intensity of the correlation can be quantitatively assessed using correlation coefficients. Here, the use of Spearman's rank coefficient ρ is highly informative, as it represents a measure for the monotonicity of a correlation. More precisely, if one finds a strictly monotone function (represented by $\rho = 1$), the bi-objective perspective on DSM would be redundant, i.e., Pareto curves would become points. For the 2017 time series, we find $\rho = 0.64$, whereas for 2030, we find $\rho = 0.76$. Thus, for both 2017 and 2030 monotone trends can be observed. However, there is also a potential for balancing trade-offs between the two objectives, which we expect to become particularly important in the presence of efficiency losses.

6.2.3. Frequency analysis

Details concerning fluctuation patterns in the considered time series can be gained by frequency analysis. For this purpose, we conduct discrete Fourier transforms. The resulting spectra are depicted in Figure 6.2. Our analysis focuses on fluctuations on an hourly-to-daily scale, which is typically considered in DSM due to, in general, limited storage capacities. Note that the lower frequencies are nevertheless of interest for different technologies addressing, e.g., seasonal storage [229]. Confining to the DSM-relevant frequency components, one finds the most important frequency components at $\frac{1}{24}\,\mathrm{h}^{-1}$ and $\frac{1}{12}\,\mathrm{h}^{-1}$. In other words, all time series are characterized by the superposition of two sine waves, one with a period length of one day (day/night fluctuations) and one with a period length of half a day (intra-day fluctuations). Noticeably and independently from the considered year, in case of the residual load share, the day/night fluctuation is more distinct, leading to a temporal course with typically one local minimum per day, which matches the peak in generation from photovoltaics at noon. In contrast, in the case of the electricity price, the intra-day fluctuation is more distinct, which typically leads to two local maxima a day, that occur around 8:00 am and 8:00 pm respectively. This two-peak-behavior directly follows from the market clearing procedure itself, considering that the peak in generation from photovoltaics at noon is superimposed with a daily flat plateau in the network-wide electricity consumption between 8:00 am and 8:00 pm. That is, at the beginning and the

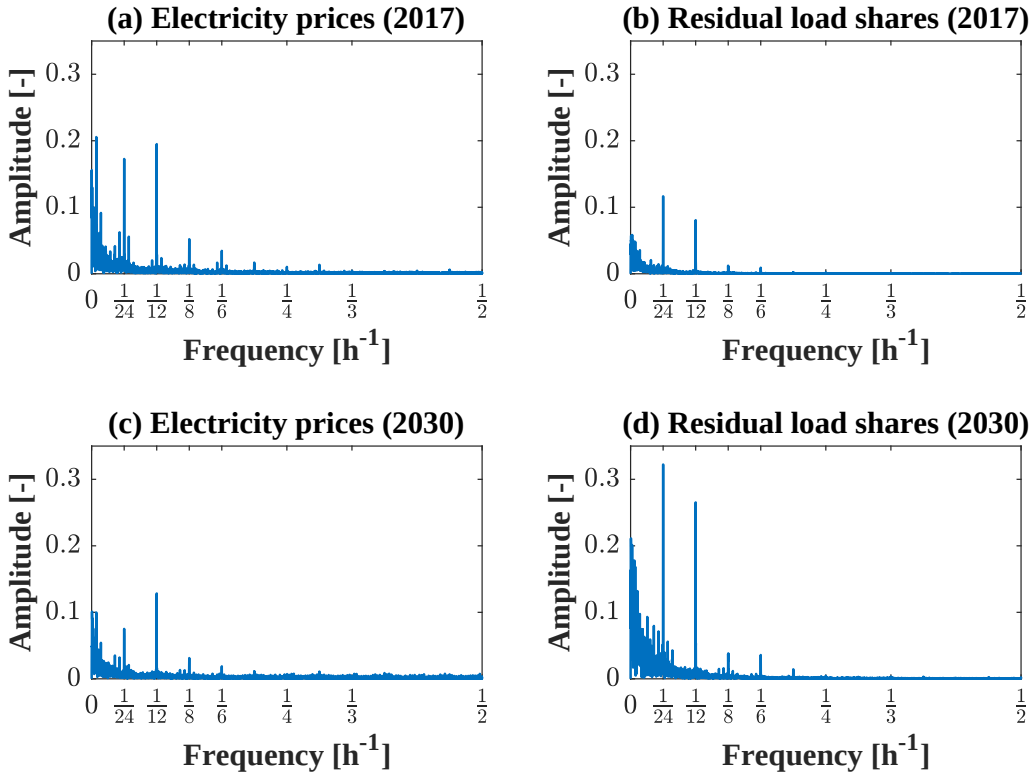

Figure 6.2.: Frequency spectra of the fluctuation patterns of the considered normalized time series obtained by discrete Fourier transform.

end of this plateau, a high demand meets a comparably low supply from photovoltaics. In contrast, at the center of the plateau, i.e., at noon, there is much more supply from photovoltaics at low marginal costs and thus less conventional electricity generation at higher marginal costs is required to meet the demand, involving temporary daily price troughs. We again highlight that the different fluctuation patterns can be observed irrespective of the considered year, indicating that these are not caused by imperfect market behavior. Moreover, we observe that components with frequencies above $\frac{1}{6}\,\mathrm{h}^{-1} \ldots \frac{1}{4}\,\mathrm{h}^{-1}$ show a larger contribution to the spectra of the electricity prices than to the spectra of the residual load shares. This corresponds to the existence of very short-term price spreads, whereas there are no spreads to be exploited on a similar scale with regard to a reduction of the contribution to the residual load. Together, these findings represent an indication that some measures increasing the flexibility potential might influence the economic and the environmental objective in a different manner. In particular, whereas the exploitation of day/night fluctuations is possible even for processes with limited capabilities for load changes, the exploitation of short-term spreads certainly requires high plant agility, i.e., loose ramping constraints. Consequently, we expect mostly economic and almost no environmental incentives for loosening ramping constraints. Finally, we note that Figure 6.2 also illustrates the trend of increasing fluctuations of the residual load shares and decreasing (relative) fluctuations of the prices in the projected time series (cf. Table 6.1).

6.3. Generic process model

In order to evaluate the benefits from DSM through the exploitation of the fluctuations of the time series analyzed in Section 6.2, we consider a generic process that is able to vary its electricity consumption and store products. Processes like these act as virtual batteries, i.e., load increases act similarly to charging and load reductions to discharging a battery storage from the perspective of the electricity grid [123, 230]. We remark however that these virtual batteries can obviously not have net feeding into the grid at any point in time. Moreover, we require that all load reductions have to be caught up, i.e., we confine to load shiftings and not allow for a pure load shedding, etc. (cf. the classification by Gellings [231]). Depending on the considered process, the ability to shift loads is however limited by different constraints, which will be analyzed in detail in the following. To this end, we first introduce an idealized base case, referred to as ideal storage-type customer, and discuss ranges for the limiting process parameters. Afterwards, we present two extensions to account for off-design efficiency losses and enable temporary shut-downs of the process. As we consider variations of the constitutive model parameters across large ranges, we account for the vast majority of DSM-relevant processes and thus aim at generalizable conclusions not limited to a specific process. Model equations are only briefly covered herein, instead, the focus is set on the underlying assumptions. Actual mathematical formulations have been extensively discussed in the relevant literature as well as in the previous chapters of the present thesis.

6.3.1. Ideal storage-type customer

The mathematical model of the ideal storage-type customer, which we consider as a base case, is given in the early work of Daryanian et al. [62] and describes an idealized process with flexible power uptake and opportunities for product storage. Essentially, the model relies on two assumptions:

- The production rate is proportional to the power consumption irrespective of the operating point. Note that in the battery analogy, this assumption leads to an ideal behavior in the sense that neither charging nor discharging is associated with any losses.
- The total production of the process must remain constant, equaling that of a constant operation at the nominal production rate.

The ability of the process to shift loads is further limited by (i) its operating range and average utilization rate, (ii) its storage capacities, and (iii) ramping constraints:

(i) The power consumption of the considered process is limited by lower and upper bounds. Besides the size of the flexibility range, the average utilization rate $\overline{UR}$ [%], giving the ratio between the nominal and the maximum power consumption, is of crucial importance. Note here that energy-intense processes are commonly sized for $\overline{UR} \approx 1$ to minimize capital costs. This represents one of the main hindrances for load shiftings, as missed production can hardly be caught up [232]. For instance, for the chlor-alkali capacities in Germany, an average utilization rate of >90% is found, limiting the potential of DSM [141]. Thus, we herein consider a process with a

representative average utilization rate $\overline{UR} = 95\%$ and a flexibility range of 50-100%, but investigate benefits from an oversizing of the facilities, i.e., from reduced average utilization rates. In Section 6.4, we consider oversizings by up to 20%, i.e., average utilization rates $\overline{UR} = 79 \dots 95\%$. Obviously, the oversizing of production facilities is usually in conflict with total annualized costs, requiring integrated optimizations in practice (see, e.g., [233]).

(ii) To account for limited storage opportunities, the process model considers a simple buffer tank that is filled by the instantaneous production and continuously emptied by the nominal production, which can be interpreted as a constant down-stream demand that needs to be met. Note that several real processes are indeed characterized by seasonal fluctuations in the demand that might interact with the provision of load flexibility. The consideration of such aspects is however beyond the scope of the chapter. The stored amount increases in the case of overproduction and decreases in case of underproduction. We do not consider losses during storage. The stored amount is limited between zero and the maximum storage capacity. In order to avoid emptying the tank throughout the year, we force both the initial and the final tank levels to 50%. Moreover, we allow for a systematic comparison between processes of different scales by using the time, up to which the production rate at the design operating point can be maintained with a completely filled storage tank, as a measure for the storage capacity, denoted by S^{max} [h]. In the parameter studies in Section 6.4, we consider a wide range of storage capacities between $S^{max} = 3 \dots 48$ h. Note that in contrast to an oversizing of the actual production facilities, a retrofit of storage capacities is easily possible in many settings (see, e.g., [7]).

(iii) Finally, the flexibility of a process can be limited by imposing a maximum change between two successive operating points. In particular, such ramping constraints are required to ensure that transitions between two operating points can be realized without violating requirements [78]. Ramping constraints are commonly modeled using a set of linear equations [37]. As we again target the comparability between processes of different scales, we herein impose a maximum on the change in utilization rates between consecutive hours, denoted by Δ^{max} [%/h]. In Section 6.4, we investigate the influence of different ramping constraints by considering $\Delta^{max} = 5 \dots 25\%$/h. Note that whereas adjustments of the average utilization rate by oversizing as well as of the storage capacities usually involve the installation of additional equipment, ramping constraints can be affected by changes in the operating philosophy (see, e.g., [89]) beside modifications in the process design (see, e.g., [87]).

6.3.2. Off-design efficiency losses

The assumption of a linear production characteristic made above by assuming a constant efficiency might in some cases be a poor approximation of the reality (cf. the processes considered in Chapter 3). Instead, efficiency losses in case of off-design operation occur and have to be taken into account when assessing the potential of DSM [141]. In order to enable a systematic treatment, we introduce the following assumptions, leading to an extended storage-type customer model that can account for efficiency characteristics:

- The electric efficiency, i.e., the ratio between the production rate and the power consumption, is a function of the instantaneous utilization rate.
- The highest electric efficiency is reached at the design operating point.
- To account for off-design efficiency losses, we use a quadratic approximation of the generally nonlinear function, leading to a cubic function for calculating the production rate from the power consumption.

Following the assumptions, the entire loss characteristic is defined by one parameter, denoted by ζ [%], giving the relative loss in the electric efficiency at the lowest admissible utilization rate (50%) compared to the electric efficiency at nominal operation. Considering the large variety of processes in the DSM-relevant literature, we evaluate a range of potential parameters with $\zeta = 0 \dots 33\%$ in Subsection 6.4.2. Using the battery analogy, we thereby account for losses during charging and discharging. Moreover, we highlight that under the aforementioned assumptions, the constraint on satisfying the integral production target involves that any load shifting, i.e., any temporary off-design operation, increases the total electricity consumption.

In order to enable the use of efficient MILP solvers, the cubic relation for calculating the production rate is replaced by a piecewise-linear continuous approximation. More precisely, as in the case study of Chapter 5, we apply linear segmentation (cf. [219]) and use six intervals with individual slopes and intercepts. Interval bounds are found by minimizing the error between the cubic function and the piecewise linearization.

6.3.3. Shut-down opportunities

In order to circumvent temporary price peaks, it can be beneficial to shut down the entire process for a certain period. Herein, we focus on shut-downs that allow for fast warm-starts after a limited down-time. For modeling these opportunities, we make use of the following assumptions:

- During the down-time, no product is produced and negligible electricity is consumed.
- Shut-downs are only possible if the process is operated at its lower operating bound. Thereby, we prohibit abrupt shut-downs from high utilization rates that might stress the equipment disproportionately.
- Warm-starts, i.e., start-ups after down-times, are only possible within a specified limited period.

Herein, we also study the influence of a varying maximum down-time, denoted by τ^{max} [h], on the different objective functions. To capture the highly different characteristics of the numerous DSM-relevant processes, we apply $\tau^{max} = 0 \dots 12$ h in Subsection 6.4.4.

Modeling of shut-downs relies on the introduction of binary variables indicating operating modes and transitions [5, 30, 37, 234, 235]. Constraints on the maximum down-time can then be established using a set of linear inequalities [30, 36]. Note that modeling the opportunities for shut-downs under the given assumptions further involves an adjustment of the ramping constraints introduced above, which can be found in [21].

6.4. Scheduling results for the generic process model

In this section, we evaluate to what extent different measures increasing the flexibility potential on the demand side can enhance the exploitation of the different characteristics of the input time series described in Section 6.2. For this purpose, we implement the scheduling optimization problem embedding the three versions of the generic process model from Section 6.3 in GAMS version 26.1.0 (GAMS Development Corp.). The documentations of the GAMS models can be found in Appendix D. Therein, we use an hourly discretization and consider the entire year. The resulting (MI)LPs are solved using CPLEX version 12.8 (IBM Corp.). Default solver settings are applied. Pareto curves when conducting bi-objective optimizations are identified using an ϵ-constraint method, furnishing 15 equidistant Pareto optima. In the following, optimization results are discussed in detail for the historic time series from 2017. Afterwards, a brief presentation of the key insights from using the projected time series for 2030 is given. Note that throughout this section, all objective values are relative to a stationary operation for comparison purposes.

6.4.1. Effect of average utilization rates on Pareto curves

First, we analyze the impact of different average utilization rates on bi-objective optimizations by varying the oversizing of the process, i.e., by decreasing the average utilization rate below its reference value $\overline{UR} = 95\%$, when considering the ideal storage-type customer model from Subsection 6.3.1. Furthermore, we also identify an optimized average utilization rate by treating it as an additional optimization variable. Pareto curves given in Figure 6.3 correspond to different process variants, spanning a range for various storage capacities S^{max} and ramping limits Δ^{max}.

- P1 - a variant with a low storage capacity $S^{max} = 3$ h and a limited load shifting capability due to severe ramping limits $\Delta^{max} = 5\%/\text{h}$. Consequently, the flexibility potential of this variant is expected to be substantially restricted.
- P2 - a variant with a low storage capacity $S^{max} = 3$ h, but strongly loosened ramping limits $\Delta^{max} = 25\%/\text{h}$. Compared to P1, the flexibility potential of the variant is higher.
- P3 - a variant with a limited load shifting capability due to severe ramping limits $\Delta^{max} = 5\%/\text{h}$, but with a significantly increased storage capacity $S^{max} = 48$ h. Again, there is a higher flexibility potential than for P1. Note that a comparison with P2 is not intended here.
- P4 - a variant with both an increased storage capacity $S^{max} = 48$ h and loosened ramping limits $\Delta^{max} = 25\%/\text{h}$, hence exhibiting the largest flexibility potential of all variants.

Figure 6.3 confirms the anticipated larger saving potentials in the economic objective compared to the environmental one in 2017 (cf. Table 6.1). Concerning an assessment of trade-offs between the objectives, the most important result is that for the idealized case without efficiency losses, there are mostly synergetic effects between economic and environmental objectives when conducting optimizations of the production schedule. In

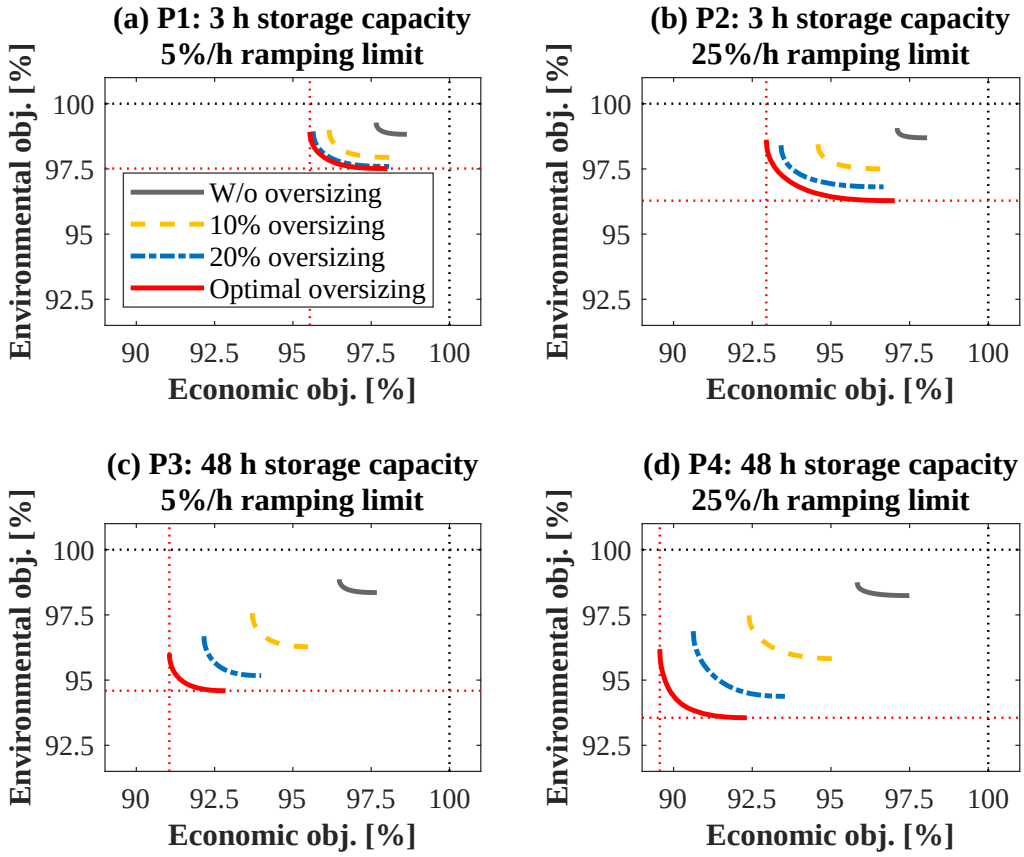

Figure 6.3.: Pareto curves for different parametrizations of the ideal storage-type customer model (cf. Subsection 6.3.1). The solid gray curve is the reference with average utilization rate $\overline{UR} = 95\%$. The dashed yellow curve corresponds to a 10% oversizing compared to the reference ($\overline{UR} = 86\%$), the dashed-dotted blue curve to 20% ($\overline{UR} = 79\%$). The solid red line is the Pareto curve for an optimal average utilization rate.

particular, in none of the considered cases, the optimization for a single objective leads to a deterioration of the other one compared to a stationary operation. Nevertheless, one sees that there is a nonnegligible space for balancing between the objectives. More precisely, optimizations for a single objective leave a substantial saving potential in the second objective of up to 3-4 percentage points unexploited. Summarizing, in case of near-ideal processes, i.e., with negligible efficiency losses, optimizations of the production schedule for economic performance, which are likely conducted from an entrepreneurial perspective, lead to improved environmental objectives, but do not exploit the full environmental potential. Furthermore, one finds a crucial importance of the average utilization rates in Figure 6.3: if these are too high, almost no saving potentials can be exploited, leading to incentives for the oversizing of production facilities. Decreasing the average utilization rate evenly affects the economic and the environmental objective, so that Pareto curves are almost parallel to each other for a fixed process variant. Along the same lines, we find a nearly constant optimal average utilization rate along the corresponding red Pareto curve in Figure 6.3. These findings are very important, as we can conclude that the average utilization rate that is favored from an economic perspective and thus highly relevant for

investment decisions, is also favored from an environmental perspective.

Finally, comparing the variants P1...P4, there is a strong dependence of both the achievable savings as well as the shape of the Pareto curves on the parametrization of the process variant. Whereas the first finding is an obvious result of the different flexibility potentials of the variants as discussed at the beginning of the subsection, the second finding is not as intuitive. For instance, we find that process variants with loosened ramping constraints exhibit Pareto curves, which extend over larger ranges of objective values and thus indicate more distinct trade-offs between the objectives. Moreover, in Figure 6.3 (b), the anchor points of the Pareto curves with lower average utilization rates do not always dominate the corresponding anchor points at higher average utilization rates, as is however the case for all other variants. These findings indicate that measures adjusting the storage capacities and those adjusting the ramping limits affect economic and environmental objectives in DSM in a different manner, which has been anticipated in Subsection 6.2.3 and will be discussed in more detail in Subsection 6.4.3.

6.4.2. Influence of off-design efficiency losses on Pareto curves

Before analyzing the influence of storage capacities and ramping limits, we first study the effect of off-design efficiency losses. For this purpose, bi-objective optimizations using the process variants introduced in Subsection 6.4.1 are repeated using a fixed oversizing of 20% (i.e., $\overline{UR} = 79\%$). In contrast, we now vary the intensity of the off-design efficiency losses (cf. Subsection 6.3.2). Comparing the Pareto curves under consideration of off-design efficiency losses in Figure 6.4 to their respective references without losses, one finds a significant influence of the loss intensity on the achievable savings through an almost parallel shifting of the Pareto curves. For instance, when considering rather low losses of 10% at the lower operating bound, the saving potential in both objectives for all process variants is reduced by $30 \ldots 50\%$. Moreover, high loss intensities bear a severe risk that optimizations for economic objectives that still yield promising cost savings lead to impaired environmental performance. In particular, applying the contribution to the residual load as the environmental objective as done in this chapter, an optimal spot market participation can lead to a net increase in the environmental objective. Most likely, if applying other objectives, e.g., the carbon footprint of the electricity consumption, similar observations will be made.

We furthermore remark that the findings presented above, i.e., that increased storage capacities and loosened ramping constraints do not affect economic and environmental objectives evenly, are crucial for assessing the risk of a net increase in the environmental objective. In particular, in the case of processes with low storage capacities and a high load shifting agility (i.e., P2), economic optimizations are more likely to lead to an increase in the environmental objective even for large low-loss operating ranges. In spite of that, our results indicate that the presence of off-design efficiency losses is not significant for the intended bi-objective assessment of measures increasing the flexibility potential, as we find rather even influences on the two objectives represented by nearly parallel Pareto curves.

6.4.3. Parameter study on storage and ramping constraints

As discussed in the previous subsection, it is sufficient to consider an ideal process without off-design efficiency losses for studying the influence of storage capacities and ramping

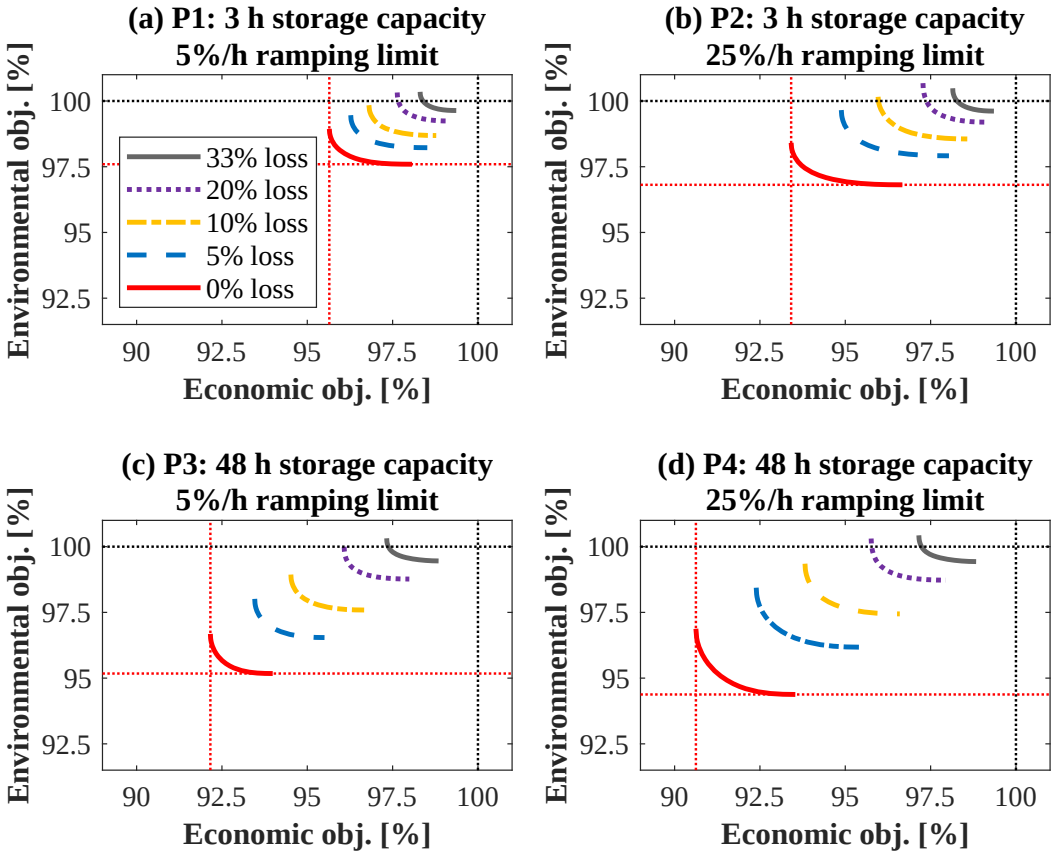

Figure 6.4.: Pareto curves for different parametrizations of the storage-type customer model when assuming varying curves for quadratic off-design efficiency losses (cf. Subsection 6.3.2) and using a fixed oversizing of 20%. The solid red curve is the reference without losses. The dashed-dotted blue curve corresponds to 5%, the dashed yellow curve to 10%, the dotted purple line to 20%, and the solid gray line to 33% losses at the lower operating bounds compared to the reference.

limits on the economic and environmental objectives. Therefore, we conduct a detailed parameter study on these influences on the results of single-objective optimizations using a fixed average utilization rate $\overline{UR} = 79\%$, which corresponds to an oversizing of 20% compared to the reference and thus enables promising saving potentials. In Figure 6.5, both objectives exhibit only weak gradients at the upper bounds of the considered parameter ranges, indicating only minor additional improvements from increasing storage capacities to above $S^{max} = 48$ h and loosening ramping limits to more than $\Delta^{max} = 25\%/\text{h}$. In fact, when discarding both storage and ramping constraints, the additional improvements are substantially smaller than what can be achieved by exhausting the parameter ranges from Figure 6.5. Thereby, our results indicate low additional economic and environmental values of providing seasonal storage capacities in the order of several hundreds of hours. Furthermore, it can be clearly seen that the two process parameters do not affect the two objectives evenly, representing a noticeable difference to the average utilization rate. In particular, we find that the influence of loosening ramping constraints, i.e., the influence of increasing the agility of a process, is much more distinct in the case of the economic

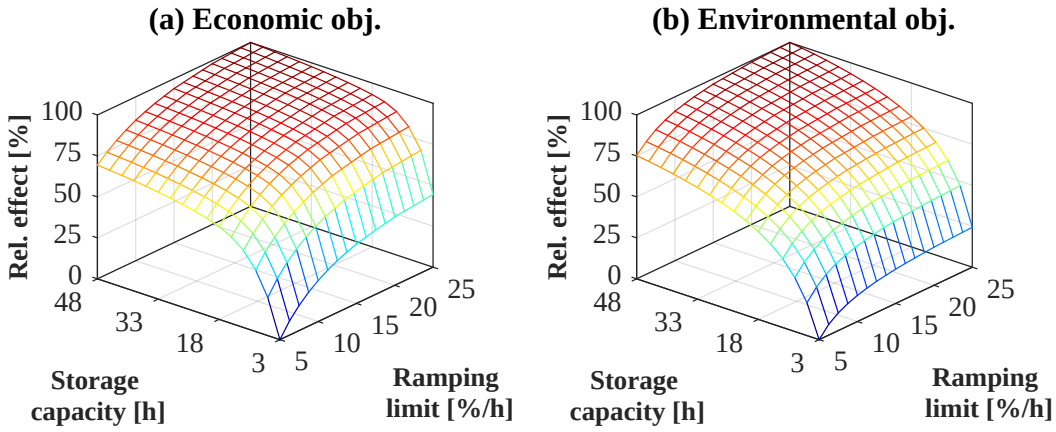

Figure 6.5.: Parameter study for single-objective optimizations assuming an oversizing of 20% compared to the reference with an average utilization rate $\overline{UR} = 95\%$. The relative effect, i.e., the additional savings, from increasing the storage capacity S^{max} and loosening the ramping constraints Δ^{max} is scaled between its minimum ($S^{max} = 3$ h, $\Delta^{max} = 5\%/\text{h}$) and maximum ($S^{max} = 48$ h, $\Delta^{max} = 25\%/\text{h}$).

objective. For instance, comparing a process with low storage capacities and limited load shifting capabilities (P1 with $S^{max} = 3$ h and $\Delta^{max} = 5\%/\text{h}$) to a process with substantial storage capacities and loose ramping limits (P4 with $S^{max} = 48$ h and $\Delta^{max} = 25\%/\text{h}$), one finds improvements in economic and environmental savings of factors of 2.15 and 2.34 respectively. In the case of the economic objective, ∼44% of this improvement can be obtained by only loosening the ramping constraints while maintaining low storage capacities (P2), whereas this value is only ∼24% for the environmental objective. If in contrast only increasing storage capacities while maintaining the severe ramping limits (P3), one achieves ∼69% of the maximum improvement in the economic objective, but even ∼75% of the maximum improvement in the environmental one.

Presumably, this behavior is caused by the different fluctuation patterns of the input time series, which are analyzed in Subsection 6.2.3 by means of discrete Fourier transform. Apparently, different spectra require different process capabilities for exploitation. More precisely, the identified electricity price spreads on short time scales (period lengths below 6 h) can only be exploited by processes with substantial load shifting capabilities, i.e., with loose ramping limits. As spreads in the residual load share on a similar scale are much less distinct, the impact of loosening ramping constraints is less significant in the case of the environmental objective. Following this argumentation, the very low sensitivity of the environmental objective to ramping limits above 10 . . . 15% is not surprising. Along the same lines, the differences in the spectra of the input time series also explain the course of the sensitivity of the objectives to storage limits. More precisely, the very distinct frequency component with a period length of 12 h in the price time series requires comparably low storage capacities for exploitation, resulting in high sensitivities of the economic objective in this range. As the frequency component with a period length of 24 h is less distinct, the sensitivity drastically decreases at higher storage capacities. As in the case of the environmental objective, the frequency component with a period length of 24 h is the dominant one, this behavior does not occur. With regard to a bi-objective assessment

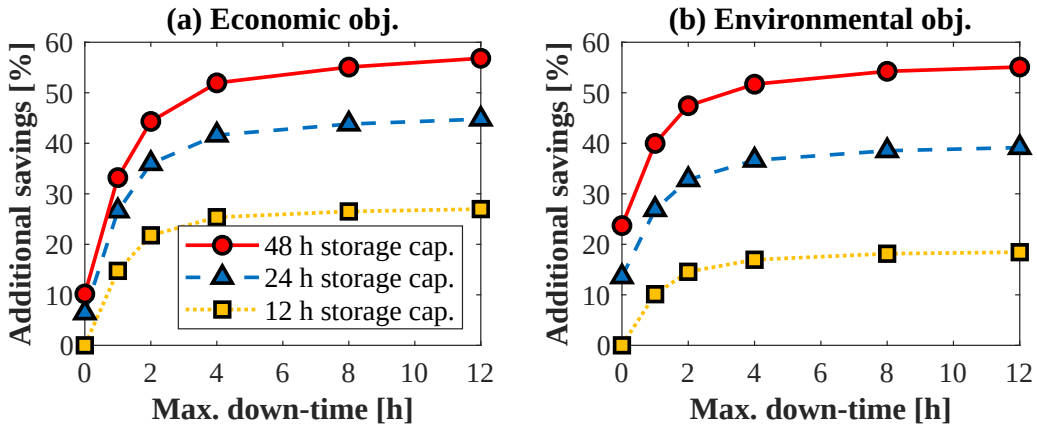

Figure 6.6.: Analysis of additional savings enabled by temporary shut-downs (cf. Subsection 6.3.3) as a function of the maximum down-time. A fixed oversizing of 20% (i.e., an average utilization rate $\overline{UR} = 79\%$) is used and a ramping constraint of 10%/h is imposed. Line styles indicate the storage capacity S^{max}. Yellow squares (□) with dotted lines: $S^{max} = 12$ h, blue upwards-pointing triangles (△) with dashed lines: $S^{max} = 24$ h, and red circles (○) with solid lines: $S^{max} = 48$ h. Values are scaled to the case with $S^{max} = 12$ h and without shout-downs.

of different measures increasing the flexibility potential on the demand side, the findings presented in this subsection are highly relevant, as they unveil a systematic problem: current price time series set promising monetary incentives for increasing the load shifting capabilities of a process, although this will not noticeably affect the ability of the process to contribute to a reduction of the residual load. Even more severely, there are no adequate monetary incentives for increasing the storage capacities of a process up to the level that would be favored from an environmental perspective.

6.4.4. Improvements from shut-down opportunities

Finally, we assess possible improvements from temporary shut-downs of the process (cf. Subsection 6.3.3) by performing single-objective optimizations. For this purpose, we again consider an ideal process without efficiency losses using a fixed oversizing of 20% (i.e., $\overline{UR} = 79\%$) and a fixed ramping constraint of $\Delta^{max} = 10\%/\text{h}$. In the study, we vary both the storage capacity S^{max} and the maximum down-time τ^{max}. As can be seen in Figure 6.6, temporary shut-downs are incentivized from both the economic and the environmental perspective, as they allow for avoiding peak hours in electricity prices as well as in residual load shares. Nevertheless, the differences between the two objectives are apparent. In particular, we find that the additional savings through temporary shut-downs are substantially more distinct from the economic perspective and predominate the benefits from a further increase in the storage capacities. For instance, when considering a storage capacity of 12 h as a reference, savings enabled by shutting down the process for a maximum of only 1 h are more than twice as high as the savings from doubling the storage capacities without shut-down opportunities and still 50% higher than from quadrupling the storage capacities. Similar observations however cannot be made for the environmental objective. Here, the benefits of increased storage capacities are more important.

As in Subsection 6.4.3, the reason for the described differences in the behavior in Figure 6.6 (a) and (b) lies in the different fluctuation patterns of the electricity price and the residual load share time series. In particular, electricity price spreads occur on faster time scales, leading to a higher "value" of short down-times of few hours compared to further increased storage capacities that are not required for short-term buffering. For the contribution to the residual load, in contrast, the frequency components with large period lengths correspond to the only important spreads, explaining the higher importance of storage. Note that the differences in the influence on economic and environmental objectives are not as distinct for temporary shut-downs as for loosened ramping constraints, which hardly affect environmental savings (cf. Subsection 6.4.3). This finding stems from the fact that shut-downs also widen the operating range, which is equally beneficial from both the economic and the environmental perspective as discussed in Subsection 6.4.1.

6.4.5. Key insights from the 2030 study using the projected time series

In this subsection, we only qualitatively describe the key observations from the scheduling optimizations considering the projected time series for 2030 and omit a detailed presentation of quantitative results. Corresponding figures can however be found in Appendix E. In particular, we are herein interested in the question of whether the drawn conclusions will remain valid in the future, where electricity markets will be characterized by an even stronger penetration of generation from intermittent renewable sources. Furthermore, the considered setting also allows for assessing whether an imperfectly competitive market behavior, which characterizes the historic 2017 price time series but not the projected 2030 time series, distorts the observations.

Comparing the results for 2030 with those for 2017, the most noticeable difference is that the relative saving potentials are now larger for the environmental objective as anticipated in Subsection 6.2.1 by comparing the widths of the distributions. Note that this also has a considerable effect on the shape of the Pareto curves. The differences become most recognizable when considering off-design efficiency losses - here, optimizations for the environmental objective now bear a risk of increasing costs and not vice versa as in 2017. Furthermore, the Pareto curves seem to shrink, which is in good agreement with the higher correlation coefficient in 2030 (cf. Subsection 6.2.2). Note that both effects are likely a direct consequence of the decreasing relative economic saving potentials in 2030. However, as discussed above, these can at least partially be explained by an assumed ideally competitive market avoiding extreme price peaks. In the present thesis, we do not intend to draw a final conclusion whether an increase in the renewable generation will cause an increase in the relative price spreads in addition to the inevitable rise in both the average prices and the absolute spreads. In contrast, we strongly emphasize that these findings do not affect the prioritization of different possible measures increasing the flexibility potential at the demand side and should thus be discussed elsewhere.

The observations made in Subsections 6.4.3 and 6.4.4 are indeed still visible when considering the projected time series for 2030, although they appear slightly weakened. We thus conclude that parts of the high-frequency electricity price fluctuations that set incentives for loosening ramping constraints and creating shut-down opportunities but not for increasing storage capacities, presumably stem from nonideal market behavior. However,

the statements made concerning uneven influences of different measures increasing the flexibility potential on the two objectives remain unaffected. In particular, this includes the finding that measures loosening ramping limits or allowing for temporary shut-downs are also in the future still significantly incentivized from an economic perspective, although they continue to exhibit only minor capabilities for improving the environmental objective. Consequently, our observations are primarily caused by the prevalent market mechanisms governing the price setting that impose the differences in time scales for the fluctuations in electricity mixes and prices (cf. Subsection 6.2.3) irrespective of the actual partially nonideal strategic behavior of market participants.

6.5. Summary and recommendations for future research

We present a systematic assessment of measures increasing the flexibility potential on the demand side with regard to both economic and environmental objectives. In particular, we conduct a massive simulative study focusing on the German electricity sector using both historic electricity time series data as well as projected time series data for the future. Therein, the economic perspective is represented by achievable cost savings for the electricity purchase through the exploitation of day-ahead spot market price spreads, whereas environmental impacts are measured in a simplifying yet tangible manner by reductions in the contribution to the integral residual load through the exploitation of time-variable electricity mixes. We first analyze the input time series to be exploited and find different contributions of certain frequency components to the spectra. In particular, the electricity price time series is characterized by fluctuations on shorter time scales, typically comprising two minima a day. In contrast to that, the share of renewable electricity generation is mostly characterized by a day/night fluctuation.

We further present a model formulation for a generic process that acts as a battery from a grid level perspective to perform single- and bi-objective scheduling optimizations exploiting the fluctuations in the time series. The generic model comprises general formulations for DSM activities for load shiftings, includes the possibility for temporary shut-downs, and allows for taking off-design efficiency loss characteristics into account. By varying the constitutive process parameters, we consider the vast majority of DSM-relevant processes from literature, which allows for a systematic treatment not limited to a specific application. In particular, we thereby find that measures loosening ramping limits or allowing for temporary shut-downs are significantly incentivized from an economic perspective, although they exhibit only minor capabilities for reducing the residual load. The latter is achieved much more effectively through substantially increased storage capacities, which are in turn not monetarily incentivized in an adequate manner. Measures optimizing the operating range of processes - which are commonly designed for nearly full utilization - by an oversizing of existing production facilities however affect economic and environmental objectives evenly. The considered case studies using both historic and projected future time series lead to similar observations, indicating that the presented findings ultimately originate from the prevalent price setting mechanism of the electricity market and not from the strategic behavior of market participants.

Although our analysis suggests that comparable observations are likely when considering other countries, whose electricity markets apply similar pricing schemes and whose energy sectors are characterized by comparably high penetrations of intermittent renewable elec-

tricity sources as the German one, we recommend additional studies using both historic and projected future time series to validate the findings. Along these lines, we are highly interested in extending the scope of the simulative study towards market environments that account for network congestion in the price setting by applying nodal prices, such as several US-American electricity markets. In particular, this should address the question of whether regional differences in the renewable electricity generation, which induce regional price signals due to limited transmission capacities, then also lead to different regional prioritizations of measures increasing the flexibility potential on the demand side.

Moreover, future research perspectives lie in methodological developments that could further increase the generalizability and hence the significance of the results. In particular, future research should revisit the applied objective functions and consider studying the impacts of participating in real-time markets, which are commonly characterized by more distinct price peaks in comparison to day-ahead markets. In addition to that, a fair quantification of the environmental benefits of load shiftings from hours with high renewable shares in the electricity mix to ones with lower shares remains an open challenge. Furthermore, a thorough study of the interactions between seasonally varying product demands and the provision of load flexibility is recommended. Along these lines, we are also interested in assessing the impacts of changing raw material requirements and products in case of load variations as well as the potentials of switching the energy sources. Note that whereas the first is presumably highly process-specific and therefore it would be hard to derive generalizable conclusions, the latter will apply for numerous processes - for instance, consider power-to-heat concepts and gas-fired but electrically boosted furnaces, e.g., for glass production. Here, temporarily increasing/decreasing the electricity consumption might be motivated either by economic or environmental considerations that do not necessarily need to result in similar operating schemes. Finally, we see clear benefits if incorporating design aspects into the scheduling optimizations by considering annualized costs and environmental impacts, but at the same time, acknowledge that these will be highly process-specific again and in some cases also hard to quantify, particularly when assessing the environmental impacts.

7. Conclusion and discussion of future research perspectives within the context of industrial demand side management

Industrial DSM is nowadays considered as an important measure to align the network-wide electricity consumption with a volatile generation and thereby enhance the transformation of the energy system towards a high penetration of intermittent renewable electricity sources. As at the same time, high shares of volatile electricity generation induce price fluctuations at liberalized electricity markets, there is also an economic motivation for flexibilizing the operation of energy-intense processes. In order to raise the economic potentials, model- and optimization-based approaches have become the method of choice. In particular, numerous works have formulated the provision of load flexibility as a scheduling problem with time-variable electricity prices using quasi-stationary linearized process models to assess economic benefits from DSM for a large variety of processes. Moreover, there have been several recent works that argue to account for the transient process behavior during load changes in scheduling optimizations for processes with slow dynamics and to apply low-dimensional dynamic models for that purpose. In addition to that, authors have started to also take compensation payments for the provision of ancillary services into account when assessing the monetary advantages from DSM. Finally, there is a low yet increasing number of publications that target the quantification of environmental benefits when performing DSM by either assessing the environmental impacts of economically optimized schedules or directly conducting scheduling optimizations for environmental objectives. Against this backdrop of a vast amount of literature focusing on DSM and using model-based approaches for this purpose, we herein identify promising further research perspectives beyond the described state-of-the-art and propose model-based methodologies to address them systematically.

First, we propose an algorithm relying on a reduced-space scheduling formulation, which exposes only the actual decision variables to the optimizer, to allow for long planning horizons with fine discretizations in discrete-time scheduling although considering nonlinear process models. In the original version of the algorithm, we assign one DoF to multiple generally nonconsecutive intervals with similar electricity prices and thereby apply a tailored time grid relying on fewer decision variables. A wavelet-based analysis of the solution applying a given time grid is conducted to iteratively refine the grid in a systematic manner. We then extend the algorithm by a sounder sensitivity-based refinement strategy in the case of scheduling problems comprising only continuous DoFs. To this end, we perform optimizations directly in the space of coefficients of the wavelet transform. In this case, dimensionality reductions correspond to truncations of the wavelet transform, which

allows for calculating the marginal effect of using specific additional decision variables in the optimization problem. Both versions of the algorithm are applied to the operational optimization of processes comprising turbomachines that are characterized by strongly nonlinear efficiency characteristics, for which we use ANNs as surrogate models. In the two case studies, namely a CAES plant and an ASU, we demonstrate that the use of the proposed algorithms allows for low-dimensional approximations of the original problem considering the full dimensionality. In particular, we show that formulations relying on substantially reduced numbers of decision variables can still exploit almost the full economic potential. These reductions in the dimensionalities of the optimization problems ultimately lead to significant savings in computational times when performing deterministic global optimizations, circumventing the worst-case exponential scaling of solution times with the horizon length if considering the full dimensionality.

Second, we address the continued demand for highly accurate yet computationally tractable dynamic models for distillation columns by presenting a hybrid mechanistic/data-driven modeling approach, enhancing the classical compartmentalization with surrogate modeling using ANNs. Thereby, the proposed ANN-based compartment enables substantial reductions in the number of differential variables without significantly influencing the quality of dynamic responses and at the same time, omits time-consuming flash calculations on each stage. Due to substantial improvements in the computational efficiency in dynamic optimizations compared to the use of full-order stage-to-stage models, we apply the proposed reduced hybrid model within a single-layer eNMPC framework. Therein, we also make use of a suboptimal fast update method confining to a limited number of major SQP iterations without interim linesearch. This combination finally allows for making decisions about energy trading in real-time directly at the control layer in the case of processes with large time constants. We consider the control of a nitrogen plant participating in a continuous intra-day electricity market as a case study and demonstrate that the proposed eNMPC framework combining reduced hybrid models with suboptimal update methods enables adjustments of the production schedule in real-time based on newly updated price forecasts. Moreover, we find that applying the eNMPC framework allows for economic savings compared to stationary process operation strategies that are close to the maximum achievable savings when assuming no model/plant mismatch and perfect price forecasts. Finally, we also show that the proposed single-layer eNMPC framework can effectively reject disturbances with comparably slow dynamics.

Third, we propose the formulation of a mathematical program to identify optimal bidding strategies at pay-as-bid markets for ancillary services. Namely, we herein consider the participation in the German market for PRL, where participants offer the provision of reserve capacities for maintaining the grid frequency within desired bounds and receive valuable compensation payments in return. To allow for optimal decision-making accounting for revenues from both the PRL market and the day-ahead spot market, we propose a two-stage stochastic MINLP minimizing the expected weekly production costs and balancing between higher probabilities of acceptance of the bids and higher potential compensation payments. To this end, the first stage is formulated as an MINLP and corresponds to the identification of the optimal bidding strategy at the PRL market, i.e., what amounts to offer and what prices to ask given a forecast of the development of the maximum compensation payments. Depending on the results of the PRL market auction, a subordinated MILP scheduling optimization is conducted to market the remaining flexibility at the spot electricity market in the second stage. The two-stage problem is solved

using a tailored enumeration approach, which decomposes the MINLP formulation of the deterministic equivalent into finite numbers of NLPs identifying the optimal ask prices and MILPs determining the corresponding production schedules. Optimal bidding strategies are identified for a flexible industrial aluminum electrolysis cell using an autoregressive forecasting model for the upcoming maximum accepted ask prices. We identify a high economic advantage when simultaneously marketing the operational flexibility at both the PRL and the day-ahead market. Additionally conducted parameter studies identify important influences of the process characteristics and the quality of the price forecasts for both the spot and the PRL market on the optimal bidding strategies.

Fourth, we target a systematic approach towards a bi-objective assessment of measures increasing the flexibility potential on the demand side, enabling generalizable conclusions not limited to a specific application and providing thorough explanations for the observations. To this end, we assess if measures that enhance the exploitation of fluctuating electricity prices, promote the penetration of renewables by allowing to consume electricity primarily in periods with a high renewable share. We conduct a simulative study focusing on the German electricity sector using both historic electricity time series data as well as projected time series data for the future. Therein, the economic perspective considers cost savings through the exploitation of day-ahead spot market price spreads, whereas environmental impacts are assessed by quantifying the reductions in the contribution to the integral residual load. We analyze the fluctuating input time series and find different contributions of certain frequency components to the spectra. In particular, whereas the electricity price time series is characterized by fluctuations on a sub-daily scale with a period length of 12 h, typically comprising two minima a day, the share of renewable electricity generation is mostly characterized by a daily fluctuation with a period length of 24 h and a peak at noon. Furthermore, we present a generic model formulation comprising general formulations for DSM activities for load shiftings, including the possibility for temporary shut-downs, and allowing for the consideration off-design efficiency losses. Using this model, we perform single- and bi-objective scheduling optimizations exploiting the fluctuations in the time series. By varying the constitutive process parameters, we find that there are significant monetary incentives for increasing the agility of a process, although this does not enable improvements in environmental objectives. In contrast, substantially increased storage capacities are not monetarily incentivized in an adequate manner, although they are highly crucial for reducing the contribution to the integral residual load. The considered simulative studies apply both historic and projected future time series and lead to similar observations, indicating that these findings are a direct result of the prevalent price setting mechanism of the electricity market and do not stem from a nonideal behavior of market participants.

Certainly, there are opportunities for further developments in the proposed methodologies as discussed at the end of the respective chapters. However, there are three open challenges, representing future research directions within the broader context of the present thesis, that we consider to be of high importance, so that we like to discuss them in the following. Systematically addressing them will certainly require sophisticated model-based approaches that might partially rely on or - at least to some extent - benefit from the methods presented in this thesis.

(i) Integrating demand side management into the design perspective

Our results as well as many others from the literature show that performing DSM can offer substantial economic benefits by making operational decisions in response to time-variable electricity prices. This gives rise to the question to what extent these adjusted operating schemes could affect decisions made on superordinated layers that address long-term strategic (investment) decisions. Along these lines, few authors have started to account for investments in retrofits of existing processes to enhance the exploitation of time-variable electricity prices, for instance, by allowing for the extension of buffering capacities (cf. [7]) or the purchase of additional electrolyzer stacks (cf. [233]). However, we see great potential when also accounting for these circumstances once designing novel processes. More precisely, future process designs should account for a flexible operating scheme and thus carefully evaluate if additional investments in flexibility would amortize within the lifetime of the process. Against this backdrop, we also refer to the work of Burre et al. [229] who argue for a unifying perspective on DSM, electric energy storage, and e-production under the heading "Power-to-X"'. That is, we emphasize that although processes for storing surplus electricity (either in form of desired products or to release it in times of shortages) are explicitly meant to be operated flexibly, they are commonly designed considering stationary operating conditions (e.g., [236–238]).

Integrating the operational perspective into design optimizations will however lead to complex large-scale optimization problems. Note that this will particularly hold if processes are characterized by nonnegligible time constants that require the use of dynamic models and/or strongly nonlinear behavior, leading either to large-scale mixed-integer dynamic optimization (MIDO) problems or MINLPs to be solved. Along these lines, approaches to reduce the dimensionality, such as the methods for time series aggregation referred to in Chapter 2 or the algorithms proposed in Chapter 3, might play an important role. We emphasize that there is nevertheless still also a continued demand for efficient decomposition methods to solve the integrated optimization of design and operation. Note that this will apply even more, as the long-term prediction of electricity prices, which is essential with regard to investment decisions, involves substantial uncertainties. Ideally, these uncertainties should be taken into account by using two-stage stochastic or, in some cases, potentially also robust problem formulations, which makes it even more difficult to solve the corresponding optimization problems.

(ii) Bringing developed operating schemes into industrial practice

Whereas the first perspective is concerned with further developments in model formulations and solution approaches, the second one rather addresses the remedy of hindrances for the implementation of the developed methods in industrial practice. Note that in the following, we will set the focus of the discussion on changes in the control strategy, as these will have more direct influences on the actual process behavior than adjustments in the superordinated production planning/scheduling. Here, the combination of reduced hybrid mechanistic/data-driven dynamic models with fast update methods for eNMPC proposed in Chapter 4 has been proven in silico to be highly efficient for a real-time capable economic decision-making (with potential further methodological developments outlined at the end of the chapter). However, the applicability of such approaches essentially relies on the availability of suitable (reduced hybrid) models. Revisiting our approach for hybrid

mechanistic/data-driven modeling as well as the others referenced throughout the thesis, one finds that in almost all cases, the hybrid modeler has been in the comfortable situation of having the "real" model for data generation. Undeniably, in industrial settings, this will not always be the case, as the construction of high-fidelity full-order models is tedious and expensive. Consequently, the required data will necessarily originate from measurements, involving either costly experiments or interventions of the operation of a process. The structure of the hybrid models therefore needs to match the available data, that is, e.g., only those relations can be replaced where both the input and the output of the data-driven model can be measured or estimated. Moreover, in cases where the only data source is an existing monitored process, one finds a severe conflict of interest between model identification (that aims at covering the entire operating range) and a safe operation of the process (that preferably stays in regions where the model is accurate). Opportunities for the adaptive identification of data-driven model parts without conscious disruptions of the operation of the process thus appear highly promising to facilitate the implementation of the proposed hybrid model structures in industrial practice.

As stated above, the derivation of high-fidelity full-order process models, which would enable offline opportunities to generate the required training data for the surrogate models, is nowadays still considered as a substantial intervention of established workflows and therefore mostly avoided. Against this backdrop, the increasing importance of digitalization and concepts like digital twins however indicate an upcoming paradigm change. In particular, the benefits from having a sound model-based decision-support throughout the lifetime of a plant are increasingly recognized as a measure to ensure the competitiveness of processes [239–241]. This means that high-fidelity full-order dynamic models might eventually become readily available, allowing for generating accurate process data in-silico without performing costly experiments or disrupting the plant's operation, which would substantially facilitate hybrid mechanistic/data-driven approaches to reduced dynamic modeling for control purposes.

(iii) Designing future electricity markets that set the right incentives

Finally, we would like to argue for a more system-wide assessment of industrial DSM that carefully evaluates its benefits from the perspective of the entire electricity sector and compares it to different alternative flexibility measures for the electricity grid, such as DSM of households, the use of electric energy storage systems, and the provision of conventional or dispatchable renewable back-up producers. Here, interdisciplinary researchers need to identify those technical solutions involving the lowest macroeconomic costs while fulfilling the primary objectives. Among these, the reduction of the emissions associated with the electricity generation by substitution of fossil sources by renewable ones and the stabilization of the electricity grid, i.e., the maintenance of the security of supply through the remedy of both temporal and spatial shortages, certainly belong to the most important objectives. Note that due to the severeness of the challenge to transform the electricity sector towards very high shares of electricity generation from mostly intermittent renewable sources, we rather expect that a large variety of technological alternatives will indeed be required to meet this challenge.

Once those options involving the lowest macroeconomic costs are identified, their system-wide value needs to be aligned with microeconomic benefits, i.e., desired technological alternatives must be incentivized, so that entrepreneurial investments become profitable.

This implies a critical assessment if contradictions exist, which prevent adequate incentives, as is shown in Chapter 6 for the provision of storage capacities that allow for load shiftings on daily time scales. In the case that market principles of balancing demand and supply that cause temporal price spreads do not set adequate investment incentives, an alternative lies in an active incentivization, as is currently done for ancillary services that balance short-term frequency fluctuations (cf. Chapter 5). Along these lines, there is a strong need to reconsider specific aspects of the energy economics that originate from times with an electricity generation exclusively from dispatchable power plants. This will certainly hold for the legislative framework that currently de facto penalizes the provision of load flexibility by industrial processes [242]. Likewise, the current price setting schemes require a revision. For instance, Schäfer et al. [243] find that the intra-day trading of rectangular energy blocks, which is economically clearly incentivized by even sub-hourly price spreads, substantially contributes to the prevalent frequency fluctuations and thus impairs the grid stability. In this context, model-based methods for industrial DSM (and hence also for the operational optimizations of other "Power-to-X" processes as discussed above) become highly relevant, as they enable an assessment how given market designs influence entrepreneurial decisions, which then in turn allows for simulating the effects on electricity grids and thus finally also for evaluating given market designs or ultimately for deriving improved ones.

Appendix A.

Additional information on the wavelet-based grid adaptation

A.1. Grid point deletion and insertion algorithms

Recall that $\boldsymbol{\lambda}_i^{s*}$ denotes a vector containing all wavelet coefficients $\lambda_{i,a,b}^{s*}$. Then, a suitable criterion for grid point deletion that uses a threshold $0 \leq \epsilon^d \leq 1$ reads:

$$|\lambda_{i,a,b}^{s*}| < \epsilon^d \cdot \|\boldsymbol{\lambda}_i^{s*}\|_2 \quad \forall a, b.$$

If a wavelet coefficient $\lambda_{i,a,b}^{s*}$ is selected for deleting a grid point, the corresponding basis function $\psi_{a,b}$ will be discarded when deriving the matrix $\boldsymbol{B}^{s+1}$ for the next iteration (cf. Appendix A.2).

For grid point insertion, only the boundary coefficients are considered. The coefficient $\lambda_{i,a,b}^{s*}$ is called a boundary coefficient if $\lambda_{i,a+1,2b}^{s*} = 0 \vee \lambda_{i,a+1,2b+1}^{s*} = 0$ holds. Let there be vectors $\tilde{\boldsymbol{\lambda}}_i^{s*}$ containing all boundary coefficients in descending order and $\tilde{\boldsymbol{\lambda}}_{i,o}^{s*}$ containing the o highest boundary coefficients in descending order. Then, we apply the following criterion for insertion:

$$\left\|\tilde{\boldsymbol{\lambda}}_{i,o-1}^{s*}\right\|_2 < \epsilon^i \cdot \left\|\tilde{\boldsymbol{\lambda}}_i^{s*}\right\|_2 < \left\|\tilde{\boldsymbol{\lambda}}_{i,o}^{s*}\right\|_2 .$$

that corresponds to selecting the o highest boundary coefficients, i.e., the entries of $\tilde{\boldsymbol{\lambda}}_i^{s*}$ up to index o, for grid point insertion. If a wavelet coefficient $\tilde{\lambda}_{i,a,b}^{s}$ is selected for inserting grid points, the two basis functions $\psi_{a+1,2b}$ and $\psi_{a+1,2b+1}$ will be included when deriving the matrix $\boldsymbol{B}^{s+1}$ for the next iteration (cf. Appendix A.2).

Selecting the same coefficient for both grid point deletion and insertion during one iteration is excluded. Furthermore, we only allow for deletion of those boundary coefficients where $\lambda_{i,a+1,2b}^{s*} = 0 \wedge \lambda_{i,a+1,2b+1}^{s*} = 0$ holds.

A.2. Derivation of the assignment matrix

The matrix $\boldsymbol{B}^{s+1}$ is constructed in a recursive manner, starting from $\boldsymbol{B}^{s+1} = 1$ and $a = 0$ and iterating from $b = 0$ to $b = 2^a - 1$ each time. A wavelet basis function $\psi_{a,b}$ that shall be included for $\boldsymbol{B}^{s+1}$ adds both a row and a column to the matrix. A discarded basis function $\psi_{a,b}$ in contrast only adds a new row. For instance, if $\psi_{0,0}$ shall be used, the matrix is adjusted to:

$$\boldsymbol{B}^{s+1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and we set $a = 1$. This time, we assume that $\psi_{1,0}$ shall be used, whereas $\psi_{1,1}$ shall remain unused. The first adjustment then results in:

$$\boldsymbol{B}^{s+1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The second step gives:

$$\boldsymbol{B}^{s+1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

The procedure is repeated N times. Thereby, a matrix $\boldsymbol{B}^{s+1}$ is finally obtained with 2^N rows and as many columns as wavelet basis functions that shall be used in the next iteration.

Appendix B.

Unit models for the nitrogen plant case study

B.1. Compressor and turbine

Air is assumed an ideal gas with an isentropic coefficient $\gamma = 1.401$. The electrical energy consumption is calculated from the isentropic relation and further accounts for losses in efficiencies in case of off-design operation. We assume isentropic efficiencies $\eta_C = 0.8$ for compressors and $\eta_T = 0.9$ for turbines. Besides, we assume an electrical efficiency of the engine $\eta_E = 0.9$. All efficiencies are assumed to depend quadratically on the air flow and decrease by $\Delta\eta = 0.025$ at the operating limits of $\pm 10\%$. Incoming feed air enters the main air compressor at atmospheric pressure and exits at 10 bar. The bypass flow through the turbine is expanded from 9 bar to 6.6 bar.

B.2. Intermediate chiller

After leaving the main air compressor, the feed stream is chilled to the ambient temperature of 25° C.

B.3. Counter current heat exchanger

Steady-state material and energy balances are assumed for both heat exchanger zones. The heat exchanger zone 1, where no phase change takes place, is discretized into five well-mixed finite volumes. In contrast, the heat exchanger zone 2 is assumed to be only one lumped volume. The heat transfer coefficients kA are adjusted, so that minimum driving temperature differences are kept for all discrete volumes at nominal operation.

B.4. Liquefaction cycle

In the liquifier, part of the product stream is cooled down to its dew temperature and then completely liquified. For calculating the electricity demand, we assume that the heat removal is realized via a refrigeration cycle with an energy efficiency ratio of $\zeta = 2.5$, i.e., the electricity demand equals 40% of the heat removal.

The storage tank itself is modeled via a dynamic balance equation:

$$\dot{M} = L_{in} - L_{out}.$$

For the re-vaporization, we assume that the energy requirement is negligible.

B.5. Distillation column

The column consists of 40 theoretical stages. The incoming feed air enters the column at the bottom stage. Operating pressures are 6.6 bar at the bottom stage and 6.4 bar at the top. The reboiler operates at a pressure of 2.5 bar. The reflux from the condenser is assumed to be saturated liquid. Proportional controllers are used for levels in the column sump as well as in the integrated reboiler and condenser.

B.5.1. Thermodynamic models

When solving the flash calculations, the chemical potential of the liquid phase is calculated using the NRTL model for activity coefficients and the extended Antoine equation for vapor pressures. The vapor phase is modeled as an ideal gas. Moreover, we use the DIPPR 106 equation for the enthalpy of vaporization. All thermodynamic parameters are retrieved from the Aspen Plus version 8.8 database (Aspen Technology, Inc.).

B.5.2. Compartmentalization

We use the following compartmentalization approach: stages 1-9 (counted from top to bottom) form compartment 1, stages 10-19 form compartment 2, stage 20-29 form compartment 3, and stages 30-39 form compartment 4. The feed stage (stage 40) and the reboiler are modeled separately. The condenser, which has negligible hold-up, is assumed to be part of compartment 1. For each compartment, we generate 250,000 samples of input/output data.

B.5.3. Surrogate modeling

For training of the ANNs replacing the stage-to-stage calculations determining the intra-compartment input/output behavior, we apply single-layer feed-forward network structures with a minimum of 5 neurons in the hidden layer. We iteratively increase the number of neurons until a minimum performance criterion is undercut or a maximum number of 50 neurons is reached. As a performance criterion, we use the RMSE. The standard fitting procedure of MATLAB R2018a's Neural Network Toolbox (Mathworks, Inc.) is applied using 70% of the samples for training, and each 15% for testing and validation. The Levenberg-Marquardt backpropagation is used as training algorithm.

Following the described approach, the ANNs summarized in the table below are obtained for modeling the input/output correlations inside the compartments:

Table B.1.: Network sizes (i.e., number of hidden neurons) for input/output correlations inside each compartment and root mean squared error (RMSE) for the test set during the fitting procedure.

Compartment	**Output**	**#Neurons [-]**	**RMSE [-]**	**Unit**
1	L_9	8	7.88E^{-4}	mol/s
1	Q	8	2.09	W
1	$\log(1-x_9^{N_2})$	25	2.48E^{-4}	[-]
1	$\log(x_9^{Ar})$	25	2.18E^{-5}	[-]
1	$\log(1-y_1^{N_2})$	50	4.43E^{-3}	[-]
1	$\log(y_1^{Ar})$	50	4.39E^{-3}	[-]
1	h_1^V	10	1.06E^{-4}	J/mol
2	L_{19}	10	4.50E^{-4}	mol/s
2	V_{10}	10	8.74E^{-4}	mol/s
2	$\log(1-x_{19}^{N_2})$	25	1.73E^{-4}	[-]
2	$\log(x_{19}^{Ar})$	25	3.00E^{-5}	[-]
2	$\log(1-y_{10}^{N_2})$	40	7.58E^{-4}	[-]
2	$\log(y_{10}^{Ar})$	40	1.96E^{-3}	[-]
3	L_{29}	20	9.12E^{-4}	mol/s
3	V_{20}	20	9.96E^{-4}	mol/s
3	$\log(1-x_{29}^{N_2})$	20	2.01E^{-4}	[-]
3	$\log(x_{29}^{Ar})$	20	1.06E^{-4}	[-]
3	$\log(1-y_{20}^{N_2})$	30	3.68E^{-4}	[-]
3	$\log(y_{20}^{Ar})$	30	1.61E^{-4}	[-]
4	L_{39}	30	1.08E^{-3}	mol/s
4	V_{30}	30	1.54E^{-3}	mol/s
4	$\log(1-x_{39}^{N_2})$	20	7.83E^{-5}	[-]
4	$\log(x_{39}^{Ar})$	20	3.51E^{-4}	[-]
4	$\log(1-y_{30}^{N_2})$	25	1.68E^{-4}	[-]
4	$\log(y_{30}^{Ar})$	25	1.01E^{-4}	[-]
4	h_{39}^L	12	9.27E^{-2}	J/mol

Appendix C.

Regression model for forecasting the development of the market for primary balancing power

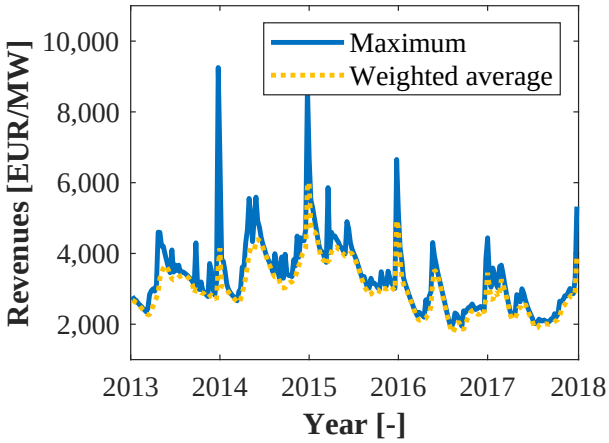

Figure C.1.: Historic accepted ask prices for primary balancing power for the period 2013-2017. The solid blue line shows the maximum accepted ask price for each week, the dotted yellow line additionally gives the weighted average of all accepted ask prices. The data is retrieved from `www.regelleistung.net`.

In Figure C.1, historical data for the maximum accepted ask price for PRL in the period 2013-2017 is plotted. We assume that the maximum accepted ask price $\widetilde{A}^{PRL,k}$ for week k consists of three components and hence reads:

$$\widetilde{A}^{PRL,k} = f^{PRL,k}_{Short-term} \cdot f^{PRL,k}_{Season} \cdot \widetilde{A}^{PRL,k}_{Base}$$

where $\widetilde{A}^{PRL,k}_{Base}$ is the base component for the maximum accepted ask price in week k and is calculated as the average over the previous 52 weeks:

$$\widetilde{A}^{PRL,k}_{Base} = \frac{1}{52} \cdot \sum_{l=1}^{52} \widetilde{A}^{PRL,k-l}.$$

The factor $f^{PRL,k}_{Season}$ further accounts for the yearly repeating trend that can be observed in in Figure C.1 (high prices around Christmas and new year, low prices during summer). It

is calculated by averaging the ratios between actual prices and base prices in each calender week of the year. Having data for both actual prices and base prices for the past three years, this factor is calculated by:

$$f_{Season}^{PRL,k} = \frac{1}{3} \cdot \sum_{m=1}^{3} \frac{\widetilde{A}^{PRL,k-52 \cdot m}}{\widetilde{A}_{Base}^{PRL,k-52 \cdot m}}$$

where we assume a year to consist of exactly 52 weeks in the notation for the sake of readability.

The factor $f_{Short-term}^{PRL,k}$ finally represents short-term market effects. We assume that the logarithm of this factor can be expressed by a stationary zero-mean ARMA(P_{ARMA},Q_{ARMA}) model:

$$y^{PRL,k} = \log\left(f_{Short-term}^{PRL,k}\right)$$

$$y^{PRL,k} = \epsilon^{PRL,k} + \sum_{p=1}^{P_{ARMA}} \phi_p \cdot y^{PRL,k-p} + \sum_{q=1}^{Q_{ARMA}} \theta_q \cdot \epsilon^{PRL,k-q}$$

$$\epsilon^{PRL,k} \sim \mathcal{N}\left(0, \sigma^2\right).$$

The current value of $y^{PRL,k}$ is hence the sum of a zero-mean white noise $\epsilon^{PRL,k}$ with variance σ^2, a weighted sum over P_{ARMA} previous (autoregressive) values and a moving average of Q_{ARMA} previous noise terms. The parameters ϕ_p and θ_q are estimated to minimize the error between model forecasts and measured values. The model order $(P_{ARMA}, Q_{ARMA})^T$ is commonly controlled by using an information criterion (e.g., the Akaike information criterion as done in Chapter 5) to prevent overfitting.

One-step-ahead forecasts, i.e., forecasts $\hat{y}_{k+1}$, which will be needed for optimal bidding at the PRL market, can be made by using the equation below:

$$\hat{y}^{PRL,k+1} = \sum_{p=1}^{P_{ARMA}} \phi_p \cdot y^{PRL,k+1-p} + \sum_{q=1}^{Q_{ARMA}} \theta_q \cdot \left(y^{PRL,k+1-q} - \hat{y}^{PRL,k+1-q}\right).$$

The variance of the one-step-ahead forecast is simply the model variance σ^2, i.e.,

$$var\left(\hat{y}_{k+1}\right) = \sigma^2,$$

which enables forecasts of the probability that a bid with a given ask price is accepted.

Appendix D.

Documentations of GAMS models for the generic process

D.1. Ideal storage-type customer (LP)

Symbols

Sets

Name	Domains	Description
t	*	Hours in a year (1 . . . 8760)

Parameters

Name	Domains	Description
c	t	Hourly day-ahead price [EUR per MWh]
r	t	Hourly percentage of the consumed electricity consumption that is generated from nonrenewable sources [%]
P_{nom}		Nominal power uptake [MW]

Variables

Name	Domains	Description
P	t	Power uptake in hour t [MW]
Δ	t	Rate of change in power in hour t [MW]
S	t	Stored surplus energy at beginning of hour t [MWh]
C		Electricity costs for the production [EUR]
R		Total contribution to the residual load [MWh]

Equations

Name	Domains	Description
Costs		Yearly costs
ResLoad		Yearly electricity consumption from nonrenewable sources
Storage	t	Accumulation of surplus energy
RateOfChange	t	Ramping of power uptake
Start		Initial condition for storage
Target		Production target

Equation definitions

Costs

$$C = \sum_t (c_t \cdot P_t)$$

ResLoad

$$R = \sum_t \left(\frac{r_t}{100} \cdot P_t \right)$$

Storage

$$S_{t+1} = S_t + (P_t - P_{nom}) \qquad \forall t$$

RateOfChange

$$\Delta_{t+1} = P_{t+1} - P_t \qquad \forall t$$

Start

$$S_1 = 0$$

Target

$$P_{nom} \cdot 8760 = \sum_t P_t$$

D.2. Extended storage-type customer accounting for off-design efficiency losses (MILP)

Symbols

Sets

Name	Domains	Description
t	*	Hours in a year (1 . . . 8760)
s	*	Segments for efficiency loss (1 . . . 6)

Parameters

Name	Domains	Description
c	t	Hourly day-ahead price [EUR per MWh]
r	t	Hourly percentage of the consumed electricity consumption that is generated from nonrenewable sources [%]
P_{nom}		Nominal power uptake [MW]
a	s	Intercept of segment [MW]
b	s	Slope of segment [-]
lb	s	Lower bound of segment [MW]
ub	s	Upper bound of segment [-]

Variables

Name	Domains	Description
P	t	Power uptake in hour t [MW]
Δ	t	Rate of change in power in hour t [MW]
S	t	Stored surplus energy at beginning of hour t [MWh]
C		Electricity costs for the production [EUR]
R		Total contribution to the residual load [MWh]
P_{seg}	s, t	Power uptake in each segment [MW]
P_{eff}	t	Effective power uptake accounting for efficiency losses [MW]
z	s, t	Binary variable for efficiency losses (indicates if segment is active)

Equations

Name	Domains	Description
Costs		Yearly costs
ResLoad		Yearly electricity consumption from nonrenewable sources
Storage	t	Accumulation of surplus energy
RateOfChange	t	Ramping of power uptake
Start		Initial condition for storage
Target		Production target
Section1	t	Selection of exactly one z per hour
Section2	t	Determination of individual Ps in each segment
BigM1	s, t	Sets lower bounds for power uptake depending on z
BigM2	s, t	Sets upper bounds for power uptake depending on z
Efficiency	t	Determination of effective power uptake

Equation definitions

Costs

$$C = \sum_t (c_t \cdot P_t)$$

ResLoad

$$R = \sum_t \left(\frac{r_t}{100} \cdot P_t \right)$$

Storage

$$S_{t+1} = S_t + (P_{eff,t} - P_{nom}) \qquad \forall t$$

RateOfChange

$$\Delta_{t+1} = P_{t+1} - P_t \qquad \forall t$$

Start

$$S_1 = 0$$

Target

$$P_{nom} \cdot 8760 = \sum_t P_{eff,t}$$

Section1

$$\sum_s z_{s,t} = 1 \qquad \forall t$$

Section2

$$\sum_s P_{seg,s,t} = P_t \qquad \forall t$$

BigM1

$$z_{s,t} \cdot lb_s \leq P_{seg,t,s} \qquad \forall s,t$$

BigM2

$$z_{s,t} \cdot ub_s \geq P_{seg,t,s} \qquad \forall s,t$$

Efficiency

$$P_{eff,t} = \sum_s (z_{s,t} \cdot a_s + b_s \cdot P_{seg,s,t}) \qquad \forall t$$

D.3. Ideal storage-type customer including shut-down opportunities (MILP)

Symbols

Sets

Name	Domains	Description
t	*	Hours in a year (1 . . . 8760)

Parameters

Name	Domains	Description
c	t	Hourly day-ahead price [EUR per MWh]
r	t	Hourly percentage of the consumed electricity consumption that is generated from nonrenewable sources [%]
P_{nom}		Nominal power uptake [MW]
P_{max}		Maximum power uptake [MW]
Δ_{max}		Maximum rate of change [MW]

Variables

Name	Domains	Description
P	t	Power uptake in hour t [MW]
Δ	t	Rate of change in power in hour t [MW]
S	t	Stored surplus energy at beginning of hour t [MWh]
C		Electricity costs for the production [EUR]
R		Total contribution to the residual load [MWh]
y	t	Binary variable indicating if process is running
z_{on}	t	Binary variable indicating transition to on
z_{off}	t	Binary variable indicating transition to off

Equations

Name	Domains	Description
Costs		Yearly costs
ResLoad		Yearly electricity consumption from nonrenewable sources
Storage	t	Accumulation of surplus energy
RateOfChange	t	Ramping of power uptake
Start		Initial condition for storage
Target		Production target
BigM1	t	Sets lower bounds for power uptake depending on y
BigM2	t	Sets upper bounds for power uptake depending on y
BigM3	t	Sets lower bounds for rate of change depending on y
BigM4	t	Sets upper bounds for rate of change depending on y
Transitions	t	Coupling of mode variable and transition variables
MaxDowntime	t	Shut down limited to a certain period (here 1 h)

Equation definitions

Costs

$$C = \sum_t (c_t \cdot P_t)$$

ResLoad

$$R = \sum_t \left(\frac{r_t}{100} \cdot P_t \right)$$

Storage

$$S_{t+1} = S_t + (P_t - P_{nom}) \qquad \forall t$$

RateOfChange

$$\Delta_{t+1} = P_{t+1} - P_t \qquad \forall t$$

Start

$$S_1 = 0$$

Target

$$P_{nom} \cdot 8760 = \sum_t P_t$$

BigM1

$$y_t \cdot \frac{P_{max}}{2} \leq P_t \qquad \forall t$$

BigM2

$$y_t \cdot P_{max} \geq P_t \qquad \forall t$$

BigM3

$$-y_{t+1} \cdot \Delta_{max} - (1 - y_{t+1}) \cdot \frac{P_{max}}{2} \leq \Delta_{t+1} \qquad \forall t$$

BigM4

$$y_t \cdot \Delta_{max} + (1 - y_t) \cdot \frac{P_{max}}{2} \geq \Delta_{t+1} \qquad \forall t$$

Transitions

$$z_{on,t} - z_{off_t} = y_{t+1} - y_t \qquad \forall t$$

MaxDowntime

$$y_{t+1} \geq z_{off,t} \qquad \forall t$$

Appendix E.

Results of the 2030 study using the projected time series

E.1. Effect of average utilization rates on Pareto curves

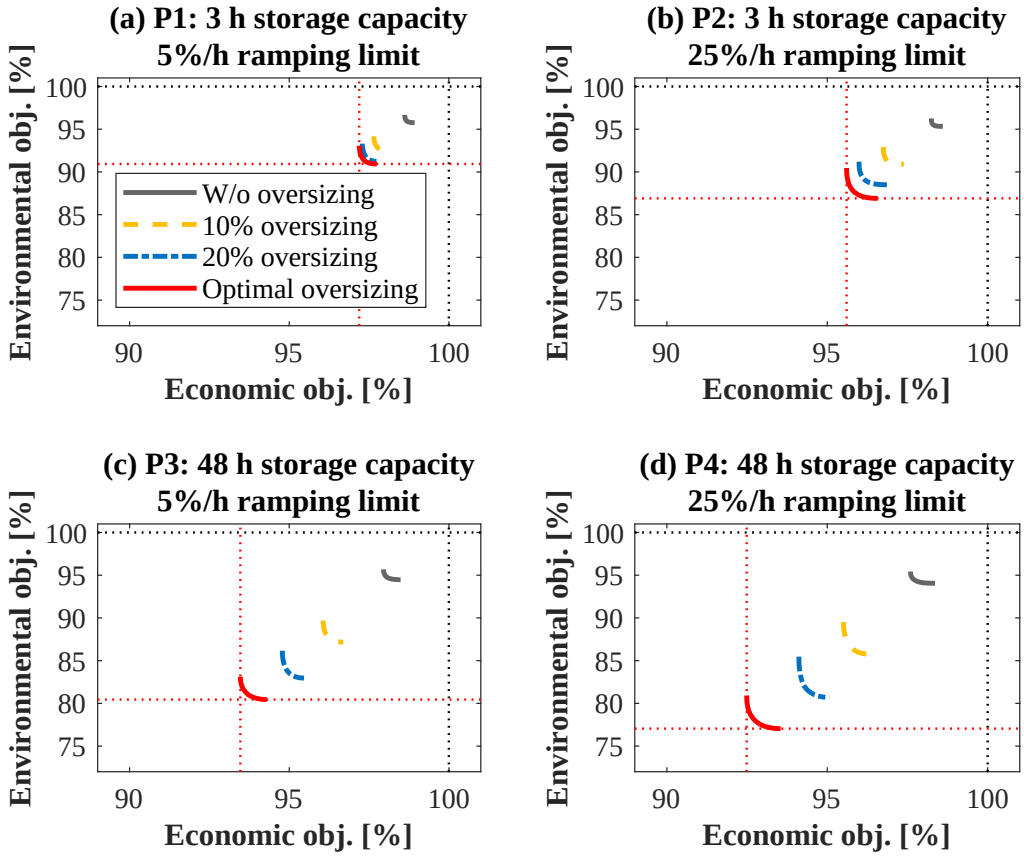

Figure E.1.: Pareto curves for different parametrizations of the ideal storage-type customer model (cf. Subsection 6.3.1). The solid gray curve is the reference with average utilization rate $\overline{UR} = 95\%$. The dashed yellow curve corresponds to a 10% oversizing compared to the reference ($\overline{UR} = 86\%$), the dashed-dotted blue curve to 20% ($\overline{UR} = 79\%$). The solid red line is the Pareto curve for an optimal average utilization rate.

E.2. Influence of off-design efficiency losses on Pareto curves

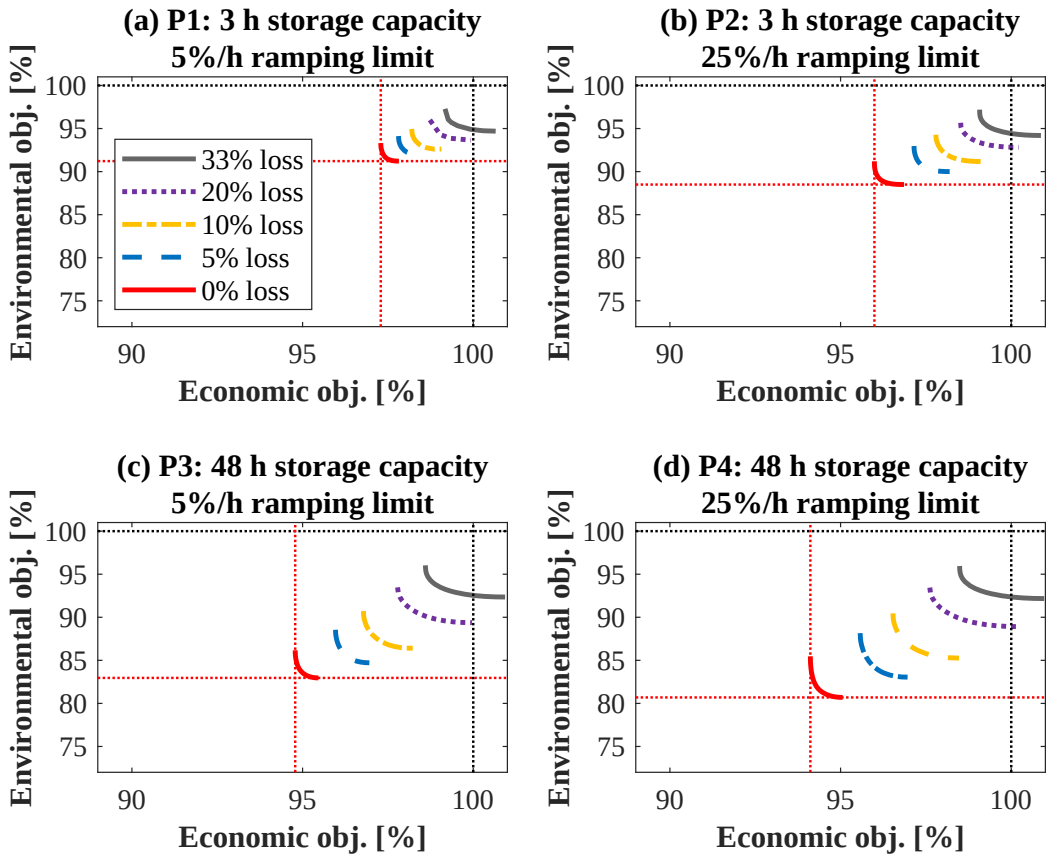

Figure E.2.: Pareto curves for different parametrizations of the storage-type customer model when assuming varying curves for quadratic off-design efficiency losses (cf. Subsection 6.3.2) and using a fixed oversizing of 20%. The solid red curve is the reference without losses. The dashed-dotted blue curve corresponds to 5%, the dashed yellow curve to 10%, the dotted purple line to 20%, and the solid gray line to 33% losses at the lower operating bounds compared to the reference.

E.3. Parameter study on storage and ramping constraints

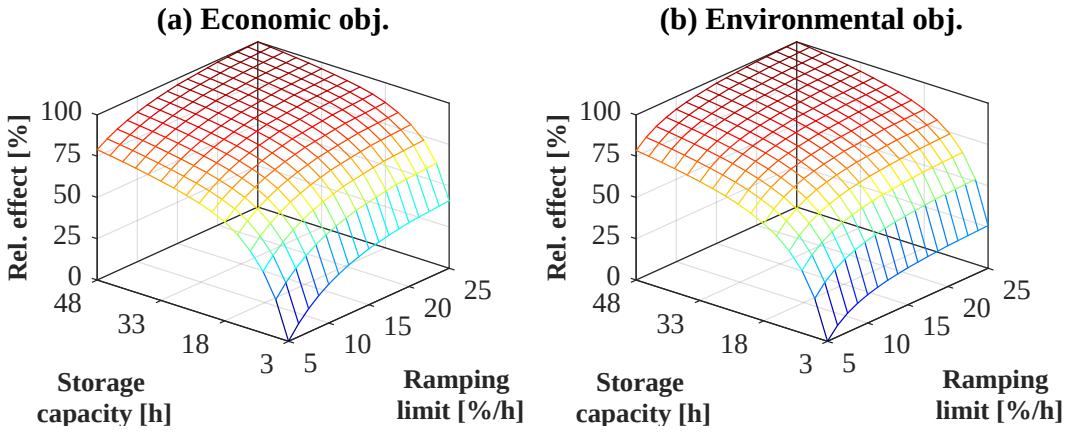

Figure E.3.: Parameter study for single-objective optimizations assuming an oversizing of 20% compared to the reference with an average utilization rate $\overline{UR} = 95\%$. The relative effect, i.e., the additional savings, from increasing the storage capacity S^{max} and loosening the ramping constraints Δ^{max} is scaled between its minimum ($S^{max} = 3$ h, $\Delta^{max} = 5\%$/h) and maximum ($S^{max} = 48$ h, $\Delta^{max} = 25\%$/h).

E.4. Improvements from shut-down opportunities

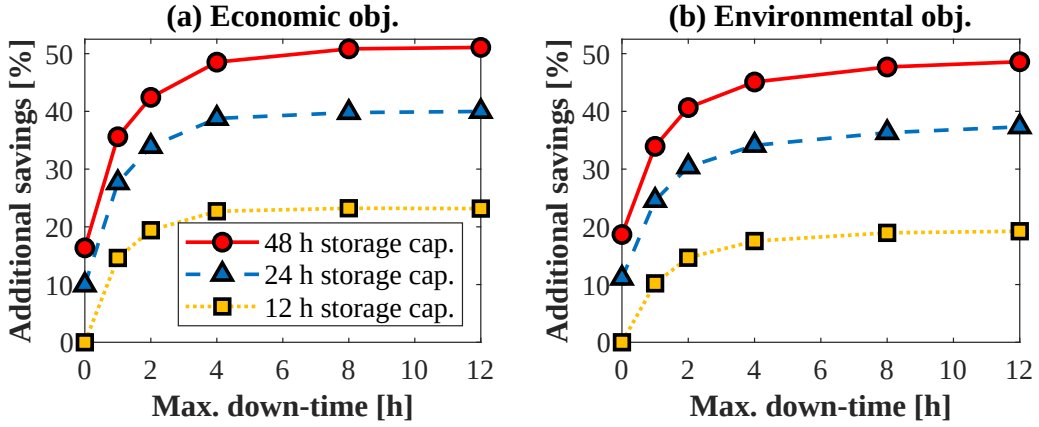

Figure E.4.: Analysis of additional savings enabled by temporary shut-downs (cf. Subsection 6.3.3) as a function of the maximum down-time. A fixed oversizing of 20% (i.e., an average utilization rate $\overline{UR} = 79\%$) is used and a ramping constraint of 10%/h is imposed. Line styles indicate the storage capacity S^{max}. Yellow squares (□) with dotted lines: $S^{max} = 12$ h, blue upwards-pointing triangles (△) with dashed lines: $S^{max} = 24$ h, and red circles (○) with solid lines: $S^{max} = 48$ h. Values are scaled to the case with $S^{max} = 12$ h and without shout-downs.

Bibliography

[1] A. Mitsos, N. Asprion, C. A. Floudas, M. Bortz, M. Baldea, D. Bonvin, A. Caspari, and P. Schäfer. Challenges in process optimization for new feedstocks and energy sources. *Computers & Chemical Engineering*, 113:209–221, 2018.

[2] P. D. Lund, J. Lindgren, J. Mikkola, and J. Salpakari. Review of energy system flexibility measures to enable high levels of variable renewable electricity. *Renewable and Sustainable Energy Reviews*, 45:785–807, 2015.

[3] M. Paulus and F. Borggrefe. The potential of demand-side management in energy-intensive industries for electricity markets in Germany. *Applied Energy*, 88(2):432–441, 2011.

[4] Q. Zhang and I. E. Grossmann. Enterprise-wide optimization for industrial demand side management: Fundamentals, advances, and perspectives. *Chemical Engineering Research and Design*, 116:114–131, 2016.

[5] M. G. Ierapetritou, D. Wu, J. Vin, P. Sweeney, and M. Chigirinskiy. Cost minimization in an energy-intensive plant using mathematical programming approaches. *Industrial & Engineering Chemistry Research*, 41(21):5262–5277, 2002.

[6] M. H. Karwan and M. F. Keblis. Operations planning with real time pricing of a primary input. *Computers & Operations Research*, 34(3):848–867, 2007.

[7] S. Mitra, J. M. Pinto, and I. E. Grossmann. Optimal multi-scale capacity planning for power-intensive continuous processes under time-sensitive electricity prices and demand uncertainty. Part I: Modeling. *Computers & Chemical Engineering*, 65: 89–101, 2014.

[8] Q. Zhang, I. E. Grossmann, C. F. Heuberger, A. Sundaramoorthy, and J. M. Pinto. Air separation with cryogenic energy storage: Optimal scheduling considering electric energy and reserve markets. *AIChE Journal*, 61(5):1547–1558, 2015.

[9] D. P. Xenos, I. M. Noor, M. Matloubi, M. Cicciotti, T. Haugen, and N. F. Thornhill. Demand-side management and optimal operation of industrial electricity consumers: An example of an energy-intensive chemical plant. *Applied Energy*, 182:418–433, 2016.

[10] R. C. Pattison, C. R. Touretzky, T. Johansson, I. Harjunkoski, and M. Baldea. Optimal process operations in fast-changing electricity markets: Framework for scheduling with low-order dynamic models and an air separation application. *Industrial & Engineering Chemistry Research*, 55(16):4562–4584, 2016.

[11] R. Adamson, M. Hobbs, A. Silcock, and M. J. Willis. Integrated real-time production scheduling of a multiple cryogenic air separation unit and compressor plant. *Computers & Chemical Engineering*, 104:25–37, 2017.

[12] I. Lotero, A. Gopalakrishnan, and T. Roba. On improving the online performance of production scheduling: Application to air separation units. *Computers & Chemical Engineering.*, 2017.

[13] S. Misra, M. Kapadi, R. D. Gudi, and R. Srihari. Energy-efficient production scheduling of a cryogenic air separation plant. *Industrial & Engineering Chemistry Research*, 56(15):4399–4414, 2017.

[14] M. T. Kelley, R. C. Pattison, R. Baldick, and M. Baldea. An MILP framework for optimizing demand response operation of air separation units. *Applied Energy*, 222: 951–966, 2018.

[15] S. Zhao, M. P. Ochoa, L. Tang, I. Lotero, A. Gopalakrishnan, and I. E. Grossmann. Novel formulation for optimal schedule with demand side management in multiproduct air separation processes. *Industrial & Engineering Chemistry Research*, 58(8): 3104–3117, 2019.

[16] A. Caspari, C. Offermanns, P. Schäfer, A. Mhamdi, and A. Mitsos. A flexible air separation process: 2. Optimal operation using economic model predictive control. *AIChE Journal*, 65(11):e16721, 2019.

[17] A. Obermeier, C. Windmeier, E. Esche, and J.-U. Repke. A discrete-time scheduling model for power-intensive processes taking fatigue of equipment into consideration. *Chemical Engineering Science*, 195:904–920, 2019.

[18] C. Tsay, A. Kumar, J. Flores-Cerrillo, and M. Baldea. Optimal demand response scheduling of an industrial air separation unit using data-driven dynamic models. *Computers & Chemical Engineering*, 126:22–34, 2019.

[19] C. A. Babu and S. Ashok. Peak load management in electrolytic process industries. *IEEE Transactions on Power Systems*, 23(2):399–405, 2008.

[20] X. Wang, N. H. El-Farra, and A. Palazoglu. Optimal scheduling of demand responsive industrial production with hybrid renewable energy systems. *Renewable Energy*, 100: 53–64, 2017.

[21] L. C. Brée, K. Perrey, A. Bulan, and A. Mitsos. Demand side management and operational mode switching in chlorine production. *AIChE Journal*, 65(7):e16352, 2019.

[22] J. I. Otashu and M. Baldea. Demand response-oriented dynamic modeling and operational optimization of membrane-based chlor-alkali plants. *Computers & Chemical Engineering*, 121:396–408, 2019.

[23] L. C. Brée, A. Bulan, R. Herding, J. Kuhlmann, A. Mitsos, K. Perrey, and K. Roh. Techno-economic comparison of flexibility options in chlorine production. *Industrial & Engineering Chemistry Research*, 59(26):12186–12196, 2020.

[24] P. M. Castro, L. Sun, and I. Harjunkoski. Resource-task network formulations for industrial demand side management of a steel plant. *Industrial & Engineering Chemistry Research*, 52(36):13046–13058, 2013.

[25] X. Zhang, G. Hug, Z. Kolter, and I. Harjunkoski. Industrial demand response by steel plants with spinning reserve provision. In *2015 North American Power Symposium (NAPS)*, pages 1–6. IEEE, 2015.

[26] H. Hadera, I. Harjunkoski, G. Sand, I. E. Grossmann, and S. Engell. Optimization of steel production scheduling with complex time-sensitive electricity cost. *Computers & Chemical Engineering*, 76:117–136, 2015.

[27] S. Zhao, I. E. Grossmann, and L. Tang. Integrated scheduling of rolling sector in steel production with consideration of energy consumption under time-of-use electricity prices. *Computers & Chemical Engineering*, 111:55–65, 2018.

[28] X. Zhang and G. Hug. Bidding strategy in energy and spinning reserve markets for aluminum smelters' demand response. In *2015 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5, 2015.

[29] R. Vujanic, S. Mariéthoz, P. Goulart, and M. Morari. Robust integer optimization and scheduling problems for large electricity consumers. In *2012 American Control Conference (ACC)*, pages 3108–3113, 2012.

[30] S. Mitra, I. E. Grossmann, J. M. Pinto, and N. Arora. Optimal production planning under time-sensitive electricity prices for continuous power-intensive processes. *Computers & Chemical Engineering*, 38:171–184, 2012.

[31] A. Ghobeity and A. Mitsos. Optimal time-dependent operation of seawater reverse osmosis. *Desalination*, 263(1):76–88, 2010.

[32] A. Jiang, J. Wang, L. T. Biegler, W. Cheng, C. Xing, and Z. Jiang. Operational cost optimization of a full-scale SWRO system under multi-parameter variable conditions. *Desalination*, 355:124–140, 2015.

[33] C. M. Williams, A. Ghobeity, A. J. Pak, and A. Mitsos. Simultaneous optimization of size and short-term operation for an RO plant. *Desalination*, 301:42–52, 2012.

[34] H. Hadera, J. Ekström, G. Sand, J. Mäntysaari, I. Harjunkoski, and S. Engell. Integration of production scheduling and energy-cost optimization using mean value cross decomposition. *Computers & Chemical Engineering*, 129:106436, 2019.

[35] K. Seo, T. F. Edgar, and M. Baldea. Optimal demand response operation of electric boosting glass furnaces. *Applied Energy*, 269:115077, 2020.

[36] C. K. Simoglou, P. N. Biskas, and A. G. Bakirtzis. Optimal self-scheduling of a thermal producer in short-term electricity markets by MILP. *IEEE Transactions on Power Systems*, 25(4):1965–1977, 2010.

[37] S. Mitra, L. Sun, and I. E. Grossmann. Optimal scheduling of industrial combined heat and power plants under time-sensitive electricity prices. *Energy*, 54:194–211, 2013.

[38] M. Kia, M. S. Nazar, M. S. Sepasian, A. Heidari, and P. Siano. Optimal day ahead scheduling of combined heat and power units with electrical and thermal storage considering security constraint of power system. *Energy*, 120:241–252, 2017.

[39] E. G. Tsimopoulos and M. C. Georgiadis. Strategic offers in day-ahead market co-optimizing energy and reserve under high penetration of wind power production: An MPEC approach. *AIChE Journal*, 65(7):e16495, 2019.

[40] E. Lizarraga-Garcia, A. Ghobeity, M. Totten, and A. Mitsos. Optimal operation of a solar-thermal power plant with energy storage and electricity buy-back from grid. *Energy*, 51:61–70, 2013.

[41] M. J. Wagner, A. M. Newman, W. T. Hamilton, and R. J. Braun. Optimized dispatch in a first-principles concentrating solar power production model. *Applied Energy*, 203: 959–971, 2017.

[42] H. Lund, G. Salgi, B. Elmegaard, and A. N. Andersen. Optimal operation strategies of compressed air energy storage (CAES) on electricity spot markets with fluctuating prices. *Applied Thermal Engineering*, 29(5):799–806, 2009.

[43] C. Olk, D. U. Sauer, and M. Merten. Bidding strategy for a battery storage in the German secondary balancing power market. *Journal of Energy Storage*, 21:787–800, 2019.

[44] I. E. Grossmann. Enterprise-wide optimization: A new frontier in process systems engineering. *AIChE Journal*, 51(7):1846–1857, 2005.

[45] E. Kondili, C. C. Pantelides, and R. W. H. Sargent. A general algorithm for short-term scheduling of batch operations - I. MILP formulation. *Computers & Chemical Engineering*, 17(2):211–227, 1993.

[46] A. D. Dimitriadis, N. Shah, and C. C. Pantelides. RTN-based rolling horizon algorithms for medium term scheduling of multipurpose plants. *Computers & Chemical Engineering*, 21:1061–1066, 1997.

[47] C. T. Maravelias and I. E. Grossmann. Minimization of the makespan with a discrete-time state-task network formulation. *Industrial & Engineering Chemistry Research*, 42(24):6252–6257, 2003.

[48] S. Velez and C. T. Maravelias. Multiple and nonuniform time grids in discrete-time MIP models for chemical production scheduling. *Computers & Chemical Engineering*, 53:70–85, 2013.

[49] J. M. Pinto and I. E. Grossmann. A continuous time mixed integer linear programming model for short term scheduling of multistage batch plants. *Industrial & Engineering Chemistry Research*, 34(9):3037–3051, 1995.

[50] G. Schilling and C. C. Pantelides. A simple continuous-time process scheduling formulation and a novel solution algorithm. *Computers & Chemical Engineering*, 20: 1221–1226, 1996.

[51] M. G. Ierapetritou and C. A. Floudas. Effective continuous-time formulation for short-term scheduling. 1. Multipurpose batch processes. *Industrial & Engineering Chemistry Research*, 37(11):4341–4359, 1998.

[52] M. G. Ierapetritou and C. A. Floudas. Effective continuous-time formulation for short-term scheduling. 2. Continuous and semicontinuous processes. *Industrial & Engineering Chemistry Research*, 37(11):4360–4374, 1998.

[53] C. T. Maravelias and I. E. Grossmann. New general continuous-time state-task network formulation for short-term scheduling of multipurpose batch plants. *Industrial & Engineering Chemistry Research*, 42(13):3056–3074, 2003.

[54] H. Lee and C. T. Maravelias. Combining the advantages of discrete- and continuous-time scheduling models: Part 1. Framework and mathematical formulations. *Computers & Chemical Engineering*, 116:176–190, 2018.

[55] H. Lee and C. T. Maravelias. Combining the advantages of discrete- and continuous-time scheduling models: Part 2. Systematic methods for determining model parameters. *Computers & Chemical Engineering*, 128:557–573, 2019.

[56] H. Lee and C. T. Maravelias. Combining the advantages of discrete- and continuous-time scheduling models: Part 3. General algorithm. *Computers & Chemical Engineering*, 139:106848, 2020.

[57] C. A. Floudas and X. Lin. Continuous-time versus discrete-time approaches for scheduling of chemical processes: A review. *Computers & Chemical Engineering*, 28 (11):2109–2129, 2004.

[58] A. Sundaramoorthy and C. T. Maravelias. Computational study of network-based mixed-integer programming approaches for chemical production scheduling. *Industrial & Engineering Chemistry Research*, 50(9):5023–5040, 2011.

[59] C. T. Maravelias. General framework and modeling approach classification for chemical production scheduling. *AIChE Journal*, 58(6):1812–1828, 2012.

[60] I. Harjunkoski, C. T. Maravelias, P. Bongers, P. M. Castro, S. Engell, I. E. Grossmann, J. Hooker, C. Méndez, G. Sand, and J. Wassick. Scope for industrial applications of production scheduling models and solution methods. *Computers & Chemical Engineering*, 62:161–193, 2014.

[61] P. M. Castro, I. Harjunkoski, and I. E. Grossmann. New continuous-time scheduling formulation for continuous plants under variable electricity cost. *Industrial & Engineering Chemistry Research*, 48(14):6701–6714, 2009.

[62] B. Daryanian, R. E. Bohn, and R. D. Tabors. Optimal demand-side response to electricity spot prices for storage-type customers. *IEEE Transactions on Power Systems*, 4(3):897–903, 1989.

[63] Q. Zhang, I. E. Grossmann, A. Sundaramoorthy, and J. M. Pinto. Data-driven construction of convex region surrogate models. *Optimization and Engineering*, 17 (2):289–332, 2016.

[64] Q. Zhang, A. Sundaramoorthy, I. E. Grossmann, and J. M. Pinto. A discrete-time scheduling model for continuous power-intensive process networks with various power contracts. *Computers & Chemical Engineering*, 84:382–393, 2016.

[65] Q. Zhang, M. F. Morari, I. E. Grossmann, A. Sundaramoorthy, and J. M. Pinto. An adjustable robust optimization approach to scheduling of continuous industrial processes providing interruptible load. *Computers & Chemical Engineering*, 86:106–119, 2016.

[66] Q. Zhang, J. L. Cremer, I. E. Grossmann, A. Sundaramoorthy, and J. M.Pinto. Risk-based integrated production scheduling and electricity procurement for continuous power-intensive processes. *Computers & Chemical Engineering*, 86:90–105, 2016.

[67] Q. Zhang, M. F. Morari, I. E. Grossmann, A. Sundaramoorthy, and J. M. Pinto. An adjustable robust optimization approach to scheduling of continuous industrial processes providing interruptible load. *Computers & Chemical Engineering*, 86:106–119, 2016.

[68] Q. Zhang, A. M. Bremen, I. E. Grossmann, and J. M. Pinto. Long-term electricity procurement for large industrial consumers under uncertainty. *Industrial & Engineering Chemistry Research*, 57(9):3333–3347, 2018.

[69] H. Teichgraeber and A. R. Brandt. Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison. *Applied Energy*, 239:1283–1293, 2019.

[70] C. E. Lythcke-Jorgensen, M. Münster, A. V. Ensinas, and F. Haglind. A method for aggregating external operating conditions in multi-generation system optimization models. *Applied Energy*, 166:59–75, 2016.

[71] K. Poncelet, H. Höschle, E. Delarue, A. Virag, and W. D'haeseleer. Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems. *IEEE Transactions on Power Systems*, 32(3):1936–1948, 2017.

[72] B. Bahl, J. Lützow, D. Shu, D. E. Hollermann, M. Lampe, M. Hennen, and A. Bardow. Rigorous synthesis of energy systems by decomposition via time-series aggregation. *Computers & Chemical Engineering*, 112:70–81, 2018.

[73] L. Kotzur, P. Markewitz, M. Robinius, and D. Stolten. Impact of different time series aggregation methods on optimal energy system design. *Renewable Energy*, 117:474–487, 2018.

[74] N. Baumgärtner, B. Bahl, M. Hennen, and A. Bardow. RiSES3:Rigorous Synthesis of Energy Supply and Storage Systems via time-series relaxation and aggregation. *Computers & Chemical Engineering*, 127:127–139, 2019.

[75] N. Baumgärtner, D. Shu, B. Bahl, M. Hennen, D. E. Hollermann, and A. Bardow. DeLoop: Decomposition-based Long-term operational optimization of energy systems with time-coupling constraints. *Energy*, 198:117272, 2020.

[76] S. Pineda and J. M. Morales. Chronological time-period clustering for optimal capacity expansion planning with storage. *IEEE Transactions on Power Systems*, 33 (6):7162–7170, 2018.

[77] M. J. Palys and P. Daoutidis. Using hydrogen and ammonia for renewable energy storage: A geographically comprehensive techno-economic study. *Computers & Chemical Engineering*, 136:106785, 2020.

[78] M. Baldea and I. Harjunkoski. Integrated production scheduling and process control: A systematic review. *Computers & Chemical Engineering*, 71:377–390, 2014.

[79] S. Sass and A. Mitsos. Optimal operation of dynamic (energy) systems: When are quasi-steady models adequate? *Computers & Chemical Engineering*, 124:133–139, 2019.

[80] J. Du, J. Park, I. Harjunkoski, and M. Baldea. A time scale-bridging approach for integrating production scheduling and process control. *Computers & Chemical Engineering*, 79:59–69, 2015.

[81] M. T. Kelley, R. C. Pattison, R. Baldick, and M. Baldea. An efficient MILP framework for integrating nonlinear process dynamics and control in optimal production scheduling calculations. *Computers & Chemical Engineering*, 110:35–52, 2018.

[82] C. Tsay and M. Baldea. Integrating production scheduling and process control using latent variable dynamic models. *Control Engineering Practice*, 94:104201, 2020.

[83] L. S. Dias, R. C. Pattison, C. Tsay, M. Baldea, and M. G. Ierapetritou. A simulation-based optimization framework for integrating scheduling and model predictive control, and its application to air separation units. *Computers & Chemical Engineering*, 113:139–151, 2018.

[84] M. Z. Jamaludin and C. L. E. Swartz. Dynamic real-time optimization with closed-loop prediction. *AIChE Journal*, 63(9):3896–3911, 2017.

[85] H. Li and C. L. E. Swartz. Dynamic real-time optimization of distributed MPC systems using rigorous closed-loop prediction. *Computers & Chemical Engineering*, 122:356–371, 2019.

[86] J. Miller, W. L. Luyben, P. Belanger, S. Blouin, and L. Megan. Improving agility of cryogenic air separation plants. *Industrial & Engineering Chemistry Research*, 47 (2):394–404, 2008.

[87] Y. Cao, C. L. E. Swartz, M. Baldea, and S. Blouin. Optimization-based assessment of design limitations to air separation plant agility in demand response scenarios. *Journal of Process Control*, 33:37–48, 2015.

[88] R. C. Pattison, C. R. Touretzky, I. Harjunkoski, and M. Baldea. Moving horizon closed-loop production scheduling using dynamic process models. *AIChE Journal*, 63(2):639–651, 2017.

[89] P. Schäfer, L. F. Bering, A. Caspari, A. Mhamdi, and A. Mitsos. Nonlinear dynamic optimization for improved load-shifting agility of cryogenic air separation plants. In M. R. Eden, M. G. Ierapetritou, and G. P. Towler, editors, *13th International Symposium on Process Systems Engineering (PSE 2018)*, volume 44 of *Computer Aided Chemical Engineering*, pages 547–552. Elsevier, 2018.

[90] S. Engell. Feedback control for optimal process operation. *Journal of Process Control*, 17(3):203–219, 2007.

[91] R. Amrit, J. B. Rawlings, and L. T. Biegler. Optimizing process economics online using model predictive control. *Computers & Chemical Engineering*, 58:334–343, 2013.

[92] M. Ellis, H. Durand, and P. D. Christofides. A tutorial review of economic model predictive control methods. *Journal of Process Control*, 24(8):1156–1178, 2014.

[93] M. Diehl, R. Amrit, and J. B. Rawlings. A Lyapunov function for economic optimizing model predictive control. *IEEE Transactions on Automatic Control*, 56(3): 703–707, 2011.

[94] M. Heidarinejad, J. Liu, and P. D. Christofides. Economic model predictive control of nonlinear process systems using Lyapunov techniques. *AIChE Journal*, 58(3): 855–870, 2012.

[95] D. Angeli, R. Amrit, and J. B. Rawlings. On average performance and stability of economic model predictive control. *IEEE Transactions on Automatic Control*, 57(7): 1615–1626, 2012.

[96] R. Huang, V. M. Zavala, and L. T. Biegler. Advanced step nonlinear model predictive control for air separation units. *Journal of Process Control*, 19(4):678–685, 2009.

[97] R. Huang and L. T. Biegler. Economic NMPC for energy intensive applications with electricity price prediction. In I. A. Karimi and R. Srinivasan, editors, *11th International Symposium on Process Systems Engineering*, volume 31 of *Computer Aided Chemical Engineering*, pages 1612–1616. Elsevier, 2012.

[98] A. Caspari, J. M. M. Faust, P. Schäfer, A. Mhamdi, and A. Mitsos. Economic nonlinear model predictive control for flexible operation of air separation units. *IFAC-PapersOnLine*, 51(20):295–300, 2018.

[99] A. Caspari, C. Tsay, A. Mhamdi, M. Baldea, and A. Mitsos. The integration of scheduling and control: Top-down vs. bottom-up. *Journal of Process Control*, 91: 50–62, 2020.

[100] A. Caspari, Y. M. Pérez, C. Offermanns, P. Schäfer, A.-M. Ecker, A. Peschel, F. Schliebitz, G. Zapp, A. Mhamdi, and A. Mitsos. Economic nonlinear model predictive control of multi-product air separation processes. In A. A. Kiss, E. Zondervan, R. Lakerveld, and L. Özkan, editors, *29th European Symposium on Computer Aided Process Engineering*, volume 46 of *Computer Aided Chemical Engineering*, pages 1–6. Elsevier, 2019.

[101] A. Caspari, C. Offermanns, P. Schäfer, A. Mhamdi, and A. Mitsos. A flexible air separation process: 1. Design and steady-state optimizations. *AIChE Journal*, 65 (11):e16705, 2019.

[102] E. D. Gilles and B. Retzbach. Reduced models and control of distillation columns with sharp temperature profiles. *19th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes*, 1980.

[103] W. Marquardt. Nonlinear model reduction for binary distillation. In C. McGreavy, editor, *Dynamics and Control of Chemical Reactors and Distillation Columns*, IFAC Symposia Series, pages 123–128. Pergamon, 1988.

[104] A. Kienle. Low-order dynamic models for ideal multicomponent distillation processes using nonlinear wave propagation theory. *Chemical Engineering Science*, 55(10): 1817–1828, 2000.

[105] A. Caspari, C. Offermanns, A.-M. Ecker, M. Pottmann, G. Zapp, A. Mhamdi, and A. Mitsos. A wave propagation approach for reduced dynamic modeling of distillation columns: Optimization and control. *Journal of Process Control*, 91:12–24, 2020.

[106] Y. S. Cho and B. Joseph. Reduced-order steady-state and dynamic models for separation processes. Part II. Application to nonlinear multicomponent systems. *AIChE Journal*, 29(2):270–276, 1983.

[107] W. E. Stewart, K. L. Levien, and M. Morari. Simulation of fractionation by orthogonal collocation. *Chemical Engineering Science*, 40(3):409–421, 1985.

[108] P. Seferlis and A. N. Hrymak. Optimization of distillation units using collocation models. *AIChE Journal*, 40(5):813–825, 1994.

[109] A. Benallou, D. E. Seborg, and D. A. Mellichamp. Dynamic compartmental models for separation processes. *AIChE Journal*, 32(7):1067–1078, 1986.

[110] R. R. Horton, B. W. Bequette, and T. F. Edgar. Improvements in dynamic compartmental modeling for distillation. *Computers & Chemical Engineering*, 15(3):197–201, 1991.

[111] J. Lévine and P. Rouchon. Quality control of binary distillation columns via nonlinear aggregated models. *Automatica*, 27(3):463–480, 1991.

[112] A. Linhart and S. Skogestad. Reduced distillation models via stage aggregation. *Chemical Engineering Science*, 65(11):3439–3456, 2010.

[113] Y. Fu and X. Liu. An advanced control of heat integrated air separation column based on simplified wave model. *Journal of Process Control*, 49:45–55, 2017.

[114] X. Liu, Y. Zhou, L. Cong, and J. Zhang. Nonlinear wave modeling and dynamic analysis of internal thermally coupled distillation columns. *AIChE Journal*, 58(4): 1146–1156, 2012.

[115] Y. Cao, C. L. E. Swartz, J. Flores-Cerrillo, and J. Ma. Dynamic modeling and collocation-based model reduction of cryogenic air separation units. *AIChE Journal*, 62(5):1602–1615, 2016.

[116] Y. Cao, C. L. E. Swartz, and J. Flores-Cerrillo. Optimal dynamic operation of a high-purity air separation plant under varying market conditions. *Industrial & Engineering Chemistry Research*, 55(37):9956–9970, 2016.

[117] Y. Cao, C. L. E. Swartz, and J. Flores-Cerrillo. Preemptive dynamic operation of cryogenic air separation units. *AIChE Journal*, 63(9):3845–3859, 2017.

[118] S. Khowinij, M. A. Henson, P. Belanger, and L. Megan. Dynamic compartmental modeling of nitrogen purification columns. *Separation and Purification Technology*, 46(1):95–109, 2005.

[119] S. Bian, S. Khowinij, M. A. Henson, P. Belanger, and L. Megan. Compartmental modeling of high purity air separation columns. *Computers & Chemical Engineering*, 29(10):2096–2109, 2005.

[120] Z. Chen, M. A. Henson, P. Belanger, and L. Megan. Nonlinear model predictive control of high purity distillation columns for cryogenic air separation. *IEEE Transactions on Control Systems Technology*, 18(4):811–821, 2010.

[121] J. I. Otashu and M. Baldea. Scheduling chemical processes for frequency regulation. *Applied Energy*, 260:114125, 2020.

[122] A. W. Dowling, R. Kumar, and V. M. Zavala. A multi-scale optimization framework for electricity market participation. *Applied Energy*, 190:147–164, 2017.

[123] J. I. Otashu and M. Baldea. Grid-level "battery" operation of chemical processes and demand-side participation in short-term electricity markets. *Applied Energy*, 220:562–575, 2018.

[124] A. Sternberg and A. Bardow. Power-to-What? - Environmental assessment of energy storage systems. *Energy & Environmental Science*, 8:389–400, 2015.

[125] M. Matzen and Y. Demirel. Methanol and dimethyl ether from renewable hydrogen and carbon dioxide: Alternative fuels production and life-cycle assessment. *Journal of Cleaner Production*, 139:1068–1077, 2016.

[126] S. Deutz, D. Bongartz, B. Heuser, A. Kätelhön, L. Schulze-Langenhorst, A. Omari, M. Walters, J. Klankermayer, W. Leitner, A. Mitsos, S. Pischinger, and A. Bardow. Cleaner production of cleaner fuels: wind-to-wheel – environmental assessment of CO2-based oxymethylene ether as a drop-in fuel. *Energy & Environmental Science*, 11:331–343, 2018.

[127] J. C. Koj, C. Wulf, and P. Zapp. Environmental impacts of power-to-X systems - A review of technological and methodological choices in life cycle assessments. *Renewable and Sustainable Energy Reviews*, 112:865–879, 2019.

[128] P. Finn and C. Fitzpatrick. Demand side management of industrial electricity consumption: Promoting the use of renewable energy through real-time pricing. *Applied Energy*, 113:11–21, 2014.

[129] M. Kopsakangas-Savolainen, M. K. Mattinen, K. Manninen, and A. Nissinen. Hourly-based greenhouse gas emissions of electricity – cases demonstrating possibilities for households and companies to decrease their emissions. *Journal of Cleaner Production*, 153:384–396, 2017.

[130] D. L. Summerbell, D. Khripko, C. Barlow, and J. Hesselbach. Cost and carbon reductions from industrial demand-side management: Study of potential savings at a cement plant. *Applied Energy*, 197:100–113, 2017.

[131] M. T. Kelley, R. Baldick, and M. Baldea. Demand response operation of electricity-intensive chemical processes for reduced greenhouse gas emissions: Application to an air separation unit. *ACS Sustainable Chemistry & Engineering*, 7(2):1909–1922, 2019.

[132] N. Baumgärtner, R. Delorme, M. Hennen, and A. Bardow. Design of low-carbon utility systems: Exploiting time-dependent grid emissions for climate-friendly demand-side management. *Applied Energy*, 247:755–765, 2019.

[133] A. Cozad, N. V. Sahinidis, and D. C. Miler. Learning surrogate models for simulation-based optimization. *AIChE Journal*, 60(6):2211–2227, 2015.

[134] A. M. Schweidtmann and A. Mitsos. Deterministic global optimization with artificial neural networks embedded. *Journal of Optimization Theory and Applications*, 180 (3):925–948, 2019.

[135] A. Mitsos, B. Chachuat, and P. Barton. McCormick-based relaxations of algorithms. *SIAM Journal on Optimization*, 20(2):573–601, 2009.

[136] D. Bongartz and A. Mitsos. Deterministic global optimization of process flowsheets in a reduced space using McCormick relaxations. *Journal of Global Optimization*, 69 (4):761–796, 2017.

[137] D. Bongartz and A. Mitsos. Deterministic global flowsheet optimization: Between equation-oriented and sequential-modular methods. *AIChE Journal*, 65(3):1022–1034, 2019.

[138] A. M. Schweidtmann, W. R. Huster, J. T. Lüthje, and A. Mitsos. Deterministic global process optimization: Accurate (single-species) properties via artificial neural networks. *Computers & Chemical Engineering*, 121:67–74, 2019.

[139] A. M. Schweidtmann, D. Bongartz, W. R. Huster, and A. Mitsos. Deterministic global process optimization: Flash calculations via artificial neural networks. In A. A. Kiss, E. Zondervan, R. Lakerveld, and L. Özkan, editors, *29th European Symposium on Computer Aided Process Engineering*, volume 46 of *Computer Aided Chemical Engineering*, pages 937–942. Elsevier, 2019.

[140] C. Nentwich, J. Winz, and S. Engell. Surrogate modeling of fugacity coefficients using adaptive sampling. *Industrial & Engineering Chemistry Research*, 58(40):18703–18716, 2019.

[141] F. Klaucke, T. Karsten, F. Holtrup, E. Esche, T. Morosuk, G. Tsatsaronis, and J.-U. Repke. Demand response potentials for the chemical industry. *Chemie Ingenieur Technik*, 89(9):1133–1141, 2017.

[142] M. Schlegel, K. Stockmann, T. Binder, and W. Marquardt. Dynamic optimization using adaptive control vector parameterization. *Computers & Chemical Engineering*, 29(8):1731–751, 2005.

[143] B. B. Hubbard. *The world according to wavelets: the story of a mathematical technique in the making*. AK Peters, Ltd., 1996.

[144] S. G. Mallat. Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$. *Transactions of the American Mathematical Society*, 315(1):69–87, 1989.

[145] A. Prata, J. Oldenburg, A. Kroll, and W. Marquardt. Integrated scheduling and dynamic optimization of grade transitions for a continuous polymerization reactor. *Computers & Chemical Engineering*, 32(3):463–476, 2008.

[146] A. Hartwich and W. Marquardt. Dynamic optimization of the load change of a large-scale chemical plant by adaptive single shooting. *Computers & Chemical Engineering*, 34(11):1873–1889, 2010.

[147] X. Luo, J. Wang, M. Dooner, and J. Clarke. Overview of current development in electrical energy storage technologies and the application potential in power system operation. *Applied Energy*, 137:511–536, 2015.

[148] H. Lund and G. Salgi. The role of compressed air energy storage (CAES) in future sustainable energy systems. *Energy Conversion and Management*, 50(5):1172–1179, 2009.

[149] I. E. Grossmann and S. Lee. Generalized convex disjunctive programming: Nonlinear convex hull relaxation. *Computational Optimization and Applications*, 26(1):83–100, 2003.

[150] D. Bongartz, J. Najman, S. Sass, and A. Mitsos. MAiNGO - McCormick-based algorithm for mixed-integer nonlinear global optimization. `http://permalink.avt.rwth-aachen.de/?id=729717`.

[151] G. P. McCormick. Computability of global solutions to factorable nonconvex programs: Part I - Convex underestimating problems. *Mathematical Programming*, 10 (1):147–175, 1976.

[152] A. Tsoukalas and A. Mitsos. Multivariate McCormick relaxations. *Journal of Global Optimization*, 59(2):633–662, 2014.

[153] M. Tawarmalani and N. V. Sahinidis. A polyhedral branch-and-cut approach to global optimization. *Mathematical Programming*, 103(2):225–249, 2005.

[154] D. Kraft. Algorithm 733: TOMP - fortran modules for optimal control calculations. *ACM Transactions on Mathematical Software*, 20(3):262–281, 1994.

[155] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006.

[156] O. Exler and K. Schittkowski. A trust region SQP algorithm for mixed-integer nonlinear programming. *Optimization Letters*, 1(3):269–280, 2007.

[157] A. Flores-Quiroz, R. Palma-Behnke, G. Zakeri, and R. Moreno. A column generation approach for solving generation expansion planning problems with high renewable energy penetration. *Electric Power Systems Research*, 136:232–241, 2016. ISSN 0378-7796.

[158] C. L. Lara, D. S. Mallapragada, D. J. Papageorgiou, A. Venkatesh, and I. E. Grossmann. Deterministic electric power infrastructure planning: Mixed-integer programming model and nested decomposition algorithm. *European Journal of Operational Research*, 271(3):1037–1054, 2018.

[159] A. Flores-Quiroz, J. M. Pinto, and Q. Zhang. A column generation approach to multiscale capacity planning for power-intensive process networks. *Optimization and Engineering*, 20:1001–1027, 2019.

[160] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using wavelet transform. *IEEE Transactions on Image Processing*, 1(2):205–220, 1992.

[161] R. Kender, B. Wunderlich, I. Thomas, A. Peschel, S. Rehfeldt, and H. Klein. Pressure-driven dynamic simulation of start up and shutdown procedures of distillation columns in air separation units. *Chemical Engineering Research and Design*, 147:98–112, 2019.

[162] L. T. Biegler. *Nonlinear programming*. Society for Industrial and Applied Mathematics, 2010.

[163] M. Schlegel and W. Marquardt. Detection and exploitation of the control switching structure in the solution of dynamic optimization problems. *Journal of Process Control*, 16(3):275–290, 2006.

[164] M. Schlegel and W. Marquardt. Adaptive switching structure detection for the solution of dynamic optimization problems. *Industrial & Engineering Chemistry Research*, 45(24):8083–8094, 2006.

[165] E. Aydin, D. Bonvin, and K. Sundmacher. Toward fast dynamic optimization: An indirect algorithm that uses parsimonious input parameterization. *Industrial & Engineering Chemistry Research*, 57(30):10038–10048, 2018.

[166] A. M. Sahlodin, H. A. J. Watson, and P. I. Barton. Nonsmooth model for dynamic simulation of phase changes. *AIChE Journal*, 62(9):3334–3351, 2016.

[167] K. A. Khan and P. I. Barton. Generalized derivatives for hybrid systems. *IEEE Transactions on Automatic Control*, 62(7):3193–3208, 2017.

[168] P. Stechlinski, M. Patrascu, and P. I. Barton. Nonsmooth differential-algebraic equations in chemical engineering. *Computers & Chemical Engineering*, 114:52–68, 2018.

[169] P. Barton, K. Khan, P. Stechlinski, and H. Watson. Computationally relevant generalized derivatives: theory, evaluation and applications. *Optimization Methods and Software*, 33(4–6):1030–1072, 2018.

[170] A. Raghunathan and L. Biegler. Mathematical programs with equilibrium constraints (MPECs) in process engineering. *Computers & Chemical Engineering*, 27(10):1381–1392, 2003.

[171] A. U. Raghunathan, M. Soledad Diaz, and L. T. Biegler. An MPEC formulation for dynamic optimization of distillation operations. *Computers & Chemical Engineering*, 28(10):2037–2052, 2004.

[172] B. T. Baumrucker, J. G. Renfro, and L. T. Biegler. MPEC problem formulations and solution strategies with chemical engineering applications. *Computers & Chemical Engineering*, 32(12):2903–2913, 2008.

[173] A. Caspari, L. Lüken, P. Schäfer, Y. Vaupel, A. Mhamdi, L. T. Biegler, and A. Mitsos. Dynamic optimization with complementarity constraints: Smoothing for direct shooting. *Computers & Chemical Engineering*, 139:106891, 2020.

[174] W. Marquardt. Wave propagation phenomena in chemical engineering processes. *Chemie Ingenieur Technik*, 61(5):362–377, 1989.

[175] R. Carta, G. Tola, A. Servida, and M. Morbidelli. Performance of collocation models for simulating transient multistage separation units. *Computers & Chemical Engineering*, 19(11):1141–1151, 1995.

[176] A. Linhart and S. Skogestad. Computational performance of aggregated distillation models. *Computers & Chemical Engineering*, 33(1):296–308, 2009.

[177] M. J. Willis, C. D. Massimo, G. A. Montague, M. T. Tham, and A. J. Morris. Artificial neural networks in process engineering. *IEE Proceedings D - Control Theory and Applications (V*, 138(3):256–266, 1991.

[178] 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*, 114:111–121, 2018.

[179] V. Venkatasubramanian. The promise of artificial intelligence in chemical engineering: Is it here, finally? *AIChE Journal*, 65(2):466–478, 2019.

[180] M. L. Thompson and M. A. Kramer. Modeling chemical processes using prior knowledge and neural networks. *AIChE Journal*, 40(8):1328–1340, 1994.

[181] M. von Stosch, R. Oliveira, J. Peres, and S. F. de Azevedo. Hybrid semi-parametric modeling in process systems engineering: Past, present and future. *Computers & Chemical Engineering*, 60:86–101, 2014.

[182] K. McBride and K. Sundmacher. Overview of surrogate modeling in chemical process engineering. *Chemie Ingenieur Technik*, 91(3):228–239, 2019.

[183] B. Roffel, B. H. L. Betlem, and J. A. F. de Ruijter. First principles dynamic modeling and multivariable control of a cryogenic distillation process. *Computers & Chemical Engineering*, 24(1):111–123, 2000.

[184] W. R. Huster, A. M. Schweidtmann, and A. Mitsos. Working fluid selection for organic rankine cycles via deterministic global optimization of design and operation. *Optimization and Engineering*, 21:517–536, 2020.

[185] C. M. Bishop. *Pattern recognition and machine learning (information science and statistics)*. Springer, 2006.

[186] S. S. Haykin. *Neural networks and learning machines*. Pearson Education, 3rd edition, 2009.

[187] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

[188] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

[189] J. Eason and S. Cremaschi. Adaptive sequential sampling for surrogate model generation with artificial neural network. *Computers & Chemical Engineering*, 68:220–232, 2014.

[190] S. S. Garud, I. A. Karimi, and M. Kraft. Smart Sampling Algorithm for surrogate model development. *Computers & Chemical Engineering*, 96:103–114, 2017.

[191] Y. Tian, J. Zhang, and J. Morris. Modeling and optimal control of a batch polymerization reactor using a hybrid stacked recurrent neural network model. *Industrial & Engineering Chemistry Research*, 40(21):4525–4535, 2001.

[192] Z. Wu, A. Tran, D. Rincon, and P. D. Christofides. Machine learning-based predictive control of nonlinear processes. Part I: Theory. *AIChE Journal*, 65(11):e16729, 2019.

[193] G. Y. Zhu, M. A. Henson, and L. Megan. Low-order dynamic modeling of cryogenic distillation columns based on nonlinear wave phenomenon. *Separation and Purification Technology*, 24(3):467–487, 2001.

[194] A. Caspari, A. M. Bremen, J. M. M. Faust, F. Jung, C. D. Kappatou, S. Sass, Y. Vaupel, R. Hannemann-Tamás, A. Mhamdi, and A. Mitsos. DyOS - A framework for optimization of large-scale differential algebraic equation systems. In A. A. Kiss, E. Zondervan, R. Lakerveld, and L. Özkan, editors, *29th European Symposium*

on Computer Aided Process Engineering, volume 46 of *Computer Aided Chemical Engineering*, pages 619–624. Elsevier, 2019.

[195] R. W. H. Sargent and G. R. Sullivan. The development of an efficient optimal control package. In J. Stoer, editor, *Optimization Techniques*, pages 158–168. Springer, 1978.

[196] D. Kraft. On converting optimal control problems into nonlinear programming problems. In K. Schittkowski, editor, *Computational Mathematical Programming*, pages 261–280. Springer, 1985.

[197] P. Gill, W. Murray, and M. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Review*, 47(1):99–131, 2005.

[198] R. Hannemann, W. Marquardt, U. Naumann, and B. Gendler. Discrete first- and second-order adjoints and automatic differentiation for the sensitivity analysis of dynamic models. *Procedia Computer Science*, 1(1):297–305, 2010. ICCS 2010.

[199] M. Schlegel, W. Marquardt, R. Ehrig, and U. Nowak. Sensitivity analysis of linearly-implicit differential–algebraic systems by one-step extrapolation. *Applied Numerical Mathematics*, 48(1):83–102, 2004.

[200] B. Øksendal. *Stochastic differential equations: An introduction with applications*. Springer, 2003.

[201] G. Fridgen, L. Häfner, C. König, and T. Sachs. Providing utility to utilities: The value of information systems enabled flexibility in electricity consumption. *Journal of the Association for Information Systems*, 17(8):555–580, 2016.

[202] M. Diehl, H. Bock, and J. Schlöder. A real-time iteration scheme for nonlinear optimization in optimal feedback control. *SIAM Journal on Control and Optimization*, 43(5):1714–1736, 2005.

[203] M. Diehl, H. J. Ferreau, and N. Haverbeke. Efficient numerical methods for nonlinear MPC and moving horizon estimation. In L. Magni, D. M. Raimondo, and F. Allgöwer, editors, *Nonlinear model predictive vontrol: Towards new challenging applications*, pages 391–417. Springer, 2009.

[204] J. E. Cuthrell and L. T. Biegler. On the optimization of differential-algebraic process systems. *AIChE Journal*, 33(8):1257–1270, 1987.

[205] V. M. Zavala and L. T. Biegler. The advanced-step NMPC controller: Optimality, stability and robustness. *Automatica*, 45(1):86–93, 2009.

[206] L. Würth, R. Hannemann, and W. Marquardt. Neighboring-extremal updates for nonlinear model-predictive control and dynamic real-time optimization. *Journal of Process Control*, 19(8):1277–1288, 2009.

[207] X. Yang and L. T. Biegler. Advanced-multi-step nonlinear model predictive control. *Journal of Process Control*, 23(8):1116–1128, 2013.

[208] A. Helbig, O. Abel, and W. Marquardt. Structural concepts for optimization based control of transient processes. In F. Allgöwer and A. Zheng, editors, *Nonlinear model predictive control*, pages 295–311. Birkhäuser, 2000.

[209] R. Scattolini. Architectures for distributed and hierarchical model predictive control – A review. *Journal of Process Control*, 19(5):723–731, 2009.

[210] V. M. Zavala, C. D. Laird, and L. T. Biegler. A fast moving horizon estimation algorithm based on nonlinear programming sensitivity. *Journal of Process Control*, 18(9):876–884, 2008.

[211] R. Schneider, R. Hannemann-Tamás, and W. Marquardt. An iterative partition-based moving horizon estimator with coupled inequality constraints. *Automatica*, 61:302–307, 2015.

[212] A. Linhart and S. Skogestad. An aggregation model reduction method for one-dimensional distributed systems. *AIChE Journal*, 58(5):1524–1537, 2012.

[213] W. R. Huster, Y. Vaupel, A. Mhamdi, and A. Mitsos. Validated dynamic model of an organic rankine cycle (ORC) for waste heat recovery in a diesel truck. *Energy*, 151:647–661, 2018.

[214] Y. Vaupel, W. R. Huster, F. Holtorf, A. Mhamdi, and A. Mitsos. Analysis and improvement of dynamic heat exchanger models for nominal and start-up operation. *Energy*, 169:1191–1201, 2013.

[215] J. R. Birge and F. Louveaux. *Introduction to stochastic programming*. Springer, 2nd edition, 2011.

[216] C. Li and I. E. Grossmann. A finite ϵ-convergence algorithm for two-stage stochastic convex nonlinear programs with mixed-binary first and second-stage variables. *Journal of Global Optimization*, 75:921–947, 2019.

[217] G. E. P. Box and G. M. Jenkins. *Time series analysis: Forecasting and control*. Prentice Hall PTR, 3rd edition, 1994.

[218] H. Akaike. Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, and G. Kitagawa, editors, *Selected papers of Hirotugu Akaike*, pages 199–213. Springer, 1998.

[219] C. A. Floudas. *Nonlinear and mixed-integer optimization: Fundamentals and applications*. Topics in Chemical Engineering. Oxford University Press, 1995. ISBN 9780195356557.

[220] T. Fischer and C. Krauss. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2): 654–669, 2018.

[221] S. Mei, W. Wei, and F. Liu. On engineering game theory with its application in power systems. *Control Theory and Technology*, 15(1):1–12, 2017.

[222] Y. Shao and V. M. Zavala. Space-time dynamics of electricity markets incentivize technology decentralization. *Computers & Chemical Engineering*, 127:31–40, 2019.

[223] P. Schäfer, N. Hansmann, S. Ilieva, and A. Mitsos. Model-based bidding strategies for simultaneous optimal participation in different balancing markets. In A. A. Kiss, E. Zondervan, R. Lakerveld, and L. Özkan, editors, *29th European Symposium on Computer Aided Process Engineering*, volume 46 of *Computer Aided Chemical Engineering*, pages 1639–1644. Elsevier, 2019.

[224] S. Mouret, I. E. Grossmann, and P. Pestiaux. A new Lagrangian decomposition approach applied to the integration of refinery planning and crude-oil scheduling. *Computers & Chemical Engineering*, 35(12):2750–2766, 2011.

[225] X. Li, Y. Chen, and P. I. Barton. Nonconvex generalized Benders decomposition with piecewise convex relaxations for global optimization of integrated process design and operation problems. *Industrial & Engineering Chemistry Research*, 51(21):7287–7299, 2012.

[226] D. S. Kirschen and G. Strbac. *Fundamentals of power system economics*. Wiley, 2004.

[227] F. Sensfuß, M. Ragwitz, and M. Genoese. The merit-order effect: A detailed analysis of the price effect of renewable electricity generation on spot market prices in Germany. *Energy Policy*, 36(8):3086–3094, 2008.

[228] T. Kern, F. Böing, and S. von Roon. Merit Order Netz-Ausbau 2030 - Preiszeitreihe Basisszenario. `https://www.ffe.de/themen-und-methoden/speicher-und-netze/521-merit-order-netz-ausbau-2030-mona-2030`, 2018.

[229] J. Burre, D. Bongartz, L. C. Brée, K. Roh, and A. Mitsos. Power-to-X: Between electricity storage, e-production, and demand side management. *Chemie Ingenieur Technik*, 92(1–2):74–84, 2020.

[230] N. Depree, R. Düssel, P. Patel, and T. Reek. The virtual battery - Operating an aluminium smelter with flexible energy input. In E. Williams, editor, *145th TMS Annual Meeting and Exhibition*, Light Metals, pages 571–576. Springer, 2016.

[231] C. W. Gellings. The concept of demand-side management for electric utilities. *Proceedings of the IEEE*, 73(10):1468–1470, 1985.

[232] F. Ausfelder, C. Beilmann, M. Bertau, S. Bräuninger, A. Heinzel, R. Hoer, W. Koch, F. Mahlendorf, A. Metzelthin, M. Peuckert, L. Plass, K. Räuchle, M. Reuter, G. Schaub, S. Schiebahn, E. Schwab, F. Schüth, D. Stolten, G. Teßmer, K. Wagemann, and K.-F. Ziegahn. Energy storage technologies as options to a secure energy supply. *Chemie Ingenieur Technik*, 87(1–2):17–89, 2015.

[233] K. Roh, L. C. Brée, K. Perrey, A. Bulan, and A. Mitsos. Flexible operation of switchable chlor-alkali electrolysis for demand side management. *Applied Energy*, 255:113880, 2019.

[234] N. V. Sahinidis and I. E. Grossmann. Reformulation of multiperiod MILP models for planning and scheduling of chemical processes. *Computers & Chemical Engineering*, 15(4):255–272, 1991.

[235] M. Erdirik-Dogan and I. E. Grossmann. Simultaneous planning and scheduling of single-stage multi-product continuous plants with parallel lines. *Computers & Chemical Engineering*, 32(11):2664–2683, 2008.

[236] J. Burger, E. Ströfer, and H. Hasse. Production process for diesel fuel components poly(oxymethylene) dimethyl ethers from methane-based products by hierarchical optimization with varying model depth. *Chemical Engineering Research and Design*, 91(12):2648–2662, 2013.

[237] G. Wang, A. Mitsos, and W. Marquardt. Conceptual design of ammonia-based energy storage system: System design and time-invariant performance. *AIChE Journal*, 63 (5):1620–1637, 2017.

[238] D. Bongartz, J. Burre, and A. Mitsos. Production of oxymethylene dimethyl ethers from hydrogen and carbon dioxide - Part I: Modeling and analysis for OME1. *Industrial & Engineering Chemistry Research*, 58(12):4881–4889, 2019.

[239] N. Kockmann. Digital methods and tools for chemical equipment and plants. *Reaction Chemistry & Engineering*, 4(9):1522–1529, 2019.

[240] A. Bamberg, L. Urbas, S. Bröcker, N. Kockmann, and M. Bortz. What makes the digital twin an ingenious companion? *Chemie Ingenieur Technik*, 92(3):192–198, 2020.

[241] N. Asprion. Modeling, simulation, and optimization 4.0 for a distillation column. *Chemie Ingenieur Technik*, 92(7):879–889, 2020.

[242] E.-M. Ländner, A. Märtz, M. Schöpf, and M. Weibelzahl. From energy legislation to investment determination: Shaping future electricity markets with different flexibility options. *Energy Policy*, 129:1100–1110, 2019.

[243] B. Schäfer, C. Beck, K. Aihara, D. Witthaut, and M. Timme. Non-Gaussian power grid frequency fluctuations characterized by Lévy-stable laws and superstatistics. *Nature Energy*, 3:119–126, 2018.